ZDM Mathematics Education (2015) 47:1195–1209
DOI 10.1007/s11858-015-0710-7

ORIGINAL ARTICLE

Reverse-scaffolding algebra: empirical evaluation of design architecture

Kiera Chase[1] · Dor Abrahamson[1]

Accepted: 3 July 2015 / Published online: 17 July 2015

Abstract Scaffolding is the asymmetrical social co-enactment of natural or cultural practice, wherein a more able agent implements or performs for a novice elements of a challenging activity. What the novice may not learn, however, is *how* the expert's co-enactments support the activity. Granted, in many cultural practices novices need not understand underlying process. But where process is content, such as mathematics, scaffolding is liable to undermine tenets of reform-oriented pedagogy. We point to tensions between traditional conceptualizations of scaffolding and discovery-based pedagogical methodology for mathematics education. Focusing on co-enactment as a critical feature of scaffolding activities, we introduce "reverse scaffolding", wherein experts enact for novices only what they *know* to do rather than what they *do not know* to do. We demonstrate our approach by discussing a novel technological learning activity, Giant Steps for Algebra, wherein students construct models of realistic narratives. We argue for the method's potential via reporting on findings from mixed-methods analyses of a quasi-experimental implementation with 40 students.

Keywords Constructivism · Design-based research · Early algebra · Technology

Telling a kid a secret he can find out himself is not only bad teaching, it is a crime. Have you ever observed how keen 6 year olds are to discover and reinvent things and how you can disappoint them if you betray some secret too early? Twelve year olds are different; they got used to imposed solutions, they ask for solutions without trying. (Freudenthal 1971, p. 424).

1 Objective

The didactical metaphor of scaffolding has become so ubiquitous in the rhetoric of education researchers and practitioners, that its meaning has become diffuse, its theoretical rationale unquestioned, and its pedagogical operationalization vague (Pea 2004). We submit that scaffolding is a victim of its own popularity: its adoption by multifarious and even competing theories of learning has rendered untenable any consistent definition of what exactly scaffolding means. And yet precisely due to its consequent murkiness, the idea of scaffolding might serve as a prism onto a range of educational theories: If all educational scholars would each define what they mean when they say "scaffolding", we may get a rainbow of clearly juxtaposed theories of learning.

Still, we suspect, not all theorists have ready definitions for scaffolding. We certainly did not. Yet by way of implementing our own ill-defined notion of scaffolding in the form of well-defined learning activities, we could begin to ascend from intuitive to articulated notions of "scaffolding". This exercise led us to realize that our own conceptualization of scaffolding differs from the standard conceptualization in ways that might be pedagogically significant. We came to call our own conceptualization "reverse scaffolding".

The idea of reverse scaffolding was conceived as a response to what we view as an enduring dilemma in mathematics education. On the one hand, it is harmful to

✉ Dor Abrahamson
dor@berkeley.edu

1 University of California, 4533 Tolman Hall, Berkeley, CA 94720-1670, USA

show a child how to solve a problem (Kamii and Dominick 1998). Imposed methods are liable to remain opaque to a child who has not had opportunities to explore the problem space, recognize the limitations of familiar methods, and determine relevant embedded properties, patterns, intermediary states, and goal functions that would satisfy solution criteria. On the other hand, once students have reinvented critical principles of cultural-historical techniques, educators may intervene by introducing artifacts that implement those principles more efficiently, thus relieving the student to pursue more advanced problems. By way of enabling the student to arrive at the principles themselves, the cultural tools that implement these principles could become transparent to the child rather than opaque. These artifacts would thus implement for the students' what they already *know* to do, not what they *do not know* to do. Thus whereas in direct scaffolding cultural mediation fades *out*, in reverse scaffolding it fades *in* (see Fig. 1).

This paper presents and discusses findings from an empirical evaluation study designed to investigate whether reverse scaffolding might effectively serve as a pedagogical design framework for discovery-based learning. Our empirical context of inquiry into this research problem is the comprehensive process of developing and evaluating a technological environment for early, presymbolic algebra, Giant Steps for Algebra (GS4A). Our design solution was to create conditions for students to reinvent, acknowledge, and articulate principles of early algebra in the context of constructing diagrammatic models for story-based

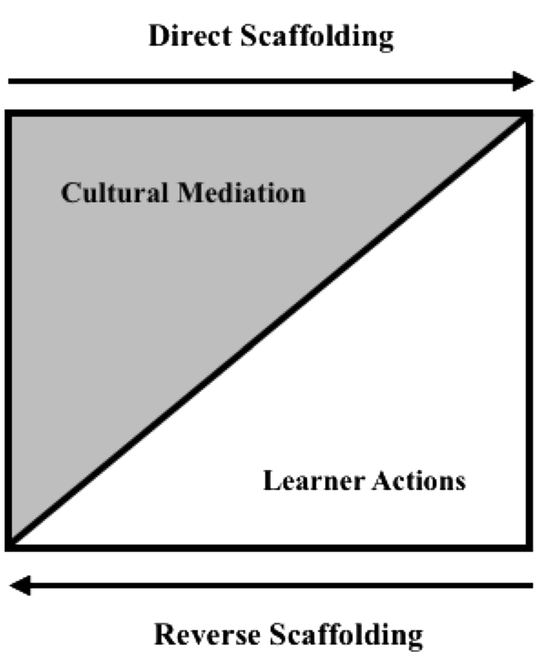


Fig. 1 Fade out vs. fade in: in direct scaffolding, the mediating cultural agent or artifact enacts for learners what they *do not* know to do. In reverse scaffolding, the agent performs what they *do* know to do

word problems that involve some unknown variable. Only later would the environment "take over" by automating the implementation of these discovered principles. For example, the child who is constructing a model of an algebra story might make evident that she is toiling to keep constant the size of the variable quantity across all its appearances in the model. In response, the system elicits from the child an articulation of this principle and thereafter keeps this variable constant, relieving the child of this "busy work". As we later elaborate, this particular construction principle, namely that a variable quantity should be of consistent size across all its appearances in a diagrammatic model, is both foundational and implicit to any algebraic solution procedure. We have identified three such principles and marked them as necessary proto-conceptual knowledge en route to adopting normative algebraic practices. We name these emerging principles "situated intermediary learning objectives" (SILOs).

The research design of this evaluation study was to compare the learning gains of students who participated in two instructional conditions for GS4A, reverse scaffolding as an experimental condition and a baseline condition as a control. We envisioned the study as potentially contributing to an increasing body of empirical work that both suggests the effectiveness of pedagogical practices oriented on having students discover subject matter content and outlines heuristics for implementing these practices using technology (Holmes et al. 2014; Kapur 2014; Schneider et al. 2015). More broadly, we hope to contribute to an ongoing dialogue between scholars who support discovery-based learning and those who pooh-pooh its viability (Alfieri et al. 2011; Kirschner et al. 2006).

1.1 Introduction of theoretical constructs

The reverse-scaffolding pedagogical architecture emerged in the context of conducting an educational design-based research project that investigated mathematical cognition, teaching, and learning through developing and evaluating a new form of activity for early algebra (Abrahamson et al. 2014). Central to the rationale of reverse scaffolding is the construct of transparency, which we now elaborate.

The construct of *transparency* captures the psychological relation between an individual and an artifact s/he is using toward the accomplishment of some goal, be it a bicycle, an abacus, or a quadratic equation. More precisely, we say that an artifact is transparent to an individual when he or she has developed an understanding for how embedded features of an artifact function to promote the accomplishment of the goal (Meira 1998). Transparency is cited most often in the case of concrete mechanisms, such as appreciating how gears work. Yet transparency equally obtains in the case of formal procedures for handling non-substantive objects, such as protocols for generating and manipulating symbolic notation. Algorithms, like gears, enable the implementation of their normative purpose through a concatenation of interrelated functions. Understanding an algorithm thus consists of rendering transparent each of these functions as well as their systemic interrelations. The pedagogical architecture of reverse scaffolding was envisioned with the purpose of creating opportunities for students to develop transparency for the algebraic conceptual system.

In our experimental activity, students would develop transparency for algebra by working in a computer microworld, where they use non-symbolic virtual elements to model word problems. The world problems comprise a protagonist (a giant) who travels along a straight path from a designated point of departure to a destination point where she buries treasure. Her travel consists both of "giant steps" (the unknown variable) and meters (the known unit). The giant travels twice from the point of departure to the treasure site, but these two journeys consist of different combinations of "steps" and meters, thus setting up what amounts to a diagrammatic representation of an algebraic set of two equations with one variable. Students are to solve this system through diagrammatic reasoning.

As they engage in the modeling task, the students are to figure out a set of construction heuristics, such as recognizing that all giant steps are equal in size. As we will explain, this system of construction heuristics bears an interesting epistemological status. The heuristics are informal, context-bound, pragmatic know-how bearing potential for generalization as formal mathematical rules. We came to articulate these principles through applying methods of interaction analysis (Jordan and Henderson 1995) and grounded theory (Strauss and Corbin 1990) to data gathered in a pilot study (Abrahamson et al. 2014; Abrahamson and Chase 2015). We named these construction heuristics "situated intermediary learning objectives" (SILOs). SILOs, within GS4A, are the rules that participants must follow and integrate, first implicitly and later explicitly, in order to solve the situated problems. SILOs make algebra transparent—SILOs are what we believe a person knows when we say that the person knows early, presymbolic algebra. Thus, for students to develop subjective transparency of the GS4A model, they must implement and acknowledge the SILOs that make this model work. Though SILOs are not concepts, they are proto-conceptual—the constellation of SILOs embodies the meaning indexed by mathematical concepts.

GS4A was designed for students to develop transparency for diagrammatic algebra solution processes. And yet the GS4A activity flow does not follow regular scaffolding methodology but rather what we are calling reverse scaffolding. We propose reverse scaffolding (RS) as a pedagogical technique for guided reinvention of mathematical concepts in technological environments, where instructional-interaction decisions are implemented in software procedures. RS design architecture presents students with a situation bearing a problem and encourages the students to solve the problem by building a model of the situation. In the course of constructing these models, students are steered to reflect on structural properties of the model. Once students figure out how to generate and manage a structural property of a model, the software "takes over" by automatically enacting and maintaining this property for the students, so as to simplify the students' further inquiry and problem solving of more demanding items. Crucially, RS design performs for students only what they already know to do. RS interface actions are thus designed to promote student agency in constructing transparent conceptual systems. The purpose of the study reported in this paper was to evaluate RS empirically.

We hypothesized that students who participated in RS activities, would have to determine the SILOs themselves and would thus develop transparency for the algebra conceptual system. The empirical study reported herein evaluated whether they would better develop transparency as compared to students for whom the SILOs were *pre*-implemented in software as automatic, undisclosed interface supports. After a theoretical overview of scaffolding, below, we will describe a study that evaluated this hypothesis.

Section 2, below, will outline the history of scaffolding as a general approach to educational practice and then elaborate further on transparency. Section 3 will present our design for early algebra, GS4A. Following Sect. 4, Methods, Sect. 5 offers Results and discussion, and we end with Sect. 6, Conclusions and Sect. 7, Implications.

2 Theoretical approach

"Scaffolding" was initially adopted by educational researchers as a metaphor that aims to clarify the nature of certain actions that experts perform during didactical interactions with novices: "scaffolding" would describe and categorize the various forms of supportive strategies that instructors employ during problem-solving sessions. And yet it was understood early on that the practice of scaffolding had not been invented by professional educators. Instead, the didactical practice of scaffolding drew on ubiquitous ethnographic observations of adult–child interactions in naturalistic settings, where the adult was acting compassionately and unreflectively to help the child achieve a goal and thus eventually learn to do so independently. What was viewed as ecologically authentic and culturally sanctioned out of school would be captured and implemented within school.

2.1 "Scaffolding": historical evolution of a pedagogical construct

Although Vygotsky never quite used the metaphor of a scaffold, it is often attributed to him, perhaps due to an association of scaffolding with his construct, the *zone of proximal development*. ZPD is oft quoted as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky 1978, p. 86). Whereas Bruner (1986) generally disagreed with Vygotsky's thesis on knowledge acquisition, he did agree "that there is at least one deep parallel in all forms of knowledge acquisition—precisely the existence of a Zone of Proximal Development and the procedures for aiding the learner in entering and processing across it" (p. 78). It was Wood et al. (1976) who first called this activity "scaffolding" in their report on a laboratory study of adults helping young children build pyramids out of chinked blocks.

Since then, numerous researchers have investigated scaffolding within educational contexts. Cazden (1981) examined teacher–pupil scaffolding interactions, focusing on the use of interrogation to guide learning. Smit et al. (2013) found additional scaffolding tactics, specific to those contexts, enumerating didactical efforts to hand over agency to the learner. Holton and Clarke (2006) discovered that self-scaffolding strategies increase and sustain student agency. In summary, scaffolding is the asymmetrical social co-enactment of natural or cultural practice, wherein a more able agent implements or performs for a novice elements of a challenging activity.

The notion of scaffolding is not exclusively inherent to the actions of a co-present teacher. Rather, aspects of the instructional rationale and interactions that scaffold learning may be cumulatively layered and distributed onto other classroom instructional infrastructure, such as technological resources. In particular, scaffolding practices are distributed over informative, functional, and interactive tools or media that complement, emulate, and possibly enhance the variety of customized supports that co-present human agents provide (Meira 1998).

As such, when educators invest and distribute their pedagogical efforts into a variety of interactive elements within a technological learning environment, we might still conceptualize these concrete or virtual features as bona fide scaffolds—scaffolds that are embodied, embedded, and latent to the artifacts until students engage, mobilize, and leverage them (Barab et al. 2007). In these environments, it may not be the direct "live" actions of an adult that scaffold the child's assigned task by co-enacting it but rather elements of the artifacts that mediate the co-enactment "remotely" (Quintana et al. 2004; Reiser 2004). Looking closer at the variety of technological artifacts bearing the potential to scaffold the learning process, we will now focus on symbolic elements, because these are important to processes of mathematization.

In their analyses of educational technology, Quintana et al. (2004) located scaffolding affordances in representations, specifically in features of the learning environment that highlight for learners how certain interactive features function. For example, selectable hints would appear on a computer monitor to suggest the meanings of symbolic notations and animated procedures. Here, "scaffolding" is taken to mean that the software both structures and problematizes the situation, stewarding the students toward discovering the instructional unit's target content (Reiser 2004). In discussing the automated behaviors of a computer-based algebraic tool, Pedemonte and Chiappini (2008) emphasize that students should be supported in understanding the underlying structural properties of these tools. Once they do, the tools may reduce the users' cognitive load. These systems *scaffold discovery*. Thus, constructivist parlance appears to have usurped the sociocultural term without adhering to its ideological underpinnings.

Whereas by no means have we exhausted the different contexts in which notions of "scaffolding" have been applied, we hope to have demonstrated how widespread, varied, and pervasive the term of scaffolding has become (for a broader review see Smit et al. 2013; Pea 2004). "Scaffolding" is now used almost synonymously with "supporting" or just "teaching". As Pea (2004) aptly noted, "the concept of scaffolding has become so broad in its meanings in the field of educational research and the learning sciences that it has become unclear in its significance" (p. 272). Pea goes on to discuss how this dilution in the meaning of scaffolding has created so much variance that establishing empirical boundaries has become difficult. He summarizes

> The goals of scaffolding research going forward should be to study how scaffolding processes—whether achieved in part by the use of software features, human assistance, or other material supports—are best conceived in ways that illuminate the nature of learning as it is spontaneously structured outside formal education and as it can most richly inform instructional design and educational practices. (p. 446)

Indeed, views of learning outside of the mathematics classroom suggest types of authentic instructional methodologies that flout the very rationale of scaffolding. For example, Reed and Bril (1996), cultural anthropologists of skill acquisition, have documented a pervasive parenting practice in which mothers create for their infants opportunities to develop new motor-action coordinations. Rather than model, explain, or directly help the infants achieve the target skill, the mothers instead create for the infants what the researchers call a “field of promoted action”, in which the infants discover for themselves how to negotiate the challenging situation into which they have been thrust. As such, the infants develop effective motor-action responses customized to their own musculoskeletal complex. From a distant yet complementary perspective, sports scientists informed by Nikolai Bernstein’s theories of kinesiology and biomechanics have put forth the hypothesis that athletes learn better when they must each invent for themselves their personal solutions to motor-action problems (Chow et al. 2007). It would make little sense to name these various cultural practices as “scaffolding”, because they are founded on the principle of *not*-helping rather than helping—in each of these practices the learner is set in a dedicated micro-ecology geared to promote the personal discovery of new affordances for specific actions.

Given the numerous operationalizations and pedagogical commitments of “scaffolding” as well as the conflicting precedents of non-scaffolding cultural practices, what form of scaffolding should we implement in educational software programs? Our concern is that wherever tutor or tool performs for the learner aspects of the learning activity, these scaffolding actions may not be transparent to the learner.

2.2 Transparency

The theoretical construct of *transparency* captures relations between, on the one hand, artifacts inherent to a cultural practice and, on the other hand, a social agent’s understanding of how features of these artifacts mediate the accomplishment of a particular practice (Meira 1998). In general, sophisticated artifacts bear information structures, logical relations, and activity constraints that enable competent users implicitly to offload their intentionality onto external entities in their work environment and thus promote their goals (Kirsh 2010). However, artifacts that are used for fostering content learning, it has been argued, should be transparent, because figuring out how they work is tantamount to understanding the embedded content (e.g., compare a calculator, which obscures the calculation to a novice user, to an abacus that may render the process transparent). And yet figuring out how an artifact works—that is, making sense of its embedded mechanisms and their functionalities—is a subjective process, and so Meira (1998) speaks about students developing subjective transparency of an artifact.

The theory of transparency is important for education, because we must make detailed choices regarding which specific logico-mathematical operations our pedagogical artifacts should “blackbox” vs. “glassbox” (Goldstein and Papert 1977). For example, careful consideration must be given in deciding how a digital learning tool automates aspects of students’ modeling activity, responds to their input, and provides feedback. These design decisions may bear critical consequences for the students’ chances of developing subjective transparency for those targeted conceptual structures that should emerge via interacting with these artifacts.

2.3 Summary and research questions

This study used a technological learning environment of our own design to create an empirical context in which we could revisit the pedagogical notion of scaffolding in light of the cognitive construct of transparency. The design is inspired by the reform-oriented didactical principle that instructors should foster student reinvention of mathematical practices (Gravemeijer 1999). We thus seek to gain insight into the following:

- Does the reverse scaffolding activity architecture bear pedagogical utility? In particular, does this architecture enable student development of subjective transparency for targeted mathematical notions? What is the qualitative character of learning in RS designs?

To address this research question, we sought to create as our empirical context a learning environment wherein participants would re-invent situated practices that we view as formative of our target domain knowledge, early algebra. We soon realized that we ourselves would need to design our empirical context, because we could not find any existing environment that agreed with our thesis as to the potential impact of reverse scaffolding. Indeed, whereas many learning artifacts scaffold the enactment of what students do *not yet know* to do, the artifact we created reverse-scaffolds in the sense that it relieves students of what they *know to do*. The design principle of reverse scaffolding for conceptual transparency is elaborated in the next section, where we describe our design for early algebra.

3 Giant Steps for Algebra: an experimental design solution for students to develop subjective transparency of early algebraic structures and solution procedures

The investigation reported in this paper was conducted as a design-based research study (Confrey 2005). The instructional activity at the core of this study revolved around a problematic realistic situation. The situation consisted of several narratives describing the location of buried treasure, which the student is then tasked to locate. We gave the participants a set of materials with which to model the situation and thus solve the problem of locating the treasure. It was conjectured that algebraic knowledge could emerge from this activity as pre-formal knowledge (van Reeuwijk 1995). In particular, we expected students to apply and articulate a set of simple heuristics by which to monitor the quality of their model. These would be hands-on construction know-how, that is, implicit technical criteria for evaluating whether the virtual model they are building preserves relevant information structures embedded in the problem narrative. Whereas—or perhaps *because*—these heuristics are at best intermediary bits of pragmatic knowledge solicited from naïve worldly aptitudes and applied in a highly specified context of a realistic problem, they could potentially constitute kernels of informal symbol sense from whence formal algebraic knowledge would later sprout (Arcavi 1994; Bartolini Bussi and Mariotti 2008; Noss and Hoyles 1996; Radford 2003). Our design emphasis on having children construct knowledge by constructing artifacts is also inspired by earlier work on the relation between mechanical bricolage and conceptual learning (Papert 1980).

In a pilot study comprising activities with concrete materials (Abrahamson et al. 2014) we engaged students in the guided solution of our problem set. We collected and analyzed data in an attempt to find patterns in students' model constructions and verbal utterances. In line with our hypothesis, students' modeling actions appeared to make manifest certain proto-algebraic logico-quantitative construction heuristics or principles. These *situated* and *intermediary learning objectives* (SILOs) that students apparently developed became our "blueprint" for articulating an emerging pedagogical perspective, because we realized that our didactical efforts—including design, facilitation, assessment, and evaluation—could all focus on nurturing these SILOs (Abrahamson and Chase 2015). In this section, we will explain the design problem, rationale, and solution.

3.1 Design problem: rethinking visualizations for algebraic equivalence

When teachers begin to introduce algebraic problem solving into mathematics classrooms, they very often rely on natural phenomena or cultural artifacts to evoke and exemplify new concepts (Jones 2010). The most common notion used in early algebra is the balance-scale (Vlassis 2002). The balance-scale evokes the embodied felt sense and inferential logical structure of two commensurate weights resting on a beam on either side of a fulcrum, at equal distances. These phenomenological resources are offered to students as the normative means of visualizing the symbolic algebraic proposition, that is, as the relational equivalence of two expressions across the equal sign (Herscovics and Linchevski 1994; Jones et al. 2012; Molina and Ambrose 2008). By adding to, or removing from both sides of the scale the same weight tokens, students are further to visualize what will later become the standard procedure to solve for x, that is, manipulating algebraic propositions to determine the values of unknown variables.

The general appeal of this model notwithstanding, students continue to struggle to demonstrate proficiency in algebra courses (Oakes et al. 2004), so that algebra has been referred to in the USA as the "gate-keeper" course for college admittance and completion (Moses and Cobb 2001). We therefore wondered whether the balance-scale is a useful manipulative for developing subjective transparency for algebraic problem solving, and our design problem became to search for alternative experiential entry points into algebra content.

3.2 Design solution: a storyline model of algebraic equivalence

Dickinson and Eade (2004) proposed a two-sided number line as a preferred entry model into algebraic reasoning (see Fig. 2). This alternative visualization of equivalence between two algebraic expressions appears to bear certain useful affordances for making sense of the algebraic rationale and solution procedures. Specifically, the visualization bears inherent logico-figural constraints, such as mirroring equal yet unknown variable quantities above and below the common line, that facilitate an offloading of source information in forms conducive to reducing complexity. This model, we assumed, could create opportunities for students to develop subjective transparency for algebraic propositions, because the number-line model renders highly salient the logical relations between variable and integers, both within- and between expressions (we invite the reader to try the problem in Fig. 3 so as to appreciate the model's appeal). A student's activity of generating and solving number-line models would necessitate and thus implicate an evolving skill of performing core algebra operations, such as maintaining the variable's consistent quantity across the diagrammatic display or expressing the value of

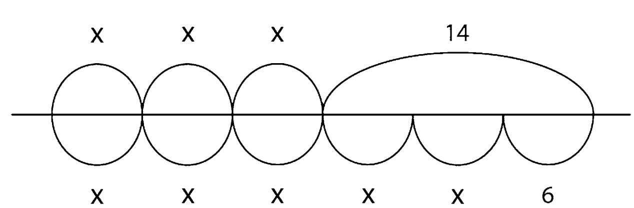


Fig. 2 Number-line instantiation of "$3x + 14 = 5x + 6$" (Dickinson and Eade 2004)

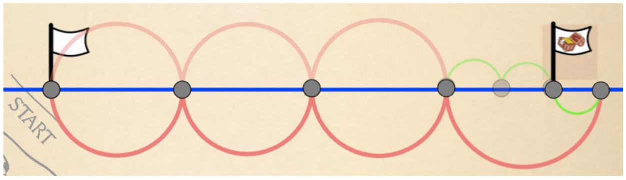


Fig. 3 A sample narrative in the GS4A instructional activity (*on the left*) and its virtual model (*on the right*). On both Day 1 (*above-the-line*) and Day 2 (*below the same line*), the giant travels from the flag (*on the left*) toward a treasure site (*on the right*). *Large red* loops represent giant steps, *small green* loops represent meters (color figure online)

this variable in terms of known unit quantities. We therefore chose to use this number-line algebra visualization for our experimental instructional unit.

The activity includes a story problem as well as resources for modeling the story toward solving the problem (Walkington et al. 2013). Per the embodied-design framework (Abrahamson 2009, 2014), the design solution seeks to engage and leverage students' tacit knowledge about simple ambulatory motion (walking) as well as their naïve sense for spatial relations. Figure 3 shows a sample narrative (on the left) and a model of this narrative in the computer microworld (on the right). We call this microworld "Giant Steps for Algebra" (GS4A).

> "A giant has stolen the elves' treasure. Help the elves find their treasure! Here is what we know. On the first day, a giant walked 3 steps and then another 2 meters, where she buried treasure. On the next day, she began at the same point and wanted to bury more treasure in exactly the same place, but she was not sure where that place was. She walked 4 steps and then, feeling she'd gone too far, she walked back one meter. Yes! She found the treasure!"

In designing GS4A, our goal was to provide a learning activity that would introduce students to the 'nuts and bolts' of pre-formal algebra. GS4A would leverage students' existing mathematical know-how in creating opportunities to reinvent procedural strategies that render algebraic structures transparent. At present, GS4A does not extend into activities that employ symbolic notations.

3.3 Rationale for technological implementation and learning assessment

Earlier, we introduced the idea of a SILO (situated intermediary learning objective).

A SILO is a form of pragmatic know-how, an implicit construction skill that people develop as they engage in goal-oriented practices using resources in their environment. In the case of designed learning environments, students can be steered to achieve particular SILOs that will later form the cognitive basis of curricular domain content—SILOs then serve as referents of formal rules, grounding schemas of mathematical concepts.

In GS4A, students tinker with virtual resources to build diagrammatic models of problematic situations that are communicated to them textually. These pictorial and often imprecise models that students build on the screen may appear naïve. However the models are proto-algebraic, in the sense that they reify a coordinated system of construction skills that serve to express relations among known and unknown quantities ultimately to determine the value of the unknown quantities. As such, the models bear algebraic structure and may thus foster algebraic practice.

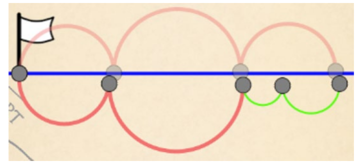
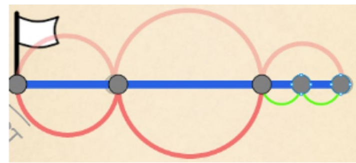
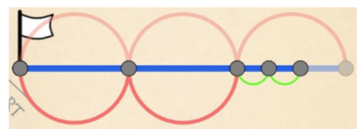
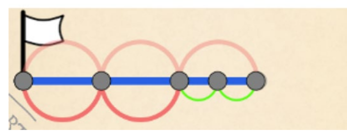
And yet SILOs are therefore only proto-conceptual, in the sense that they have not yet been articulated and expressed in formal semiotic register and have not as yet been generalized as an all-encompassing system of reasoning about known relations between unknown quantities. Hence SILOs are situated, intermediary learning objectives.

Pilot studies (Abrahamson et al. 2014) had suggested the following three SILOs for GS4A:

1. *Consistent measures*. All variable units (giant steps) and all fixed units (meters) are respectively uniform in size both within and between expressions (days);
2. *Equivalent expressions*. The two expressions (Day 1 and Day 2) are of identical magnitude—they share the "start" and the "end" points, so that they subtend precisely the same linear extent (even if the total distances traveled differ between days, e.g., when a giant oversteps and then goes back);
3. *Shared frame of reference*. The variable quantity (giant steps) can be described in terms of the unit quantity (meters).

Moving from the earlier pilot design cycle into the larger cycle reported herein, we hypothesized that it could be worthwhile to use these three SILOs to inform the design of the students' learning progression through an activity. The design principle was to allow learners to progress through an activity sequence while enabling them to achieve each SILO along the way. Borrowing the notion of "levels" from popular computer games—the gradual rewarding of manifest competence with increased power that is linked to increased skill—in GS4A we *level transparency*. With each new level the computer takes over the maintenance of a specific SILO, relieving the user of enacting this specific SILO. It is thus that our pedagogical architecture is *reverse* scaffolding. The learning environment provides a form of scaffolding, in the sense that it motivates, parses, and levels students' learning progression toward enacting a goal cultural practice. But it is reverse, because the technological

Table 1 Leveling transparency: matched SILOs and levels in the Giant Steps for Algebra technological design under the reverse-scaffolding study condition

Level	System constraints, user activity, behavior criterion, questions	Interface	SILO achieved
1. Free form	System offers no support in coordinating units or expressions		1. Consistent measures
Activity	User builds all parts of the model manually		
Criterion	User expresses frustration in equalizing units		
Questions	1–3		
2. Fixed meters	System generates meter units in predetermined size and maintains uniform size automatically		2. Equivalent expressions
Activity	User builds variables manually		
Criterion	User models uniform variable units within and between Days 1 and 2		
Questions	4–6		
3. Stretchy	System monitors for manual adjustment to the size of *any* of the variable-unit instances and accordingly adjusts the size of *all* variable units		3. Shared frame of reference
Activity	User adjusts size of variable unit to equalize the two propositions		
Criterion	User reads off the value of a variable unit in terms of the number of known units (meters) it subtends, e.g., one giant step is 2 m long		
Questions	7–9		

supports fade in rather than fade out. Students are not given training wheels—rather, they have to reinvent the wheels.

Table 1 presents the GS4A set of SILOs in relation to the interaction features of their corresponding activity level. Participants playing GS4A transition from each interaction level to the next upon demonstrating, via electronic actions, mastery over one of the SILOs. At each new level, the technology offloads the enactment of a SILO from the user. That is, the microworld automatically generates and maintains for the student the specific technical details and functional relations corresponding to that specific SILO. For example, when a participant determines that meters should be of uniform length in the model, we evaluate that he has achieved SILO 1, Consistent Measures. The participant then enters Level 2, in which the computer generates and maintains uniform meters. Leveling transparency is thus an activity-sequencing mechanism for operationalizing our pedagogical framework of reverse scaffolding.

Our evaluation study was a quasi-experimental research design, in which we compared learning gains of student groups participating in essentially the same activity (same materials, same sequence of word problems) only that the technology was set up to offer either reverse scaffolding or baseline conditions. In the baseline condition, our control, the interface automates construction operations from the outset—the very operations that reverse-scaffolding students must discover on their own are proffered in the baseline ab initio as intact interaction features. For example, meters appear in uniform size as do giant steps; when students stretch or shrink one of the giant steps, all steps auto-adjust their length accordingly across the entire model, thus resuming uniform size. Therefore, students participating in the baseline condition need never struggle to achieve the SILOs. They would use these features of an automatically functioning algebraic system without knowing what they are, and consequently the features would remain opaque to the students. According to the theory of transparency, these students would never understand the conceptual system embodied in the interface.

We hypothesized that students participating in the reverse-scaffolding condition would demonstrate greater learning gains than those in the baseline condition. The quasi-experimental study design would include assessment items for gathering data relevant to evaluating this hypothesis.

4 Methods

4.1 Participants

Twenty Grade 4 students (9.5–11.3 years old; 9 male, 11 female) and twenty Grade 9 students (14.4–16.1 years old; 9 male, 11 female) volunteered to participate individually in our task-based semi-structured interviews (Clement 2000; Ginsburg 1997)—21 in the experimental group (reverse scaffolding) and 19 in the control group (baseline). In discreet consultation with their teachers, students were each assigned either a “low”, “middle”, or “high” mark to characterize their prior mathematics achievement, and this served to control for mathematical proficiency across conditions.

4.2 Research design and assessment items

The rationale of the experimental design was to measure and compare learning under two conditions: reverse scaffolding (experimental group) and baseline (control group). Participants in the experimental group worked on problems according to the “leveling transparency” architecture, in which they had to “earn” automatic interface functionalities (see Table 1). In contrast, participants in the control group engaged the same problems with all automatic functionalities from the outset.

Upon completing the activity, all participants responded to five post-activity assessment items. These items were designed to evaluate participants’ subjective transparency of pre-formal algebra concepts, as operationalized in the three SILOs. The assessment items fall into two categories: (a) New-Context problems, in which we measured for the application of learned skills (transfer; see Fig. 4); and (b) In-Context problems that targeted the three SILOs directly within the familiar GS4A setting (see below).

New-Context assessment item “Turtle Years”, which built on the cultural notion of “dog years”, required of participants to compare *time* units (see Fig. 4a). New-Context problem “Two Buildings” used spatial units as in the GS4A context yet differed in that the modeled quantities here were oriented vertically rather than horizontally (see Fig. 4b). These items were designed to measures participants’ application of the SILOs in new problem contexts.

In-Context assessment items consisted of three screenshots from the familiar GS4A context. The screenshots showed work-in-progress of some hypothetical student. Each of the three images explicitly violated one or more of the SILOs. Participants were told that the hypothetical student had failed to complete the problem, and they were asked to determine why. These test items were designed to measure participants’ achievement of each SILO. We then calculated a Total post-activity assessment score by summing the New-Context and In-Context results.

4.3 Data collection procedures

The experiments took place on the school’s campus. All participants agreed to be videotaped for the duration of the intervention and post-activity assessment items, lasting approximately 1 h. All participants answered the same series of questions. In both activities, first the intervention and then the post-activity assessment, the researcher–interviewer followed a pre-designed semi-structured interview protocol to interact with the participant (Ginsburg 1997). Specifically, the interviewer often asked the participants to explain their models so that their reasoning could be made explicit.

4.4 Data analysis

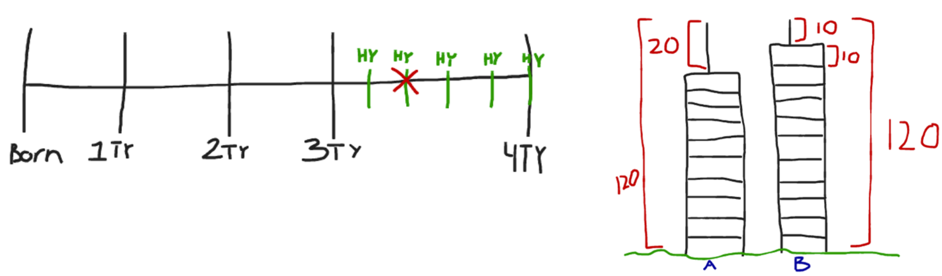
Data analysis consisted of first evaluating for a main effect of the intervention by comparing post-intervention achievements of the experimental and control groups (see Sect. 4.4.1, below). Once we had determined a main effect and

Turtle Years

My turtle, named Yurtle, is being tricky and won’t tell me how old she is. Help me figure out how old she is in human years. Yesterday she told me that she has lived 3 turtle years and 2 human years. She also told me that her 4th turtle birthday will be in 3 human years.

Two Buildings

Two buildings are built next to each other and are exactly the same height. One building is 10 floors and has a spire that is 20 feet on top of it. The other building is 11 floors and has a spire that is 10 feet on top of it. How tall are the buildings?



a b

Fig. 4 Two New-Context items were used in the post-intervention assessment of all participants: **a** Turtle years; and **b** Two buildings. Participants received these texts. The images were created by the first author for the purpose of this *figure* to illustrate correct responses

Table 2 New-Context post-activity assessment items coding scheme

SILO/solution	Turtle Years	Two buildings	Possible points
1. Consistent measures	All variable units [turtle years (TYs)] and all fixed units [human years (HYs)] are respectively uniform in size both within and between expressions of age	All variable units (floors) and all fixed units (feet) are respectively uniform in size both within and between expressions of height	1
2. Equivalent expressions	The two expressions (Age 1 is 3TYs + 2HYs; and Age 2 is 4TYs − 3HYs) are of identical magnitude—they share the "start" and the "end" points, so that they subtend precisely the same temporal extent	The two expressions (building A and building B) are of identical magnitude—they share the "start" and the "end" points, so that they subtend precisely the same vertical extent	1
3. Shared frame of reference	The variable quantity (TYs) can be described in terms of the unit quantity (HYs)	The variable quantity (floors) can be described in terms of the unit quantity (feet)	1
Correct solution	17 Human Years	120 Feet	1

wished better to understand *how* the experimental condition led to higher achievement, we further performed qualitative analyses (see Sect. 4.4.2, below).

4.4.1 Post-intervention quantitative analysis

We scored all participants' responses to the post-activity assessment items. In order to create the coding scheme, the first author applied micro-genetic analysis techniques (Siegler 2006) to review the entire corpus of video footage of participants working on the post-activity assessment items. The analyst identified moments during the problem-solving process in which the participants demonstrated, either through their utterances, their drawings, or their gestures one or more of the SILOs. A coding scheme was compiled for measuring participants' achievement of each SILO (see Table 2).

The New-Context assessment items—Turtle Years and Two buildings—were each scored on a scale of 0–4. Therefore, individual students' total scores in the two New-Context category ranged from 0 to 8. The In-Context assessment items—the SILO violation problems—were each scored on a scale of 0–3. Therefore, individual students' total scores in the three In-Context category ranged from 0 to 9. The composite Total score, therefore, ranged from 0 to 17.

A second analyst independently scored 21 % of this data corpus. Results from an inter-rater reliability test were Kappa = 0.822 ($p < 0.001$), 95 % CI (0.646, 0.998), almost perfect agreement. All remaining disagreements between the raters were resolved.

Next, we conducted an ANOVA to determine whether there were a difference between study- and control-group scores in each category (New-Context, In-Context, Total).

4.4.2 Intervention qualitative analysis

Having completed an evaluation of the study's main effect, we turned to the corpus of video data from the intervention itself (a total of approximately 50 h). We watched all video footage with the goal of identifying and characterizing moments during the intervention activity in which participants achieved one of the SILOs. Our rationale was that observing consistent behavioral patterns across the study conditions would help us understand *how* differences in the conditions led to the different learning gains. In particular, we hoped to evaluate our research hypothesis that RS bears pedagogical value.

The qualitative analysis reported herein focuses on two participants, one from each study group, who were considered by their teacher as high performing. This choice was informed by the observation that these students, while representative of their groups, were particularly articulate and therefore enabled us better to capture nuanced differences between the groups with respect to student development of subjective transparency.

Table 3 ANCOVA results

	df	*F*	Adjusted R^2	*P*
New-context	1	3.99	0.19	0.05
In-context	1	3.92	0.15	0.05
Total	1	6.57	0.27	0.01

5 Results and discussion

5.1 Main effect and discussion

The ANOVA revealed significant differences between mean group scores for the three post-activity assessment items (New-Context items, In-Context, and Total items) across reported math ability levels. An ANCOVA that controlled for math ability levels was then conducted to compare these results across the experimental and control conditions. This test found a significant difference between the RS and baseline conditions (see Table 3). The results show that for the New-Context and In-Context categories, respectively 19 % (Adjusted $R^2 = 0.19$) and 15 % (Adjusted $R^2 = 0.15$) of the difference is due to condition, and this is significant at the level $p = 0.05$. For the Total, 27 % (Adjusted $R^2 = 0.27$) is due to condition, and this is significant at the level $p = 0.01$.

Whereas this study should be replicated in similar and new domains, these findings at the very least suggest the plausibility of the reverse-scaffolding design architecture as an instructional methodology. Further studies could continue to investigate the hypothetical construct of SILO so as better to understand its emergent mediating effect on student performance in mathematical tasks as well as implications of this effect for the design of instruction and assessment.

5.2 Data excerpts: qualitative analysis of the emergence of SILOs in guided problem-solving interactions with educational technology

Qualitative analyses of the videotapes enabled the research team to develop deeper understandings of the experimental activity. In this subsection we present excerpts from two contrasting sample interviews, one with Lucy (experimental group: reverse scaffolding), and one with Mary (control group: baseline). (Both names are pseudonyms). These samples are arguably comparable in that the participants were of similar age (9 [9], 9 [11]) and both were considered by their teacher as being on the high end of proficiency. Furthermore, these students scored similarly on the post-activity assessment items, with Lucy scoring a Total of 15 points, and Mary a Total of 14. We will present the emergence of SILO 1, Consistent Measures, for each participant, with a focus on the specific case of the integer unit "meter".

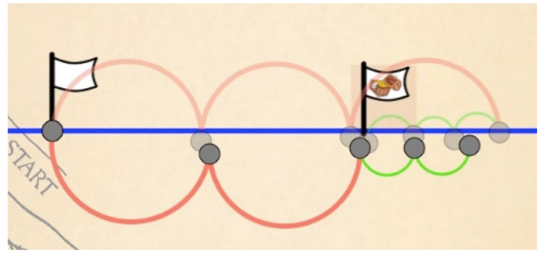


Fig. 5 Screenshot from Lucy's work-in-progress on the "$3x - 3 = 2x + 2$" narrative. Day 1, *above the line*, models "$3x - 3$" as running from *left* to *right* with three *large red arcs* ($3x$) and then retracing toward the left with three *small green arcs* ("-3"). Day 2, *below the line*, models "$2x + 2$" as two *large red* arcs ($2\times$) followed by two *small green arcs* ("+2") (color figure online)

Lucy (reverse-scaffolding study condition) is working within the instructional phase on Question 2. This item is composed of a Day-1-and-Day-2 narrative corresponding to the formal proposition "$3x - 3 = 2x + 2$". Lucy is at Level 1 of the activity regime, and so the software is not yet providing her with any automation for the SILOs. With respect specifically to SILO 1, Consistent Measures, at Level 1 Lucy must proactively build and maintain consistent measures herself as she interacts with the interface. In just over 4 min, Lucy has completed modeling both the Day 1 and Day 2 travel narratives (see Fig. 5).

Figure 5 shows a screenshot from Lucy's work in progress. Note that the steps and meters are respectively of equal size both within- and between-days. At this point Lucy realized a problem. Whereas the narrative describes the giant as arriving at the same destination on both days, the diagram currently depicts Day 1 and Day 2 journeys as arriving at different destinations. In particular, note how the "$2x + 2$" model fragment (Day 2, below the line) extends farther to the right as compared to the "$3x - 3$" model fragment (Day 1, above the line). Lucy believed she could amend this violation by adjusting the sizes either of the giant steps or the meters.

In the transcription below, parenthetical texts, such as "(Day 1)" serve to clarify for the reader the interlocutors' communicative intent as suggested by their non-verbal multi-modal utterance, including screen actions as well as various deictic gestures (e.g., pointing toward elements on the shared visual display).

Lucy began by adjusting the meters on Day 2.

Res.: Well I noticed that you made meters smaller on this day (Day 2).

Lucy: Yeah. So I should make them smaller on this day (Day 1) too probably.

Res.: Yeah? Why don't you try that. I think that could probably make sense. Why do you think that you should probably do that?

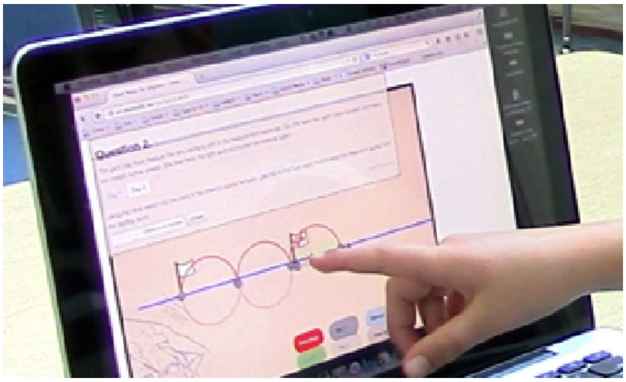


Fig. 6 Lucy indicates on the computer monitor where she anticipates Day 1 and Day 2 travel destinations will be co-located once she adjusts the meters

Lucy: Because then they will all be the same size, and then you will be, umm… I'll be able to the see whether it's still on that point (Lucy points to where she expects the treasure flag will be after the adjustments, a point on the screen slightly to the right of its current location, see Fig. 6).

We interpret Lucy's multimodal utterances as indicating an achievement of SILO 1—she apparently knows that meters should be of consistent size both within and across days. Lucy has thus constructed this function for a purpose: By repairing the model in accord with the SILO she will have access to information pertinent to solving the problem at hand. To Lucy, the structural feature of consistent measures is an essential property of her model bearing purposeful function.

We now turn to Mary, a participant in the baseline condition. Mary is working on the same item as Lucy, "$3x - 3 = 2x + 2$". However, Mary's interface a priori produces meters of fixed size that are therefore automatically consistent both within and across Day 1 and Day 2, and the interface also recalibrates all giant steps when any of them is resized. Mary has been working on this problem for almost 10 min. She has built both Day 1 and Day 2 model fragments yet is unable to determine a solution (see Fig. 7). Mary has just erased her diagram and is starting over. We join the researcher–student dyad as Mary begins modeling the first 3 giant steps (large red arcs) and is adjusting their size: she stretches, shrinks, and finally leaves them slightly extended as compared to her erased diagram.

Mary: Maybe the giant got bigger (as compared to the Question 1).
Res.: Yeah, maybe the giant got bigger, or smaller, I don't know. It's a different giant, that's all I know.

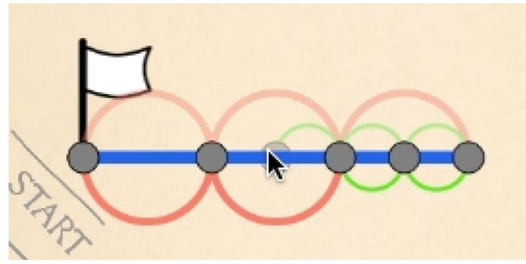


Fig. 7 Working on "$3x - 3 = 2x + 2$", Mary has modeled the Day 1 and Day 2 narratives as diagram fragments. The two journeys do not end at the same location, and so Mary has not solved the problem

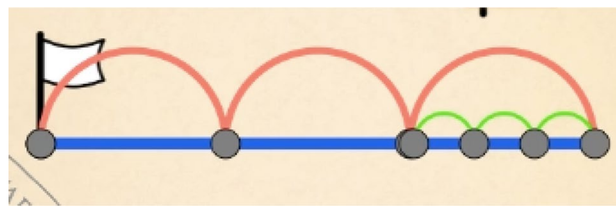


Fig. 8 Mary is still working on modeling the "$3x - 3 = 2x + 2$" story. She has now stretched the giant steps arbitrarily, so that they have each become 3 microworld meters in length, whereas the correct solution is 5 m

Mary: So Day 1 and Day 2 are different giants?
Res.: No no no, it's a different giant from Question 1.
Mary: (Slightly stretches the giant steps again, then creates three small green arcs running back toward the left for the "-3", see Fig. 8) Ohhh, so…

Res.: What happened?
Mary: So the meters are always the same.
Res.: Does that make sense?
Mary: Yeah, so the meters are always the same, you can move the giant steps but not the meters.

Recall that Mary is working under a condition where all measures are always uniform in size—all giant steps are always equal to each other, and all meters are always equal to each other—only that whereas the giant steps are adjustable (they all adjust simultaneously) the meters are not adjustable (they are of fixed size). Mary's insight into the consistent measure of the fixed meters unit in this interface is apparently facilitated by way of juxtaposition with the variable size of the giant-step. Her inflection indicates surprise at this discovery, and she reiterates her discovery thrice, as if memorizing it as a new rule. To Mary, the feature of consistent measures is not an essential property of the model bearing a purposeful function—it is a feature per se.

In each of the episodes the participant has arrived at the conclusion that meters, within this microworld, should remain consistent in size both within and across days. What differs between the two episodes is how this knowledge surfaces, and how it is used. Lucy constructs a model in which her meters are fairly consistent. When she is unable to determine a solution, she can imagine how adjusting the meters uniformly will solve the problem without violating her established measuring practice. Through exercising agency in ensuring that corresponding measure units are equal, Lucy has developed subjective transparency for a goal structural property of effective constructions in this microworld: She discovered that corresponding units *should be* of uniform size.

Mary, too, is reasoning about the consistency of meter units, and yet she does so only as a feature of the interface—not as how things *should be* but how things just *are*. Mary, like Lucy, appreciates that the meter units behave differently from the variable units, and yet she conceptualizes the relation between meters and variables not as an instrumental function advancing the solution process but as some arbitrary interface feature she has detected. Thus Mary did not experience an opportunity to appreciate the potential utility of this embedded feature; she did not develop subjective transparency for this received feature. Per Freudenthal, Mary is a victim of a didactical crime: by giving Mary all the tools she would need to solve a problem, she never understood the rationale of these tools. In fact we robbed Mary of an opportunity for discovery by imposing upon her an opaque mathematical artifact.

6 Conclusions

Students can reinvent mathematical knowledge through engaging in modeling-based activities. To do so, they should attend to the structure of their own constructions; they should apprehend this structure as reifying their tacit knowledge—knowledge that is brought forth as material form through the dialectics of shaping construction resources to tell a story; they should come to see latent structural features of their own spontaneous models of problem situations as properties of purposeful functions. Tasks, construction resources, facilitation, and activity flow can and should be designed to enable this apprehension of structure. In particular, students can develop subjective transparency of mathematical concepts by creating, articulating, and generalizing the structural properties of the models they use to solve situated problems. Our findings support previous arguments for learning through discovery (Martin and Schwartz 2005; Noss and Hoyles 1996; Radford 2003) and raise questions for skeptics (Alfieri et al. 2011; Kirschner et al. 2006). Moreover, we have offered and validated a pedagogical design architecture for discovery-based mathematics learning in technological environments (see also Holmes et al. 2014; Schneider et al. 2015).

We set out to investigate whether students can learn mathematical content under conditions where we assign them a task and provide them with relevant resources but do not tell them how these resources should all come together to get the task done. We called this minimal-interventional approach "reverse scaffolding". The phrase "reverse scaffolding" implies a form of instruction in which the expert does not perform for the novice what the novice cannot yet do but only what the novice *can* already do.

To evaluate our proposed reverse scaffolding, we conducted an empirical study that compared the implementation of an instructional activity under both reverse-scaffolding and control protocols. Statistical analyses of the two groups' mean scores on post-intervention assessments demonstrated a positive main effect: participants in the reverse-scaffolding condition significantly outperformed those in the control condition.

Qualitative analyses of the intervention process implicated psychological mechanisms apparently causative of the main effect: reverse-scaffolding students struggled more than baseline students to manage structural properties of the modeling system, so that these properties became evident to them—they developed more transparent structural understanding of the modeling procedure as well as greater facility in articulating the emerging conceptual system. In particular, we submit, the construction principles that students discovered by tinkering with elements of the virtual models were tantamount to the core mathematical content of the instructional design.

Our results must be considered only as a preliminary proof of concept within the context of a larger ongoing design-based research process. Additionally, whereas our experimental technology embeds some of the human tutor's facilitation actions in the form of "intelligent" interactive software, our current build cannot as yet eliminate the human tutor—we still rely heavily on a human tutor to monitor that the participants indeed build subjective transparency as they progress through the activity levels. In the interest of eventually scaling up the project, that is, creating empirically evaluated and internationally accessible online learning tools for early algebra, it would be necessary to continue refining the software's interactions as well as to extend students' work into the symbolic semiotic register.

7 Implications

We have put forth a framework for building educational technology that facilitates students' development of subjective transparency—the framework calls to embed within the interactive technology an activity sequence designed to foster incremental development of transparency. Leveling transparency, as an activity-design principle for creating technology-based learning activities, could stand to significantly inform the design of pedagogical tools for discovery-based learning.

References

Abrahamson, D. (2009). Embodied design: constructing means for constructing meaning. *Educational Studies in Mathematics, 70*(1), 27–47.

Abrahamson, D. (2014). Building educational activities for understanding: an elaboration on the embodied-design framework and its epistemic grounds. *International Journal of Child-Computer Interaction, 2*(1), 1–16.

Abrahamson, D., & Chase, K. (2015). Interfacing practices: domain theory emerges via collaborative reflection. *Reflective Practice, 16*(3), 372–389.

Abrahamson, D., Chase, K., Kumar, V., & Jain, R. (2014). Leveling transparency via situated, intermediary learning objectives. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, K. O'Connor, T. Lee & L. D'Amico (Eds.), Proceedings of "Learning and Becoming in Practice", the 11th International Conference of the Learning Sciences (ICLS) 2014 (Vol. 1, pp. 23–30). Boulder, CO: International Society of the Learning Sciences.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18.

Arcavi, A. (1994). Symbol sense: informal sense-making in formal mathematics. *For the Learning of Mathematics, 14*(3), 24–35.

Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E.-J., & Herring, S. C. (2007). Situationally embodied curriculum: relating formalisms and contexts. *Science Education, 91*, 750–782.

Bartolini Bussi, M. G., & Mariotti, M. A. (2008). Semiotic mediation in the mathematics classroom: artefacts and signs after a Vygotskian perspective. In L. D. English, M. G. Bartolini Bussi, G. A. Jones, R. Lesh & D. Tirosh (Eds.), *Handbook of international research in mathematics education, 2nd revised edition* (pp. 720–749). Mahwah, NJ: Lawrence Erlbaum Associates.

Bruner, J. S. (1986). *Actual minds: Possible worlds*. Cambridge, MA: Harvard University Press.

Cazden, C. B. (1981). Performance before competence: assistance to child discourse in the Zone of Proximal Development. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 3*(1), 5–8.

Chow, J. Y., Davids, K., Button, C., Shuttleworth, R., Renshaw, I., & Araújo, D. (2007). The role of nonlinear pedagogy in physical education. *Review of Educational Research, 77*(3), 251–278.

Clement, J. (2000). Analysis of clinical interviews: foundations and model viability. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547–589). Mahwah: Lawrence Erlbaum Associates.

Confrey, J. (2005). The evolution of design studies as methodology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 135–151). Cambridge: Cambridge University Press.

Dickinson, P., & Eade, F. (2004). Using the number line to investigate the solving of linear equations. *For the Learning of Mathematics, 24*(2), 41–47.

Freudenthal, H. (1971). Geometry between the devil and the deep sea. *Educational Studies in Mathematics, 3*(3/4), 413–435.

Ginsburg, H. P. (1997). *Entering the child's mind*. New York: Cambridge University Press.

Goldstein, I., & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. *Cognitive Science, 1*(1), 84–123.

Gravemeijer, K. P. E. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning, 1*(2), 155–177.

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics, 27*(1), 59–78.

Holmes, N. G., Day, J., Park, A. K., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science, 42*(4), 523–538.

Holton, D., & Clarke, D. (2006). Scaffolding and metacognition. *International Journal of Mathematical Education in Science and Technology, 37*(2), 127–143.

Jones, I. (2010). *Presenting equality statements as diagrams.* Paper presented at the Proceedings of the 6th Congress of the European Society for Research in Mathematics Education.

Jones, I., Inglis, M., Gilmore, C., & Dowens, M. (2012). Substitution and sameness: two components of a relational conception of the equals sign. *Journal of Experimental Child Psychology, 113*(1), 166–176.

Jordan, B., & Henderson, A. (1995). Interaction analysis: foundations and practice. *Journal of the Learning Sciences, 4*(1), 39–103.

Kamii, C. K., & Dominick, A. (1998). The harmful effects of algorithms in grades 1-4. In L. J. Morrow & M. J. Kenney (Eds.), *The teaching and learning of algorithms in school mathematics, 1998 yearbook* (pp. 130–140). Reston: NCTM.

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Kirsh, D. (2010). Thinking with external representations. *AI & SOCIETY, 25*, 441–454.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*(4), 587–625.

Meira, L. (1998). Making sense of instructional devices: the emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education, 29*(2), 129–142.

Molina, M., & Ambrose, R. (2008). From an operational to a relational conception of the equal sign. Thirds graders' developing algebraic thinking. *Focus on Learning Problems in Mathematics, 30*(1), 61–80.

Moses, R. P., & Cobb, C. E. (2001). *Radical equations: civil rights from Mississippi to the Algebra Project*. Boston: Beacon Press.

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: learning cultures and computers*. Dordrecht: Kluwer.

Oakes, J., Joseph, R., & Muir, K. (2004). Access and achievement in mathematics and science: inequalities that endure and change. In J. Banks & M. Banks (Eds.), *Handbook of research on multicultural education*. San Francisco: Jossey-Bass.

Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. NY: Basic Books.

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

Pedemonte, B., & Chiappini, G. (2008). Algebra on Numerical SETs: a system for teaching and learning algebra. *International Journal of Continuing Engineering Education and Life-Long Learning, 18*(5/6), 627–639.

Schneider, B., Bumbacher, E., & Blikstein, P. (2015). Discovery versus direct instruction: Learning outcomes of two pedagogical models using tangible interfaces. In T. Koschmann, P. Häkkinen, & P. Tchounikine (Eds.), *"Exploring the material conditions of learning: opportunities and challenges for CSCL," the Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference* (Vol. 1, pp. 364–371). Gothenburg, Sweden: ISLS.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J. S., Fretz, E., Duncan, R. G., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337–386.

Radford, L. (2003). Gestures, speech, and the sprouting of signs: a semiotic-cultural approach to students' types of generalization. *Mathematical Thinking and Learning, 5*(1), 37–70.

Reed, E. S., & Bril, B. (1996). The primacy of action in development. In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development* (pp. 431–451). Mahwah: Lawrence Erlbaum Associates.

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

Schneider, B., Bumbacher, E., & Blikstein, P. (2015). Discovery versus direct instruction: Learning outcomes of two pedagogical models using tangible interfaces. In T. Koschmann, P. Häkkinen, & P. Tchounikine (Eds.), "Exploring the material conditions of learning: opportunities and challenges for CSCL", the Proceedings of the Computer Supported Collaborative Learning (CSCL) Conference (Vol. 1, pp. 365-371). Gothenburg, Sweden: ISLS.

Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (6 ed., Vol. 2, Cognition, perception, and language, pp. 464–510). Hoboken, NJ: Wiley.

Smit, J., van Eerde, H. A. A., & Bakker, A. (2013). A conceptualisation of whole-class scaffolding. *British Educational Research Journal, 39*(5), 817–834.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques*. Newbury Park: Sage Publications.

van Reeuwijk, M. (1995). *The role of realistic situations in developing tools for solving systems of equations.* Paper presented at the annual conference of the American Educational Research Association, San Francisco.

Vlassis, J. (2002). The balance model: hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3), 341–359.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press. **(Original work published 1930)**.

Walkington, C., Petrosino, A., & Sherman, M. (2013). Supporting algebraic reasoning through personalized story scenarios: how situational understanding mediates performance and strategies. *Mathematical Thinking and Learning, 15*(2), 89–120.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100.