Journal of Educational Psychology

Embodied Geometric Reasoning: Dynamic Gestures During Intuition, Insight, and Proof

Mitchell J. Nathan, Kelsey E. Schenck, Rebecca Vinsonhaler, Joseph E. Michaelis, Michael I. Swart, and Candace Walkington

Online First Publication, October 26, 2020. http://dx.doi.org/10.1037/edu0000638

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

http://dx.doi.org/10.1037/edu0000638

Embodied Geometric Reasoning: Dynamic Gestures During Intuition, Insight, and Proof

Mitchell J. Nathan and Kelsey E. Schenck
University of Wisconsin–Madison

Rebecca Vinsonhaler
University of Texas–Austin

Joseph E. Michaelis
University of Illinois–Chicago

Michael I. Swart
University of Wisconsin–Madison

Candace Walkington
Southern Methodist University

Grounded and embodied cognition (GEC) serves as a framework to investigate mathematical reasoning for proof (reasoning that is logical, operative, and general), insight (gist), and intuition (snap judgment). Geometry is the branch of mathematics concerned with generalizable properties of shape and space. Mathematics experts ($N = 46$) and nonexperts ($N = 44$) were asked to judge the truth and to justify their judgments for four geometry conjectures. Videotaped interviews were transcribed and coded for occurrences of gestures and speech during the proof production process. Analyses provide empirical support for claims that geometry proof production is an embodied activity, even when controlling for math expertise, language use, and spatial ability. *Dynamic depictive gestures* portray generalizable properties of shape and space through enactment of transformational operations (e.g., dilation, skewing). Occurrence of dynamic depictive gestures and nondynamic depictive gestures are associated with proof performance, insight, and intuition, as hypothesized, over and above contributions of spoken language. Geometry knowledge for proof may be embodied and accessed and revealed through actions and the transformational speech utterances describing these actions. These findings have implications for instruction, assessment of embodied knowledge, and the design of educational technology to facilitate mathematical reasoning by promoting and tracking dynamic gesture production and transformational speech.

Educational Impact and Implications Statement

How do mathematical intuitions arise, and how can they help with advanced forms of reasoning such as geometry proofs? One idea is that intuitions arise from body movements that allow people to directly experience mathematical ideas and relationships. We analyzed videotaped interviews of 46 mathematics experts and 44 nonexperts and found they are each more likely to show correct mathematical intuitions and generate mathematically valid proofs when they produced gestures while speaking. The research findings contribute to theories of *embodied cognition* by showing that people can tap into nonverbal ways of mathematical thinking. This work is important for education in STEM (science, technology, engineering, and mathematics) because it demonstrates that embodied cognition applies beyond basic mathematics such as counting and computation to conceptual forms of reasoning involved in geometry proofs.

Keywords: expert–nonexpert differences, geometry, gesture, mathematical cognition, proof

Mitchell J. Nathan and Kelsey E. Schenck, Department of Educational Psychology, University of Wisconsin–Madison; Rebecca Vinsonhaler, Department of Curriculum and Instruction, University of Texas–Austin; Joseph E. Michaelis, Department of Computer Science, University of Illinois–Chicago; Michael I. Swart, Wisconsin Center for Education Research, University of Wisconsin–Madison; Candace Walkington, Department of Teaching and Learning, Southern Methodist University.

The authors acknowledge the important contributions of Oh Hoon Kwon, John Binzak, and John McGinty in carrying out this investigation. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A160020 to University of Wisconsin–Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Preliminary findings of this work have been presented at the 2018 Annual Meeting of the American Educational Research Association.

Correspondence concerning this article should be addressed to Mitchell J. Nathan, Department of Educational Psychology, University of Wisconsin–Madison, 1025 West Johnson Street, Madison, WI 53706. E-mail: MNathan@wisc.edu

Making meaning of mathematical ideas and notational systems is of central importance to education (e.g., Schoenfeld, 1992). It is in this context that scholars increasingly have turned to principles of grounded and embodied cognition (GEC; Barsalou, 2008; Glenberg & Robertson, 2000; Nathan, 2008). Lakoff and Núñez (2000) proposed a theoretical account of how embodied cognition can explain many of the most significant mathematical developments in history. In their account, grounding is achieved through conceptual metaphors that link mathematics to physicality and movement, which plays a central role explaining that mathematical ideas ultimately come to have meaning by being grounded in sensory-motor processes.

In addition to language-based processes, such as metaphor, scholars have observed ways that people engaged in mathematical activity spontaneously use their bodies as a direct means to explore and express their reasoning (Alibali & Nathan, 2012; Chu & Kita, 2011, 2016; Edwards, 2003; Marghetis, Edwards, & Núñez, 2014; McNeill, 1992). For example, Yoon and colleagues (2011) documented how gestures during a conceptual calculus activity supported the generation of new insights regarding the relationships of a function to its derivative and its antiderivative. Others have shown the central role of gestures for advancing learning of a range of mathematical ideas, including equations (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), symmetry (e.g., Valenzeno, Alibali, & Klatzky, 2003), ordinal numbers (Sinclair & de Freitas, 2014), multiplicative reasoning (Abrahamson & Trninic, 2015), graphs (Bieda & Nathan, 2009), beginning algebra (Alibali & Nathan, 2012; Nemirovsky & Ferrara, 2009), geometry (Smith, 2018), and complex numbers (Soto-Johnson & Troup, 2014), among others.

An emerging literature on GEC (Barsalou, 2008; Shapiro, 2010) suggests that reasoning is connected to body-based processes, including gesture (Alibali & Nathan, 2012; Edwards, Ferrara, & Moore-Russo, 2014; Goldin-Meadow, 2005) and physical and simulated action (Beilock & Goldin-Meadow, 2010; Hostetter & Alibali, 2019). Still, there is a need for studies that directly investigate the role of body-based processes in complex reasoning activities, and for theoretical advancements that provide testable hypotheses for when and how body-based processes affect reasoning. Investigations of mathematical reasoning, in particular, are of interest, in part because of the theoretical appeal of embodied accounts of highly generalized, symbolic, and abstract forms of thinking (e.g., Abrahamson & Lindgren, 2014; Lakoff & Núñez, 2000).

To date, empirical studies of embodied mathematical cognition tend to focus on the mathematics of numbers and operations, such as arithmetic, probability, and algebra (Abrahamson, 2009; Alibali, Church, Kita, & Hostetter, 2014; Goldin-Meadow, Cook, & Mitchell, 2009; Howison, Trninic, Reinholz, & Abrahamson, 2011; Marghetis, Núñez, & Bergen, 2014; Ottmar & Landy, 2017), or shape identification (Smith, King, & Hoyte, 2014), with relatively little attention paid to more advanced areas topics such as proof practices (Nathan, 2014; though see Marghetis, Edwards, & Núñez, 2014). Yet proof is fundamental to the discourse practices of mathematicians, serving as the primary method by which mathematicians test claims, construct knowledge, and disseminate their research (Lakatos, 2015). This is why justification and proof are key topics in mathematics education (National Council of Teachers of Mathematics, 2000; Stylianides, 2007; Yackel & Hanna, 2003).

Furthermore, much of the work on embodied mathematics emphasizes early mathematical development among young children (e.g., Butterworth, 1999). As children mature into more abstract and generalized thinkers, there is a need to understand whether and how embodiment plays a significant role in their reasoning and learning. Consequently, there is value in extending research on embodiment to the topic of geometry proofs, a precollege strand of mathematics education (Pelavin & Kane, 1990) that focuses directly on articulating logically supported generalized truths about space and shape.

The central objective of this paper is to investigate people's geometric reasoning and informal proof practices within a GEC framework. We specifically examine the roles that participants' gestures and concurrent speech play in predicting the quality of their geometric reasoning. The main objective is to identify theoretically motivated associations between body-based processes and geometric thinking for improving our empirical understanding of the embodied nature mathematical reasoning. This investigation is important for understanding the scope and predictive power of theories of GEC. Theoretically motivated advancements in our knowledge of the nature of embodied mathematical reasoning are important for developing effective, evidence-based approaches to education that can inform the design of students' learning environments and teacher professional learning experiences.

Theoretical Framework

Grounded and Embodied Cognition

Although there is a diversity of theories of GEC, they generally share certain tenets. One is cognition and computation are not the exclusive result of operations performed with amodal symbols; rather, reasoning and computation are necessarily carried out by recruiting perceptual and motor processes (Gibson, 1976, 1979); and *offline cognition*—processes, such as planning, that are performed when one cannot directly access task-relevant inputs and outputs—is achieved via simulation of perceptuo-motor experiences, situated actions, and bodily states, which play a causal or constitutive role in intellectual processes (Barsalou, 2008; Shapiro, 2019; Wilson, 2002).

A second tenet is based on evidence that suggests some aspects of mathematical cognition are grounded and embodied (de Freitas & Sinclair, 2014; Nathan, 2014). Numbers, for example, are understood within a spatial frame that applies cross-culturally (Fischer, 2012). Algebraic symbol manipulation, despite its apparent abstract nature, is sensitive to spatial grouping (Landy & Goldstone, 2007). Gestures—spontaneous arm and hand movements that speakers produce when communicating—depict learners' mental simulations and conceptual metaphors of mathematical objects and operations (Alibali & Nathan, 2012). Typically, gesture taxonomies identify four main categories (McNeill, 1992): (a) deictic or indexical gestures, such as pointing, index an object and can provide information about its location; (b) iconic gestures convey semantic content through visual similarity using hand shape or motion; (c) metaphoric gestures convey semantic content via metaphorical mapping; and (d) beat gestures, simple, rhythmic motions that do not clearly express semantic content but generally align with speech prosody. Gesture studies scholars—especially those studying mathematical cognition—often combine iconic and metaphoric gestures into a broader category of *representational gestures*, to emphasize that these each depict meaning (Alibali, Heath, & Myers, 2001; Kita, 2000). Representational gestures produced during proof construction suggest that experts' mathematical reasoning is inherently embodied, not merely an aid to communicate mathematical ideas (Marghetis, Edwards, & Núñez, 2014). Scholars suggest "that gesture and other bodily movement is essential . . . in the intellectual construction of mathematics;" and along with words, symbols, diagrams, and objects, "the mathematician's body may be a constitutive part of his or her situated proving" (Marghetis & Núñez, 2013, p. 229).

Gesture as Action and Simulated Action

Research shows that gesture production reliably predicts some forms of mathematical thinking, and that engaging students in gesture production can enhance learning. For example, primary grade children learning about mathematical equivalence produce gestures that indicate they have learned to the attend to spatial properties that distinguish the two sides of the equation delineated by the equal sign, even before they can verbally articulate that distinction (Alibali & Goldin-Meadow, 1993). In addition to their communicative function, gestures perform simulated actions. There is substantial empirical support for the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008, 2019). As evidence, speakers gesture more often when their speech is based on imagery, and the form of a gesture parallels the form of the underlying mental simulation, especially when describing imagery to mentally transform or manipulate objects (Hostetter, Alibali, & Bartholomew, 2011). Experts gesture less frequently than novices in some studies (Chu & Kita, 2011), presumably because experts have had more practice performing and verbalizing the tasks (e.g., Provost, Johnson, Karayanidis, Brown, & Heathcote, 2013). As simulated actions, gestures can *act on* entities, as though physically rotating an imagined triangle. Furthermore, because these actions on imagined objects are not bound to physical constraints, simulated actions can *transform* an entity, such as growing and shrinking the triangle. In this vein, scholars have identified a specific class of representational gestures, called d*ynamic depictive gestures*, that portray transformations on imagined objects (Garcia & Infante, 2012; Newcombe & Shipley, 2012), enabling simulated actions to support students' imagination and enact mathematical generalizations.

Dynamic Gestures Explore Generalized Properties

Dynamic gestures specifically enact spatiotemporal transformations of imagined entities that allow people to physically experience the operations, generalization, and chain of logical inference that support production of valid analytical proof schemes (Harel & Sowder, 2005; Pier et al., 2019). *Nondynamic* (or *static*) *gestures* primarily identify properties of objects (e.g., shape, location), whereas dynamic gestures perform operative actions that allow agents to explore the range of an object's properties (e.g., how it rotates or skews; Chu & Kita, 2011; Garcia & Infante, 2012; Hegarty, Mayer, Kriz, & Keehner, 2005; Newcombe & Shipley, 2012). Figure 1 visually compares these properties of dynamic and nondynamic depictive gestures during student interviews while reasoning about the truth of the midsegment conjecture, *The segment that joins the midpoints of two sides of any triangle is parallel to the third side*. Nondynamic and dynamic gestures activate different cognitive processes. In Figure 1a, the participant uses a dynamic gesture to continually adjusts the angle of one side (right arm) as it meets a second side (left hand) as she realizes in midsentence that her initial evaluation that the conjecture was false was inaccurate, and that, in fact, the midsegment will always remain parallel to the third side (here, the base), In Figure 1b, another participant uses a nondynamic gesture to form two sides of the triangle that meet at a vertex and makes no further adjustments to the shape.

People with lower spatial reasoning skills produced a higher proportion of static gestures and thus conveyed less dynamic

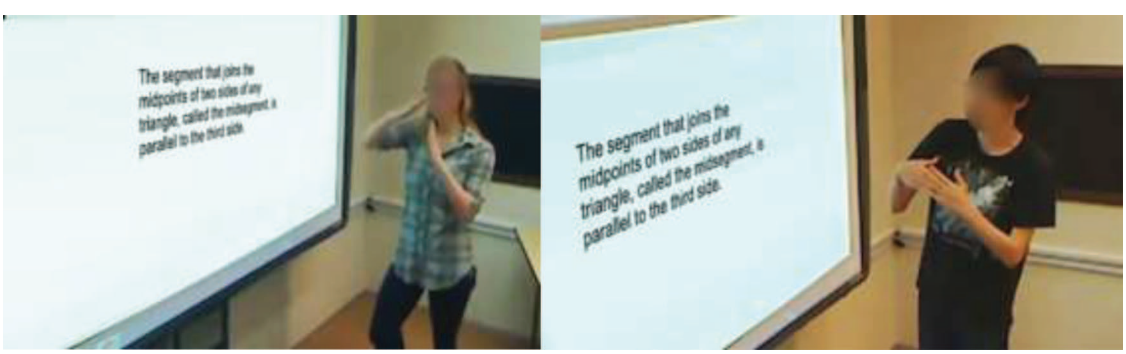

Figure 1. Participants reasoning about the midsegment conjecture, *The segment that joins the midpoints of two sides of any triangle is parallel to the third side*. (a) Left panel shows how a dynamic depictive gesture enacts generalizable relationships of a mathematical object (triangle). The participant continually adjusts the angle of one side (right arm) as it meets a second side (left hand) as she realizes in midsentence that her initial evaluation that the conjecture was false was inaccurate, and that, in fact, the midsegment will always remain parallel to the third side (here, the base). (b) Right panel shows a nondynamic gesture makes a static property joining two sides of the imagined triangle meeting at a vertex with no further changes to the shape. The photographs are published with the consent of the experimental participants. See the online article for the color version of this figure.

information about object manipulation than those with higher spatial reasoning skills (Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013). In another study, participants who produced dynamic gestures were more likely to generate correct insights and showed higher rates of valid proofs for mathematical conjectures (Pier et al., 2019). These findings suggest dynamic gestures may be associated with cognitive processes that are especially helpful for mathematical generalizations and proofs.

Action-cognition transduction proposes that the advantages for mathematical reasoning incurred by dynamic gesture production come about because the actions performed on either real or imagined entities engage the sensorimotor system to recreate those cognitive states that simulate the important properties, relationships, and behaviors of the object (Nathan & Walkington, 2017). The motor programs generated to process and control those movements necessarily anticipate the outcomes of those potential actions, producing predictive expectations that naturally support inferences about those future states. This feedforward architecture offers an account of an action-based inferential mechanism that considers all such plausible future states of an object (Nathan, 2017; Nathan & Martinez, 2015; Wolpert, Doya, & Kawato, 2003) and forms the basis for embodied reasoning about generalized spatial properties (Nathan & Walkington, 2017). A body-based account of inference making is much-needed since people otherwise struggle to formulate the generalizations that foster mathematically valid proofs.

Geometric Reasoning Using Dynamic Gesture Production and Transformational Speech

Geometric proof is a valuable area for improving education and extending research on mathematical cognition. Geometric proofs typically address universal statements about space and shape, which are an important area for the study of generalized and abstract thought Lehrer & Chazan (1998/2012). Proof in this domain does not readily lend itself to "canned" procedures or mathematical algorithms, such as long division, that might enable people to generate a valid answer with little understanding of the math involved (Koedinger, 1998). Thus, the study of proof is especially intriguing for GEC because of its central role of conceptual understanding of generalizations and abstractions.

Despite proofs' centrality to mathematical practice and educational objectives, students struggle to construct viable and convincing mathematical arguments and provide valid generalizations of mathematical ideas (Dreyfus, 1999; Martin, McCrone, Bower, & Dindyal, 2005). Students often overgeneralize what they can conclude from specific examples (e.g., Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009). Even when presented with valid deductive proofs, students may find them unconvincing (Chazan, 1993), and they fail to appreciate the utility of deductive reasoning for communicating generalized arguments based on logical inference (Harel & Sowder, 1998). High school textbooks, largely operating from a frame of philosophical rationalism, typically emphasize the role of proofs for establishing certainty (e.g., de Villiers, 1998/2012) rather than as a means for developing mathematical intuition and sense of inquiry and wonder (Gravemeijer, 1998/2012; Lehrer et al., 1998/2012; Lockhart, 2009) that draws on students' spatial skills (Goldenberg, Cuoco, & Mark, 1998/2012). Proof used in this way has served to disconnect students' proof practices from the construction of mathematical knowledge (Herbst, 2002).

Expanded notions of proof. Scholars looking beyond proof as a *product* have explored proving as a form of disciplinary discourse (e.g., Balacheff, 1991; Knuth, 2002). Harel and Sowder (1998), arguing for a broader view than traditional high school textbooks, define proving as "the process employed by an individual to remove or create doubts about the truth of an observation" (p. 241). Harel and Sowder (2005) propose a taxonomy of proof schemes that strives to capture this broader conception. *Transformational* proofs, in particular, make use of mental or physical operations to demonstrate the validity of conjectures (Clements & Battista, 1992). Transformational proofs, part of the deductive analytic proof scheme (Harel & Sowder, 2005), have three defining characteristics: They are *general*, showing the argument is true for all members of an object class; they use *operational thought*, where the prover progresses systematically through a goal structure, anticipating the outcomes of proposed transformations; and they exhibit a chain of *logical inference*, with conclusions following from valid premises.

Embodied perspectives on geometry reasoning. Several investigations have examined geometric reasoning from an embodied lens to explore fruitful avenues for improving our understanding of mathematical cognition. Studies of mathematicians' gestures suggest that experts' proof practices are "fundamentally . . . embodied" (Marghetis, Edwards, & Núñez, 2014, p. 228). Primary and secondary grade students' understanding of angles improves when manipulating their body position and movement (Petrick & Martin, 2012; Shoval, 2011; Smith et al., 2014). Use of global-positioning system devices and mapping software enabled students to figure out where to stand and walk to make the geometric constructions for a marching band, that enriched subsequent geometric reasoning with pencil-and-paper formats (Ma & Hall, 2018).

Pier and colleagues (2019) found that students' transformational speech and dynamic gesture independently contributed to students' proof performance. Like physical and simulated transformations, transformational speech was defined as verbal descriptions of goal-directed manipulations of mathematical objects through conditional statements ("if . . . then . . .") and inferences.

Theoretically Motivated Hypotheses

There is a need to investigate the psychological mechanisms that underlie embodied proof performance. The literature points toward some productive avenues but lacks some of the rigor and clarity that can reliably test hypotheses about the stated promise of embodied mathematical thinking. Findings of many of these studies are summarized in the model of Figure 2. As reviewed, studies implicate embodied processes exhibited by representational gestures—and specifically dynamic gestures—as observable manifestations of simulated actions that mediate valid mathematical proof production (e.g., Nathan & Walkington, 2017; Pier et al., 2019; Walkington et al., 2014; Williams-Pierce et al., 2017). As shown in Figure 2, spatiotemporal processes (top pathway) on their own can yield mathematical intuitions about the truth status of a conjecture, but do not support the production of a valid verbal justification of that judgment without verbal processes. Studies also point to the important role of language processes—especially

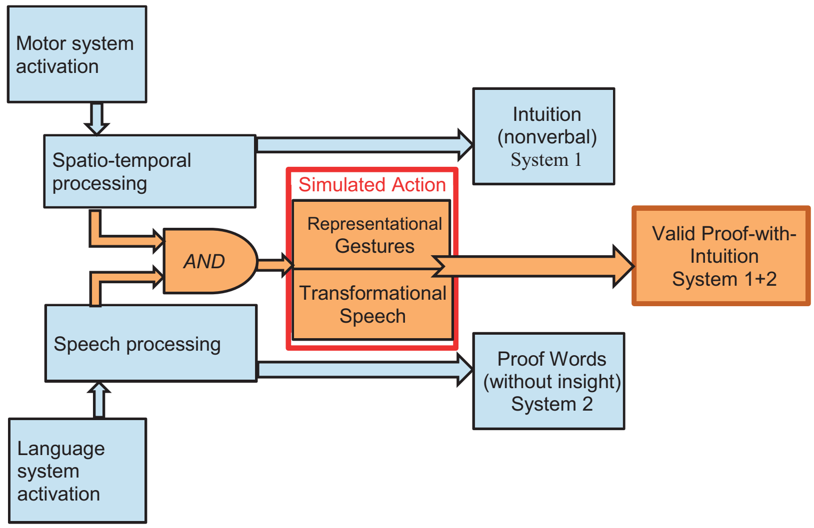

Figure 2. Logic model. Spatiotemporal processes (top pathway) combine with language processes (bottom pathway) to generate a proof with intuition, which is hypothesized to be mediated by simulated action, as exhibited by the speaker's dynamic gestures and transformational speech. Moderators (not shown) include math expertise, prior content knowledge, spatial skills, verbal skills. See the online article for the color version of this figure.

action-based transformational speech—for supporting logical and causal inference (Knuth et al., 2009; Pier et al., 2019). Language processes can generate proofs (bottom pathway), but may be generated without an intuitive understanding of why or how they are true. To produce a mathematically valid proof-with-intuition, both spatiotemporal and language processes must be engaged and, further, must activate key mediators in language (transformational speech) and action (representational gestures) to assure that the proof that is generated meets the three criteria for a valid mathematical proof that is logical, operational, and generalizable. The research also suggests that these influences are moderated by one's mathematical expertise (Goldin-Meadow, 2010), suggesting that even though experts' proofs may be mediated by body-based processes, experts may be less dependent on gesture production to access implicit knowledge and convey that knowledge as they formulate mathematically valid proofs. The present study investigates these issues by examining the role of transformational speech and representational gestures across a common set of multiple mathematical conjectures.

We frame this investigation in terms of a central research question: *Is geometric reasoning associated with participants' simulated actions?* As one looks across the research literature, proof performance is often reported in a variety of ways without a standard outcome measure. Consequently, there is value in stipulating precise measures with which proof performance is assessed. The present study uses three outcome measures, listed in increasingly more complex levels of reasoning: Intuition, insight, and mathematically valid proof. We are especially interested in the production of representational gestures and transformational speech as instantiation of simulated actions. We hypothesize that gestures and transformational speech acts each make unique contributions to models of participants' mathematical reasoning.

Our measure of intuition makes the fewest assumptions, and simply requires that, upon their initial encounter with a conjecture, participants accurately state whether the statement is true or false. Intuitions, we posit, are nonverbal "snap judgments" about the veracity of a mathematical conjecture, and thus primarily mediated by embodied processes. Because correct intuitions do not depend on a person's ability to articulate a generalizable property of shapes, we expect that representational gestures will be the strongest predictors, rather than the more restrictive category of dynamic depictive gestures. We also expect that linguistic factors are not likely to be reliable predictors of intuition performance. We summarize this in the following hypothesis:

> *H1:* Intuitions about the veracity of a geometry conjecture are reliably associated with one's representational gestures

Insight provides a measure of participants' grasp of the "essential meaning," or *gist*, about the mathematical ideas that come into play while forming a proof, "irrespective of exact words, numbers, or pictures." (cf. Reyna, 2012, p. 333). We expect insight performance to be associated with one's speech and one's generation of representational gestures. The second hypothesis states:

> *H2:* Insights about the veracity of a geometry conjecture are reliably associated with one's representational gestures and one's transformational speech.

To be considered a mathematically valid proof, as noted, one's argument must include three characteristics (Harel & Sowder, 1998, 2005; Pier et al., 2019): *generality*, *operational thought,* and *logical inference*. We expect to see valid proof performance associated with *dynamic depictive gesture* production as an instantiation of the simulated actions needed to support operational thinking and generalization. We also expect valid proofs to be associated with the production of transformational speech acts as explanatory records of a logical chain of inference.

H3: Valid proof performance will be reliably associated with one's dynamic gestures and transformational speech.

A second research question poses, *Does the strength of the hypothesized relationships between geometric reasoning and simulated action depend on participants' mathematics expertise?* Those with greater expertise in math will have greater knowledge of planar geometry and more familiarity with proof practices (Inglis & Alcock, 2012; Koedinger & Anderson, 1990). Yet there are conflicting accounts about the nature of embodied behaviors for experts. On one hand, researchers have noted that the mathematical reasoning and proof practices exhibited by experts are inherently embodied (Marghetis & Núñez, 2013), suggesting that gesture production should be high for experts, perhaps even higher than for nonexperts (Marghetis, Edwards, & Núñez, 2014). Others (Chu & Kita, 2011) have found that experts in some tasks (e.g., spatial visualization) gesture *less* frequently than novices, possibly because of the greater refinement of their skilled performance (e.g., Provost et al., 2013). Thus, we have two competing hypotheses regarding the observable simulated actions for experts in mathematics.

H4: a. Valid proof performance by experts will be more strongly associated with gesture production than nonexperts.

H4: b. Valid proof performance by experts will be less strongly associated with gesture production than nonexperts.

The following research method is used to explore these hypotheses.

Method

Participants

Ninety adult students were recruited from a large university in the Midwestern United States. Participants included 85 undergraduate students and five graduate students. Graduate students were included in the recruitment of expert participants when recruitment efforts limited to undergraduate students became unlikely to yield at least 42 expert participants as indicated by the power analysis. All participants were English-speaking students, but for some English was not their first language. As compensation, each participant received a $25 gift card to an online retailer. Some participants also received partial course credit (extra credit) if their course instructors offered this credit option. Experts ($n = 46$) were math majors with advanced course work beyond linear algebra that included studies of formal proofs. All graduate students were included in the expert group. In this group, there were 31 males and 15 females with 52.2% of students identifying as native English speakers. Nonexperts were undergraduate non-STEM education majors ($n = 44$). This group included six males and 38 females with 97.7% of students identifying as native English speakers. Further descriptive statistics for both expert and nonexpert groups can be found in Table 1.

Power Analysis

Our a priori power analysis used $\beta = 0.80$, $\alpha = .05$, and an effect size of $f = 0.41$ for the effect of expert/novice on proof performance (based on data from Nathan & Walkington, 2017). The analysis used G*Power's (Faul, Erdfelder, Buchner, & Lang, 2009) ANOVA Repeated Measures, with correlations among a participant solving repeated geometry proofs estimated at 0.6 based on previous data (Nathan et al., 2014; Nathan & Walkington, 2017). G*Power returned a minimum sample of 26 per group. However, we additionally took into account having power to detect small/medium partial mediation effects ($d = 0.26–0.39$; Fritz & MacKinnon, 2007), which led to a desired sample size of 42 per group (84 total) when accounting for the design effect.

Materials

Conjectures. The four conjectures used in this study were selected from a variety of secondary mathematics textbooks and chosen so that each explored general properties of two-dimensional (planar) objects. Three conjectures involved proper-

Table 1
Descriptive Statistics for Experts and Nonexperts

Statistic	Expert			Nonexperts		
	Male (n = 31)	Female (n = 15)	Total (n = 46)	Male (n = 6)	Female (n = 38)	Total (n = 44)
Average age (*SD*)	22.1 (3.21)	21.6 (2.14)	21.9 (2.90)	20.3 (0.76)	20.68 (2.37)	20.6 (2.22)
Percent native English speakers	48.4%	60.0%	52.2%	100%	97.4%	97.7%
Percent ethnicity – White	45.2%	40.00%	43.48%	83.33%	89.47%	88.64%
Percent ethnicity – Asian	45.2%	53.33%	47.83%	0.00%	7.89%	6.82%
Percent ethnicity – Other	9.6%	6.67%	8.70%	16.67%	2.63%	5.13%
Average geometry test score (*SD*)	0.94 (0.08)	0.89 (0.08)	0.92 (0.08)	0.82 (0.09)	0.79 (0.08)	0.79 (0.08)
Average spatial score (*SD*)	0.82 (0.16)	0.76 (0.22)	0.80 (0.19)	0.53 (0.20)	0.46 (0.22)	0.47 (0.22)
Average verbal fluency score (*SD*)	18.1 (5.35)	21.5 (13.05)	19.2 (8.76)	16.2 (4.73)	18.0 (5.65)	17.8 (5.56)
Likelihood of correct proof	41.13%	36.67%	39.67%	4.17%	11.18%	10.23%
Likelihood of correct insight	85.48%	81.67%	84.24%	25.00%	49.34%	46.02%
Likelihood of correct intuition	79.03%	83.33%	80.43%	54.17%	61.18%	60.23%
Likelihood of representational gesture (per trial)	79.03%	90.00%	82.61%	37.50%	68.42%	64.20%
Likelihood of nondynamic gesture (per trial)	32.26%	40.00%	34.78%	29.17%	45.39%	43.18%
Likelihood of dynamic gesture (per trial)	46.77%	50.00%	47.83%	8.33%	23.03%	21.02%

Note. *SD* = standard deviation.

ties of triangles, and one conjecture concerned properties of parallelograms. Additionally, three conjectures were true statements, and one conjecture about triangles was a false statement. Table 2 shows the text for each conjecture, its truth, relevant mathematical insights regarding the conjecture, and example proofs.

One of the conjectures also identified a technical term in the conjecture printed in blue font with an underline that linked to a glossary explaining the term's meaning. Participants were informed about the glossary term prior to the interviews. Participants were invited but not required to click the link on the conjecture slide. Judgments were made by researchers on the team with math instruction experience as well as experience with the target population of participants as to which terms would be appropriate to include or not. These decisions were made within the broader goal of presenting our expert and novice participants with conjectures that would be appropriately difficult to require active reasoning and participation, and appropriately scaffolded to give all participants a chance to attempt a response without first giving up.

Dependent, independent, and control variables. The dependent variables (DVs) in this study are intuition, insight, and proof performance. The independent variables (IVs) in this study are expertise, gesture production, and transformational speech production. The remaining measures are control variables: spatial ability, verbal phonemic fluency, general geometry knowledge, and several demographic measures (self-report of math course experience, such as the last math class completed, current math class enrollment, math course history with final grades, age, sex, native language, and race/ethnicity). Measures for these are explained below.

Spatial ability. Spatial ability was assessed using the *Paper Folding Task* (Ekstrom, French, Harman, & Derman, 1976). This spatial measure is included because it has been to shown in prior studies to have been predictive of gesture production (Hostetter & Alibali, 2008) as well as predictive of performance on measures of geometry intuition, insight, and proof (Walkington, Woods, Nathan, Chelule, & Wang, 2019). The Paper Folding Task is a 20-item (2×10 items per section) multiple-choice assessment. Participants were given 3 min to complete each section of 10 questions. Scores were computed by giving one point for each correct answer and subtracting 0.25 for each incorrect answer. For the Paper Folding Task, only the first section was used for calculating spatial ability scores. The historical reliability for the Paper Folding Task ranges from 0.75 to 0.89 (Kane et al., 2004; Kozhevnikov & Hegarty, 2001; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001), and for our sample, the reliability for the first section was 0.79.

Verbal phonemic fluency. Verbal phonemic fluency was assessed using a standard task where participants to name as many words as they could think of in 60 s that begin with the letter "s" (desRosiers & Kavanagh, 1987). This verbal phonemic fluency measure is included because it has been to shown to have been predictive of gesture production in prior work (Hostetter & Alibali, 2008). Responses were given aloud and recorded through the video/audio recording software. This task was scored by counting the number of unique words uttered within the 60 s time limit. Counts of unique words excluded proper nouns and words whose variations differed only in plurality or verb tense. For example, if a person uttered "save," then the words "saving," "savings," and "saved" would not further be counted. The retest reliability for this measure has been assessed 0.88 (desRosiers & Kavanagh, 1987).

General geometry knowledge. Knowledge about geometry properties was assessed with 12 items clustered into three multiple-choice questions, developed in a prior study (Nathan & Walkington, 2017; $r = .56$ with performance on conjectures similar to those considered here). Items asked about the properties of triangles, parallelograms, and circles. Participants needed to identify correct answers in the multiple-choice questions given, and participants could select more than one response. Each question had four answer choices, and each answer choice was scored as correct or incorrect for one point each. Scores were summed across

Table 2
Truth Value, Insights, and Proof for Each of the Four Mathematical Conjectures

Conjecture label	Conjecture text	Truth	Insight	Proof
ParallelogramArea	The area of a parallelogram is the same as the area of a rectangle with the same length and height	True	1. States a parallelogram is a rectangle tilted over or pushed over. 2. States area of a parallelogram and a rectangle have the same formula.	1. Shows cutting off a triangle from the parallelogram, or rearranging the area makes them congruent. 2. State all rectangles are parallelograms and therefore the formula for area is the same.
MidsegmentTriangle	The segment that joins the midpoints of two sides of any triangle is parallel to the third side	True	1. True because the two triangles are similar 2. True because it is scaled	1. Shows base sliding up and says Similar Triangles or scaled so angles are the same 2. Explicitly says SAS and that corresponding angles are congruent
AAA	Given that you know the measure of all three angles of a triangle, there is only one unique triangle that can be formed with these three angle measurements	False	States similar triangles or infinite/many triangles	1. Gives specific counterexample 2. Visually shows scaling or discusses scaling and similar triangles
Circumscribed	A circle can be circumscribed about any triangle	True	1. Any three points on a plane make a triangle. 2. The circumcircle always passes through all three vertices of a triangle	1. Demonstrate with vertices as points along the circumcircle. 2. Show with the perpendicular bisectors of each side of the triangle.

all three clusters for a final score out of 12. The geometry knowledge test was ultimately not used in the models because of issues with internal consistency (α = .54). Results for proof, insight, and intuition were the same with or without this test included.

Demographic information. This survey included a self-report of math course experience, such as the last math class completed, current math class enrollment, and math course history with final grades. Additional demographic information was collected about age, sex, native language, and race/ethnicity. After data collection was complete, we collapsed the race/ethnicity responses into three categories for the analyses: White, Asian, and Other. This decision was made because of the high number of participants who self-identified as White (65.56%) and Asian (27.78%). The "Other" category represents the remaining options (6.66%). Because of differences in the expert and nonexpert groups, several of these variables were included in the final prediction models as control variables.

Equipment

The experiment was run in a room operating a projecting computer, an interactive white board, and two web cams that linked to a second computer. The interactive functionality allowed the researcher to advance PowerPoint slides, which contained the conjecture text or navigate through additional material (i.e., the glossary) by tapping on the white board rather than returning to the computer each time. The webcams were mounted about 6 feet high on the wall adjacent and to the left of the interactive white board. When running the experiment, the researcher would stand right below the left camera, out of the frame of the shot. This made it so the participant was more likely to be directing their response toward the camera. Figure 1 depicts the room set up.

Coding

Videos of the experimental sessions were organized and transcribed verbatim using the Transana software system (Woods & Fassnacht, 2012), which links transcript locations to video locations through time codes and supports qualitative coding and SQL Boolean search functions. Time stamps were added to split the full transcripts into the four conjectures for coding; this resulted in 360 video clips to be coded (90 participants $\times$ 4 conjectures). Two members of the research team coded each transcript using the coding scheme described below and illustrated in Figure 3.

Coding scheme. The transcripts were coded for six categories: conjecture comprehension, intuition, insight, gesture, correct true/false judgment, and *mathematical validity of the proof* (see Figure 3).

Comprehension. Participants were coded as not comprehending the conjecture if they stated they did not understand or provided explicit evidence that they were describing objects or operations not related to the conjecture. If a transcript was coded as "did not understand," the researcher did not continue coding the transcript for that conjecture. This criterion led us to exclude 17 of 360 transcripts (four expert, and 13 nonexpert) from further coding.

Intuition. Intuition was measured by the accuracy of their immediate true/false evaluation of each conjecture (Zander, Öllinger, & Volz, 2016; Zhang, Lei, & Li, 2016). If the participant correctly answered true/false or had an immediate shift to the correct true/false judgment, it was coded as 1. All other responses were coded as 0.

Insight. Insight was coded for the presence of key mathematical ideas for each conjecture, as specified by our team of mathematicians and math educators (see Table 2). If the participant demonstrated initial correct mathematical insight for the conjecture, it was coded as 1. All other responses, including instances where participants switched from incorrect insights to correct insights, were coded as 0.

Gesture. In this study, gestures produced during the interviews were coded as representational or not, if they represented or depicted some feature or operation of a mathematical object or idea. Representational gesture codes effectively combine the traditional category of iconic gestures, which depict visual similarity, with the category of metaphorical gestures, which use iconic depictions for abstract referents, such as metaphorically expressing numbers of greater value as "large." Representational gestures were further subdivided with one of two mutually exclusive codes, either nondynamic depictive (representational) gestures or dynamic depictive (representational) gestures. Our category of nondynamic depictive gestures (Figure 1b) is exclusive of any dynamic depictive gestures. As an example, tracing the outline of a mathematical object, such as a triangle, would be coded as representational (depicting parts of the triangle) but as nondynamic, because no transformation was performed on the triangle, such as dilation or skewing. Dynamic depictive gestures, the second category, are representational gestures that depict motion-based transformations of a mathematical entity to test out the generalizability of a conjecture (Figure 1a). The occurrence of at least one dynamic depictive gesture was coded as 1 in addition to the code for representational gesture. Participant transcripts that did not include any gesture or made gestures that were not depictive, were given a 0.

Correct true/false judgment. This code was used to describe whether the participant correctly identified the conjecture as always true or always false by the end of their reasoning.

Mathematical validity of the proof. For the purpose of this study, proof validity was coded for a verbalized proof (including speech and gesture) that correctly identifies the conjecture as always true or always false and, in addition, contains evidence of all three criteria stipulated by Harel and Sowder (2005): a logical chain of reasoning, such that conclusions are drawn from valid premises; generalizable arguments, showing the argument is true for a class of mathematical objects; and evidence of operational thinking, so that the transcript exhibited evidence of progression through a goal structure, anticipating the outcomes of operations.

Interrater reliability. To establish interrater reliability, a researcher who was not involved in the original coding process or the development of the coding scheme coded a random selection of 10% of the participants' videos. Overall interrater reliability for these codes is κ = .911. Individual interrater reliability measures for comprehension, intuition, insight, proof process, gesture, and correctness are shown in Table 3.

To assess the validity of the interrater reliability given the small sample size, we calculated Shaffer's *rho* for each code (Eagan et al., 2017; https://app.calcrho.org/). For this analysis, we used a *kappa* threshold of 0.65. Results for each interrater reliability measure can be found in Table 3. Overall, all Shaffer's *rho* statistics were less than

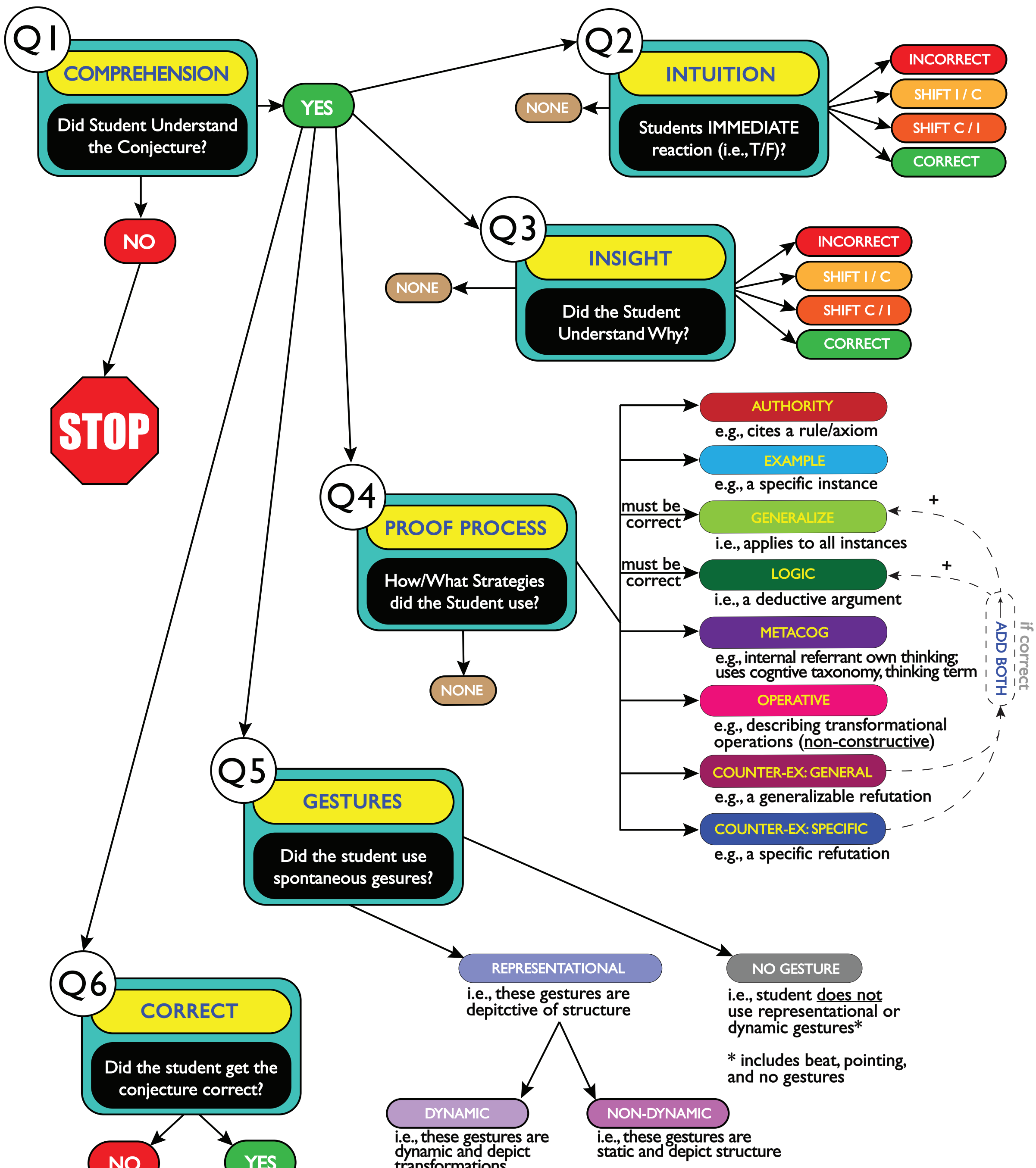

Figure 3. Coding system flowchart. See the online article for the color version of this figure.

0.05, which indicates that our sample size was sufficient to estimate the interrater reliability at a threshold of at least 0.65.

Coh-Metrix coding. To further code the content of each participant's verbal reasoning, we used metrics from Coh-Metrix, a validated text data-mining tool that produces measures of several linguistic indices, including situation model cohesion, connectives, lexical diversity, and syntactic complexity (McNamara, Graesser, McCarthy, & Cai, 2014; www.cohmetrix.com). Three variables were of particular interest: verbs, first-person pronouns, and a measure of intentional cohesion. Verb use is reported as the incidence score of verbs per 1,000 words. Similarly, first-person pronoun use is reported as the incidence score of first-person pronouns per 1,000 words. The intentional participle to intentional verb use is a ratio score, a relative measure that compares the incidence of intentional participles per 1000 words to the incidence of intentional verbs per 1,000 words, that describes the situation

Table 3
Interrater Reliability for Participant Transcript Coding

Code	Cohen's κ	Shaffer's ρ
Omnibus	0.911	0.01
Comprehension	1.000	0.00
Correct	1.000	0.00
Intuition	1.000	0.00
Insight	1.000	0.00
Proof process	0.778	0.04
Dynamic gesture	0.958	0.00
Nondynamic gesture	0.948	0.01

Note. Shaffer's ρ calculated with a κ threshold of 0.65.

model and internal cohesion created by the participants during the conjecture proof. An earlier, backward stepwise regression analysis of transcripts identified these three variables from among a set of 21 as significant contributors to students' verbal proofs (Schenck et al., 2020). To prepare each transcript for the Coh-Metrix automated analysis, we removed nonverbal cues, inaudible speech, transcript notation (parentheses, timestamps, participant numbers, etc.), and researcher speech.

Procedure

A researcher escorted the participant to the experiment room and turned on the video recording software. Participants began their session by completing the Paper Folding Task. Instructions were read out loud by the researcher as participants followed along in the written instructions. Participants sat at a table and provided answers to the items on the worksheet until all items were completed or the 3 min were up. Participants were then instructed to stand on the opposite side of the table facing the video camera and the experimenter. Participants were introduced to the verbal phonemic fluency task. Responses were given aloud and videotaped.

While participants remained standing where they were, the researcher introduced the conjecture task. Specifically, participants were instructed to read the math conjecture presented on the interactive white board, then report whether the conjecture was true or false, and finally provide an explanation why the conjecture is true or false (or why they *believed* the conjecture to be true or false). All participants completed a set of 12 conjectures total. The first four conjectures were designed to be common among all participants (see Table 2), whereas the remaining eight conjectures were tailored to suit participants' expertise levels and thus not held constant between participants. Thus, only the common four conjectures are used for this analysis. The order of these four conjectures was counterbalanced across participants by using a Latin square to create four conjecture orders. Using Power Point software, conjectures were projected in a black sans-serif font with a white background on a large interactive screen. Participants stood about 3 feet away from the center of the screen. Researchers administrating the task tapped the screen to advance the slide to the next conjecture after a full response was recorded.

Researchers were given a script for their interactions with participants. They were instructed to prompt participants to give a full response, and then afterward to invite participants to reiterate their explanation. The goal of this final prompt was to give participants an additional chance to summarize their thoughts into a fully formed answer. However, because of experimenter error, not all participants were uniformly given this chance to reiterate their explanation. Thus, for our Coh-Metrix analyses we only analyzed the initial response from each participant. To test whether there were systematic differences between those prompted only once and those prompted to reiterate their answers, we conducted post hoc Welch two-sample *t* tests to see whether the mean values of the three Coh-Metrix variables significantly differed between participants' initial (i.e., single) or final (either single or double) reports.

Although prompting participants to reiterate did often result in lengthier final transcripts, there was not a statistically significant difference in the mean scores between the initial (single) or final (either single or double) reports in participants' intentional situational model cohesion (initial reports, $M = 4.32$, $SD = 1.01$; final reports, $M = 8.97$, $SD = 0.99$), $t(37) = -1.44$, $p = .159$; verb use (initial reports, $M = 1.03$, $SD = 1.01$; final reports, $M = 1.41$, $SD = 1.00$), $t(235) = -1.899$, $p = .059$; or first-person pronoun use (initial reports, $M = 1.46$, $SD = 1.09$; final reports, $M = 2.00$, $SD = 0.968$), $t(190) = -0.653$, $p = .514$. Because there were no reliable differences in the Coh-Metrix parameters, we included all 360 observations in subsequent analyses. (*NB*). We conducted *t* tests on the outcomes variables [intuition, insight, and transformational proof] and the gesture variables [nondynamic and dynamic gestures]. None of these comparisons showed significant differences.

After giving videotaped responses to the four conjectures, participants completed surveys of their General Geometry Knowledge and demographic information. Finally, participants received their compensation and were given a copy of their consent form along with a summary of the experiment's goals. The summary also explicitly stated not to share the details about the experiment with any friends who are potentially enrolled to participate in the future.

Results

Descriptive statistics illuminated demographic and performance differences between the expert and nonexpert groups in our study (see Table 1). For example, there were significant differences in ethnicity, with fewer Whites in the expert group than in the nonexpert group, $\chi^2(2) = 84.765$ ($p < .001$), and with experts including a smaller percentage of native English speakers, $\chi^2(1) = 95.739$ ($p < .001$). Experts were more likely to produce the correct transformational proof, $\chi^2(1) = 39.752$ ($p < .001$), insight, $\chi^2(1) = 56.503$ ($p < .001$), and intuition, $\chi^2(1) = 16.722$ ($p < .001$), than nonexperts. There were also significant differences between expert and nonexpert groups in spatial ability, $t(322) = -15.76$, $p < .001$, and prior geometry knowledge, $t(322) = -15.50$, $p < .001$, with experts scoring higher than nonexperts on these two measures. There was not a significant different between expert and nonexpert groups in verbal fluency, $t(322) = -1.63$, $p = .104$.

Correlations among all of the key factors are presented in Table 4. We used logistic regression for binary outcomes (0/1) on accuracy of proof, insight, intuition, nondynamic gesture, and dynamic gesture. We used a Firth logistic regression for accuracy of transformation proof as this outcome had low prevalence in our data. This type of model uses a penalized likelihood method rather than the maximum likelihood method used in standard logistic regres-

Table 4
Correlations for Key Factors

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Expert																		
2. Ethnicity	**.480**																	
3. Age	**.240**	**.369**																
4. Sex	**.544**	.244	.155															
5. Native English speaker	**.520**	**.792**	**.357**	.335														
6. Spatial score	**.641**	**.422**	**.149**	.434	**.479**													
7. Verbal score	.096	**.177**	.018	.081	**.154**	**−.021**												
8. Geometry knowledge	**.634**	**.131**	**.119**	**.474**	**.167**	**.550**	**.179**											
9. Representational gesture	**.209**	.077	.019	.025	**.134**	**.203**	.004	**.112**										
10. Dynamic gesture	**.281**	.040	.042	.102	**.121**	**.300**	.060	**.256**	.437									
11. Nondynamic gesture	**.086**	.032	−.024	.122	**.003**	**−.109**	−.055	**−.149**	**.478**	**−.582**								
12. Comparative connectives	.040	−.012	.002	−.035	.032	.053	.111	.051	.023	**.133**	−.112							
13. Situation model-cohesion	**.241**	**.143**	.102	**.104**	**−.213**	**.211**	.093	**.234**	**.167**	.064	.098	.012						
14. Verbs	.063	−.042	.113	.076	−.035	.095	.012	.054	**.204**	.079	.114	−.117	−.103					
15. First-person pronouns	−.104	.078	.049	−.108	.023	−.067	.034	−.129	−.067	−.073	.024	.018	−.098	**.256**				
16 Second-person pronouns	**.244**	−.054	.078	**.156**	.011	**.189**	−.010	**.254**	**.172**	**.278**	−.117	.067	−.012	.015	**−.222**			
17. Intuition	**.222**	**.183**	.121	.081	**.141**	**.150**	−.058	**.156**	**.235**	**.164**	**.053**	−.101	**.132**	.067	−.111	.092		
18. Insight	**.402**	**.204**	.078	.178	**.237**	**.315**	.003	**.240**	**.269**	**.271**	**−.021**	.022	**.216**	.024	**−.173**	**.174**	**.622**	
19. Proof	**.339**	.126	.042	.190	**.201**	**.352**	.002	**.274**	**.348**	**.623**	**−.294**	.081	**.215**	.076	**−.175**	**.253**	**.348**	**.422**

Note. Bolded correlations are significant at $p < .010$ (two-tailed).

sion (Firth, 1993; Heinze & Schemper, 2002). The models for transformational proof were fit using the *logistf* command in the R software package *logistf* (Heinze, Polner, Dunkler, & Southworth, 2018). Linear mixed models for intuition, insight, nondynamic gesture, and dynamic gesture were fit using the *glmer* command in the R software package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). We included participant ID and conjecture as random effects in these models. We determined best fit model selection using the *anova()* function to make comparisons between models to test for significant reductions in deviance using a chi-square distribution. Predictors were kept in the model only if they significantly improved the fit of the model by reducing the deviance. Dropped predictors included age, native English speaker status, and all interaction terms. Finally, we interpreted odd ratios as "small" (odds ratio [OR] = 1.68 or 0.60 if reversed), "medium" (OR = 3.47 or 0.29 if reversed), and "large" ($OR > 6.71$ or <0.15 if reversed). These interpretations correspond to Cohen's d = 0.2, 0.5, and 0.8 as "small," "medium," and "large" (Chen, Cohen, & Chen, 2010). Odds ratios, rather than effect sizes, are reported here as our dependent variables are dichotomous.

Intuition

The results of Model 1 (see Table 5) showed that expertise (p = .003) and ethnicity-Asian (p = .049) were both significantly associated with correct intuition. Expertise was associated with an increase in the relative odds of producing correct intuition of 3.05, whereas ethnicity-Asian was associated with an increase in the relative odds of 2.76.

When factors coding for nondynamic and dynamic representational gestures were added to form Model 2, nondynamic representational gesture was significant (p = .004), as was ethnicity-Asian (OR = 3.05, p = .042) and expertise (OR = 2.28, p = .030). The occurrence of at least one nondynamic representational gesture was associated with an increase in the relative odds of producing correct intuition of 2.30. These results provide support for the hypothesis that nondynamic representational gestures reliably predict intuitions about the veracity of a geometry conjecture (H1).

Insight

The results for the initial fixed effects model for insight (Model 1, Table 6) showed that expertise was significantly associated with the production of correct mathematical insights ($p < .001$). Expertise was associated with an increase in the relative odds of producing correct mathematical insight of 9.15.

After nondynamic and dynamic gestures were included in Model 2, nondynamic gesture was significantly associated with correct insight, with nondynamic gesture associated with an increase in the relative odds by 2.21 (p = .019). Expertise remained significant in the models, with expertise associated with an increase in the relative odds of correct mathematical insights of 6.47 ($p < .001$).

Model 3 (see Table 6) added the three transformational speech variables (situation model-intentional cohesion, verbs, and first-person pronouns) to Model 2. Results showed that although expertise continued to be significantly associated with correct insight (OR = 5.69, p = .001), nondynamic gesture (p = .059) was replaced by dynamic gesture (OR = 2.11, p = .042). Additionally,

Table 5
Results of the Logistic Regression Predicting Intuition

Variable		β	*SE*	*OR*	*p* value
Model 0: Null model					
Random component – Participant ID variance	0.33		0.58		.042***
Random component – Conjecture variance	0.00		0.00		.054
(Intercept)		0.94	0.14	2.56	.000***
Model deviance (*df*)	434.2 (357)				
Model 1: Main effects					
Random component – Participant ID variance	0.00		0.05		.051**
Random component – Conjecture variance	0.00		0.00		.055
(Intercept)		0.12	0.62	1.12	.850
Expert[a]		1.12	0.37	3.05	.003**
Verbal		−0.02	0.02	0.98	.120
Spatial		0.01	0.06	1.01	.847
Ethnicity1 (White)[b]		0.78	0.47	2.17	.100
Ethnicity2 (Asian)[b]		1.01	0.54	2.76	.049*
Sex[c]		−0.27	0.32	0.76	.396
Model deviance (*df*)	409.6 (351)				
Model 2: Main effects with gesture					
Random component – Participant ID variance	0.00		0.00		.051
Random component – Conjecture variance	0.00		0.00		.055
(Intercept)		−0.31	0.64	0.73	.626
Expert[a]		0.82	0.38	2.28	.030*
Verbal		−0.02	0.02	0.98	.204
Spatial		−0.03	0.06	0.97	.679
Ethnicity1 (White)[b]		0.68	0.49	1.97	.164
Ethnicity2 (Asian)[b]		1.12	0.55	3.05	.042*
Sex[c]		−0.03	0.33	0.97	.926
Dynamic gestures[d]		0.27	0.32	1.31	.385
Nondynamic gestures[e]		0.83	0.30	2.30	.004**
Model deviance (*df*)	397.2 (349)				

Note. *SE* = standard error; *OR* = odds ratio. The raw regression coefficient (β) can be transformed into odds ratios by exponentiating the coefficient.
[a] Nonexpert participant is the reference group. [b] "Other" ethnicity is the reference group. [c] Female is the reference group. [d] No dynamic gesture is the reference group. [e] No nondynamic gesture is the reference group.
* $p < .05$. ** $p < .01$. *** $p < .001$.

results showed a positive association between a more cohesive situation model and correct mathematical insight ($OR = 1.50$, $p = .040$). Combined, these results support the hypothesis that correct mathematical insights about the truth of a geometry conjecture are reliably associated with representational gestures and transformational speech (H2). However, the results also provide evidence that dynamic gestures—a more specialized form of representational gesture—are also associated with correct insights.

Proof Performance

The initial fixed effect model for transformational proof (Model 1, Table 7) showed that expertise ($p = .001$) and spatial ability score ($p < .001$) were significantly associated with correct transformational proof. Expertise was associated with an increase in the relative odds of producing correct transformational proofs of 3.60, whereas spatial scores were associated with an increase in the relative odds of 1.32.

When gestures were added to form Model 2 (see Table 7), dynamic gestures became significantly associated with transformational proof ($p < .001$), whereas both expertise and spatial ability lost significance ($p = .372$ and $p = .112$, respectively). The production of at least one dynamic gesture was associated with an increase in the relative odds of producing correct transformational proofs of 12.94, exceeding the threshold for a large odds ratio.

When the three transformational speech variables (situation model-intentional cohesion, verbs, and first-person pronouns) were added to form Model 3 (see Table 7), greater situation model cohesion ($OR = 1.54$, $p = .007$) and the reduction of first-person pronouns ($OR = 0.51$, $p = .002$) were significantly associated with transformational proof. More frequent situation model cohesion was associated with an increase in the relative odds of producing a mathematically valid transformational proof. More frequent use of first-person singular pronouns was associated with a decrease in the relative odds of generating a valid transformational proof. This result can otherwise be interpreted as showing that less frequent mentions of first-person singular pronouns—fewer instances of talking about oneself, and more occasions to talk about other entities, such as the mathematical objects under scrutiny—was associated with valid transformational proofs. The occurrence of dynamic gesture continued to have the largest, most significant association with proof performance. Even when controlling for speech content, expertise, and spatial reasoning, the production of at least one dynamic gesture during their verbal proof was associated with an increase in the relative odds of producing a correct

Table 6
Results of the Logistic Regression Predicting Insight

Variable		β	*SE*	*OR*	*p* value
Model 0: Null model					
Random component – Participant ID variance	1.83		1.35		.037*
Random component – Conjecture variance	0.11		0.33		.042*
(Intercept)		0.89	0.26	2.44	.001***
Model deviance (*df*)	435.9 (357)				
Model 1: Main effects					
Random component – Participant ID variance	0.53		0.73		.033*
Random component – Conjecture variance	0.11		0.33		.045*
(Intercept)		−0.94	0.83	0.39	.258
Expert[a]		2.21	0.49	9.15	.000***
Verbal		−0.03	0.02	0.97	.170
Spatial		0.10	0.08	1.10	.199
Ethnicity1 (White)[b]		0.96	0.66	2.61	.146
Ethnicity2 (Asian)[b]		0.71	0.72	2.04	.318
Sex[c]		−0.46	0.41	0.63	.265
Model deviance (*df*)	388.8 (351)				
Model 2: Main effects with gesture					
Random component – Participant ID variance	0.44		0.67		.039*
Random component – Conjecture variance	0.11		0.33		.044*
(Intercept)		−1.37	0.84	0.25	.104
Expert[a]		1.87	0.49	6.47	.000***
Verbal		−0.02	0.02	0.98	.248
Spatial		0.04	0.08	1.05	.530
Ethnicity1 (White)[b]		0.82	0.66	2.27	.211
Ethnicity2 (Asian)[b]		0.82	0.71	2.26	.251
Sex[c]		−0.16	0.41	0.85	.696
Dynamic gestures[d]		0.63	0.36	1.87	.083
Nondynamic gestures[e]		0.79	0.34	2.21	.019*
Model deviance (*df*)	375.0 (349)				
Model 3: Main effects with gesture and speech					
Random component – Participant ID variance	0.48		0.69		.041*
Random component – Conjecture variance	0.05		0.23		.038*
(Intercept)		−1.11	0.86	0.33	.196
Expert[a]		1.74	0.50	5.69	.001***
Verbal		−0.03	0.02	0.97	.205
Spatial		0.04	0.08	1.04	.615
Ethnicity1 (White)[b]		0.82	0.68	2.26	.231
Ethnicity2 (Asian)[b]		0.85	0.75	2.34	.256
Sex[c]		−0.15	0.42	0.86	.726
Dynamic gestures[d]		0.75	0.37	2.11	.042*
Nondynamic gestures[e]		0.67	0.36	1.96	.059
Situation model-cohesion		0.41	0.20	1.50	.040*
Verbs		−0.11	0.15	0.90	.482
First-person pronouns		−0.26	0.15	0.77	.075
Model deviance (*df*)	364.0 (346)				

Note. *SE* = standard error; *OR* = odds ratio. The raw regression coefficients (β) can be transformed into odds ratios by exponentiating the coefficient.
[a] Nonexpert participant is the reference group. [b] "Other" ethnicity is the reference group. [c] Female is the reference group. [d] No dynamic gesture is the reference group. [e] No nondynamic gesture is the reference group.
* $p < .05$. *** $p < .001$.

transformation proof of 210.61 ($p < .000$), showing a large odd ratio. These results give clear evidence that valid proof performance is strongly associated with dynamic gestures and transformational speech (H3).

Expertise

To examine the difference in gesture production between experts and nonexperts, we conducted three chi-square tests of independence with Yates' continuity correction. First, we analyzed whether expert status and the production of representational gestures were independent of one another. A significant relationship was found, $\chi^2(1) = 14.75$ ($p < .001$), with experts more likely to produce representational gestures (36%) than nonexperts (17%). Second, we investigated whether expert status and the production of nondynamic representational gesture were independent. Though the proportion of nondynamic gesture was higher in nonexperts (65%) than in experts (57%), this difference was not significant, $\chi^2(1) = 2.34$ ($p = .127$). Third, we examined the relationship

Table 7
Results of the Logistic Regression Predicting Transformational Proof

Variable	β	*SE*	*OR*	*p* value
Model 1: Main effects				.000***
(Intercept)	−3.43	0.77	0.03	
Expert[a]	1.28	0.41	3.60	.001***
Verbal	−0.02	0.02	0.98	.321
Spatial	0.28	0.08	1.32	.000***
Ethnicity1 (White)[b]	0.25	0.55	1.28	.643
Ethnicity2 (Asian)[b]	−0.33	0.58	0.72	.569
Sex[c]	−0.21	0.32	0.81	.497
Model deviance (*df*)	431.8 (353)			
Model 2: Main effects with gesture				
(Intercept)	−6.74	1.70	0.00	.000***
Expert[a]	0.47	0.52	1.60	.372
Verbal	0.01	0.02	1.01	.813
Spatial	0.16	0.10	1.17	.112
Ethnicity1 (White)[b]	−0.12	0.73	0.89	.866
Ethnicity2 (Asian)[b]	0.18	0.77	1.20	.819
Sex[c]	0.41	0.41	1.51	.317
Dynamic gestures[d]	2.56	0.36	12.94	.000***
Nondynamic gestures[e]	2.92	1.41	18.54	.162
Model deviance (*df*)	415.9 (351)			
Model 3: Main effects with gesture and speech				
(Intercept)	−6.57	1.68	.001	.000***
Expert[a]	0.22	0.54	1.24	.691
Verbal	0.01	0.02	1.01	.819
Spatial	0.15	0.10	1.16	.142
Ethnicity1 (White)[b]	0.03	0.74	1.03	.964
Ethnicity2 (Asian)[b]	0.40	0.79	1.49	.613
Sex[c]	0.25	0.43	1.28	.555
Dynamic gestures[d]	5.35	1.36	210.61	.000***
Nondynamic gestures[e]	2.51	1.37	12.30	.012
Situation model-cohesion	0.43	0.17	1.54	.007**
Verbs	0.34	0.20	1.40	.089
First-person pronouns	−0.67	0.23	0.51	.002**
Model deviance (*df*)	384.0 (348)			

Note. *SE* = standard error; *OR* = odds ratio. The raw regression coefficients (β) can be transformed into odds ratios by exponentiating the coefficient.
[a] Nonexpert participant is the reference group. [b] "Other" ethnicity is the reference group. [c] Female is the reference group. [d] No dynamic gesture is the reference group. [e] No nondynamic gesture is the reference group.
** $p < .01$. *** $p < .001$.

between expertise and dynamic representational gesture production. This relationship was significant, $\chi^2(1) = 27.34$ ($p < .001$), with experts more likely to produce dynamic gestures (79%) than nonexperts (52%).

In summary, experts perform significantly more representational gestures and, more specifically, experts produce more dynamic representational gestures than nonexperts. As the relative odds of transformational proof increases in relation to the production of at least one dynamic gesture (see previous section), these combined results support the hypothesis that experts' gestures, specifically dynamic representational gestures, may be more strongly associated with valid proof performance than nonexperts' gestures (H4a).

Gesture

The best fitting model for nondynamic gesture indicated that expertise ($p = .004$), sex ($p < .001$), and an increase in verbs ($p < .001$) and second-person pronouns ($p = .035$) were significantly associated with nondynamic gesture (see Table 8). Expertise was associated with an increase in the relative odds of producing at least one nondynamic gesture of 1.19. Males were associated with a decrease in the relative odds of nondynamic gesture of 0.81. Additionally, an increase of verbs and second-person pronouns were associated with an increase in the relative odds of nondynamic gesture of 1.10 and 1.05, respectively.

The same model was also found to be the best fitting model for dynamic gesture (see Table 9). Results showed that dynamic gesture was significantly associated with expertise ($p = .031$), an increase in spatial ability scores ($p = .002$), an increase in comparative connectives use ($p = .031$), and an increase in second-person pronouns ($p < .001$) during proof production. Expertise was associated with an increase of the relative odds of producing a dynamic gesture of 1.15. Increased spatial ability scores were associated with an increase in the relative odds of producing at least one dynamic gesture of 1.04. Similarly, an increase in comparative connectives and second-person pronouns was associated

Table 8
Results of the Logistic Regression Predicting Nondynamic Gesture

Variable		β	*SE*	*OR*	*p* value
Random component – Participant ID variance	0.84		0.92		.023*
Random component – Conjecture variance	0.01		0.03		.048*
(Intercept)		0.59	0.06	1.81	.000
Expert[a]		0.18	0.06	1.20	.004**
Spatial		0.02	0.01	1.02	.077
Sex[b]		−0.21	0.05	0.81	.000***
Comparative connectives		0.01	0.02	1.01	.702
Verbs		0.09	0.02	1.10	.000***
Second-person pronouns		0.05	0.02	1.05	.035*
First-person pronouns		−0.04	0.02	0.96	.123
Model deviance (*df*)	354.4 (350)				

Note. *SE* = standard error; *OR* = odds ratio. The raw regression coefficients (β) can be transformed into odds ratios by exponentiating the coefficient.
[a] Nonexpert participant is the reference group. [b] Female is the reference group.
* $p < .05$. ** $p < .01$. *** $p < .001$.

with an increase of the relative odds of dynamic gesture by 1.05 and 1.10, respectively.

Discussion

The analysis of one's gestures during mathematical reasoning tasks offer insights into the relationship between body-based processes and cognitive processes that can advance our understanding of mathematical cognition and math education. It was with these goals in mind that we investigated the role of gesture production during geometric reasoning.

Summary of Findings

Our investigation was guided by a primary research question: *Is geometric reasoning associated with participants' simulated actions?* Secondarily, we wanted to know: *Does the strength of the relationship between geometric reasoning and simulated action depend on participants' mathematical expertise?* We address these two questions in turn with respect to our findings.

Simulated actions offer an alternative to computational accounts of intellectual behavior. Performance on mental rotation tasks (Shepard & Cooper, 1982; Shepard & Metzler, 1971) illustrates this phenomenon by showing first that response times are strongly correlated with the angular displacement, as though during mental rotation participants are continuously turning the objects just as one would do it manually; and second, that performing physical rotation interfered with mental rotation performance, and even slowed mental rotation when a manual rotation tasks was directed to be performed more slowly (Wexler, Kosslyn, & Berthoz, 1998).

As Hostetter and Alibali (2008, p. 497; also see Hostetter & Alibali, 2019) describe,

> Thinking about a particular concept, for example, involves a perceptual and motor simulation of the properties associated with that concept, even when no exemplar of the concept is present in the current perceptual environment (Barsalou, 1999).

Our investigation of the interrelationship of geometric reasoning and simulated action centered on evidence of gesture and speech

Table 9
Results of the Logistic Regression Predicting Dynamic Gesture

Variable		β	*SE*	*OR*	*p* value
Random component – Participant ID variance	1.03		1.02		.048*
Random component – Conjecture variance	0.45		0.67		.026*
(Intercept)		0.08	0.06	1.09	.188
Expert[a]		0.14	0.07	1.15	.031*
Spatial		0.04	0.01	1.04	.002**
Sex[b]		−0.10	0.06	0.91	.086
Comparative connectives		0.05	0.02	1.05	.031*
Verbs		0.03	0.02	1.03	.182
Second-person pronouns		0.09	0.02	1.10	.000***
First-person pronouns		−0.01	0.02	0.98	.554
Model deviance (*df*)	380.1 (350)				

Note. *SE* = standard error; *OR* = odds ratio. The raw regression coefficients (β) can be transformed into odds ratios by exponentiating the coefficient.
[a] Nonexpert participant is the reference group. [b] Female is the reference group.
* $p < .05$. ** $p < .01$. *** $p < .001$.

production during geometric reasoning. In support of Hypothesis H1, intuition accuracy was reliably associated with representational production. As observed (Tables 6, 7 and 8), participants who performed representational gestures and produced transformational speech were more likely to report the correct intuition (H1), verbalize the correct insight (H2), and produce a mathematically valid proof (H3) than participants who did not perform any representational gestures. As predicted, static representational gestures— gestures such as making and tracing shapes that could be used to identify nondynamic properties of the objects in question—were associated with higher intuition and insight performance. The relationship of representational gestures with performance showed effect sizes in the medium to large range. Proof performance was most strongly associated with the production of dynamic gestures—those that simulated operations that could be used to explore generalized properties of objects—with very large odds ratios ($OR = 12.94$, without speech; $OR = 210.61$, with speech in the model). These odds ratios were even greater than those for spatial ability and expertise. Indeed, the inclusion of dynamic representational gestures led to variables for both expertise and spatial ability dropping out of the model for proof.

As expected, we also found that transformational speech was reliably associated with insight (H2) and proof performance (H3). Recall that transformational speech categories were used to tag utterances that captured participants' point of view and that identified the cohesion of one's situation models regarding the actions performed on mathematical objects, although verb use did not significantly contribute to the models of geometric reasoning. Notably, gesture production was reliably associated with insight and proof, even when controlling for speech content, indicating that gestures and speech each make independent contributions to modeling performance. That gestures may carry important information about how students think and what they know above and beyond what students say has important implications for assessment, which we explore further below. Taken together, these findings suggest that geometric reasoning is associated with participants' simulated actions, as exhibited by representational gesture and transformational speech.

We sought to also understand how expertise modulated these observed effects. As expected, experts outperformed nonexperts in each of the three performance measures, intuition, insight and proof. Furthermore, experts were more likely to produce representational gestures than nonexperts, and, in particular, experts were more likely to produce dynamic representational gestures. Thus, in our competing hypotheses, we found that geometry performance for experts was more strongly associated with gesture production than for nonexperts, in support of Hypothesis H4a, and in conflict with H4b. These findings, coupled with those showing that dynamic gesture production may, to some extent, take the place of expertise, suggests that interventions that elicit mathematically relevant dynamic representational gestures might benefit nonexperts. We revisit this point later in this section.

Limitations of the Current Study

This study has several limitations that are important to take into account when interpreting the reported findings. First of all, this was a correlational study. Consequently, we must be cautious in drawing any causal inferences about these statistical associations. These relationships provide valuable empirically based hypotheses for exploring the causal link between gesture production and mathematical reasoning. One of the paradigms for this is to inhibit or otherwise engage gesture production and examine the effect. This research has shown mixed results, with several studies showing significantly degraded performance on reasoning (e.g., Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow et al., 2001), but others showing no reliable impact (Walkington, Chelule, Woods, & Nathan, 2019).

A second limitation is how we conceptualized expertise. To have similar levels of maturity, we drew from a population of undergraduates enrolled in education courses, and a small number of graduate student ($n = 5$) to fill out the expert group. Although this represents a highly selective group of young adults who have gained admission to a competitive public university, it is certainly possible to study the views of people far more experienced in proof practices (e.g., Marghetis, Edwards, & Núñez, 2014). Similarly, it is very likely that there are similar age people for whom academic tasks such as geometric reasoning is even less familiar. A broader sampling might well reveal even greater group differences that would likely generalize better to the population at large.

Implications and Conclusions

The evidence presented suggests that some facets of mathematical thinking are embodied, and that people use body-based processes such as gestures and transformational speech as an important aspect of their task performance. Still, the correlational nature of this study cannot endorse a causal claim about the role that gestures play in mathematical thinking. The patterns of gesture production may be a manifestation of other processes that are themselves revealed through these movements and not causally related to or constituents of the reasoning processes themselves. This suggests that intervention studies that both prompt and restrict gesture production are important.

Gesture scholars, particularly McNeill (1992), Kendon (2004), and Goldin-Meadow (2005), describe systems of gesture identification. This study contributes to gesture studies research by expanding the contexts in which people invoke dynamic and nondynamic gestures, and extending the applicability of these forms to geometry, beyond their earlier instantiations in calculus (Garcia & Infante, 2012), and mental rotation (Göksun et al., 2013). For example, when students work collaboratively while engaged in geometry proof tasks, dynamic gestures again appear to play a significant role in successful reasoning, even when these embodied representations and operations are distributed across the hands and arms of multiple students (Walkington, Chelule, et al., 2019). A recent review (Williams-Pierce et al., 2017) noted that an embodied perspective on proof practices in mathematics might extend our understanding of mathematical cognition in two important ways. First, the actions made by students can influence students' mathematical reasoning through action-cognition transduction, which, by inducing cognitive states through actions, can improve students' understanding of the mathematical ideas. Findings from other studies have provided evidence that suggest gestures influence one's geometric reasoning. In one study, making specified body shapes led to improved understanding of angular measure for elementary grade students (Smith et al., 2014). In a 2-week classroom intervention, high school students engaged in body movements showed greater learning gains than those engaged in mathematically comparable non–body-based activities in tests of understanding similarity (Smith, 2018). In an intervention involving a movement based video game, mathematically relevant actions fostered greater geometric reasoning for high school students than matched, but mathematically irrelevant actions, but only for those students who were already predisposed to producing the gestures (Walkington, Nathan, & Wang, 2020). The current study further contributes to this emerging body of work by showing that spontaneously produced gestures are implicated as well.

Second, it is valuable to attend to students' body movements such as their gestures when students engage in proof production because these movements contribute to a richer assessment of student thinking than is provided by speech alone. This is especially true of students' use of dynamic gestures, which indicate ways students employ transformational proof schemes as they reason about the generalizability of mathematical conjectures. Gestures appear to carry important information about how students think and what they know, above and beyond what students say. One compelling finding in this regard is how young children's gestures can signal their "leading edge" of cognitive development. Those children whose gestures and speech were discordant while addressing the cognitive disequilibrium of a Piagetian water conservation task were found to be more receptive to training in the conservation concept (Church & Goldin-Meadow, 1986). Others (e.g., Pier et al., 2019) have shown statistically that gestures make unique contributions to models of mathematical performance. In this current study we also observed ways that gestures independently contributed to the models of geometry performance, even when controlling for expertise, spatial reasoning, and speech content. This suggests that neglecting gestures when assessing student performance may underpredict students' conceptual understanding. An implication from this is that attending to information conveyed through learners' gestures offers important information to educators for making valid formative assessments. Introducing this idea, and the perceptual training that may need to support this, could provide a rich new channel in which teachers can assess student understanding and provide adaptive instruction.

References

Abrahamson, D. (2009). Embodied design: Constructing means for constructing meaning. *Educational Studies in Mathematics, 70,* 27–47. http://dx.doi.org/10.1007/s10649-008-9137-1

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 358–376). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139519526.022

Abrahamson, D., & Trninic, D. (2015). Bringing forth mathematical concepts: Signifying sensorimotor enactment in fields of promoted action. *ZDM, 47,* 295–306. http://dx.doi.org/10.1007/s11858-014-0620-0

Alibali, M. W., Church, R. B., Kita, S., & Hostetter, A. B. (2014). Embodied knowledge in the development of conservation of quantity: Evidence from gesture. In L. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives on gesture and embodiment in mathematics* (pp. 27–50). Charlotte, NC: Information Age Press.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523. http://dx.doi.org/10.1006/cogp.1993.1012

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44,* 169–188. http://dx.doi.org/10.1006/jmla.2000.2752

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,* 247–286. http://dx.doi.org/10.1080/10508406.2011.611446

Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In A. J. Bishop, S. Mellin-Olsen, J. Van Dormolen (Eds.), *Mathematical knowledge: Its growth through teaching* (pp. 173–192). Dordrecht, the Netherlands: Springer Netherlands. http://dx.doi.org/10.1007/978-94-017-2195-0_9

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

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., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4* (R package version 1.1–8). Retrieved from http://CRAN.R-project.org/package=lme4

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21,* 1605–1610. http://dx.doi.org/10.1177/0956797610385353

Bieda, K. N., & Nathan, M. J. (2009). Representational disfluency in algebra: Evidence from student gestures and speech. *ZDM, 41,* 637–650. http://dx.doi.org/10.1007/s11858-009-0198-0

Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. New York, NY: The Free Press.

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

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics- Simulations and Computations, 39,* 860–864. http://dx.doi.org/10.1080/03610911003650383

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140,* 102–116. http://dx.doi.org/10.1037/a0021790

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 257–270. http://dx.doi.org/10.1037/xlm0000168

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23,* 43–71. http://dx.doi.org/10.1016/0010-0277(86)90053-3

Clements, D., & Battista, M. (1992). Geometry and spatial reasoning. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420–464). New York, NY: Macmillan Publishing Company.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes, 27,* 594–610. http://dx.doi.org/10.1080/01690965.2011.567074 de Freitas, E., & Sinclair, N. (2014). *Mathematics and the body: Material entanglements in the classroom*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139600378

desRosiers, G., & Kavanagh, D. (1987). Cognitive assessment in closed head injury: Stability, validity and parallel forms for two neuropsychological measures of recovery. *International Journal of Clinical Neuropsychology, 9,* 162–173.

de Villiers, M. (1998). An alternative approach to proof in dynamic geometry. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 369–393). Mahwah, NJ: Erlbaum.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics, 38,* 85–109. http://dx.doi.org/10.1023/A:1003660018579

Eagan, B., Rogers, B., Serlin, R., Ruis, A. R., Arastoopour Irgens, G., & Shaffer, D. W. (2017, June). *Can we rely on IRR? Testing the assumptions of inter-rater reliability*. Paper presented at the 12th International Conference on Computer Supported Collaborative Learning, Philadelphia, PA.

Edwards, L. D. (2003, April). *A natural history of mathematical gesture*. Paper presented at the American Educational Research Association Annual Conference, Chicago, IL.

Edwards, L., Ferrara, F., & Moore-Russo, D. (2014). *Emerging perspectives on gesture and embodiment in mathematics* (International perspectives on mathematics education - Cognition, Equity & Society series). Charlotte, NC: Information Age Press.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests* (rev. ed.). Princeton, NJ: Educational Testing Service.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80,* 27–38. http://dx.doi.org/10.1093/biomet/80.1.27

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing, 13,* S161–S164. http://dx.doi.org/10.1007/s10339-012-0477-5

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science, 18,* 233–239.

Garcia, N., & Infante, N. E. (2012). Gestures as facilitators to proficient mental modelers. In L. R. Van Zoest, J.-J. Lo, & J. L. Kratky (Eds.), *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 289–295). Kalamazoo, MI: Western Michigan University.

Gibson, J. J. (1976). The myth of passive perception: A reply to Richards. *Philosophy and Phenomenological Research, 37,* 234–238. http://dx.doi.org/10.2307/2107194

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43,* 379–401. http://dx.doi.org/10.1006/jmla.2000.2714

Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing, 14,* 153–162. http://dx.doi.org/10.1007/s10339-013-0549-1

Goldenberg, E. P., Cuoco, A. A., & Mark, J. (1998/2012). A role for geometry in general education. In R. Lehrer & D. Chazan (Eds.) *Designing learning environments for developing understanding of geometry and space* (pp. 17–58). Mahwah, NJ: Erlbaum.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2,* 1–19. http://dx.doi.org/10.1515/langcog.2010.001

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20,* 267–272. http://dx.doi.org/10.1111/j.1467-9280.2009.02297.x

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12,* 516–522. http://dx.doi.org/10.1111/1467-9280.00395

Gravemeijer, K. P. (1998/2012). From a different perspective: Building on students' informal knowledge. In R. Lehrer & D. Chazan (Eds.) *Designing learning environments for developing understanding of geometry and space* (pp. 45–66). Mahwah, NJ: Erlbaum.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education, III,* 234–283.

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* (pp. 805–842). Reston, VA: National Council of Teachers of Mathematics.

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

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation, 5,* 333–356. http://dx.doi.org/10.1207/s15427633scc0504_3

Heinze, G., Polner, M., Dunkler, D., & Southworth, H. (2018). logistf: Firth's bias-reduced logistic regression (R package version 1.2–3) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/logistf/logistf.pdf

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine, 21,* 2409–2419. http://dx.doi.org/10.1002/sim.1047

Herbst, P. G. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics, 49,* 283–312. http://dx.doi.org/10.1023/A:1020264906740

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15,* 495–514. http://dx.doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review, 26,* 721–752. http://dx.doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1448–1454). Austin, TX: Cognitive Science Society.

Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The Mathematical Imagery Trainer: From embodied interaction to conceptual learning. In G. Fitzpatrick, C. Gutwin, B. Begole, W. A. Kellogg, and D. Tan (Eds.), *Proceedings of the annual meeting of the association for computer machinery special interest group on computer human interaction, human factors in computing systems* (pp. 1989–1998). New York, NY: ACM Press.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43,* 358–390. http://dx.doi.org/10.5951/jresematheduc.43.4.0358

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217. http://dx.doi.org/10.1037/0096-3445.133.2.189

Kendon, A. (2004). *Gesture: Visible action as utterance*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511807572

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511620850.011

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

Knuth, E., Choppin, J., & Bieda, K. (2009). Middle school students' production of mathematical justifications. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K–16 perspective* (pp. 153–170). New York, NY: Routledge.

Koedinger, K. R. (1998). Conjecturing and argumentation in high-school geometry students. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 319–347). London, UK: Lawrence Erlbaum.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14,* 511–550. http://dx.doi.org/10.1207/s15516709cog1404_2

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29,* 745–756. http://dx.doi.org/10.3758/BF03200477

Lakatos, I. (2015). *Proofs and refutations: The logic of mathematical discovery*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781316286425

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

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 720–733. http://dx.doi.org/10.1037/0278-7393.33.4.720

Lehrer, R., & Chazan, D. (1998/2012). *Designing learning environments for developing understanding of geometry and space*. Mahwah, NJ: Erlbaum.

Lehrer, R., Jacobson, C., Thoyre, G., Kemeny, V., Strom, D., Horvath, J., . . . Koehler, M. (1998/2012). Developing understanding of geometry and space in the primary grades. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 169–200). Mahwah, NJ: Erlbaum.

Lockhart, P. (2009). *A mathematician's lament: How school cheats us out of our most fascinating and imaginative art form*. New York, NY: Bellevue Literary Press.

Ma, J. Y., & Hall, R. (2018). Learning a part together: Ensemble learning and infrastructure in a competitive high school marching band. *Instructional Science, 46,* 507–532. http://dx.doi.org/10.1007/s11251-018-9455-3

Marghetis, T., Edwards, L. D., & Núñez, R. (2014). More than mere handwaving: Gesture and embodiment in expert mathematical proof. In L. D. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives on gesture and embodiment in mathematics* (pp. 227–246). Charlotte, NC: Information Age.

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in Cognitive Science, 5,* 299–316. http://dx.doi.org/10.1111/tops.12013

Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 67,* 1579–1596. http://dx.doi.org/10.1080/17470218.2014.897359

Martin, T. S., McCrone, S. M. S., Bower, M. L. W., & Dindyal, J. (2005). The interplay of teacher and student actions in the teaching and learning of geometric proof. *Educational Studies in Mathematics, 60,* 95–124. http://dx.doi.org/10.1007/s10649-005-6698-0

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511894664

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

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). Visuospatial working memory, executive functioning and spatial abilities. How are they related? *Journal of Experimental Psychology: General, 130,* 621–640. http://dx.doi.org/10.1037/0096-3445.130.4.621

Nathan, M. J. (2008). An embodied cognition perspective on symbols, gesture, and grounding instruction. *Symbols and Embodiment: Debates on Meaning and Cognition, 18,* 375–396. http://dx.doi.org/10.1093/acprof:oso/9780199217274.003.0018

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

Nathan, M. J. (2017). One function of gesture is to make new ideas. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Gesture studies: Vol. 7. Why gesture? How the hands function in speaking, thinking and communicating* (pp. 175–196). Philadelphia, PA: John Benjamins Publishing Company.

Nathan, M. J., & Martinez, C. V. (2015). Gesture as model enactment: The role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice, 1,* 4–37. http://dx.doi.org/10.1080/23735082.2015.1006758

Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications, 2,* 9. http://dx.doi.org/10.1186/s41235-016-0040-5

Nathan, M. J., Walkington, C., Boncoddo, R., Pier, E. L., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof. *Learning and Instruction, 33,* 182–193. http://dx.doi.org/10.1016/j.learninstruc.2014.07.001

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *Educational Studies in Mathematics, 70,* 159–174. http://dx.doi.org/10.1007/s10649-008-9150-4

Newcombe, N. S., & Shipley, T. F. (2012). Thinking about spatial thinking: New typology, new assessments. In J. S. Gero (Ed.), *Studying visual and spatial reasoning for design creativity* (pp. 179–192). New York, NY: Springer.

Ottmar, E., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on learning. *Journal of the Learning Sciences, 26,* 51–78. http://dx.doi.org/10.1080/10508406.2016.1250212

Pelavin, S. H., & Kane, M. (1990). *Changing the odds ratio: Factors increasing access to college*. New York, NY: College Board Publications.

Petrick, C., & Martin, T. (2012, April). *Mind your body: Learning mathematics through physical action*. Paper presented at the Annual meeting of the American Educational Research Association, Vancouver, Canada.

Pier, E. L., Walkington, C., Clinton, V., Boncoddo, R., Williams-Pierce, C., Alibali, M. W., & Nathan, M. J. (2019). Embodied truths: How dynamic gestures and speech contribute to mathematical proof practices. *Contemporary Educational Psychology, 58,* 44–57. http://dx.doi.org/10.1016/j.cedpsych.2019.01.012

Provost, A., Johnson, B., Karayanidis, F., Brown, S. D., & Heathcote, A. (2013). Two routes to expertise in mental rotation. *Cognitive Science, 37,* 1321–1342. http://dx.doi.org/10.1111/cogs.12042

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making, 7,* 332–359.

Schenck, K. E., Michaelis, J. E., Swart, M. I., Xia, F., Nathan, M. J., & Walkington, C. A. (2020, April). *Mathematical knowledge is embodied: Synergistic contributions of gesture and speech during geometry proof production*. Paper presented at the 2020 Annual Meeting of the American Educational Research Association, San Francisco, CA.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York, NY: MacMillan.

Shapiro, L. (2019). *New problems of philosophy: Embodied cognition* (2nd ed.). New York, NY: Routledge/Taylor & Francis Group. http://dx.doi.org/10.4324/9780203850664

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703. http://dx.doi.org/10.1126/science.171.3972.701

Shoval, E. (2011). Using mindful movement in cooperative learning while learning about angles. *Instructional Science, 39,* 453–466. http://dx.doi.org/10.1007/s11251-010-9137-2

Sinclair, N., & de Freitas, E. (2014). The haptic nature of gesture: Rethinking gesture with new multitouch digital technologies. *Gesture, 14,* 351–374. http://dx.doi.org/10.1075/gest.14.3.04sin

Smith, C. P. (2018). Body-based activities in secondary geometry: An analysis of learning and viewpoint. *School Science and Mathematics, 118,* 179–189. http://dx.doi.org/10.1111/ssm.12279

Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior, 36,* 95–108. http://dx.doi.org/10.1016/j.jmathb.2014.09.001

Soto-Johnson, H., & Troup, J. (2014). Reasoning on the complex plane via inscriptions and gesture. *The Journal of Mathematical Behavior, 36,* 109–125. http://dx.doi.org/10.1016/j.jmathb.2014.09.004

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38,* 289–321.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28,* 187–204. http://dx.doi.org/10.1016/S0361-476X(02)00007-3

Walkington, C., Boncoddo, R., Williams, C., Nathan, M. J., Alibali, M. W., Simon, E., & Pier, E. L. (2014). Being mathematical relations: Dynamic gestures support mathematical reasoning. In W. Penuel, S. A. Jurow, and 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.

Walkington, C., Chelule, G., Woods, D., & Nathan, M. J. (2019). Collaborative gesture as a case of extended mathematical cognition. *The Journal of Mathematical Behavior, 55,* 100683. http://dx.doi.org/10.1016/j.jmathb.2018.12.002

Walkington, C. A., Nathan, M. J., & Wang, M. (2020, April). *The effect of relevant directed arm motions on gesture usage and proving of geometry conjectures*. Paper presented at the 2020 Annual Meeting of the American Educational Research Association, San Francisco, CA.

Walkington, C., Woods, D., Nathan, M. J., Chelule, G., & Wang, M. (2019). Does restricting hand gestures impair mathematical reasoning? *Learning and Instruction, 64,* 101225. http://dx.doi.org/10.1016/j.learninstruc.2019.101225

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68,* 77–94. http://dx.doi.org/10.1016/S0010-0277(98)00032-8

Williams-Pierce, C., Pier, E. L., Walkington, C., Boncoddo, R., Clinton, V., Alibali, M. W., & Nathan, M. J. (2017). What we say and how we do: Action, gesture, and language in proving. *Journal for Research in Mathematics Education, 48,* 248–260. http://dx.doi.org/10.5951/jresematheduc.48.3.0248

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

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 358,* 593–602. http://dx.doi.org/10.1098/rstb.2002.1238

Woods, D., & Fassnacht, C. (2012). Transana (Version 2.52) [Computer software]. Retrieved from http://transana.com

Yackel, E., & Hanna, G. (2003). Reasoning and proof. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school* (pp. 227–236). Reston, VA: NCTM.

Yoon, C., Thomas, M. O., & Dreyfus, T. (2011). Gestures and insight in advanced mathematical thinking. *International Journal of Mathematical Education in Science and Technology, 42,* 891–901. http://dx.doi.org/10.1080/0020739X.2011.608861

Zander, T., Öllinger, M., & Volz, K. G. (2016). Intuition and insight: Two processes that build on each other or fundamentally differ? *Frontiers in Psychology, 7,* 1395.

Zhang, Z., Lei, Y., & Li, H. (2016). Approaching the distinction between intuition and insight. *Frontiers in Psychology, 7,* 1195. http://dx.doi.org/10.3389/fpsyg.2016.01195

Received November 4, 2019
Revision received August 7, 2020
Accepted August 8, 2020 ■