Learning and Instruction 33 (2014) 182–193

Contents lists available at ScienceDirect

Learning and Instruction

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

Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof

Mitchell J. Nathan [a,*], Candace Walkington [b], Rebecca Boncoddo [c], Elizabeth Pier [a], Caroline C. Williams [d], Martha W. Alibali [e]

[a] Department of Educational Psychology, University of Wisconsin – Madison, 1025 W. Johnson St., Madison, WI 53706, USA
[b] Department of Teaching and Learning, Southern Methodist University, PO Box 750455, Dallas, TX 75275, USA
[c] Department of Psychological Science, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050 USA
[d] Department of Curriculum and Instruction, University of Wisconsin – Madison, 1025 W. Johnson St., Madison, WI 53706, USA
[e] Department of Psychology, University of Wisconsin – Madison, 1202 W. Johnson St., Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 10 August 2013
Received in revised form
4 July 2014
Accepted 6 July 2014
Available online 26 July 2014

Keywords:
Grounded cognition
Embodied cognition
Justification & proof
Pedagogical language

ABSTRACT

Theories of grounded and embodied cognition posit that situated actions are central constituents in cognitive processes. We investigate whether grounding actions influence reasoning, and how pedagogical language influences the action–cognition relationship. Undergraduate students ($N = 120$) generated proofs for two mathematical tasks after performing either grounding or non-grounding actions. Grounding actions facilitated key mathematical insights for both tasks, but did not lead to superior proofs. Pedagogical language in the form of prompts (prospective statements) and hints (retrospective statements) accompanying grounding actions enhanced proof performance on one task but not the other. Results from transfer tasks suggested that participants learned to apply their mathematical insights to new contexts. The findings suggest that relations between action and cognition are reciprocal: actions facilitate insight, while pedagogical language strengthens the influence of task-relevant actions for proof production. Pedagogically supported grounding actions offer alternative ways of fostering mathematical reasoning.

Can performing body-based actions help learners to reason mathematically? Mathematical proof is an area of generalized, propositional reasoning that is notoriously difficult for students (Healy & Hoyles, 2000). People may understand generalizable mathematical ideas through action, as well as language (see Nathan, 2014, for a review). We conducted an experiment in which participants were directed to perform physical actions that were either relevant or irrelevant to the solution to a mathematical proof task. We considered actions produced from both a first-person and third-person point of view. Furthermore, we examined whether pedagogical language—specifically, prospective prompts and retrospective hints verbally connecting actions to the task—provided further support for participants' proof practices. In the following section, we review the literature on grounded and embodied cognition, the impact of pedagogical language, actions performed from different points of view, and the nature of students' proof practices.

1. Theoretical framework

1.1. Grounded and embodied cognition

We place this research within the framework of *grounded* and *embodied cognition* (GEC; Barsalou, 2008; Wilson, 2002). Theories of GEC posit that the body and the world are central constituents in cognitive processes (Shapiro, 2010). The general claim of GEC is that modal-specific systems for action, perception, and introspection form the basis for cognition through situated action and simulated re-enactment of prior modal experiences. One line of evidence in support of GEC comes from studies showing that action supports insight problem solving. For example, participants directed to produce eye gaze patterns that converged to the center of a diagram were most successful at solving the General and Fortress Problem, as that pattern embodies the key insight underlying the successful solution (Thomas & Lleras, 2007). Along similar lines, participants

* Corresponding author. Tel.: +1 608 263 0563.
E-mail addresses: mnathan@wisc.edu (M.J. Nathan), cwalkington@smu.edu (C. Walkington), boncoddo@ccsu.edu (R. Boncoddo), epier@wisc.edu (E. Pier), caro.williams@gmail.com (C.C. Williams), mwalibali@wisc.edu (M.W. Alibali).

http://dx.doi.org/10.1016/j.learninstruc.2014.07.001
0959-4752/

who produced body-based actions during training sessions with the Tower of Hanoi integrated their motor experiences into their mental representations of the task objects (Goldin-Meadow & Beilock, 2010).

The GEC framework can be productively applied to mathematical reasoning. Although mathematical reasoning typically addresses abstract entities far removed from situated action, many basic mathematical entities are grounded in the body, spatial systems, and situated actions (e.g., Alibali & Nathan, 2012; Lakoff & Núñez, 2000; Moyer & Landauer, 1967). Body-based actions offer a potentially powerful means to ground abstract mathematical ideas in concrete form, so their meaning can be more readily understood (Goldstone & Son, 2005; Nathan, 2008). There are many examples of such *grounding actions* in mathematical reasoning, including children counting with fingers to ground concepts of number (Carpenter & Moser, 1984) and teachers representing mathematical objects (such as a cube) with hand gestures to address student misunderstandings during instruction (Alibali et al., 2013).

Of course, much of formal mathematics emerged not simply from actions and perceptions, but from the social construction of mathematical knowledge mediated by language (e.g., Hersh, 1997; Romberg & Kaput, 1999). Action and perception may underpin early math development (e.g., the approximate number system, Dehaene, Bossini & Giraux, 1993), but by themselves, action and perception are inadequate for constructing the interconnected and self-referential system that we identify as the discipline of mathematics, which includes exact mathematics, precise definitions, and careful argumentation (Núñez, 2009; Rips, Bloomfield, & Asmuth, 2008). Language seems necessary for some mathematical reasoning, and may even *enable* exact mathematics, for example, exact representations of large numbers (Spelke & Tsivkin, 2001). Thus, language, along with action, offers a system that grounds mathematical cognition (Lakoff & Núñez, 2000) and supports its development (Vygotsky, 1986).

1.2. Point of view of actions

Human actions can exhibit a first-person or third-person *point of view*. That is, actions can be produced to convey the actor's own perspective, or the perspective of an observer. For example, McNeill (1992) observed that speakers' gestures often shift between first-person (*character*) and third-person (*observer*) viewpoints. Parrill (2011) documented ways that narrators assume a first-person perspective when their body parts mapped directly onto the corresponding body parts of the character being described, but conveyed a third-person viewpoint when the speakers' hands simply traced the character's motion. Gerofsky (2010) coded middle-school students' actions on graphs of polynomial functions as third-person when they *traced* the graphs, and first-person when students configured their bodies to actually *be* parts of the graphs. When an actor assumes a third-person perspective, they enact the actions of another character or object *as if* it were being observed.

There is little consensus on how point of view in gesture and action relates to cognition. Some studies have suggested that the "close-up" perspective manifested in first-person gestures is used more often by high-performing students, and may help explain their higher levels of engagement (Gerofsky, 2010). Others have suggested that third-person viewpoint allows for superior transfer, because it strips perceptual details from the action being simulated (Goldin-Meadow & Beilock, 2010). Because actions and gestures from both viewpoints have been associated with benefits for cognition, we compared actions from each viewpoint in this study.

1.3. Pedagogical language

Language is useful in mathematics learning and instruction in a variety of ways. For example, prompts that foster language production, such as generating self-explanations, are effective for promoting mathematics argumentation and learning (e.g., Hilbert, Renkl, Kessler, & Reiss, 2008; Rittle-Johnson, 2006) and quality mathematical argumentation (Kollar, Ufer, Reichersdorfer, Vogel, Fischer & Reiss, 2014). Language is also crucial for learning through its role in *pedagogy*—the transmission of semantic information through ostensive, referential communication (Csibra & Gergely, 2009). Although "telling" has its limitations, it can be effective for imparting information and correcting misconceptions (Carroll, 1968). Language may be especially important for making key conceptual connections in complex learning environments.

Teachers often use spontaneous gestures along with pedagogical language in a manner that facilitates students' processing of difficult classroom material (Alibali & Nathan, 2012; Roth, 2001). For example, a teacher might respond to a student's inability to visualize a general pattern about cubes by producing gestures, coordinated with speech, that depict a hypothetical surface (Alibali, et al., 2013). Thus, pedagogical gestures, which manifest simulated actions and perceptual states, both complement and reinforce pedagogical language.

Teachers also use pedagogical language to help students connect ideas, both *retrospectively*—using hints to reference past events—and *prospectively*—using prompts to alert students to related future concepts and experiences. Both hints and prompts have been shown to foster cognitive processing (Anderson, Corbett, Koedinger, & Pelletier, 1995; Gick & Holyoak, 1980; Hilbert et al., 2008), but their contributions to mathematical proof practices, specifically in conjunction with body-based action, have not been well studied.

1.4. Mathematical proof and justification

Proof is "an essential component of doing, communicating, and recording mathematics" (Schoenfeld, 1994, p. 74). In the United States, the Common Core State Standards identify constructing viable arguments and critiquing the reasoning of others as critical skills across grade levels, and standards for high school geometry specifically call for students to learn to construct proofs of theorems (Common Core State Standards Initiative, 2010). Yet students consistently struggle with constructing, interpreting, and evaluating proofs (e.g., Hoyles & Healy, 2007; Knuth, 2002).

Several researchers have proposed frameworks through which students' proof practices can be understood (e.g., Miyazaki, 2000; Stylianides, 2008). According to Harel and Sowder (2005), a valid *transformational proof* has three characteristics. First, it is general—it shows the argument is true for all mathematical objects falling into the category. Second, it involves operational thought, such that an individual who is generating the proof progresses through a goal structure, anticipating the results of transformations. Finally, it involves logical inference, in that conclusions are drawn from valid premises. Although traditional notions view proof as involving a formal, written argument in mathematical notation, more recent work on justification and proof has emphasized the importance and validity of proofs that are communicated verbally (e.g., Healy & Hoyles, 2000; Stylianou, Blanton, & Knuth, 2009).

Proof production has been described in terms of two components (Harel & Sowder, 2005): *ascertaining*, in which the solver determines the key mathematical insights that make the assertions likely to be true; and *persuading*, in which the solver articulates a logical chain of reasoning to be disseminated and publically evaluated in a full proof. Because of this recognition of the separate contributions that insight into conjectures and articulation of full proofs can play in proof practices, we looked at the differential influences of our experimental factors on insights made during participants' justifications, as well as the mathematical accuracy when stating their *full transformational proofs*.

2. Hypotheses and predictions

In this section we list the overarching hypotheses of our study, and the associated predictions that fall under each hypothesis.

2.1. Hypothesis 1: The influence of grounding actions

We hypothesize that performing task-relevant grounding actions facilitates proof production. Thus, the associated prediction 1a states that participants directed to perform grounding actions independently (i.e., without supporting hints or prompts) will be more likely to generate *key mathematical insights* than those who perform non-grounding actions. Prediction 1b states that participants directed to perform grounding actions independently will be more likely to generate valid *transformational proofs* than those who perform non-grounding actions.

2.2. Hypothesis 2: The combined influence of pedagogical language and action

Although reasoning may be grounded in actions, we hypothesize that pedagogical language will help participants grasp the relevance of the actions and inform their reasoning. We predict (Prediction 2a) that pedagogical language linking task-relevant grounding actions to the task domain, in the form of verbal prompts and hints, will increase participants' likelihood of generating *key mathematical insights*, compared to those who perform grounding actions without hints or prompts. We further predict (Prediction 2b) that verbal prompts and hints that signal the relevance of grounding actions will increase participants' likelihood of constructing *transformational proofs,* compared to participants who receive no hints or prompts.

2.3. Hypothesis 3: Generalizability across tasks

We explore proof production in two dissimilar contexts to test the boundary conditions for our results. The Triangle task explores a mathematical conjecture about defining properties of all triangles, and it relates to concrete and spatial properties of Euclidean geometry. The Gear task addresses the parity property of number systems and thus involves more abstract properties of number theory and group theory. We hypothesize that the effects of grounding actions and language on proof production will be similar across tasks (Prediction 3).

2.4. Hypothesis 4: Transfer

We hypothesize that grounding actions will support construction of generalizable proof knowledge, resulting in positive transfer of *insights* (Prediction 4a) and *transformational proofs* (Prediction 4b) to variants of both the Triangle and Gear tasks. In addition, we predict higher levels of transfer on insights (Prediction 4c) and *transformational proofs* (Prediction 4d) for those who originally received prompts relating the grounding actions to the pre-transfer task.

3. Methods

3.1. Participants

Participants ($N = 120$) were undergraduates (M age $= 19.2$ years; 51% female) enrolled in a psychology course at a large university in the Midwestern US; they received extra course credit for participation. The average self-reported ACT Math score was 28.2 (SD = 4.2, 91st percentile), and the average SAT Math score was 703 (SD = 90, 93rd percentile). In addition, 24.2% of participants had not taken Calculus I, 66.4% had completed up to Calculus I or II, and 9.2% had taken a mathematics class above Calculus II.

3.2. Procedure

Participants took part individually in an experimental session with an interviewer and an interactive whiteboard. Sessions were video-recorded with two cameras—one providing a full shot of the whiteboard and the other a close-up view of the participant's face and gestures. Five different interviewers conducted the sessions, each utilizing a script for all interactions with participants. There were no significant effects involving interviewers, so we collapsed across interviewer in the subsequent analyses. The instructions at the beginning of the session were delivered by playing audio files to ensure consistency (see Appendix A in the supplemental materials). The interviewer's position and gestures were also standardized.

Each participant engaged in two activities, Gear and Triangle, with order counterbalanced. Each activity further included the original task and a transfer task. To allow us to make causal claims about the effects of grounding actions, and to rule out effects of actions in general (such as increased arousal or engagement), we utilized experimental conditions that involved grounding actions and control conditions that involved non-grounding actions. Participants were randomly assigned to one of three groups (see Fig. 1). In Group A, 40 participants (first column, Fig. 1) performed grounding actions related to one of the tasks with no pedagogical prompts (*grounding actions/no prompt*). They then were asked to complete Task 1 (i.e., either Triangle or Gear) while thinking aloud, and to provide justifications. After completing this initial task, Group A participants performed non-grounding actions associated with the other task, and then completed Task 2 with no pedagogical prompts, thinking aloud and providing justifications (*non-grounding actions/no prompt*). In Group B (second column Fig. 1), 40 participants performed the same sequence, except they performed *non-grounding* actions for Task 1 and *grounding* actions for Task 2. In Group C (third column Fig. 1), 40 participants not only performed grounding actions for *both* tasks, but also received pedagogical prompts for both Tasks 1 and 2 (*grounding actions/prompt*).

All participants performed one set of actions (e.g., those prior to Task 1) from a first-person viewpoint, and the other set of actions (e.g., those prior to Task 2) from a third-person viewpoint. The specific procedures for each of the conditions are described in more detail below.

3.2.1. Proof and justification tasks

Participants were asked to generate concurrent think aloud reports while they provided justifications so we could document each participant's reasoning and learning (Ericsson & Simon, 1993, see Table 1). Participants stood in front of the interactive whiteboard while engaging in their justifications, and they were not permitted to use paper or pencil. The tasks in Table 1 were presented as text projected on the whiteboard, and participants were directed to read them aloud.

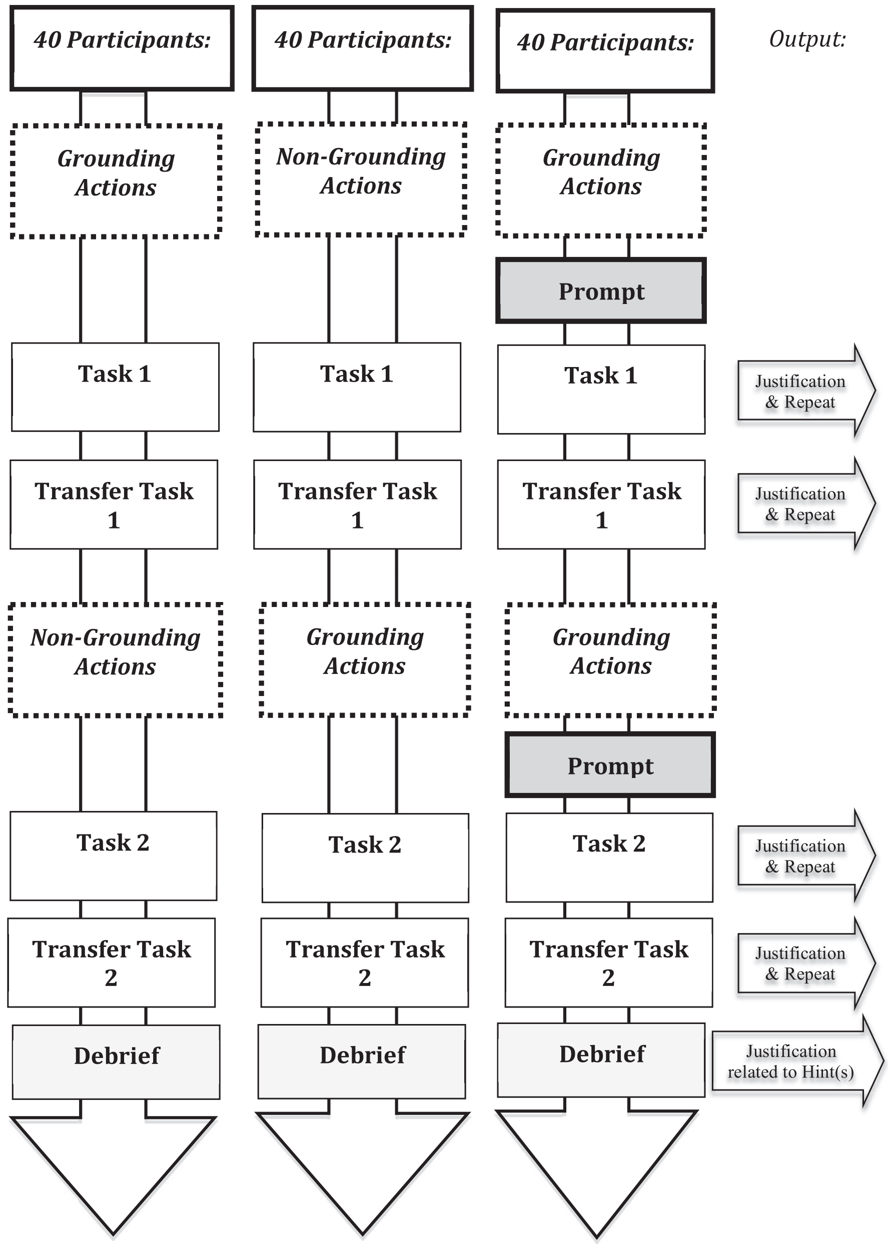


Fig. 1. Three different treatment sequences of tasks used for each condition (40 participants each).

3.2.2. Grounding versus non-grounding actions

Participants were directed to perform body-based actions that were either relevant to the task (*grounding actions*) or irrelevant (*non-grounding actions*). Participants' tapping and touching actions were performed on animated images displayed using the Promethean ActivBoard™ 595, a large (95-inch wide) interactive whiteboard that sequentially guided their specific actions. The images were scaled to each participant's body size to enhance the

Table 1
Experimental tasks.

Task	Original Task Presented/Displayed	Insight	Transformational Proof
Triangle	Mary came up with the following conjecture: For any triangle, the sum of the lengths of any two sides must be greater than the length of the remaining side. Provide a justification as to why Mary's conjecture is true or false. (*referred to as "Triangle Task"*)	Conjecture is true	True because if 2 sides were shorter than the remaining side, 2 sides together would not be able to reach to connect to the two endpoints of the remaining side (for variations on this proof, see Appendix B)
Gear	An unknown number of gears is connected in a chain. You know what direction the first gear turns, how could you figure out what direction the last gear turns? Provide a justification as to why your answer is true. (*referred to as "Gear Task"*)	Gears alternate in the direction they turn	If there is an even number of gears in the chain, the first gear will turn in the opposite direction as the last gear. If there are an odd number of gears, first and last will turn in same direction.

embodied nature of the activity. Although the images differed between the Triangle and Gear tasks, the images were identical across grounding and non-grounding conditions for each task. Participants in the grounding and non-grounding conditions were directed to perform comparable movements; however, grounding actions corresponded to the conceptual relations in the transformational proof for the associated task, whereas this was not the case for non-grounding actions.

For the grounding-action condition for the Triangle task (Fig. 2, left side), participants who performed first-person viewpoint actions touched symmetrically positioned colored dots displayed on the board (scaled to each individual's arm span) simultaneously with both hands. Participants touched a series of five dot pairs that were located progressively farther apart, eventually extending beyond each participant's reach. The two final dot pairs were customized so that each participant's chest came into contact with the whiteboard as they attempted but failed to touch the dots. Thus, participants tacitly made a series of triangles with their arms and the board, which eventually degenerated into parallel lines when the dots exceeded participants' reach, and the distance from the whiteboard to the participant's chest (the altitude of the body-based "triangle") was reduced to zero. These actions embody the key idea that when the third side of a triangle grows to the point where it is very long, the other two sides will no longer be able to connect. In the non-grounding condition, participants touched the identical dots individually with each hand, and thus did not form body-based triangles. Grounding and non-grounding actions from a third-person viewpoint were similar (Fig. 3), except that participants used their fingers (rather than their arms) to form possible and impossible triangles on a row of dots scaled to their hand length in the grounding condition, and simply tapped these dots with their hand in the non-grounding condition.

For the grounding-action condition for the Gear task (Fig. 2, right side), participants who performed first-person viewpoint actions used their dominant palm to alternate tapping the left-hand and right-hand diamonds positioned an arms-length apart. These actions embodied the key idea that linked gears in a row *alternate* in the direction they turn (e.g., right, left, right, left), with all odd-numbered gears turning in one direction and all even-numbered gears turning in the opposite direction. Because of the distance between the two dots, participants in this condition shifted their entire bodies as they tapped back-and-forth, "becoming" the chain of alternating gears. In the non-grounding condition, participants repeatedly touched one of the diamonds, rather than alternating between two diamonds. Grounding and non-grounding actions from a third-person viewpoint were similar (Fig. 3), except that participants used their index fingers to tap smaller diamonds that were a hand length apart. Thus, in the grounding-action condition, participants observed their hand in front of them alternating back-and-forth. Although action viewpoint was varied within subjects, this factor was not associated with performance differences. Consequently, we collapsed the data across viewpoints, and we do not discuss this factor further in this article.

3.2.3. Participant justifications

After completing the grounding or non-grounding actions, participants were presented with the appropriate task and asked to read it aloud and to provide a justification (Fig. 1). They were then asked right away to repeat their justification a second time to the interviewer, in an explicit attempt to better elicit the persuasion phase of the proof. When considering the accuracy of their performance, the two justifications were combined in order to capture a complete account of the participant's reasoning.

3.2.4. Transfer tasks

After the second justification of each original task, each participant was given a transfer task to assess learning (see Table 2). Following the original triangle task, participants received a second, similar conjecture about quadrilaterals. Following the original gear task, participants received a second gear task involving 11 gears connected in a circle.

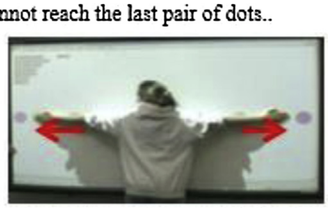
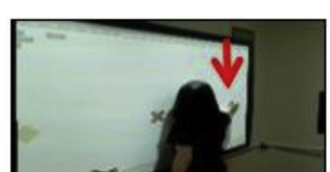
Triangle Actions	**Gear Actions**
Grounding (First-person): Participant places palms on pairs of dots. Dot locations are scaled so a participant cannot reach the last pair of dots..	*Grounding (First-person):* Participant alternates between tapping blue and yellow diamond, which are placed an arm span apart, with palm.
	
Non-Grounding (First-person): Participant taps each dot with left palm moving left to right, then taps each dot with right palm moving right to left.	*Non-Grounding (First-person):* Participant repeatedly taps the blue diamond with palm.
	

Fig. 2. Grounding and non-grounding actions performed for Triangle and Gear tasks in first-person perspective.

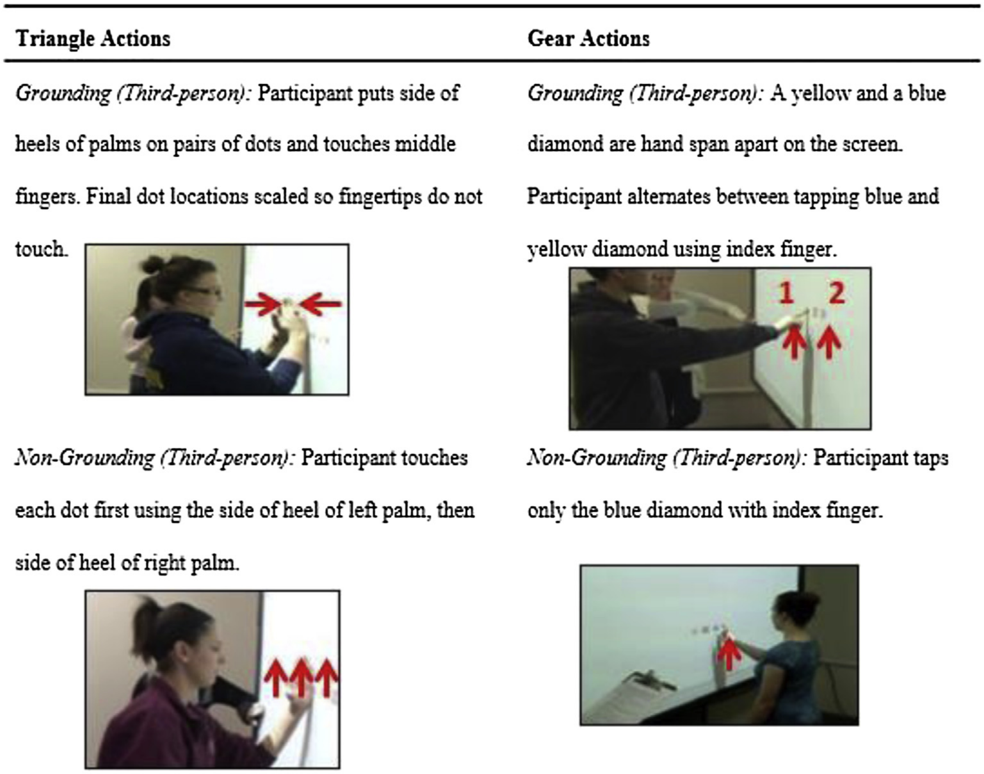


Fig. 3. Grounding and non-grounding actions performed for Triangle and Gear tasks in third-person perspective.

3.2.5. Use of pedagogical language: prospective prompts

In addition to comparing grounding versus non-grounding actions, we varied whether participants received a *prospective prompt* informing them of the relationship between their grounding actions and the task they were about to be presented with, in order to determine whether pedagogical language establishing this connection fostered superior proofs and improved learning (Fig. 1). The pedagogical prompt was identical across all tasks: "The actions you just did on the whiteboard are related to the math problem I'm about to give you. So when you're working through this math problem, think about the actions you just did".

3.2.6. Use of pedagogical language: retrospective hints

After engaging in Tasks 1 and 2, participants were debriefed. For those tasks for which they performed grounding actions (with or without a prompt), participants were informed during the debriefing of that particular task that there was a connection between their actions and the task – that is, they were given a *retrospective hint*. For example, the hint for participants who performed grounding actions from a first-person viewpoint for the Gear task was: "The actions you performed when you were tapping the diamonds with your palm were actually relevant to the solution of the first gear problem. Thinking about the problem and recalling your actions, do you now see the relationship between the problem and the actions you took with your body? Can you explain what the relationship is?" Hints for the other grounding actions had the same format, but the description of the action varied based on which of the grounding actions in Fig. 2 they had performed (see Appendix A in the supplemental materials for all hints). After receiving the hint, participants were given the opportunity to provide a third and final justification for the task(s) for which they had previously performed grounding action (*grounding actions/hint*). Participants who had performed grounding actions for both original tasks (i.e., those in Group C, who also received prospective prompts) received a hint for each in turn, and were given the opportunity to provide a third justification for each task, which they were also asked to repeat a second time. In our analyses, the performance after the hint is only directly compared to each participant's original attempt at solving the same task without a hint.

Table 2
Transfer tasks given to participants after initial triangle/gear tasks.

Task	Prompt	Insight	Transformational Proof
Triangle	Jenny came up with the following conjecture: The sum of the lengths of any three sides of a quadrilateral must be greater than the length of the remaining side. Provide a justification as to why Jenny's conjecture is true or false. (*referred to as "Triangle Transfer task"*)	Conjecture is true	True because if 3 sides were shorter than the remaining side, 3 sides together would not be able to reach to connect to the two endpoints of the remaining side (for variations on this proof, see Appendix B)
Gear	11 gears are connected together in a circle. Can the gears turn? Provide a justification as to why your answer is true. (*referred to as "Gear Transfer task"*)	Gears alternate in their turn direction	Gears will not turn, because first and eleventh gear are odd gears, and odd gears turn in the same direction.

3.3. Data analysis

Videos were clipped into segments of each participant producing a justification for each task. Segments were coded for the proof category observed, adapted from Harel and Sowder (2005) (e.g., perceptual, transformational, etc.; see Appendix B in the supplemental materials for proof types and coding criteria), with only transformational proof types considered valid. Valid transformational proofs are differentiated from invalid proof schemes, including *empirical* proofs in which a conjecture is considered true on the basis of particular examples, *authoritative* proofs in which a conjecture is validated based purely on the word of an authority, and *perceptual* proofs in which a conjecture is validated based on perceptual cues resulting from invalid mathematical transformations (e.g., this angle "looks like" it is a right angle, so it is).

We separately coded whether the justifications were *grounded*, that is, whether participants explicitly connected their justifications to the grounding actions they had performed. Thus, a subset of transformational and perceptual proofs was *grounded transformational* and *grounded perceptual* proofs, respectively. We considered a proof to be grounded if the actions participants produced were explicitly mentioned as part of the argument.

Finally, we coded utterances in participants' transcripts for whether participants recognized the *key insight* behind each task: for the Triangle and Triangle Transfer tasks, this was whether they reported that the conjecture was true, and for the Gear and Gear Transfer tasks, this was whether they acknowledged that adjacent gears would alternate in turning direction.

The primary coder was blind to participants' action/prompt condition. A second analyst coded 20% of the clips across all 4 tasks; inter-rater reliability (Cohen's *kappa*) was 0.85. As a final check, we looked at whether participants in the grounding actions/no prompt condition were aware that their actions related to the tasks. When asked at the end of the experiment, 4 participants (5%) reported making a connection between the directed movements and the Triangle task. All analyses that revealed significant differences between non-grounding and grounding actions were re-run without these participants, and results did not differ.

Data were analyzed using mixed effects logistic regression (Snijders & Bosker, 1999) using the *lmer* command (Bates & Maechler, 2010) in *R*. Model selection was conducted by using likelihood ratio tests (the *ANOVA* function in *R*) on nested models to test for significant reductions in deviance. Predictors were kept in the model only if they significantly reduced the deviance, thereby improving the fit of the model. First, we fit a logistic regression model in which the dependent measure was whether the participant generated the *key mathematical insight* behind each task (coded as 0/1). Second, we fit a logistic regression model in which the dependent measure was whether the participant constructed a *valid transformational proof* for each task (coded as 0/1).

The final models included participant as a random effect, and both task (Triangle, Gear) and action/prompt condition (non-grounding actions/no prompt, grounding actions/no prompt, grounding actions/prompt, grounding actions/hint, grounding actions/prompt and hint) as fixed-effect predictors. The grounding action/hint conditions were included to examine how each participant's performance improved or declined when given a hint and asked to provide an additional justification during debriefing. Demographic data (gender, native language, age, race/ethnicity, math course history, math achievement, etc.) were collected from participants and tested to see if they significantly improved the fit of the model. Math achievement data (ACT/SAT scores) were added as a normalized percentile; these data were missing for 5 participants and were imputed using the participants' math course history; mean imputation and omitting these participants yielded similar results. Standardized mean difference-type (d) effect sizes were calculated by taking the natural log of the odds ratio, and dividing by 1.81, as outlined in Chinn (2000); 95% confidence intervals are provided with effect sizes.

4. Results

We organize our presentation of the results in terms of each of our predictions. We first show descriptive data and outline the significant predictors in the regression models. We then discuss each of our predictions regarding grounding actions without pedagogical language (H1), grounding actions with pedagogical language (H2), generalization across tasks (H3), and transfer task performance (H4).

4.1. Recognizing key mathematical insights

Descriptive data, including means and standard errors, for participants' likelihood of generating the key mathematical insight for the Triangle and Gear tasks are presented in Table 3. In the regression model for generating the key insight, adding a predictor for students' math achievement significantly improved model fit, $\chi^2(1) = 4.37$, $p = 0.037$, as did adding predictors for task, $\chi^2(1) = 35.77$, $p < 0.001$, and for action/prompt condition, $\chi^2(4) = 17.38$, $p = 0.002$. Interactions and other demographic variables did not significantly improve the model fit.

Results of the regression for generating the key mathematical insight are shown in Table 4. The reference category (denoted as "ref.") for condition is performing grounding actions with no prompt; this is the condition to which the two other conditions in Table 4 are compared. The Odds column gives the raw coefficient exponentiated, so it is transformed to an odds ratio. The d column gives the standardized mean difference-type (d) effect sizes (Chinn, 2000) and a 95% confidence interval. Although the odds of having the insight for the Triangle task were larger than the odds of having the insight for the Gear task, this difference was not statistically significant, $\chi^2(4) = 1.26$, $p = 0.87$. Thus in terms of having the insight, participants performed similarly across tasks. Overall, as we will describe in the following sections, we found that performing grounding actions helped with producing the correct insight.

4.2. Constructing a transformational proof

Participants' mean likelihood of constructing a transformational proof for the Triangle and Gear tasks is presented in Table 3. For each task, the overall rate of constructing a transformational proof was somewhat close to 50%, suggesting tasks were neither too difficult nor too easy (Triangle = 50% success, Gear = 40.8%

Table 3
For conditions, mean participant performance (and standard deviation) across gear and triangle tasks.

Task	Condition		
	Non-Grounding	Grounding	
		No Prompt	Prompt
Insight Gear	0.55 (0.50)	0.70 (0.46)	0.68 (0.47)
Insight Triangle	0.83 (0.38)	0.92 (0.26)	0.90 (0.30)
Proof Gear	0.45 (0.50)	0.43 (0.50)	0.35 (0.48)
Proof Triangle	0.40 (0.50)	0.50 (0.51)	0.60 (0.50)
Transfer			
Insight Gear	0.53 (0.51)	0.63 (0.50)	0.50 (0.51)
Insight Triangle	0.73 (0.45)	0.95 (0.22)	0.88 (0.33)
Proof Gear	0.43 (0.50)	0.55 (0.50)	0.35 (0.48)
Proof Triangle	0.45 (0.50)	0.58 (0.50)	0.55 (0.50)

Table 4
Results for generating key mathematical insight, triangle and gear tasks.

Fixed effect	B	SE (B)	Odds	d (95% CI)	z	p	Sig
(Intercept)	1.36	0.42	3.90	0.75 (0.30, 1.21)	3.28	0.00104	**
Gear Task	(ref.)						
Triangle Task	2.10	0.37	8.18	1.16 (0.75, 1.57)	5.63	<0.001	***
Math Achievement (Normalized)	0.40	0.20	1.49	0.22 (0.0025, 0.44)	2.00	0.04525	*
Condition: Grounding, No Prompt	(ref.)						
Condition: Non-Grounding, No Prompt	−1.12	0.48	0.33	−0.62 (−1.14, −0.093)	−2.33	0.01969	*
Condition: Grounding, Prompt	−0.30	0.59	0.74	−0.16 (−0.81, 0.48)	−0.51	0.61105	

Note. '*' = $p < 0.05$, '**' = $p < 0.01$, and '***' = $p < 0.001$.

success). Note that participants had to express a complex idea involving logical inference in order to verbalize a valid proof; these values do not reflect "chance" odds of choosing a correct versus incorrect answer. Patterns differed across tasks. In the regression model for constructing a valid proof, students' math achievement significantly improved model fit, $\chi^2(1) = 40.34$, $p < 0.001$, as did adding predictors specifying which task was being proved, $\chi^2(1) = 29.39$, $p < 0.001$, and the interaction of condition and task, $\chi^2(8) = 25.301$, $p = 0.0014$. No other interactions or demographic variables were significant. Since the task-by-condition interaction was significant, we present regression results for the Triangle and Gear tasks separately. The significance of this interaction term means that the effect of action/prompt condition varied according to which task (Gear or Triangle) the participant was working on. As described in the following sections, we found that grounding actions did not lead to significant gains in production of transformational proofs *unless* accompanied by pedagogical language, and this effect was modulated by task demands.

Regression results for constructing a valid proof for the Triangle and Gear tasks are shown in Tables 5 and 6, respectively. For both tasks, prior math achievement was associated with better performance constructing valid proofs.

4.3. Hypothesis 1: Exploring the influence of grounding actions

We first test Hypothesis 1, which posits that grounding actions will lead to enhanced performance on insights (Prediction 1a) and transformational proofs (Prediction 1b). Table 4 shows that across the Gear and Triangle tasks, participants in the non-grounding-action condition were less likely to generate the key mathematical insight ($d = -0.62$, $p = 0.02$) than participants in the grounding-action condition with no prompt, supporting Prediction 1a. Table 5 shows that for the Triangle task, participants were not more likely to construct a valid proof if they performed grounding actions with no prompt, compared to non-grounding actions ($p = 0.94$), which does not support Prediction 1b. Table 6 shows that this was also true for the Gear task ($p = 0.57$), which again does not support Prediction 1b. Thus although grounding actions without prompts facilitated participants expressing key mathematical insights, they did not influence the production of valid proofs. In partial support of Hypothesis 1, we conclude that grounding actions can foster key mathematical insights for these tasks, even without conscious awareness of their relevance. However, grounding actions alone do not enhance transformational proof production.

4.4. Hypothesis 2: Exploring the combined influence of pedagogical language and action

We next investigate the evidence for Hypothesis 2, which states that pedagogical language alerting participants to the relevance of the grounding actions will improve insights (Prediction 2a) and transformational proof production (Prediction 2b).

Table 4 shows that across both tasks, receiving a prompt did not significantly affect whether participants generated the key mathematical insight ($p = 0.61$), compared to performing grounding action without a prompt. This does not support Prediction 2a. We also looked at the debriefing portion of the interview to see if receiving the retrospective hint further facilitated the insight. For participants who had not received a prompt initially, the hint did not further enhance their likelihood of generating the insight, compared to their first attempt ($p = 0.44$), which also does not support Prediction 2a. However, for participants who did initially

Table 5
Results for generating transformational proof to triangle task.

Fixed effect	B	SE (B)	Odds	d (95% CI)	z	p	Sig
(Intercept)	−0.53	0.55	0.59	−0.29 (−0.89, 0.31)	−0.97	0.33374	
Math Achievement (Normalized)	1.95	0.36	7.04	1.08 (0.68, 1.48)	5.37	<0.001	***
Condition: Grounding, No Prompt	(ref.)						
Condition: Non-Grounding, No Prompt	−0.06	0.79	0.94	−0.03 (−0.90, 0.83)	−0.08	0.93710	
Condition: Grounding, Prompt	1.32	0.80	3.76	0.73 (−0.15, 1.61)	1.65	0.09930	

Note. '*' = $p < 0.05$, '**' = $p < 0.01$, and '***' = $p < 0.001$.

Table 6
Results for generating transformational proof to gear task.

Fixed effect	B	SE (B)	Odds	d (95% CI)	z	p	Sig
(Intercept)	−0.44	0.57	0.64	−0.24 (−0.87, 0.38)	−0.77	0.44105	
Math Achievement (Normalized)	1.94	0.36	6.96	1.07 (0.68, 1.47)	5.39	<0.001	***
Condition: Grounding, No Prompt	(ref.)						
Condition: Non-Grounding, No Prompt	−0.45	0.79	0.64	−0.25 (−1.11, 0.61)	−0.57	0.56887	
Condition: Grounding, Prompt	−0.91	0.82	0.40	−0.50 (−1.39, 0.39)	−1.11	0.26644	

Note. '*' = $p < 0.05$, '**' = $p < 0.01$, and '***' = $p < 0.001$.

receive the prospective prompt to connect their actions to the task (i.e., participants in Group C), receiving the hint at the end of the session was associated with a significant increase in generating the key insight, compared to their first attempt ($d = 0.70$, $p = 0.024$), supporting Prediction 2a. Thus, adding *both* a prompt and a hint helped participants generate the key insight behind both the Triangle and Gear tasks.

Since we observed differential effects of action/prompt condition on participants' constructing proofs on the Triangle and Gear tasks, we report the findings separately for each task. The regression results for constructing a valid proof for the Triangle task are presented in Table 5. There was a trend suggesting that participants were more likely to generate valid proofs if they performed grounding actions and received a prompt, compared to performing grounding actions and receiving no prompt ($d = 0.73$, $p = 0.099$). This marginally significant effect provides weak support for Prediction 2b. Among those participants who performed grounding actions but did not receive a prompt, when they received a hint during debriefing, they were more likely to construct a valid proof ($d = 0.87$, $p = 0.0099$) than they were on their initial attempt. This supports Prediction 2b. For participants who had already received a prompt, receiving the hint during debriefing did not significantly increase the likelihood that they would construct a valid proof, relative to their first attempt ($p = 0.17$). For the Triangle task there was partial support for Hypothesis 2, which states that grounding actions coupled with language-based hints or prompts facilitated insights and valid proof production.

For the Gear task (Table 6), participants who performed grounding actions and received a prompt were not significantly more likely to construct valid proofs than were participants who performed grounding actions with no prompt ($p = 0.27$), which does not support Prediction 2b. Among those participants who performed grounding actions but did not receive a prompt, when they received a hint during debriefing, they were significantly *less* likely to construct a valid proof ($d = -0.88$, $p = 0.02$), compared to their first attempt. This directly contradicts Prediction 2b. For participants who had already received a prompt, receiving the hint during debriefing did not affect the likelihood that they would construct a valid proof, relative to their first attempt ($p = 0.32$), which also does not support Prediction 2b. Thus, for the Gear task, there was no support for Hypothesis 2. Although pedagogical language may have a weak effect on having the insight, it did not improve proof production.

4.5. Hypothesis 3: Generalizability across tasks

Although the regression tables for constructing valid proofs for the Gear and Triangle tasks were separated for interpretability, they were analyzed in a single model. The model provides a test of Hypothesis 3, which regards task differences, with Prediction 3 stating that both tasks will exhibit similar trends. The model showed that receiving a hint was significantly less effective at fostering participants' constructing valid proofs for the Gear task than for the Triangle task ($d = -1.71$, $p < 0.001$), as was receiving a prompt ($d = -1.23$, $p = 0.001$), and receiving both a hint and a prompt ($d = -2.11$, $p < 0.001$). Grounding actions with hints and/or prompts were effective for promoting proof construction for the Triangle task, but neutral or damaging to performance on the Gear task, in contradiction to Prediction 3.

4.6. Post-hoc investigation: coding of proof categories

Why were grounding actions with prompts or hints beneficial for constructing proofs for the Triangle task, but harmful for the Gear task? To explore this emergent question, we conducted a post-hoc investigation of the structure of each participant's mathematical argument, which we present prior to addressing Hypotheses 4 (transfer). Specifically, we examined how participants' arguments changed as a result of receiving a hint and making a second attempt.

Table 7 (columns 2 and 3) presents data on the 80 participants who performed grounding actions for one task, received no prompt initially, and then received a hint during the debriefing. For the Triangle task, 67.5% of participants did not alter the validity of their proofs on the second attempt following the hint, whereas for the Gear task, 82.5% of participants did not alter the validity of their proof. Receiving the hint during the debriefing *increased* the overall incidence of valid proofs on the Triangle task by 17.5% (25−7.5%). However, receiving a hint during the debriefing *decreased* the overall incidence of valid proofs on the Gear task by the same amount, 17.5%. The net gains/losses were in the same direction but smaller for the remaining 40 participants who initially received a prompt and performed grounding actions for both tasks (Table 7, columns 4 and 5). We next discuss the trends for each of the tasks in turn.

Looking at individual proof codes for the Triangle task (see Appendix B in the supplemental materials for the proof coding categories and criteria), the net gain in valid proofs seemed to occur because participants sometimes modified an invalid empirical proof or perceptual proof to be a valid, grounded transformational proof after receiving the hint. For example, one participant who initially formed an invalid proof said after the hint that: “Like at this point, they can be connected, and this point they can be connected, but then the further out I went, it wouldn't be able to connect. They're- if the two sides were longer than the base, then the base wouldn't be able to connect to the sides”. Overall, participants often, but not always, were able to draw on body-based actions to construct a valid grounded transformational proof for the Triangle task when given a prompt or hint, even when they failed to do so initially.

For the Gear task, receiving a hint was often effective for getting the participant to report the insight that acknowledges alternation (12 participants changed from a no-alternation solution to an alternation solution on their second attempt). However, performing grounding actions and receiving a hint did not support valid proof construction. The net loss in valid proofs from the first attempt to the second attempt with the hint occurred because some participants modified a valid, transformational proof to be an invalid, grounded perceptual proof following the hint. These perceptual proofs were grounded in the sense that the mathematical argument remained explicitly linked to the actions they had been directed to perform. Following the hint, some participants ($N = 14$) moved their hands in a circle as they tapped in an alternating fashion, which depicted all the gears in the chain turning the same direction. This would have made it difficult for students to form the proper abstraction needed for a successful proof (e.g.,

Table 7
Number (and %) of participants who displayed each pattern of proof validity across the two proof attempts (initial attempt and second attempt with hint). $N = 40$ Participants per task.

Proof Validity	Triangle (no prompt)	Gear (no prompt)	Triangle (prompt)	Gear (prompt)
Invalid → Invalid	9 (22.5%)	23 (57.5%)	12 (30%)	25 (62.5%)
Valid → Valid	18 (45%)	10 (25%)	24 (60%)	10 (25%)
Valid → Invalid	3 (7.5%)	7 (17.5%)	0 (0%)	4 (10%)
Invalid → Valid	10 (25%)	0 (0%)	4 (10%)	1 (2.5%)
Net increase in valid proofs	17.5%	−17.5%	10%	−7.5%

Note. “Valid” proofs are transformational proofs.

Schwartz & Black, 1999). In other words, the participants were overly focused on the immediate, perceptual characteristics of their movements (motion from side to side), rather than the abstract *idea* the actions embodied (i.e., parity).

In the other cases in which participants shifted from valid to invalid reasoning, the participants seemed to partially catch on to the connection between their actions and the Gear task, but their grounded proof was incomplete. For example, one participant who constructed a valid, non-grounded proof on his first attempt to the Gear task gave the following report upon receiving a hint, "'Cause like you have to use the same palm, so it's more like this and then the opposite direction, right? That's it, something like that. That's my justification, I guess". Overall, there was little evidence that the gear tapping actions supported proof construction, with or without hints and prompts.

4.7. Hypothesis 4: Transfer

Each original task had a matched transfer task (Table 2) that tested whether participants could generate insights and construct proofs in a new context. To investigate Hypothesis 4, we investigated whether grounding actions enhanced insight and proof on the transfer tasks (Predictions 4a–4b) and whether pedagogical language enhanced insight and proof on the transfer tasks (Predictions 4c–4d). For these analyses, each participant's action/prompt condition corresponded to the condition they had been assigned to for the original Triangle or Gear task that corresponded to each transfer task. These conditions were *non-grounding action*, *grounding action with no prompt*, or *grounding action with a prompt* (see Table 3).

In models predicting whether participants generated the key insight for each transfer task, task (triangle transfer or gear transfer, $\chi^2(1) = 27.18$, $p < 0.001$) and the participant's original action/prompt condition ($\chi^2(2) = 6.64$, $p = 0.036$) significantly improved the fit of the model, but their interaction was not significant ($\chi^2(2) = 3.49$, $p = 0.17$). The odds of generating the key insight were higher for the triangle transfer task than for the gear transfer task (odds = 5.84). Participants who performed grounding actions with no prompt were significantly more likely to generate the key insight on the transfer tasks ($d = 0.57$, $p = 0.01$), compared to participants who performed non-grounding actions, supporting Prediction 4a. However, participants who performed grounding actions and who had also received a prompt were not more likely to have the insight than participants who performed grounding actions with no prompt ($p = 0.12$), failing to support Prediction 4c.

In models predicting whether participants constructed transformational proofs for transfer tasks, neither task ($\chi^2(1) = 1.76$, $p = 0.18$), nor the participant's original action/prompt condition ($\chi^2(2) = 3.15$, $p = 0.21$), nor their interaction ($\chi^2(5) = 7.23$, $p = 0.20$) significantly improved the fit of the model. Thus, there was no evidence that performing grounding actions or receiving pedagogical language supported participants in constructing transformational proofs in new proof contexts, which does not support Predictions 4b or 4d. Overall, the results provide mixed support for Hypothesis 4: Grounding actions enabled transfer of the *insight* aspects of proof practices (Prediction 4a), but not of constructing *transformational proofs* (Prediction 4b). The null results on pedagogical language also do not support Predictions 4c or 4d.

5. Discussion

To summarize our findings, grounding actions were beneficial for generating mathematical insights, as shown in both original and transfer tasks, even without participants' conscious awareness of their relevance to the tasks (Predictions 1a and 4a). These benefits for insight were even greater when accompanied by pedagogical language in the form of prompts and hints (Prediction 2a) that signal, respectively, their prospective or retrospective relevance. We explore the implications of actions for fostering insight later in this section.

We observed mixed results when investigating the influence of actions on participants' construction of transformational proofs. Although grounding actions on their own incurred no advantage for constructing transformational proofs, grounding actions showed significant advantages for the Triangle task when accompanied by retrospective hints, and marginal benefits with prospective prompts (Prediction 2b). Yet actions with hints proved to be harmful for transformational proof production with the Gear task (counter to Prediction 3). Post-hoc analyses of the content of participants' justifications showed that after receiving the hint for the Gear task, several participants modified a valid, transformational proof to be an invalid or incomplete grounded perceptual proof, while for the Triangle task empirical or perceptual proofs were often revised to be grounded transformational. Finally, grounding actions did not promote participants' constructing transformational proofs on transfer tasks, which did not support Prediction 4b. Prompts also did not further promote participants' generating key insights or constructing transformational proofs on the transfer tasks, which does not support Predictions 4c and 4d.

5.1. Limitations of the current study

This study has several limitations that require us to qualify our conclusions. The study was conducted using a limited set of tasks, and on a specific population—college students enrolled in a psychology course. Expanding the tasks and participant pool will be valuable for understanding how generalizable the findings are.

We also used a limited number of conditions to control for various treatment influences (see Fig. 1). In the condition that experienced prompting, participants were exposed to two grounding-action treatments (one for each task), while participants in the other two conditions received only one grounding-action treatment along with one non-grounding-action treatment. Although each of the treatment groups participated in actions for each task, and the actions were carefully matched across control and treatment groups for physical engagement, there is a possibility this may have somehow biased participant performance. We also must be careful not to over-ascribe influences of the retrospective hints on performance, since these were always given alongside an additional chance to provide a justification. Some participants regressed on this second justification and it is possible that participants may have believed that we were requesting a second justification because their original justification was not appropriate. However, given that participants only regressed on the Gear task on their second attempt, and improved on the Triangle task, a systematic influence of this belief seems unlikely. Finally, participants were only able to provide appropriate responses to either task to the extent to which they understood what a "justification" or proof should entail – in other words, they may have been able to communicate a valid transformational proof, but may have determined that a different type of argument (like an example) was more convincing (e.g., Chazan, 1993).

5.2. Action, language, and cognition

In light of these findings and the limitations of the study, we consider the implications of this work on our emerging understanding of the role of the body in cognitive processing. One of the central tenets of GEC is that "the body and world [are] central constituents in cognitive processes" (Shapiro, 2010, p. 161), often in the form of situated actions (Barsalou, 2008). Our results show that actions can facilitate mathematical insights, which corroborates past research on relationships between actions and mental states (e.g., Thomas & Lleras, 2007). The current findings extend this work to the domain of mathematical reasoning. Furthermore, the experimental design we used allows us to conclude that it is *specific* actions—those we deemed *grounding actions*—that cause these benefits, rather than performing actions more generally. Even so, future research in this area should investigate the influence of action on reasoning and proving using many types of actions.

Emerging GEC models posit that actions can induce cognitive states as well as follow from cognition, suggesting a reciprocal relationship between cognition and action (Nathan, 2014). For example, teaching specific movements to children fosters their learning of a grouping procedure to solve mathematical equivalence problems (Goldin-Meadow, Cook, & Mitchell, 2009). Yet, without ostensive cues to direct one's attention to the relevant referents, the information provided by actions may sometimes be too vague to generalize or effectively apply to future tasks (Csibra & Gergely, 2009). The abstract nature of the action and its relationship to its referent may affect the utility of the action, especially if the action is not explicitly connected to the task. Thus, it is theoretically important to note that the inclusion of prospective and retrospective verbal cues designed to alert participants to the relevance of the actions to the task moderated this relationship by further enhancing the influence of grounding actions on insight, evidenced by their improving proof production for the Triangle task, while harming proof production for the Gear task. This underscores the important grounding role for cognition that is served by language—especially pedagogical language aimed at directing one's attention to relevant aspects of curricular activities.

One instructional implication is that actions can foster cognitive gains when the relevance of the actions is explicitly signaled for learners (Nathan, Walkington, Srisurichan, & Alibali, 2011; Nathan et al., 2013). In current work (Pier et al.,2014), we are also extending this investigation by examining the complex relationships among action, language, and successful proof practices by analyzing the types of speech and gestures that people use while constructing mathematical arguments.

5.3. Task demands

Grounding actions in conjunction with prompts and hints fostered proof production for the Triangle task, but harmed performance on the Gear task. We speculate that this is due to the abstract nature of the parity conjecture necessary for the Gear proof, and how this conjecture aligned with the grounding actions in our study. Other scholars (Boncoddo, Dixon, & Kelley, 2010; Schwartz & Black, 1999) have reported that people solving gear problems often follow a progression toward formalization. People initially rotate their hands to model gear motion concretely, then move to back-and-forth tapping gestures, and eventually discover the parity rule, which they abstract as alternation. By directing participants to perform the alternating tapping motions at the outset, thus placing participants in the middle of this formalization progression, we may have hindered participants' progression toward the parity rule. One instructional implication is that grounding actions may be most effective when the underlying mathematical ideas being communicated are concrete, and when the mathematical concepts align with the physical and spatial relations, as with the Triangle conjecture. In order to extend the influence of actions on cognition, concreteness fading (Goldstone & Son, 2005; McNeil & Fyfe, 2012; Nathan, 2012) may be necessary to support the development of more abstract relationships. The current results suggest it may be especially fruitful to compare the emergence of abstract versus concrete (i.e., spatiotemporal) relations, controlling for context (e.g., gears). Future investigations should explore the effect of actions on a variety of forms of mathematical reasoning across a range of contexts (cf. Walkington et al., 2014).

6. Conclusions

We conclude that grounding actions can indeed support mathematical reasoning, and that the cognition-action link is further moderated by language that directs learner attention to the applicability of the actions to the task at hand. Emerging theories of grounded and embodied cognition should look to both action and language as systems for supporting cognition in complex domains such as mathematical proof production. Curriculum programs can look to capitalize on motion and action for promoting student insight, but are likely to depend upon ostensive cues to signal for students the relevance of these bodily actions for complex reasoning.

Acknowledgments

We extend our heartfelt appreciation to Dan Klopp and Brian Ganje for their technical contributions to this project. This work was made possible through a grant from the National Science Foundation REESE Program (DRL-0816406), funding from the U.S. Department of Education Institute of Education Sciences (Award No. R305B100007) Postdoctoral Training Program in Mathematical Thinking, Learning, and Instruction, and the support of the Wisconsin Center for Education Research.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.learninstruc.2014.07.001.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: evidence from learners' and teachers' gestures. *Journal of the Learning Sciences, 21*(2), 247–286. http://dx.doi.org/10.1080/10508406.2011.611446.

Alibali, M. W., Nathan, M. J., Church, R. B., Wolfgram, M. S., Kim, S., & Knuth, E. (2013). Teachers' gestures and speech in mathematics lessons: forging common ground by resolving trouble spots. *ZDM — The International Journal on Mathematics Education, 45*, 425–440. http://dx.doi.org/10.1007/s11858-012-0476-0.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207. http://dx.doi.org/10.1207/s15327809jls0402_2.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed effects models using S4 classes. R package version 0.999375-35*. http://CRAN.R-project.org/package=lme4.

Boncoddo, R., Dixon, J. A., & Kelley, E. (2010). The emergence of a novel representation from action: evidence from preschoolers. *Developmental Science, 13*, 370–377. http://dx.doi.org/10.1111/j.1467-7687.2009.00905.x.

Carpenter, T., & Moser, J. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*(3), 179–202. http://dx.doi.org/10.2307/748348.

Carroll, J. B. (1968). On learning from being told. *Educational Psychologist, 5*, 4–10. http://dx.doi.org/10.1080/00461526809528955.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics, 24*, 359–387. http://dx.doi.org/10.1007/BF01273371.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19*(22), 3127–3131. http://dx.doi.org/10.1002/1097-0258(20001130)19:223.3.CO;2-D.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. http://dx.doi.org/10.1016/j.tics.2009.01.005.

Common Core State Standards Initiative. (2010). Common core state standards (mathematics standards). http://www.corestandards.org/the-standards/mathematics.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*, 371–396. http://dx.doi.org/10.1037/0096-3445.122.3.371.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT.

Gerofsky, S. (2010). Mathematical learning and gesture: character viewpoint and observer viewpoint in students' gestured graphs of functions. *Gesture, 10*(2–3), 321–343. http://dx.doi.org/10.1075/gest.10.2-3.10ger.

Gick, M., & Holyoak, K. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355. http://dx.doi.org/10.1016/0010-0285(80)90013-4.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: the case of gesture. *Perspectives in Psychological Science, 5*(6), 664–674. http://dx.doi.org/10.1177/1745691610388764.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*(3), 267–272.

Goldstone, R., & Son, J. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences, 14*(1), 69–110. http://dx.doi.org/10.1207/s15327809jls1401_4.

Harel, G., & Sowder, L. (2005). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*. NCTM.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4), 396–428. http://dx.doi.org/10.2307/749651.

Hersh, R. (1997). *What is mathematics, really?* USA: Oxford University Press.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*(1), 54–65. http://dx.doi.org/10.1016/j.learninstruc.2006.10.008.

Hoyles, C., & Healy, L. (2007). Curriculum change and geometrical reasoning. In P. Boero (Ed.), *Theorems in school: From history, epistemology and cognition to classroom practice* (pp. 81–115). Rotterdam: Sense.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education, 33*, 379–405. http://dx.doi.org/10.2307/4149959.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22–36. http://dx.doi.org/10.1016/j.learninstruc.2014.01.003.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction, 22*, 440–448. http://dx.doi.org/10.1016/j.learninstruc.2012.05.001.

Miyazaki, M. (2000). Levels of proof in lower secondary school mathematics. *Educational Studies in Mathematics, 41*, 47–68. http://dx.doi.org/10.1023/A:1003956532587.

Moyer, R., & Landauer, T. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519–1520. http://dx.doi.org/10.1038/2151519a0.

Nathan, M. J. (2008). An embodied cognition perspective on symbols, gesture, and grounding instruction. In M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment and meaning*. Cambridge: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199217274.003.0018.

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist, 47*(2), 125–148. http://dx.doi.org/10.1080/00461520.2012.667063.

Nathan, M. J. (2014). Grounded mathematical reasoning. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 171–183). New York: Routledge.

Nathan, M. J., Srisurichan, R., Walkington, C., Wolfgram, M., Williams, C., & Alibali, M. W. (2013). Building cohesion across representations: a mechanism for STEM integration. *Journal of Engineering Education, 102*(1), 77–116.

Nathan, M. J., Walkington, C., Srisurichan, R., & Alibali, M. W. (2011). Modal engagements in pre-college engineering: tracking math and science concepts across symbols, sketches, software, silicon, and wood. In *Proceedings of the 118th American society of engineering education annual conference and exposition*. Vancouver, CA.

Núñez, R. (2009). Numbers and arithmetic: neither hardwired nor out there. *Biological Theory, 4*(1), 68–83. http://dx.doi.org/10.1162/biot.2009.4.1.68.

Parrill, F. (2011). The relation between the encoding of motion event information and viewpoint in English-accompanying gestures. *Gesture, 11*(1). http://dx.doi.org/10.1075/gest.11.1.04par.

Pier, E. L., Walkington, C., Williams, C., Boncoddo, R., Alibali, M. W., Nathan, M. J., et al. (2014). Hear what they say and watch what they do: predicting valid mathematical proofs using speech and gesture. In W. Penuel, S. A. Jurow, & K. O'Connor (Eds.), *Learning and becoming in practice: Proceedings of the eleventh international conference of the learning sciences* (pp. 649–656). Boulder, CO: University of Colorado.

Rips, L., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*, 623–687. http://dx.doi.org/10.1017/S0140525X08005566.

Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15. http://dx.doi.org/10.1111/j.1467-8624.2006.00852.x.

Romberg, T. A., & Kaput, J. J. (1999). Mathematics worth teaching, mathematics worth understanding. In E. Fennema, & T. A. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 3–17). Mahwah, NJ: Lawrence Erlbaum.

Roth, W. M. (2001). Gestures: their role in teaching and learning. *Review of Educational Research, 71*(3), 365–392. http://dx.doi.org/10.3102/00346543071003365.

Schoenfeld, A. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior, 13*(1), 55–80. http://dx.doi.org/10.1016/0732-3123(94)90035-3.

Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 116–136. http://dx.doi.org/10.1037/0278-7393.25.1.116.

Shapiro, L. (2010). *Embodied cognition*. Routledge Press.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Sage.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition, 78*, 45–88. http://dx.doi.org/10.1016/S0010-0277(00)00108-6.

Stylianides, G. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics, 28*(1), 9–16.

Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (Eds.). (2009). *Teaching and learning proof across the grades: A K-16 perspective*. Routledge.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin and Review, 14*, 663–668. http://dx.doi.org/10.3758/BF03196818.

Vygotsky, L. S. (1986). *Thought and language (Revised edition)*. Cambridge: MIT Press.

Walkington, C., Boncoddo, R., Williams, C., Nathan, M. J., Alibali, M. W., Simon, E., et al. (2014). Being mathematical relations: dynamic gestures support mathematical reasoning. In W. Penuel, S. A. Jurow, & K. O'Connor (Eds.), *Learning and becoming in practice: Proceedings of the eleventh international conference of the learning sciences* (pp. 479–486). Boulder, CO: University of Colorado.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*, 625–636. http://dx.doi.org/10.3758/BF03196322.