Computers & Education 72 (2014) 231–248

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Integrating intuitive and novel grounded concepts in a dynamic geometry learning environment

Jonathan M. Vitale [a,*], Michael I. Swart [b], John B. Black [b]

[a] Graduate School of Education, University of California, Berkeley, 4407 Tolman Hall, Berkeley, CA 94720, USA
[b] Department of Human Development, Teachers College, 525 W. 120th Street, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 10 April 2013
Received in revised form
3 October 2013
Accepted 11 November 2013

Keywords:
Application in subject areas
Interactive learning environments
Elementary education
Pedagogical issues

ABSTRACT

The development of geometry knowledge requires integration of intuitive and novel concepts. While instruction may take many representational forms we argue that grounding novel information in perception and action systems in the context of challenging activities will promote deeper learning. To facilitate learning we introduce a *grounded integration pattern* of instruction, focusing on (1) eliciting intuitive concepts, (2) introducing novel grounding metaphors, and (3) embedding challenges to promote distinguishing between ideas. To investigate this pattern we compared elementary school children in two conditions who engaged in variations of a computer-based dynamic geometry learning environment that was intended to elicit intuitive concepts of shapes. In the *grounded integration* condition children performed a procedure of explicitly identifying defining features of shapes (e.g. right angles) with the assistance of animated depictions of spatially-meaningful gestures (e.g. hands forming right angles). In a *numerical integration* condition children identified defining features with the assistance of a numerical representation. Children in the *grounded integration* were more likely to accurately identify target shapes in a posttest identification task. We discuss the relevancy of the *grounded integration pattern* on the development of instructional tools.

1. Introduction

Geometry is a critical, yet often overlooked, branch of mathematics education. Although the National Council Of Teachers Of Mathematics (2000) stresses the value of geometry instruction, particularly poor U.S. performance on geometry items of standardized assessments, relative to other developed nations (Ginsburg, Cooke, Leinwand, Noell, & Pollock, 2005), likely reflects inadequate treatment of geometry in traditional curricula (Clements, 2004).

Overwhelmingly, traditional instructional activities rely on static images and simplistic tasks. Consequently, students are under-prepared for higher-level geometric reasoning (Clements, 2004; Clements & Battista, 1992). As a potential remedy, interactive geometry software, such as Logo, enables students to apply a constructive, inquiry-based approach to mathematics (Papert, 1980; Papert, Watt, DiSessa, & Weir, 1979). Research spanning from Piaget (Piaget & Inhelder, 1956; Piaget, Inhelder, & Szeminska, 1960) to more recent theory regarding "embodied" cognition (Lakoff & Núñez, 2000, pp. xvii, 493) suggests that young children's early knowledge of geometry is primarily visuo-spatial in nature. Therefore, the concrete representations of mathematical concepts embedded in interactive geometry software should afford a greater intuitive foothold than traditional materials.

Yet, in spite of this clear theoretical support, research on the use of concrete representations in education is inconclusive. For example, while Logo generated much early enthusiasm, a number of studies show mixed or negative findings on learning outcomes (Howe, O'Shea, & Plane, 1979; Hughes & Greenbough, 1995; Johnson, 1986; Noss & Hoyles, 1992; Simmons & Cope, 1990). In contrast, a number of recent studies highlight benefits for educational materials that are more abstract in design (Kaminski, Sloutsky, & Heckler, 2009; McNeil & Uttal, 2009).

Given the theoretical support for concrete representations, what accounts for their unevenness in practice? One possibility is that concrete representations are, in fact, too intuitive. Specifically, the additional affordances of concrete representations facilitate problem

* Corresponding author. Present address: 918 52nd Street, Oakland, CA 94608, USA.
E-mail addresses: jonvitale@berkeley.edu (J.M. Vitale), mis2125@tc.columbia.edu (M.I. Swart), black@exchange.tc.columbia.udu (J.B. Black).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2013.11.004

solving without necessarily attending to or reflecting upon deeper structure of the materials. Symbolic approaches avoid unintended consequences by introducing only those features that are relevant to learning goals, thereby facilitating transfer to structurally-similar problems (Kaminski, Sloutsky, & Heckler, 2008). In more extreme cases intuitive concepts may even introduce misconceptions that are then applied to novel materials (Ohlsson, 2009).

While in some cases abstract materials may afford a "symbolic advantage" (Koedinger, Alibali, & Nathan, 2008), in other cases the lack of familiar or intuitive representations can open a rift between the mathematics of school and everyday experience (Lave, 1988). In the approach taken here, rather than shifting instruction towards more symbolic forms to avoid intuitive concepts, we focus on introducing new visuo-spatial, or grounded, representations that can be compared with intuitive concepts, directly. In the following we describe a *grounded integration pattern* of instruction which aims to promote coordination between intuitive and novel, spatially-grounded representations of geometry. We detail how this approach is supported by recent cognitive-developmental research and how current digital technologies support its application.

1.1. The emerging cognitive science of geometry

Fundamental processes that support mathematical thinking emerge consistently early across individuals and cultures. In the numerical domain core knowledge emerges from universal abilities to judge magnitude and recognize numerosity of small collections (Feigenson, Dehaene, & Spelke, 2004), which is supported by distinct system of neural structures (Dehaene, 1997). Beyond the core knowledge that grounds intuitive concepts, environmental factors, such as SES, can promote or inhibit further development of formal mathematical ideas (Booth & Siegler, 2006). Promisingly, deficits can be addressed through motivating activities, such as board games (Ramani & Siegler, 2008; Siegler & Ramani, 2008, 2009). Clearly, instruction plays a fundamental role in supporting attainment of higher-level ideas.

Likewise, in the domain of geometry researchers are beginning to address development in terms of core knowledge. Spelke, Lee, and Izard (2010) claim that geometry concepts are grounded in two core systems associated with spatial navigation and object perception. From this perspective the concept of a shape, e.g. a *rectangle*, is derived from the experience of viewing rectangular objects, such as the face of a block, or navigating along a rectangular path. In particular, object perception fosters concepts of distance and angle, while navigation fosters concepts of distance and directional sense (i.e., left vs. right).

Consistent with Spelke et al.'s. (2010) framework, years of perceptual research suggests that individuals develop geometric ideas by exposure to common object and images. For example, Bomba and Siqueland (1983) found that 3-month-olds recognized a prototypical triangle after viewing a series of triangles, which varied randomly from the prototype, without actually seeing the prototype. In turn, internalized prototypical representations may then be applied to identify and classify novel figures.

Although prototypes facilitate development of geometric ideas at an early age, prototypical representations can hinder later development of formal concepts. In particular, individuals tend to classify images based on perceptually-salient cues, such as symmetry (Quinlan & Humphreys, 1993), dispersion (irregularity), elongation, and jaggedness (Behrman & Brown, 1968). While geometric prototypes inherently contain defining features, they also exhibit these salient secondary features (e.g. elongation of a prototypical rectangle) which may drive conceptualization.

In school-based tasks young children often apply informal labels, such as "slanty", "pointy", or "skinny" to describe common shapes (Clements, Swaminathan, Hannibal, & Sarama, 1999). While there is value in facilitating description of geometric figures with informal language, children are unlikely to recognize differences between secondary and defining features. In this case the most salient perceptual feature of a figure strongly influences its classifcation at the expense of any formal criteria. For example, in identication tasks many children mistake elongated (non-rectangular) parallelograms for rectangles, while missing squares (Burger & Shaughnessy, 1986; Clements et al., 1999).

Although irrelevant salient perceptual characteristics often influence children's thinking, there is evidence that more normative concepts are within reach. Specifically, Dehaene and Izard (2006) found that members of an Amazonian tribe, with little or no exposure to formal concepts in geometry, could distinguish between parallel and non-parallel, as well as perpendicular and non-perpendicular line segments – important concepts in Euclidean geometry. While students may have the perceptual ability to make these distinctions, instruction is necessary to promote their relevance to shape concepts.

1.2. Grounded instruction in geometry

The tendency for younger children to identify objects based on superficial characteristics is not unique to geometry. Children often classify common objects – such as taxicabs and islands – based on non-distinguishing, but salient characteristics (Keil & Baterman, 1984). On the other hand, adults or domain experts tend to classify objects based on abstract or hidden structure (Carey, 1985; Chi, Feltovich, & Glaser, 1981) and perceptually-abstract analytical systems (Mandler, 1992; Sloman, 1996). This is often described as a general concrete-to-abstract shift in development (Bruner, 1960; Piaget, 1952). This view is reflected in stage theory-based models of geometric development, such as the Van Hiele (1986) levels, which progress through visualization, analysis, abstraction, and (formal) rigor stages.

A potential pitfall of the instructional application of stage-based models is the imposition of artificial boundaries on student activities – e.g. unnecessarily delaying rigorous tasks for younger students or overlooking necessary visuo-spatial tasks for older students. In particular, research in the areas of perceptual learning (Gibson, 1969; Goldstone & Son, 2008, pp. 327–355) and grounded (embodied) cognition (Barsalou, 2008; Lakoff & Núñez, 2000, pp. xvii, 493) suggest that continued refinement of spatially-grounded representations of concepts is essential to guide higher-level knowledge (Black, Segal, Vitale, & Fadjo, 2012). Rather than shifting to abstraction, a grounded approach seeks to promote attention to increasingly precise visuo-spatial features of materials (Goldstone & Barsalou, 1998; Quinn & Eimas, 1997; Schyns, Goldstone, & Thibaut, 1998). As a geometric example, the concept of a square may be grounded in formally-relevant perceptual features – such as perpendicularity, bilateral and rotational symmetry, etc. – while learning to disregard other salient features – e.g. rotational orientation.

In their description of mathematical thinking, Lakoff and Núñez (2000, pp. xvii, 493) describe several mechanisms by which spatially-grounded knowledge develops. Specifically, "grounding metaphors" map spatial structures or actions to mathematical concepts. For example, the act of physically combining sets of objects maps to the mathematical concept of addition. In addition to the role that grounding metaphors play in the development of children's early knowledge, introducing novel metaphors provides an opportunity to extend their knowledge.

The dynamic and interactive affordances of computer-based environments provide a powerful means of introducing and reinforcing novel grounding metaphors. For example, in the software described below, we introduce the concept of parallelism through a model of hands moving in parallel. Additionally, "graphic organizers" can be layered on materials to highlight or annotate critical features (Mautone & Mayer, 2007). More implictly, task goals and constraints can be designed to guide learners towards relevant properties of the materials and away from superficial, but seductive features (Clements, Battista, & Sarama, 2001).

1.3. A grounded integration pattern of instruction in geometry

In the case of geometry, and mathematics more generally (Lakoff & Núñez, 2000, pp. xvii, 493), both intuitive and formal concepts are fundamentally visuo-spatial in nature. While this common ground provides a mechanism for developing new knowledge from old, students must also learn to distinguish between similar ideas and apply these ideas appropriately.

Students' intuitive concepts are often highly resistant to change. Traditional patterns of instruction that seek to inform without regard to student ideas tend to leave intuitive, but non-normative concepts intact (Linn & Eylon, 2011). On the other hand, methods that engage student ideas directly are more likely to lead to conceptual change. In particular the *knowledge integration pattern* – i.e., elicit ideas, add new ideas, develop criteria, and sort out ideas – has been shown to promote robust learning in the context of diverse web-based science curriculum units (Linn & Eylon, 2011).

In the context of science, where conceptual models are often quite complex, engaging students in this full pattern, many times over, may be necessary to achieve successful learning. In the context of geometry, where both intuitive and formal concepts are fundamentally grounded in simple spatial primitives, a more straightforward process may be appropriate. Here we introduce the *grounded integration pattern* as a refinement of the *knowledge integration pattern*, to promote a shift from informal to formal ideas in geometry. Specifically, this pattern incorporates three steps:

- *Elicit* intuitive concepts through familiar narratives, goals, and actions.
- *Introduce* novel grounding metaphors.
- *Embed* challenges to promote distinguishing between ideas.

The overall goal of the *grounded integration pattern* (*GIP)* is to draw upon both novel and familiar representations of concepts to engage students in demanding, critical thinking processes. In particular, the latter step explicitly aims to introduce "desirable difficulties" to promote conceptual integration (Bjork, 1994; Bjork & Linn, 2006). Activities that challenge students to distinguish between and evaluate ideas lead to better recall and more flexible application of material, potentially at the cost of some short-term efficiency (Clark & Linn, 2003; Linn, 2000; Linn, Chiu, Zhang, & McElhaney, 2010). While the inclusion of activities that deliberately increase the difficulty of the task may seem counter-intuitive, we believe it is a necessary step that is often overlooked in the design of educational tools.

Implementing the first two steps of *GIP* requires highly specific domain knowledge to determine what representations students are likely to find familiar and accessible – particularly for novel concepts. In the case of geometry ongoing cognitive development research provides relatively clear evidence for appropriate representations. The final step, "embedding challenges", on the other hand, is primarily an instructional design challenge that can be addressed in any number of ways. In the following we discuss one such approach – incorporation of parallel goals – as it was applied in a digital learning environment.

1.4. The digital geometry software – overview

We chose to implement the *grounded integration pattern* in the context of a computerized game-like environment addressing geometric concepts. Here we provide a general overview, in the following sections we detail the mapping between the *GIP* and software features. The goal of the activity is to construct a closed path to enable an animated agent (a "robot") to successfully navigate a given obstacle course (i.e., intersecting goal objects and avoiding obstacles), while also meeting formal criteria for the given geometric shape. To complete an exercise learners proceed through four phases: *planning, construction*, *feature validation,* and *testing* (Fig. 1).

In the *planning* phase learners view the layout of the obstacle course, read the name of the target shape, and read the name of features (e.g. parallel sides) necessary in their figure. Once learners determine a course of action, at their own discretion, they may proceed to the *construction* phase. Within this phase learners navigate a cursor in the form of the robot by dragging the mouse along an empty grid to construct segments and rotating the mouse about a pivot point to construct angles. Upon producing a closed figure, learners may adjust the figure via "drag-and-drop" of vertices.

When a learner is satisfied with the constructed figure, he or she may proceed to the *feature validation* phase where the presence of defining features is automatically assessed. Upon successful validation the learner proceeds to the *testing* phase where he or she runs the agent through the obstacle course, according to the designed path. Upon successful navigation, learners are rewarded by a virtual cheering audience. In cases of unsuccessful feature validation or testing learners return to the *construction* phase to revise their figure. The *testing* phase also grants learners an additional opportunity to view the arrangement of obstacles and plan a revised path.

1.4.1. Eliciting intuitive concepts in the geometry software

As detailed above children's geometric knowledge is derived from experience navigating paths and interacting with authentic objects (Spelke et al., 2010). By applying a navigational theme and representing the constructed path as a persistent and malleable object we facilitated both sources of intuitive knowledge. Moreover, in each exercise a central set of obstacles, about which the agent's path must circumscribe, is arranged to model an example of the target shape. In initial exercises central sets of obstacles are arranged to suggest a prototypical instance of the target shape (see Appendices A–C). Successful completion of these exercises may be accomplished by simply

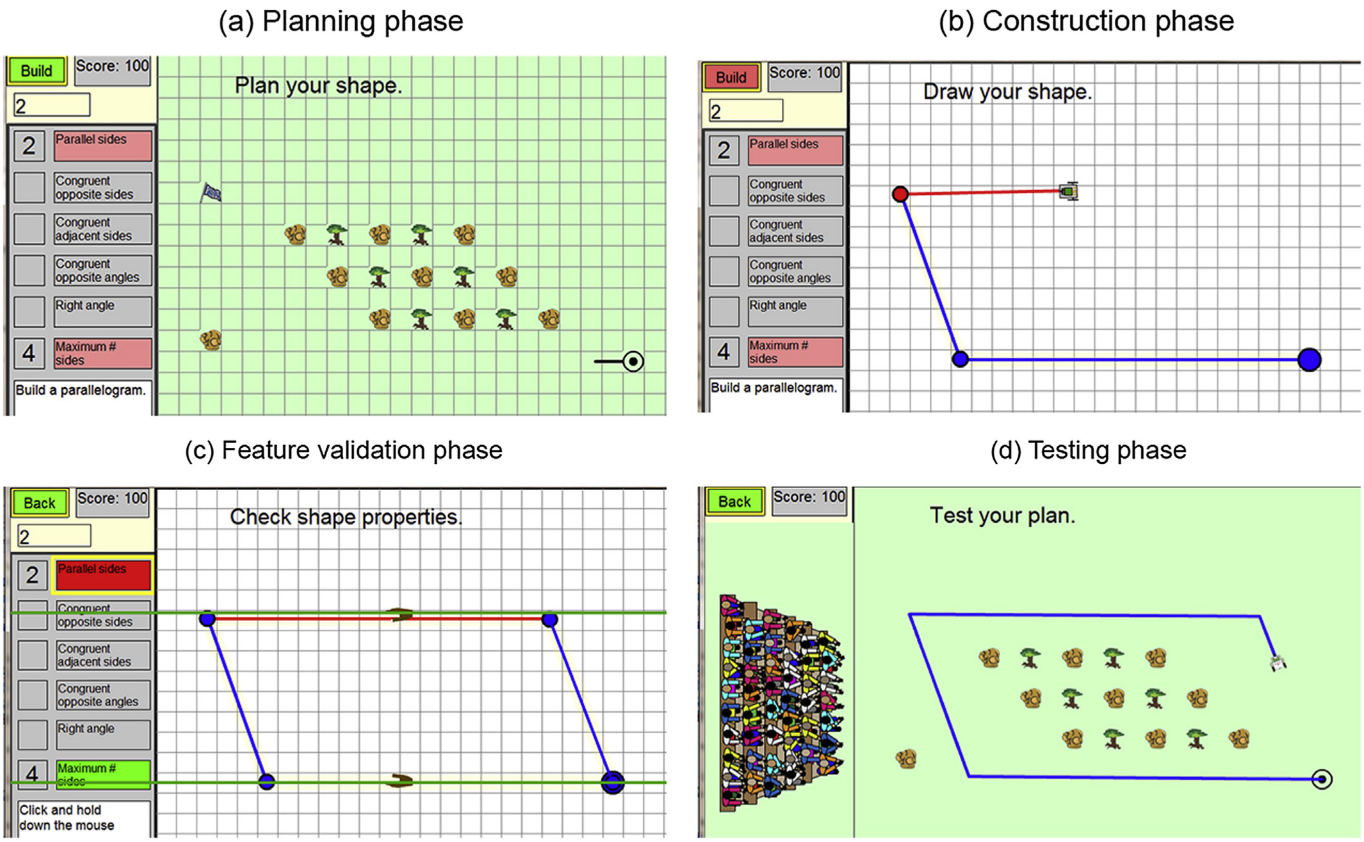


Fig. 1. Screenshots of four main phases of learning software. In this example the student is constructing a parallelogram. Image (b) displays the active construction of the top segment of the parallelogram. Image (c) displays the validation of the top and bottom segments as parallel.

reproducing a prototypical figure that meets size requirements implicit in the arrangement of obstacles. Planning is simply a matter of recognizing the specific type of quadrilateral or triangle, and gathering size information (e.g. counting grid squares).

1.4.2. Introducing novel grounding metaphors in the geometry software

To introduce novel grounding metaphors within the *feature validation* phase we provide animated models of hand gestures that depict a geometric concept. Given previous studies in which spontaneous gesture were observed to represent geometric concepts (Clements et al., 2001; Clements & Burns, 2000), there appears to be a natural relationship between our target concepts (i.e., parallel lines, congruent line segments, and right angles) and gesture. Directing learners to engage in specific gestures has been shown to improve learning outcomes (Broaders et al., 2007; Segal, 2011). In our case, to make use of available technology, gestures are depicted, rather than enforced (with a motion capture system, for example), to suggest their usage by students.

As Fig. 2 reveals, upon entering the *feature validation* phase learners are presented with an animated representation of hands that are organized and behave according to a specific spatial concept. As students manipulate these representations visual discrepancies with the underlying figure reveal potential errors in student thinking.

1.4.3. Embedding challenges in the geometry software

In addition to providing a visual representation of spatial concepts, *feature validation* also constrains students to construct valid figures by ensuring the presence of necessary features (within a margin of error). Producing a figure that both meets formal validity criteria and conforms to spatial constraints of the obstacle course (e.g. is sufficiently elongated to circumscribe all obstacles) is the central challenge of the activity.

As described above the prototypical arrangements of central obstacles in initial exercises was designed to elicit intuitive ideas about shapes. In these cases necessary features of the shape implicitly emerge as a byproduct of producing prototypical figures. For example, Fig. 1 displays an arrangement of obstacles in the form of a prototypical parallelogram, which necessarily contains two pairs of parallel sides. In later exercises, the layout of obstacles diverge from the prototype by orienting obstacles obliquely (i.e., not "upright") and eliminating salient, but non-defining features associated with prototypical instances (e.g. lateral symmetry in trapezoids). Therefore in their progress to later exercises learners are expected to shift their attention to the defining features which must be reproduced in construction.

1.5. Previous application in research

While each of the activity phases was designed to promote learning, we believe that *feature validation* represents the most critical element of our design. Specifically, the *feature validation* phase introduces novel grounding metaphors for defining features and *enforces*

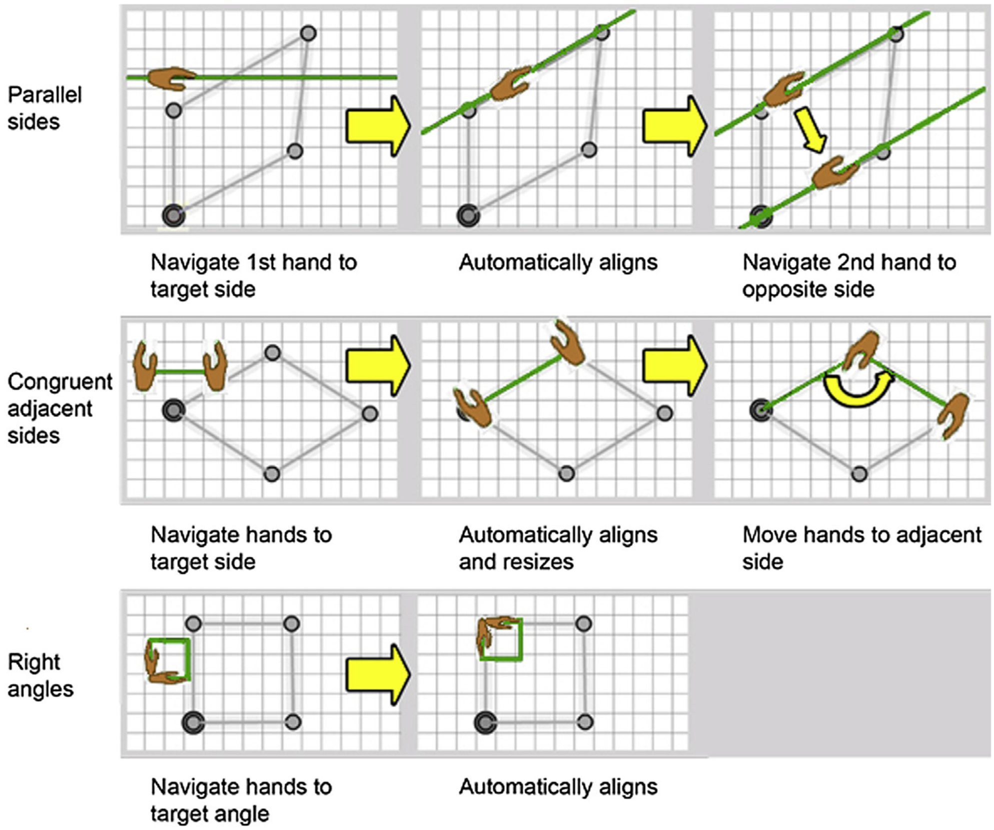


Fig. 2. Sequence of feature validation operations. Learner guides hands to the target shape component, the software adjusts the depiction (resizes, realigns), and (in the case of comparisons between multiple sides) the learner guides the newly adjusted visual depiction to the second component. Note: hand images have been enhanced for clarity.

their application in the construction of figures. Additionally, by providing a second set of criteria for success in the task, the *feature validation* phase significantly increases the difficulty of the task.

In a previously-reported pilot study we tested the effectiveness of the *feature validation* phase by comparing students with access to the full software to those who simply proceeded directly from *construction* to *testing* phases (reference omitted for blind review). Thus, students in the latter, control condition were neither provided with novel grounding metaphors, nor a constraint on the validity of their constructions. Table 1 displays the result of non-parametric tests performed to compare conditions on 6 subtests addressing identification of specific geometric figures. As predicted, students with the full software treatment were more likely to successfully identify valid instances of target shapes than children who did not perform *feature validation* in the learning task.

In addition to differences at posttest, analysis of logs generated by the learning software revealed less time-on-task for children who did not perform *feature validation*. While these students were able to construct valid figures, their construction processes often exhibited an unimpeded visual guess-and-check pattern that did not facilitate learning, echoing previous findings with Logo (Simmons & Cope, 1993). Although these results reinforce the need for desirable difficulties, they obscure the role that novel grounded metaphors played in learning.

1.6. This study

In the current study we tested whether the *grounded integration pattern* facilitates learning by comparing two groups of children who performed a series of geometry exercises in two variations of the software. In a treatment group children viewed gestural depictions of novel spatial concepts in the feature validation phase – i.e., the *grounded integration* (GI) condition, as described above. In a control group children

Table 1
Pilot test accuracy by treatment condition (reference omitted for review).

	Median % trials correct		Mann–Whitney		Effect size = $Z/\sqrt{N}$
	GI ($n = 10$)	Control ($n = 10$)	U	Z	
Trapezoid	60%	42%	26.0	-1.82^{*}	.41
Parallelogram	78%	48%	18.0	-2.42^{**}	.54
Rhombus	80%	58%	19.5	-2.31^{*}	.52
Isosceles tri/trap	73%	13%	16.5	-2.54^{**}	.57
Rectangle	75%	48%	11.0	-2.96^{**}	.66
Right tri/trap	23%	8%	27.5	-1.71	.38
Overall	**64%**	**37%**	**10.0**	**3.03^{**}**	**.68**

$^{**}p < .01$; $^{*}p < .05$.
Note: Statistical presentation differs from original for consistency with study results presented below.

viewed a numerical display of novel spatial concepts in the feature validation phase – i.e., the *numerical integration* (NI) condition. Specifically, children in the NI condition were constrained to produce valid figures, with the same precision as the GI condition, by ensuring that a numerical representation of the target feature fell within a specified range. This ensured that the participants in the NI condition were provided with a similarly rigorous, two-goal task as participants in the GI condition, while placing focus on the role of grounded representation.

We predict that the *grounded integration pattern*, as implemented in the GI software, will promote a shift from a prototypical understanding of geometric shapes towards a more formal, defining features-based conceptualization. To test this, following training, we engaged participants in a shape identification task that was explicitly constructed to elicit prototypical concepts. Children in the GI condition should be more prepared than children in the NI condition to disregard seductive, prototypical features in favor of more subtle, defining features.

2. Method

2.1. Participants

A class of sixteen third grade and three fourth grade students were recruited from an after-school program located in a low-income, predominantly Hispanic neighborhood of New York City. Study-related activities were conducted over the course of a 13 week period from March to June. Participating children were randomly assigned to either the *grounded integration* or *numerical integration* condition. The treatment condition consisted of nine children ($M = 9.3$ years, SD = .76, 67% female, 100% Hispanic). The NI condition consisted of ten children ($M = 9.3$ years, SD = .30, 50% female, 100% Hispanic). Two children in the GI condition were native Spanish speakers, but could communicate sufficiently in English and showed little difficulty understanding and completing the tasks. Additionally, one child, originally assigned to the GI condition (but not reflected in the statistics above), chose not to engage in assessment materials following the first unit and was not included in further elements of the study.

2.2. Materials and procedure

2.2.1. Software and curriculum

All study-related tasks were conducted in the context of a weekly after-school robotics course. However, in some instances in which students had no other assigned classes, we conducted multiple sessions in a week. Aside from the tools described here the children also engaged in physical construction of Lego-based robots. This supplementary activity did not affect the structure of this study, except to inform the visual design of the learning tool.

Children were randomly assigned to either condition at the start of the first class, thereby determining the version of the software they would use throughout the experiment. To avoid contamination between conditions each learning session was divided into two halves, in which the GI and NI groups alternated order from session to session. Children not currently engaging in study activities completed homework in a separate room. The primary learning software was presented as a game, which they would complete individually, at their own pace until they had completed all "levels".

During learning sessions, the experimenters (two of the authors) circulated among the children to provide assistance and motivate engagement. In both conditions experimenters attempted to guide students through procedural difficulties interacting with the software. When students encountered conceptual difficulties relating to target learning goals the experimenters reinforced guidance and feedback provided by the software without offering specific strategies. For example, if a student had difficulties constructing a parallelogram, he or she might be asked to explain, in spatial or numerical terms, what it means more sides to be parallel. In the GI condition this included asking students to demonstrate parallelism with nearby materials or gestures, and applying corrective feedback. Additionally, while the exercises were completed individually, the experimenters encouraged students to discuss strategies with nearby classmates.

Over the course of the study children completed a series of exercises across three units, focusing on three defining features of shapes: parallel sides, congruent adjacent sides, and right angles, respectively. In the first unit children completed a set of 10 construction exercises, focusing on trapezoids and parallelograms. In the second unit children completed 6 exercises, focusing on kites and rhombi. Finally, in the third unit children completed 6 exercises, focusing on rectangles and squares. Each exercise was designed, via the placement of obstacles and goals, to promote construction of a path resembling the target shape. Exercises within each unit generally progressed from the prototypical to atypical (see Appendices A–C for layouts of all exercises). All actions performed by participants were logged and stored in text files for analysis.

In addition to the software, we provided participants with a printed sheet of "thumbnail" images (50% scale) displaying the obstacle layout for each exercise to assist construction. The reduction in scale prevented a direct tracing strategy in constructing figures, while affording attention to general spatial characteristics of the arranged obstacles. Additionally, to motivate participation, children received a star sticker after completion of each exercise. Stickers were placed on a personal "stamp certificate" that could be kept after completion of the study.

In both conditions the validity of figures were assessed through the same numerical procedure. Parallel sides were validated by automatically computing the absolute difference between selected sides' percent gradient (0%-horizontal, 100%-vertical, 50%-diagonal rising left-to-right, −50%-diagonal rising right-to-left). Sides within ±3% gradient difference were considered parallel. Congruent sides were validated by computing the absolute difference between segment lengths (in pixels). Sides within ±23 pixel difference (approximately 7 mm of screen size) were considered congruent. Finally, right angles were validated by computing the difference from 90° of the internal angle at the selected vertex. Angles within 4° of 90° were considered valid right angles.

In the *numerical integration* version of the software we introduced a panel of numerical values situated below the construction space (Fig. 3). This panel displayed a series of blocks with statistical measure(s) associated with each component of the constructed figure, including sides (length, percent gradient) and vertices (degree measure of internal angle). As students moused-over blocks associated components were automatically highlighted, and vice-versa. Spatial changes to components were reflected in real-time changes to numerical representations.

During *feature validation* participants were prompted to click on boxes that corresponded to components displaying the target defining feature. For example, to validate parallel sides in the parallelogram of Fig. 3 a participant could select either the two blocks with “0% grade” (i.e., horizontal) or the two blocks with “−64% grade” and “−65% grade.”

2.2.2. Shape identification pretest

To ensure that the two conditions initiated the experiment with roughly equivalent background knowledge we provided a pencil-and-paper pretest of geometric shape knowledge. Based on an assessment developed by Burger and Shaughnessy (1986), students were provided with a reference sheet displaying ten common triangles and quadrilaterals, with an identifying letter printed next to each figure (A–J). On a separate sheet of paper students were asked to identify which of the figures were squares (1 in total), rectangles (2), parallelograms (3), rhombi (2), kites (3, including rhombi), trapezoids (2, not including parallelograms), right trapezoids (1), isosceles trapezoids (1), quadrilaterals (8), right triangles (1), isosceles triangles (1), equilateral triangles (0), scalene triangles (0), and triangles (2). Additionally, participants were asked to identify which figures exhibited parallel sides, congruent sides, congruent angles, and right angles.

2.2.3. Computerized shape identification posttest

The main assessment measure was individually administered, immediately following a student’s completion of a unit. Each unit posttest included two subtests with 30 items, for a total of six subtests and 180 items. The goal of each item was to identify the two positive examples of a target shape from four simultaneously presented figures (Fig. 4). During each item we recorded which figures were selected, as well as other interactions with the interface (e.g. “mouse-overs”, “mouse-clicks”).

To encourage more interaction, figures were initially displayed in light gray without a border, but were darkened and highlighted with a black border when “moused-over”– providing a clearer image of the figure. Upon clicking on a figure, its internal area was filled blue as a visual indicator of selection. Figures could be selected and de-selected freely in the process of decision making. Once two figures were selected, the participant could then click on a central button to advance to the next Item.

The rationale for this item structure, with four – as opposed to a simpler choice between two figures – was to make common features of the four figures explicit and comparable. Given the evidence that children, without an appropriate intervention, are more likely to classify geometric figures based on holistic similarity than defining features, this item structure allows us to identify what feature the participant most likely attended to or disregarded.

Six subtests assessed participants’ ability to identify trapezoids, parallelograms (following unit 1), rhombi, isosceles triangles/trapezoids (following unit 2), rectangles and right triangles/trapezoids (following unit 3). The triangle/trapezoid blocks were included to test transfer of defining feature concepts to novel stimuli. For each subtest a diverse set of valid instances of the target shape were constructed. From these valid instances, invalid instances were derived by altering a single feature (e.g. the slope of a parallel side). Pairs of valid–invalid pairs were combined to produce stimuli thirty sets of four figures for each subtest (see Appendices D–I for stimuli images and details).

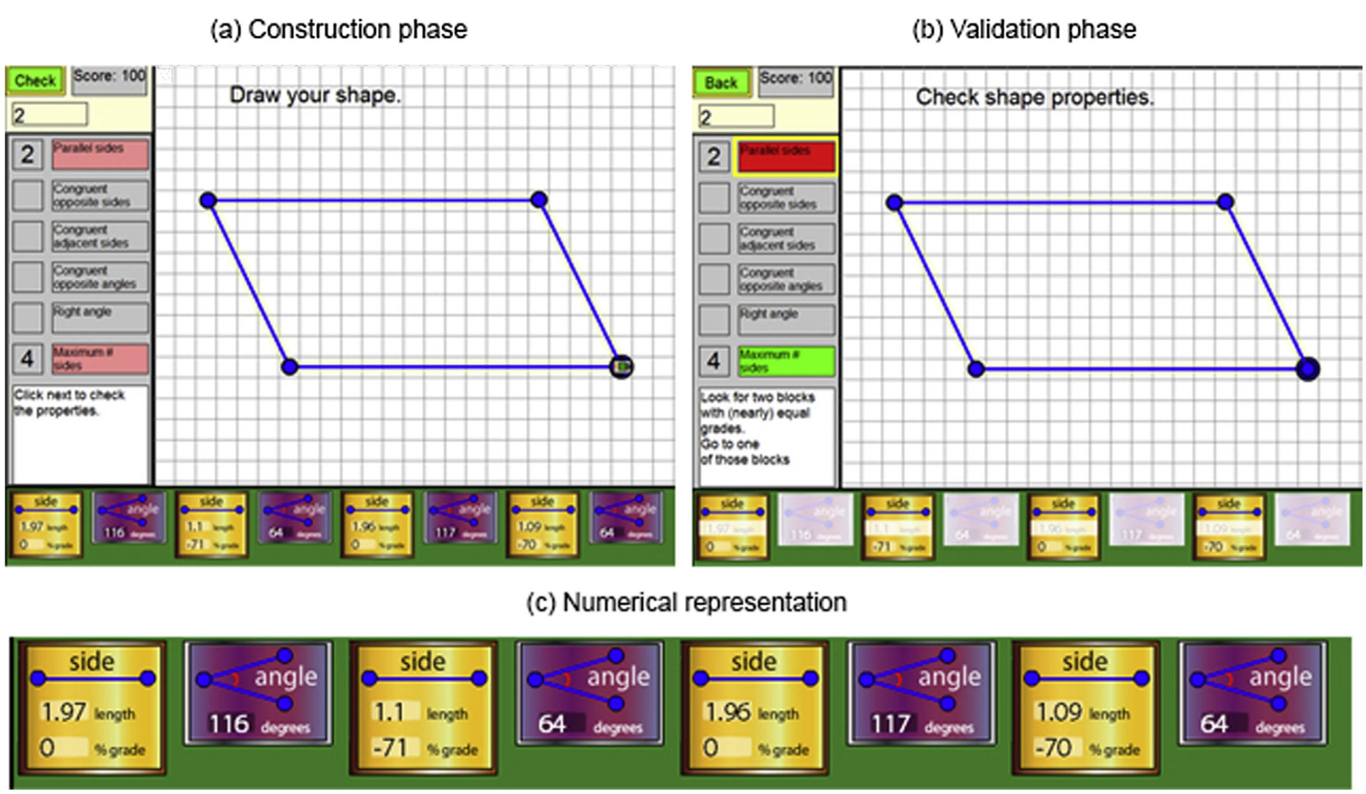


Fig. 3. Screenshot of *numerical integration* version of software. (a) Screenshot during construction phase. (b) Screenshot during parallel sides validation. Instructions state, “Look for two blocks with (nearly) equal grades. Go to one of those blocks.” (c) Zoomed-in image of numerical representation. Only “% grade” is (fully) visible during a parallel check. Likewise, only length and degree are (fully) visible during side congruency and right angle checks, respectively. Note: 1.0 length is equivalent to 8 grid boxes (23 pixels per box).

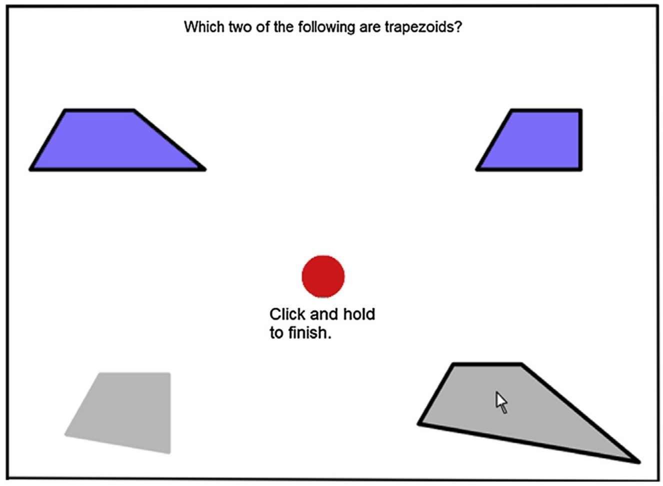


Fig. 4. Screenshot of shape identification task. Participants choose two valid examples of a shape (trapezoid) from four available figures. Here, two figures are currently selected (blue). One figure is currently being highlighted (gray, outlined, mouse cursor over). Finally, the fourth figure is neither highlighted nor selected (light gray, not outlined). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To introduce the task an instruction screen provided a written description of the formal rules of inclusion for the target shape class. For example, "A trapezoid is a four-sided figure with one pair of parallel sides", or in the case of mixed triangles and trapezoids, "A shape is right if it contains at least one right angle". In the case of subtests utilizing figures similar to those given in the learning exercises (i.e., trapezoids, parallelograms, rhombi, and rectangles), this instruction page displayed both a valid and invalid example of the shape. The participants were asked to click the sides or angles corresponding to the target defining feature (parallel sides, congruent sides, right angles) within the valid example. In the case of mixed triangle/trapezoid blocks, which were intended to be novel visual stimuli for the given target feature, no visual example was provided.

3. Results

3.1. Shape identification pretest

Because the primary target of instruction was quadrilaterals we focused on the accuracy of identifying quadrilaterals in the pretest. To compare conditions on pretest knowledge we tallied the number of true positive identifications (hits), false positives, true negatives, and false negatives (misses) each participant registered for each of nine target quadrilaterals. Because the students' textbook provided an exclusive definition of trapezoids (i.e., one and only one pair of parallel sides) we scored identification of trapezoids accordingly.

Table 2 displays the results of these tests accounting for students' accurate and inaccurate identification and non-identification of presented figures. A two-tailed t-test conducted on participants' F_1-scores (which equally penalizes both false negatives and false positives) did not reveal any difference between groups [$t(17) = .7$, $p > .1$]. Therefore, in terms of geometry knowledge, there were no obvious differences between groups prior to training.

3.2. Computerized shape identification posttest

On each posttest identification item two out of four figures were selected as representatives of a shape's class, resulting in 0, 1, or 2 figures selected correctly. While in most testing situations one-correct would be regarded as better than zero-correct, here one-correct suggests that the participant judged the figures according to the incorrect, distracter feature (e.g. choosing both "upright" figures). Therefore we coded items with both figures selected correctly as accurate and items with one or zero figures selected correctly as inaccurate. We then calculated the percentage of accurate (two-correct) items within each subtest. On any given item chance performance was 1 out of 6, or 16.7%. Fig. 5 displays the distribution of participants' accuracy scores across all six subtests.

Fig. 5 demonstrates considerable skew in the distribution of accuracy results. Additionally, in several cases (e.g. parallelograms), there was a clear distinction between those students who were successful on most items, and those who were rarely successful. In particular, within the NI condition there were a large number of children who were incorrect on nearly every trial, suggesting that they exclusively attended to the distractor feature. In preparation to compare conditions, we performed Shapiro–Wilk test of normality to each distribution, which revealed several deviations from normality (i.e., parallelogram in GI condition; parallelogram, isosceles tri/trap, rectangle, and rhombus in NI condition). Therefore to compare conditions we performed a series of non-parametric, one-tailed Mann–Whitney test. Significant results indicate that the distribution of GI participants' accuracy scores are shifted higher than the distribution of the NI participants' accuracy scores.

As Table 3 displays, across five of six subtests, tests revealed a significant positive shift in the distribution of GI participants' accuracy scores, compared to NI participants' accuracy. Only in the case of right triangles/trapezoids was there no significant difference, likely reflecting the low overall scores for participants in both conditions (median accuracy scores near- or below-chance). Therefore, our primary hypothesis, that GI accuracy scores would exceed those of NI, was generally confirmed.

Table 2
Shape identification pretest mean statistics by treatment condition (standard deviations in parentheses).

Condition	True positive	False positive	True negative	False negative	F_1-score
GI	6.2 (4.3)	5.0 (1.4)	62 (1.4)	16.8 (4.3)	.32 (.04)
NI	6.8 (4.9)	4.8 (3.7)	62 (3.7)	16.2 (4.9)	.39 (.03)

Table 3
Posttest accuracy by treatment condition.

	Median % trials correct		Mann–Whitney		Effect size = $Z/\sqrt{N}$
	GI ($n = 9$)	NI ($n = 10$)	U	Z	
Trapezoid	37%	20%	21.5	-1.92^{*}	.44
Parallelogram	27%	10%	20.5	-2.01^{*}	.46
Rhombus	56%	13%	16.0	-2.37^{*}	.54
Isosceles tri/trap	39%	7%	18.5	-2.37^{**}	.50
Rectangle	63%	5%	17.0	-2.31^{**}	.53
Right tri/trap	17%	7%	32.0	-1.07	.25
Overall	**37%**	**11%**	**14.0**	**2.53^{**}**	**.58**

$^{**}p < .01$; $^{*}p < .05$.

In the cases where participants identified shapes that were addressed during the learning curriculum (trapezoid, rectangle, rhombus, and parallelogram), GI participants clearly outperformed their NI counterparts. In the two cases where participants identified shapes that were not addressed directly during the curriculum (right and isosceles triangles and trapezoids) the results were mixed. While children in the GI condition were more accurate in identifying isosceles shapes, in identifying right angle shapes the scores of both groups were too low to distinguish between conditions. Yet, for right angle shapes the difference in medians (23% for GI vs. 8% for NI) lies in the expected direction. In future iterations of the task, stimuli with more perceptual distinction would be appropriate.

Considering the subtle characteristics of defining features, which likely require more time to identify, we compared conditions in terms of trial durations. Unlike the accuracy results application of Shapiro–Wilk's test of normality did not reveal any significant deviations from normality for trial durations. Therefore we applied two-tailed t-tests for each of the six subtests (Table 4). Results revealed that GI participants spent more time on items than NI participants overall and on four subtests (parallelograms, rhombi, isosceles triangles/trapezoids, and rectangles). A trend towards a similar result emerged for right triangles/trapezoids; although no significant difference emerged for trapezoids, perhaps due to the novelty of the task.

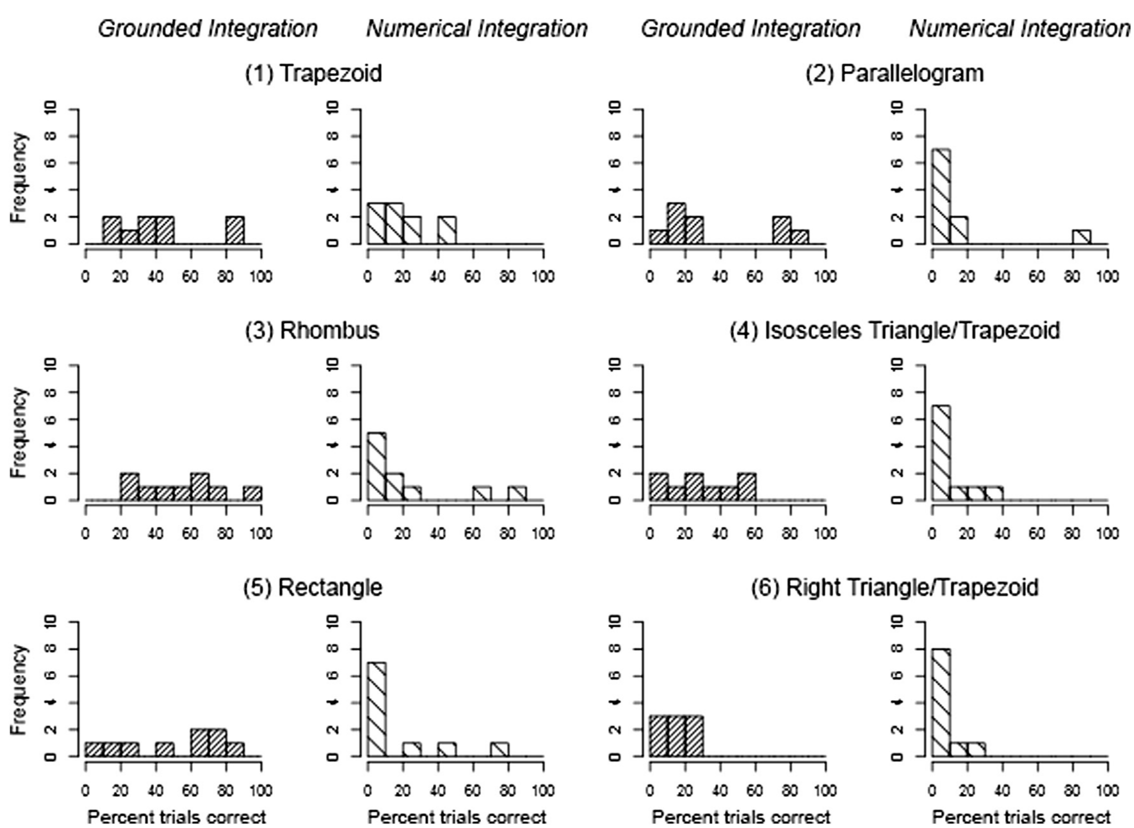


Fig. 5. Distributions of % trials correct by condition. For each subset, GI is displayed on left and NI is displayed on right. The x-axis represents the percent of items in which a participant selected both figures correctly. The y-axis represents the number of participants with the given level of accuracy.

3.3. Performance in learning environment

Every child successfully completed all 26 exercises. From log data we sought to determine whether conditions were equivalent in terms of difficulty and participant strategy. We measured duration of each exercise as a rough proxy for difficulty. Likewise, we computed a mean count of the number of "transforms" (i.e., vertex drags) during each exercise with the assumption that a more challenging exercise would require more adjustments to the shape. Finally, we calculated the number of failed validation attempts – i.e., instances in which children attempted to validate figures that did not meet criteria (Table 5).

Two-tailed t-tests revealed that GI participants, on average, took longer to complete exercises in Unit 2 and Unit 3 (Unit 2: $t[17] = 2.3$, $p < .05$, Unit 3: $t[16] = 2.3$, $p < .05$), but not in Unit 1 ($t[17] = 1.6$, *n.s.*). Similarly, GI participants showed a trend towards performing significantly more transformation operations than NI participants in Unit 2 ($t[17] = 2.0$, $p < .1$), performed significantly more transformation operations in Unit 3 ($t[16] = 3.9$, $p < .01$), but showed no significant difference from NI participants in Unit 1 ($t[17] = 1.3$, *n.s.*). Therefore, differences in both exercise duration and number of transforms emerged in later units as participants became more familiar with the learning environment.

However, in all three units large differences in number of failed validation attempts appeared between conditions (Unit 1: $t[17] = 5.3$, $p < .001$; Unit 2: $t[17] = 5.3$, $p < .001$; Unit 3: $t[16] = 3.6$, $p < .01$). This tendency for GI participants to enter the validation phase with invalid figures suggests that these children had difficulty assessing the relevant properties of their constructions. In the NI condition, on the other hand, children who quickly learned the relevant numerical rules (e.g. parallel sides have equal "% grade") were unlikely to proceed to validation with invalid figures.

While it may be the case that some participants in the GI condition had difficulty understanding concepts of spatial defining features, in many cases failed validations simply reflected strict criteria, causing a distinct perceptual challenge that was not present in the NI condition. Furthermore, given that students were not penalized for failed validation attempts, it may be the case that some participants simply did not apply a high level of vigilance to this challenging perceptual task.

One student in particular ("CJ") demonstrated a high level of conceptual knowledge but hesitancy to fully engage with the perceptual task. CJ quickly picked up relevant concepts and demonstrated knowledge routinely by assisting other participants. At the Unit 1 posttest CJ correctly identified valid trapezoids in 83% of trials and valid parallelograms in 83% of trials – 1st and 2nd highest scores among all participants, respectively. Yet, in one of the more challenging Unit 1 exercises (7) she performed 66 failed validation attempts – more than any other student. Upon observing this performance one of the experimenters asked her to demonstrate the meaning of parallel sides, which she accomplished readily with hand gestures. When questioned about this difficulty CJ indicated that because she was ahead of the other students and growing tired of the task she was not applying her full effort to assess figures in advance, but instead relying upon the validation feedback. For CJ a large number of failed validations represented some disengagement with the challenging perceptual task.

4. Discussion

The purpose of this study was to demonstrate how a *grounded integration pattern* of instruction could be effectively implemented in geometry learning software. We sought to elicit intuitive knowledge concepts by using a common navigational theme and an object-like representation of the constructed path. We introduced novel grounding metaphors in the form of gestural depictions. Finally, we embedded a challenging two-goal structure that directed students to coordinate between two sets of spatial criteria. To evaluate this design we compared students using the full, *grounded integration* software, to those using a version that applied numerical, rather than spatially-grounded, representations.

Results showed that participants in the GI condition were more likely to identify figures accurately than participants in the NI condition in five out six subtests. Considering the difficulty of the posttest, and its difference from the training task (i.e., training was a production task with structured feedback, testing was a perceptual recognition task with no feedback), this result suggests that the *grounded integration pattern* did effectively shift children's conceptual representation from prototypical to formal features.

While the small sample size does limit the generality of these results, medium to large effect sizes did emerge. Additionally, effect sizes observed here are generally similar to those produced in pilot work, where no *feature validation* phase was included in the control condition (see Table 1, above). However, it should be noted that for several subtests of the pilot study posttest post-trial feedback was provided (correct or incorrect). While, the effect of this feedback on effect sizes is unknown, by potentially training participants equally across conditions during the test feedback may have caused convergence between conditions. Generally, consist current and pilot results suggests that the two-goal structure implicit in *feature validation* facilitates learning.

In addition to accuracy, differences in posttest trial duration suggest that children in the NI condition were responding quickly to figures with high visual similarity, echoing Behrman's and Brown (1968) findings that individuals are likely to classify shapes based on salient perceptual characteristics, such as elongation. On the other hand, for children in the GI condition seeking defining features proved to be a greater perceptual challenge and required a more effortful, deliberative strategy, resulting in longer item durations.

Table 4
Posttest mean duration by treatment condition (standard deviations in parentheses).

	Duration (seconds)		t (df = 17)	d
	GI	NI		
Trapezoid	11.6 (3.8)	9.5 (1.4)	0.03	0.8
Parallelogram	11.2 (2.8)	6.8 (1.0)	4.6***	2.1
Rhombus	9.9 (2.6)	7.8 (1.5)	2.1*	1.0
Isosceles tri/trap	13.2 (2.8)	6.9 (1.4)	6.3***	2.9
Rectangle	8.9 (1.6)	6.1 (1.3)	4.2***	1.9
Right tri/trap	8.9 (3.5)	6.4 (1.7)	2.0†	0.9
Overall	**10.6 (1.4)**	**7.3 (0.9)**	**6.2*** **	**2.8**

***$p < .001$; **$p < .01$; *$p < .05$; †$p < .1$.

Table 5
Mean statistics per exercise within each unit of the learning task by treatment condition (standard deviations in parentheses).

Unit	Duration (minutes)		Transform count		Failed validation attempt count	
	GI	NI	GI	NI	GI	NI
(1) Parallel sides	19.9 (3.9)	17.3 (3.2)	36.2 (9.3)	29.7 (12.3)	10.2 (3.8)	3.3 (1.4)
(2) Congruent adjacent sides	13.9 (3.9)	10.2 (3.2)	25.1 (11.8)	15.2 (7.9)	9.7 (4.9)	1.4 (.9)
(3) Right angles	14.6 (4.1)	10.2 (3.0)	27.5 (11.6)	11.4 (4.3)	10.9 (8.0)	1.4 (.4)

a. Relevant error for each unit: (1) abs. difference in % gradients; (2) abs. difference in pixel length; (3) abs. difference from 90.

While relatively clear differences between conditions emerged for posttest measures, some differences in training performance – in particular time-on-task (in unit 2 and 3) and number of failed validations, which could call into question the validity of the posttest results. To some extent this difference in time-on-task may reflect inherent and unavoidable differences between perceptual and symbolic processes – i.e., the symbolic advantage (Koedinger et al., 2008). As the example of CJ illustrates, this perceptual difficulty can affect even high-performing students.

On the other hand, children in the NI condition often discovered numerical margins of error explicitly, and were able to apply this knowledge to construction strategies. For example, to construct right angles in the third unit NI participants could adjust a single vertex until they produced an exact 90° angle, whereas GI participants received no such direct evidence of their angle's accuracy. Rather these participants needed to commit to a particular angle, attempt to validate it, and then apply another transformation if incorrect.

These additional feedback cycles that emerged in the GI condition provided participants with an opportunity to engage in further perceptual training, which was then directly applicable to the posttest. Therefore, the differences that emerged between conditions likely reflected both a perceptual and a conceptual advantage for participants in the GI conditions. Yet, we do not claim that participants in the NI condition did not learn, but simply that the concepts that they learned were not the focus of posttest items. In a numerical task effects may have been reversed. Although we did not explore this possibility due to time restrictions, tradeoffs between numerical and spatial representations in geometry will be pursued in future research.

While this study does provide strong evidence for the *grounded integration pattern* as a whole, we did not attempt to test each step of the pattern individually. Rather our focus was on the introduction and integration of novel grounded representations of geometric concepts. The original research presented here represents an initial, but promising step into research on the use of grounded metaphors in instructional software. In the following we outline areas for future research and open questions relating to each step of the *grounded integration pattern*.

4.1. Eliciting intuitive knowledge

The first step of the *grounded integration pattern* makes an assumption that there are some (nearly) universal ideas that children maintain about specific concepts. Focusing themes and affordances of instructional materials on these ideas fosters accessibility. Determining what concepts are intuitive is a challenge that varies by subject matter. While in the case of mathematics several lines of research suggest that some ideas are derived from innate systems (Feigenson et al., 2004), we make no demand for innateness for intuitive concepts in general.

In the case of geometry, the research on intuitive concepts is emerging but relatively concrete. Spelke et al.'s (2010) assertion regarding the navigational and perceptual basis for geometry served as framework for the construction of our tool. Prior research in geometry instruction confirms that these systems, combined with insufficiently rigorous instruction, constrain student ideas and sometimes promotes non-normative concepts (Clements, 2004).

Beyond geometry, research on intuitive concepts is well-developed in many other domains. In particular a large body of research details the ideas – often misconceived – that children maintain about scientific phenomena. For example, the idea that seasons are caused by changes in distance from the Sun to the Earth is quite common among children and many adults (Vosniadou & Brewer, 1992). In many cases these ideas emerge because they are deeply grounded in personal experience – e.g. it gets warmer as you approach a heat source.

The persistence and robustness of many common misconceptions or preconceptions (Clement, 1982; Halloun & Hestenes, 1985) may even cast doubts on whether eliciting students' intuitive beliefs is beneficial. According to Ohlsson's (2009) "resubsumption" approach the risk of introducing misconception outweighs potential benefits. Rather, Ohlsson suggests that new ideas should be introduced within an unfamiliar context. As new ideas are developed, without risk of transferring misconception from prior knowledge, they are slowly transferred into an authentic context. In geometry this approach could be applied by introducing concepts of defining features in the context of unfamiliar shapes (e.g. a "dart"), and then re-applying these concepts to known shapes.

In contrast, Linn and Eylon's (2011) knowledge integration framework suggests that while students initial ideas may be non-normative, by developing strong criterion for distinguishing between normative and non-normative ideas, children can develop accurate conceptions within authentic contexts. While the study described here, and the *grounded integration pattern* in general, align with the knowledge integration approach, an alternative approach, using an unfamiliar context and figures, is worth consideration.

4.2. Introducing novel grounding metaphors

Depending on the nature of the material, concepts may be grounded in any number of modal systems – e.g. perceptual, motor, emotional (Barsalou, 2008). In this case, for students in the grounded condition the gestural depictions were primarily visual, but by the end of each learning unit most children had performed the corresponding physical gestures. Yet, the frequency of and purpose behind the use of gestures varied among children. In some cases, gestures appeared to serve as an initial reminder of the meaning of some geometric concept, but faded as children became increasingly fluent. For other children gestures were adopted as a tool for accurate figure construction – e.g. by matching their hands with components of the constructed figures. Similarly, some children, recognizing the limitations of these gestures, applied physical artifacts from the classroom – such as a straight edge or the corner of a sheet of paper (for right angles) – to increase their precision in shape construction. Finally, in some cases children showed little use of external tools or gestures.

Because we did not track student gestures, we cannot make definitive conclusions about the necessity of performing gestures in learning target concepts; however, we did observe some general patterns. In particular, children with stronger initial concepts of parallelism, congruency, and perpendicularity (e.g. CJ) more readily demonstrated corresponding gestures, and sometimes served as informal tutors. For struggling students, one or more demonstrations of corresponding hand gestures, by peer or experimenter, was often required. Taken together, these observations support Beilock and Goldin-Meadow's (2010) notion of a reciprocal relationship between conceptual thought and gesture.

Yet, in some cases learning to apply appropriate gestures did not facilitate performance in the posttest. Specifically, those children who exclusively applied gestures as an evaluative tool in the learning environment (e.g. placing hands held perpendicularly against the screen to determine the presence of a right angle) were less successful applying the same strategy during the posttest due to the smaller area of the figures. Therefore, for these students, applying gestures as tools – as one might a ruler or T-square – may have lightened the cognitive load during the learning task (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), but served as a disadvantage when the materials no longer afforded this strategy. The applicability of gestures beyond the learned context is therefore an important consideration.

Yet, given a generally positive impact of gestures, learning software may support student thinking by directing, evaluating, and providing feedback on gestures. With only the common computer mouse or a touch display, movements that are congruent to displayed representations may promote development of fundamental mathematical concepts (Segal, 2011). Beyond the computer mouse, which is limited in its potential to model all relevant concepts, emerging gesture-based technologies, such as multi-touch screens and motion-capture systems, have enormous instructional potential. In the case of geometry, appropriate gestures often require at least two hands. A motion capture system, for example, could provide feedback on gestures that incorporated students' entire arms as they are aligned in parallel.

Yet, while gesture-based technology has many promising applications to learning, we should not overlook the power of dynamic visualizations to convey meaning. Widely available "dynamic geometry systems" (DGS), such as Cabri-Geometry, typically enable students to construct figures with specific constraints (e.g. with parallel sides) and then observe how these features remain invariant over various transformations (Battista, 2001). By observing this invariant spatial feature as the while manipulating the figure, may "pick-up" on the relevant perceptual information (Gibson, 1969) and develop a corresponding mental representation. Unlike the tool described here these systems typically do not provide supplementary grounding metaphors. As such, these systems are primarily targeted to secondary students who are familiar with the fundamental concepts, and have shown promise in enhancing students' geometric proof skills (Laborde, 2001).

Yet, it may be the case that the DGS approach of illustrating concepts through interaction with abstract geometric figures may be appropriate for younger students. By eschewing a narrative these systems may buffer against highly-contextualized knowledge. For example, Goldstone and Son (2005) found that learners demonstrated greater transfer when using idealized images in a NetLogo simulation than when using more realistic images. On the other hand, contextually-abstract representations may make it difficult for students to understand the practical value of these activities. The trade-offs between rich, authentic and abstract spatial representations should be considered while developing instructional activities, and requires more thorough attention by researchers.

4.3. Embedding challenges to promote distinguishing between ideas

A prominent attribute of expertise is the ability to apply formal knowledge flexibly to any context. Yet, even domain experts apply informal – and in some cases misconceived – notions in everyday settings, such as the distinction between heat and temperature (Lewis & Linn, 1994). Therefore rather than seeking to replace intuitive ideas, a more realistic goal is to foster the ability to think critically about inconsistent concepts. Research in the area of desirable difficulties in science suggests that this type of flexible thinking can be promoted through instruction that is both challenging and motivating (Bjork & Linn, 2006).

In the case of geometry, intuitive notions about shapes emerge from common experiences and are relevant and appropriate in many circumstances. Yet, as the overall low scores in the posttest describe here demonstrate, the persistence of intuitive concepts may come at the expense of new ideas. While relatively higher accuracy scores and extended trial durations in the GI condition do suggest that these children were more likely to attend to more subtle defining features, overall low median accuracy (e.g. 23% and 27% median accuracy for parallelograms, and right triangles/trapezoids, respectively) suggests that the learning materials presented here leave room for improvement.

In terms desirable difficulties, while dual goals – relating to general spatial characteristics and defining features of the figure – were intended to elicit a confrontation of ideas, some children did develop means of circumventing this cognitive conflict. For example, in some cases students made accurate predictions for the location of vertex points during the planning phase, and then used either the grid or their fingers, held on the computer screen, to assist in placement during construction. While these strategies did facilitate accurate production of figures, these students missed an opportunity to engage in "productive failure" (Kapur, 2008). For the rectangles and square exercises, the additional symmetry of layouts also contributed to rapid, successful figure building. This may have negatively impacted transfer of the concept of right angles to novel figures as seen in the final posttest.

In particular, reflecting Vygotsky's (1978) zone of proximal development, the grid may represent a necessary scaffold for novice learners that became unnecessary as the learners gained expertise. With learning software these scaffolds can be targeted for automated removal as learners demonstrate stronger abilities. As an example, Goldstone and Son (2005) applied "concreteness fading" by shifting highly contextualized images to increasingly iconic figures over the course of the learning task to encourage attention to general spatial features. In their case the contexualized images served as a scaffold to orient students to the meaning of the task, but also provided unnecessary distraction from distinguishing features. We applied a similar, reduction of scaffolds in the shift from highly prototypical to atypical layouts of obstacles.

While gradual reduction of scaffolds is an important strategy for advancing students within learning environments, our primary approach here was to incorporate distinct challenges that are overcome by coordinating between multiple ideas, goals, or representations. These "grounded coordination challenges" (reference omitted for blind review), foster deeper reflection and development of flexible knowledge by directing learners' attention to inconsistencies between representations within the same modality. In this study children in the GI condition coordinated between animations depicting necessary spatial features and underlying figures that were constructed to fit spatial characteristics of the obstacle course. Discrepancies between representations promoted revision of ideas. On the other hand, children in the GI condition coordinated between a spatial and numerical representation, because no salient visual inconsistencies emerged, children were less likely to revise their ideas.

4.4. Conclusions and recommendations

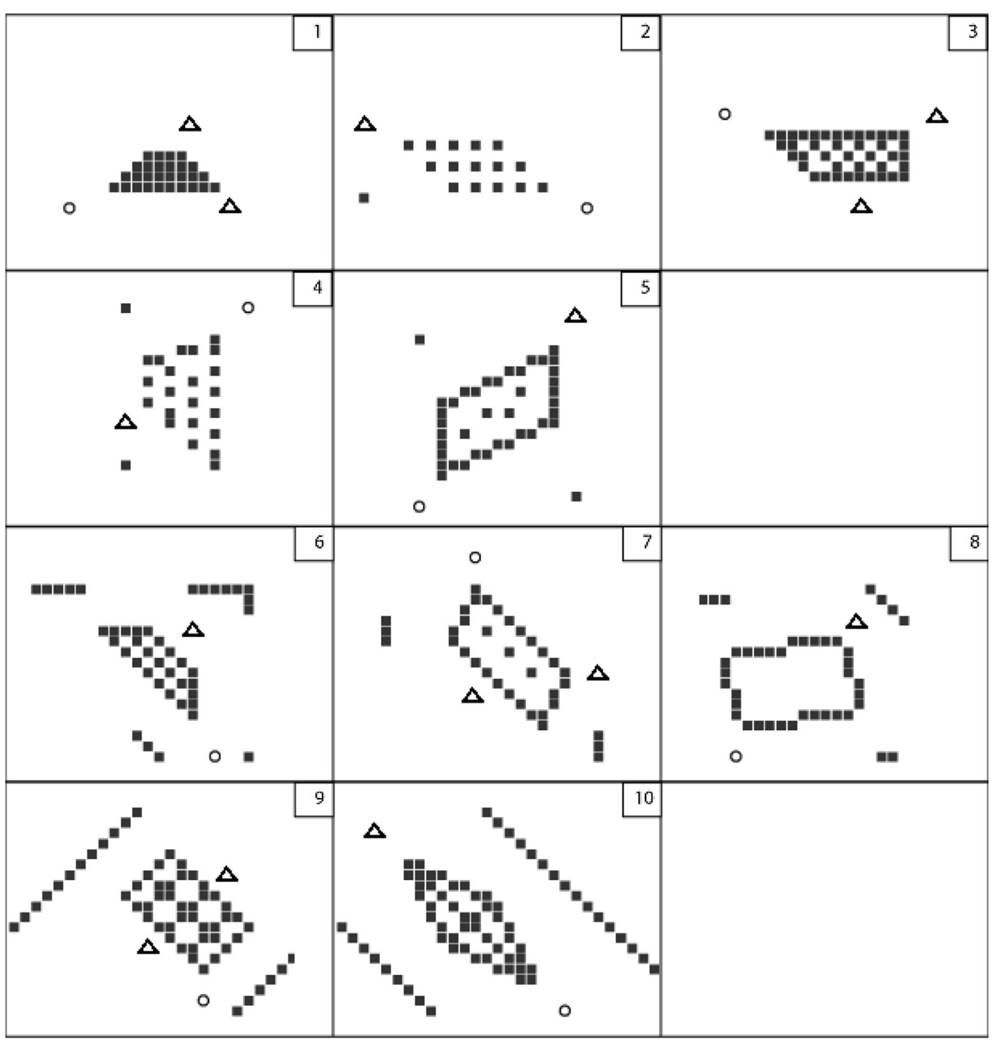
This study indicates that geometry tasks need to be challenging, yet spatially-grounded. While this may seem nearly self-evident, geometry instruction is typically quite superficial for young children or abstract for older students (Clements, 2004). For example, children often engage in activities with cut-out or physical "pattern blocks". Though these blocks provide a potential grounding for geometry concepts, they are often limited and prototypical. This experience does not adequately prepare students for higher-level geometric ideas and tasks, like deductive proofs. This is true of many other domains where young children are taught with intuitive concrete manipulatives but transitioned abruptly to symbolic contexts to engage in more rigorous activities.

While the extended game-like features of the tool explored here may not be relevant to all classroom settings, the *grounded integration pattern* offers a framework for constructing new activities in any setting. Specifically, activities should elicit intuitive concepts with accessible materials, such as concrete manipulatives, stories, or animations. Novel grounding metaphors can be animated depictions of some action, as in this study, or learners can be asked to perform an embodied representation of an action.

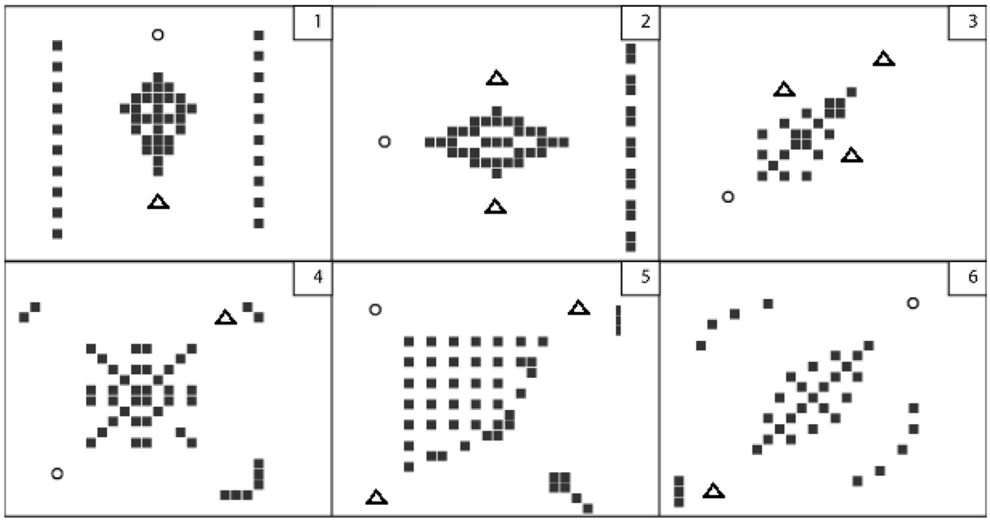
Most importantly, learners should be expected to engage in the difficult work of coordinating intuitive and novel concepts by participating in activities where intuitive concepts are salient, but formal concepts are necessary to achieve success. Digital technology can be particularly useful in constraining student ideas and guiding their integration of difficult concepts (Clark & Linn, 2003); however, in non-digital environments construction activities with frequent teacher or peer feedback may be equally as valuable. For example, if an appropriate project is chosen, origami may focus students on specific geometric concepts in service of constructing a desired figure.

While the development of accessible learning materials represents an ongoing trajectory in educational research, more research is needed to address how challenges can be embedded within technological systems to foster integration of ideas. Computer software can be structured to demand some confrontation between opposing ideas and provide related feedback; thereby posing an advantage over traditional materials. However, the characteristics of guidance that strikes the appropriate balance between challenge and accessibility require continued investigation. We encourage researchers to continue to elucidate activities that challenge and inspire children to engage deeply with learning materials.

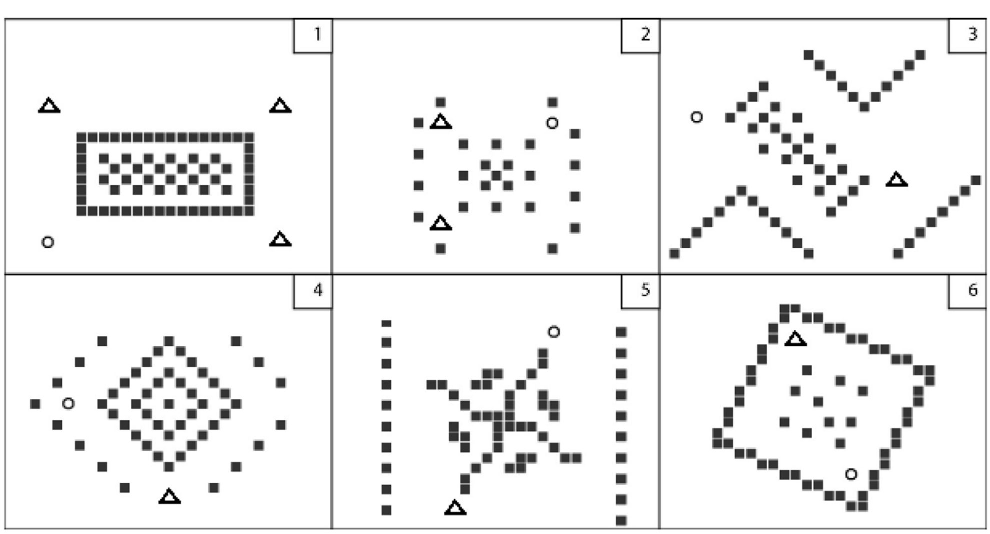
Appendix A. Unit 1 exercise layouts: trapezoids and parallelograms. Learners validate parallel sides. Open circles represent the starting and ending point for agent. Open triangles represent goal points for agents. Solid squares represent obstacles for agents.



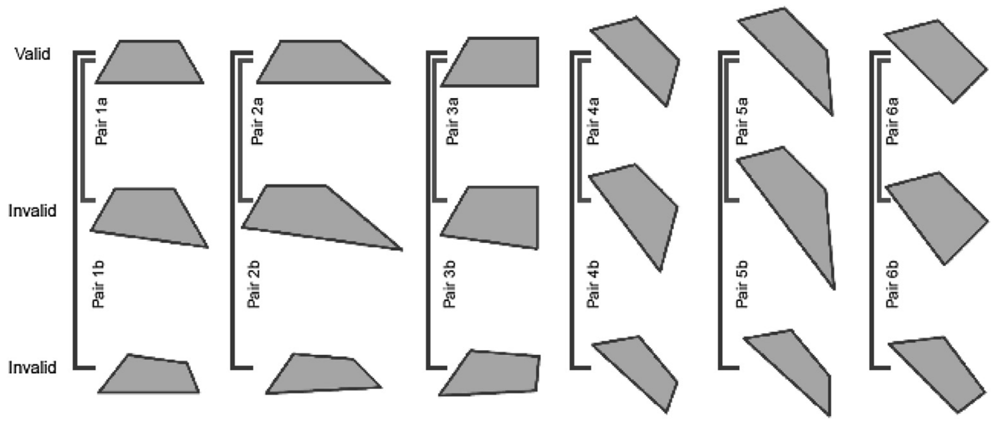
Appendix B. Unit 2 exercise layouts: kites and rhombi. Learners validate congruent adjacent sides.



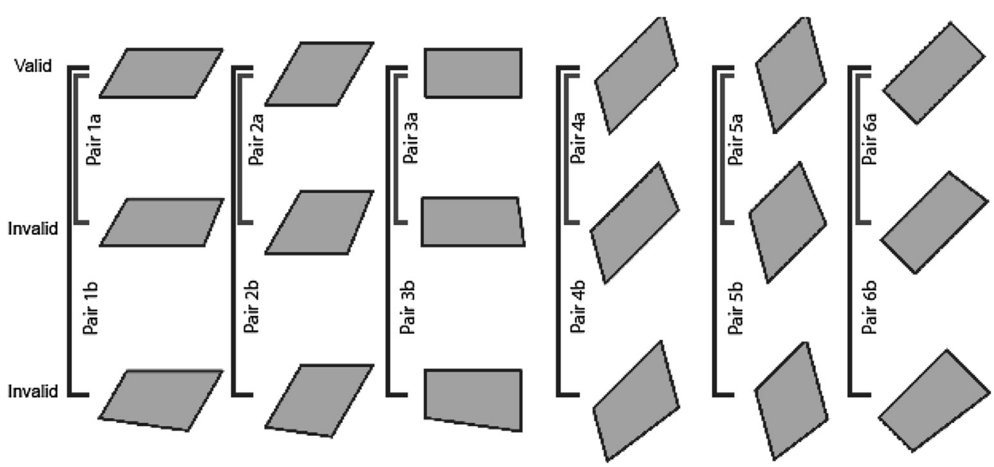
Appendix C. Unit 3 exercise layouts: rectangles and squares. Learners validate right angles and congruent adjacent sides (for squares, only).



Appendix D. Shape identification, trapezoid stimuli.

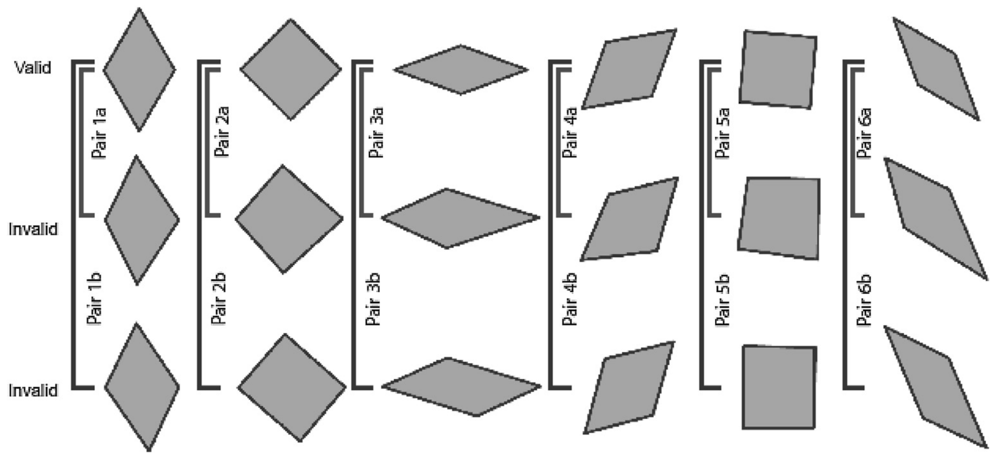


Appendix E. Shape identification, parallelogram stimuli.

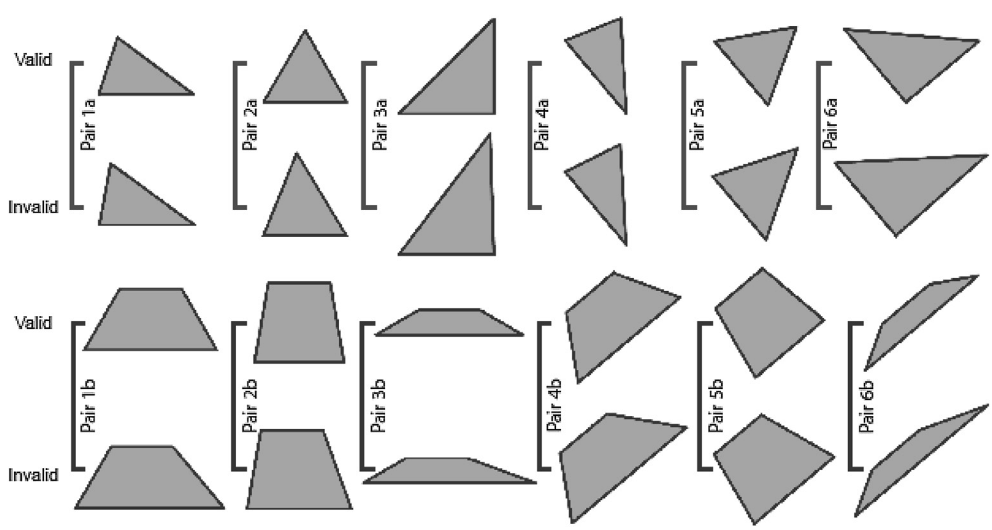


Details: For trapezoid, parallelogram, rhombus, and rectangle stimuli, the top row displays six valid instances of the target shape class. The second and third rows display an invalid derivation of the valid instance above. Each trial of the shape identification task displayed four figures, or two pairs of valid–invalid pairs. Fifteen pairs of pairs were constructed by combining each of six pairs in “set a” (top and middle rows) with every other pair. Fifteen additional pairs of pairs were constructed by combining each of “set b” (top and bottom rows) with every other pair.

Appendix F. Shape identification, rhombus stimuli.

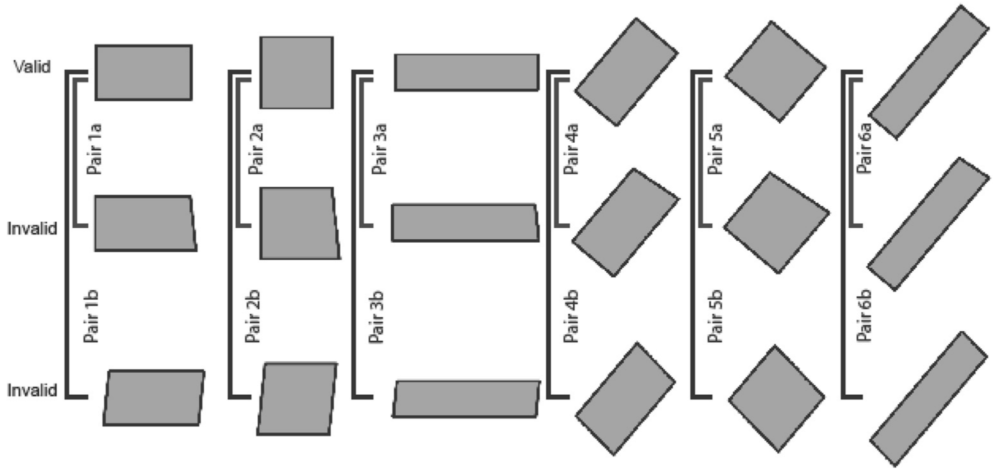


Appendix G. Shape identification, isosceles triangle/trapezoid stimuli.

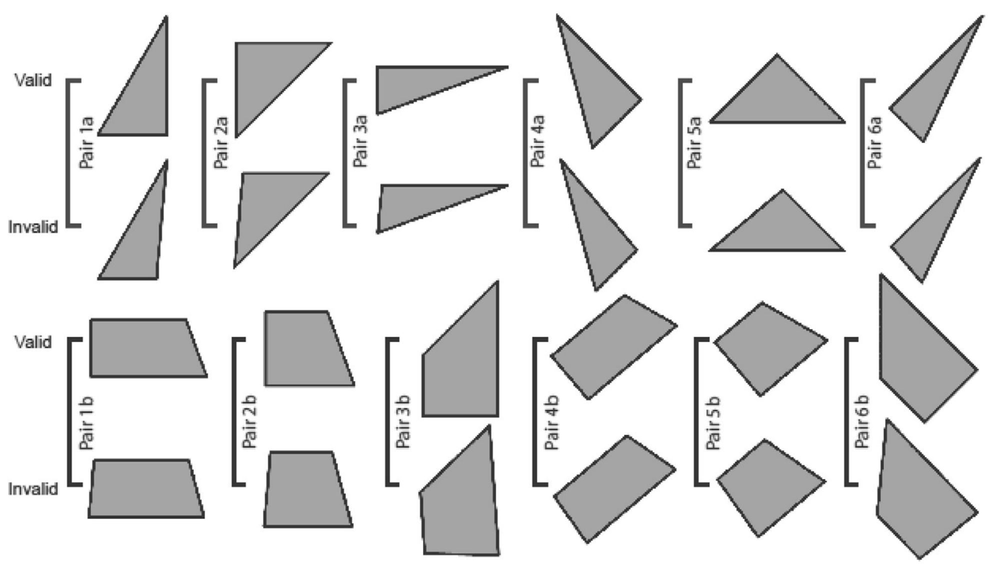


Details: For isosceles and right triangle/trapezoid stimuli, the first and third rows display six valid instances of the target triangle or trapezoid class, respectively. The second and fourth rows display an invalid derivation of the valid instance directly above. Each trial of the shape identification task displayed four figures, or two pairs of valid–invalid pairs. Fifteen pairs of (triangle) pairs were constructed by combining each of six pairs in “set a” (first and second rows) with every other pair. Fifteen additional pairs of (trapezoid) pairs were constructed by combining each of “set b” (third and fourth rows) with every other pair. Hence, each trial displayed triangles or trapezoids only.

Appendix H. Shape identification, rectangle stimuli.



Appendix I. Shape identification, right triangle/trapezoid stimuli.



References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645.

Battista, M. T. (2001). Shape makers: a computer environment that engenders students' construction of geometric ideas and reasoning. *Computers in the Schools, 17*(1–2), 105–120.

Behrman, B. W., & Brown, D. R. (1968). Multidimensional scaling of forms: a psychological analysis. *Perception & Psychophysics, 4*(1), 19–25.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21*(11), 1605–1610.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: introducing desirable difficulties. *APS Observer, 19*(3), 29–39.

Black, J. B., Segal, A., Vitale, J. M., & Fadjo, C. L. (2012). Embodied cognition and learning environment design. In D. Jonassen, & S. Land (Eds.), *Theoretical foundations of learning environments* (2nd ed.) (pp. 198–223). New York: Routledge.

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology, 35*, 294–328.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189–201.

Broaders, S. C., Wagner, S. C., Zachary, M., Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136*(4), 539–550.

Bruner, J. S. (1960). *The culture of education*. Cambridge, MA: Harvard University Press.

Burger, W. F., & Shaughnessy, J. M. (1986). Characterizing the Van Hiele levels of development in geometry. *Journal for Research in Mathematics Education, 17*(1), 31–48.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: the impact of instructional time. *Journal of the Learning Sciences, 12*(4), 451–493.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66–71.
Clements, D. H. (2004). Geometric and spatial thinking in early childhood education. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood education* (pp. 267–297). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 420–464). New York: Macmillan.
Clements, D. H., Battista, M. T., & Sarama, J. (2001). Logo and geometry. *Journal for Research in Mathematics Education. Monograph, 10*. i–177.
Clements, D. H., & Burns, B. A. (2000). Students' development of strategies for turn and angle measure. *Educational Studies in Mathematics, 41*(4), 31–45.
Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concept of shape. *Journal for Research in Mathematics Education, 30*(2), 192–212.
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
Dehaene, S., & Izard, V. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science, 311*(5759), 381–384.
Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Science, 8*(7), 307–314.
Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005). *International mathematics performance: New findings from the 2003 TIMSS and PISA*. Washington, DC: American Institutes for Research.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological Science, 12*(6), 516–522.
Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65*(2–3), 231–262.
Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*(1), 69–110.
Goldstone, R. L., & Son, J. Y. (2008). *A well grounded education: The role of perception in science and mathematics*. In *Symbols, embodiment and meaning*.
Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043. http://dx.doi.org/10.1119/1.14030.
Howe, J., O'Shea, T., & Plane, F. (1979). Teaching mathematics through LOGO programming: an evaluation study. In *Computer assisted learning: Scope, progress, and limits* (Roehampton, England).
Hughes, M., & Greenbough, P. (1995). Feedback, adult intervention, and peer collaboration in initial LOGO learning. *Cognition, 13*(4), 525–539.
Johnson, P. A. (1986). *Effects of computer-assisted instruction compared to teacher-directed instruction on comprehension of abstract concepts by the deaf* (Unpublished doctoral dissertation). Northern Illinois University.
Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 320*, 454–455.
Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: the portability of generic instantiations. *Child Development Perspectives, 3*(3), 151–155.
Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424.
Keil, F. C., & Baterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning, 23*, 221–236.
Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: evidence from algebra problem solving. *Cognitive Science, 32*(2), 366–397.
Laborde, C. (2001). Dynamic geometry environments as a source of rich learning contexts for the complex activity of proving. *Educational Studies in Mathematics, 44*(1), 151–161.
Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. In *Basic Books*. New York: Basic Books.
Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.
Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: implications for curricular improvements. *Journal of Research in Science Teaching, 31*(6), 657–677.
Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education, 22*(8), 781–796.
Linn, M. C., Chiu, J., Zhang, H., & McElhaney, K. (2010). Can desirable difficulties overcome deceptive clarity in scientific visualizations? In A. Benjamin (Ed.), *Successful remembering and successful forgetting a Festchrift in honor of Robert A Bjork* (pp. 239–262) New York: Routledge.
Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York: Routledge.
Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*, 587–604.
Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology, 99*(3), 640–652.
McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: perspectives from development and education. *Child Development Perspectives, 3*(3), 137–139.
National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
Noss, R., & Hoyles, C. (1992). Afterword: looking back and looking forward. In C. Hoyles, & R. Noss (Eds.), *Learning mathematics and Logo* (pp. 427–468). Cambridge, MA: MIT Press.
Ohlsson, S. (2009). Resubsumption: a possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*(1), 20–40.
Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.
Papert, S., Watt, D., DiSessa, A. A., & Weir, S. (1979). *Final report of the Brookline Logo Project: Parts I and II (Logo memos nos. 53 and 54)*. Cambridge, MA: M.I.T. Artificial Intelligence Laboratory.
Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. New York: Norton.
Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry*. New York: W. W. Norton.
Quinlan, P. T., & Humphreys, G. W. (1993). Perceptual frames of reference and two-dimensional shape recognition: further examination of two-dimensional shape recognition. *Perception, 22*, 1343–1364.
Quinn, P. C., & Eimas, P. D. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. *Review of General Psychology, 1*(3), 271–287.
Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*(2), 375–394.
Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*(1), 1–54.
Segal, A. (2011). *Do gestural interfaces promote thinking? Embodied interaction: Congruent gestures and direct touch promote performance in math* (Unpublished doctoral dissertation). New York: Columbia University.
Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science, 11*(5), 655–661.
Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545–560.
Simmons, M., & Cope, P. (1990). Fragile knowledge of angle in Turtle geometry. *Educational Studies in Mathematics, 21*, 375–382.
Simmons, M., & Cope, P. (1993). Angle and rotation: effects of different types of feedback on the quality of response. *Educational Studies in Mathematics, 24*(2), 163–176.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.
Spelke, E. S., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: natural geometry. *Cognitive Science, 34*(5), 863–884.
van Hiele, P. M. (1986). *Structure and insight: A theory of mathematics education*. Orlando, FL: Academic Press.
Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: a study of conceptual change in childhood. *Cognitive Psychology, 24*(4), 535–585.
Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.