# Hear What They Say and Watch What They Do: Predicting Valid Mathematical Proofs Using Speech and Gesture

Elizabeth Pier, University of Wisconsin-Madison, epier@wisc.edu
Candace Walkington, Southern Methodist University, cwalkington@smu.edu
Caroline Williams, University of Wisconsin-Madison, ccwilliams3@wisc.edu
Rebecca Boncoddo, Central Connecticut State University, boncoddo@ccsu.edu
Jessica Waala, Martha Wagner Alibali and Mitchell J. Nathan, University of Wisconsin-Madison,
waala@wisc.edu, mwalibali@wisc.edu, mnathan@wisc.edu

**Abstract:** In mathematics, practices of proof are notoriously difficult for learners to adopt. In prior work, we found that when providing verbal justifications, learners' speech patterns predict whether their justifications are mathematically sound. However, current views on the embodied nature of cognition suggest that actions and speech may co-constitute reasoning processes. The current study investigated whether the *gestures* learners use while formulating proofs also predict proof validity. 120 undergraduates provided verbal justifications for two mathematical tasks. We analyzed speech patterns in participants' justifications using text analysis software, and we coded participants' gestures as *dynamic* or *static*. Results showed that dynamic gestures were correlated with mathematically valid proofs and with specific patterns of speech. A stepwise logistic regression model found that speech and gesture separately account for unique variance in a model predicting proof validity, indicating that gesture contributes to mathematical reasoning in an abstract task domain.

## Introduction

Recent research has highlighted ways that mathematical cognition is *embodied*, or formulated through perception and action, and has specifically identified gesture as an important cognitive resource for mathematics learning. In mathematics classrooms, teachers and students spontaneously use gesture to connect disparate mathematical representations and ideas (Alibali et al., 2014; Alibali & Nathan, 2012; Nathan et al., 2013). Teachers' strategic use of gestures can support appropriate mathematical generalizations (Radford, 2003), and attending to students' gestures is important when evaluating their mathematical reasoning (Boncoddo et al., 2013; Nemirovsky & Ferrera, 2009; Williams et al., 2012). The present study is part of an emerging research program investigating how gesture and action can influence students' mathematical cognition. We identify the different types of speech and gestures that students use as they generate and communicate mathematical proofs. We examine how these speech and gesture categories relate to mathematically valid proofs in order to better understand how learners can be supported in adopting proof practices.

The ICLS theme, "Learning and Becoming in Practice," is particularly relevant for this line of investigation. Recent *in situ* classroom research has shown that teachers and students in mathematics and engineering classrooms spontaneously use gesture to enact connections across representations (Alibali et al., 2014; Nathan et al., 2013). In some cases, by using gesture and body-based action, they can effectively *become* the mathematical objects and relations they intend to convey. For example, students and teachers often use their bodies to dynamically simulate mathematical objects and relations, such as representing changes in linear slopes with an arm (Alibali & Nathan, 2012). When the body is used as a tool for depicting and understanding a fluid , moving transformation of a mathematical entity, we refer to this as a *dynamic depictive gesture* (Walkington et al., this volume).

In this work, we examine how the production both of dynamic depictive gestures and of particular speech patterns (e.g., "if… then" statements) is associated with valid proofs. Our goal is to understand the complex interplay between learners' use of gesture and speech as they produce mathematical justifications. While many studies have examined discourse patterns or gesture patterns in isolation, research on how these two modalities interact and work together to form complex patterns of interpersonal, embodied communication is more rare (though, see Radford, 2003). Here, we take a novel approach in which we use quantitative measures of speech and gesture to predict accuracy of justifications for mathematical tasks. The emergence of recent software tools for automated text analysis, coupled with the identification of dynamic gestures during mathematical reasoning from video-based classroom studies, gives us leverage in understanding how speech and gesture contribute to proof practices.

## Theoretical Framework

### The Importance of Proof Practices in Mathematics
Both the Common Core State Standards (2010) and the National Council of Teachers of Mathematics (NCTM) Standards (2000) identify constructing mathematical arguments and engaging in proof practices as crucial components of mathematical understanding. In addition, constructing valid proofs is central to both the professional domain of mathematics and mathematics education (Schoenfield, 1994), as it is the means by which mathematicians communicate key ideas and novel understandings to each other. However, student difficulties with proof are well documented (e.g., Harel & Sowder; Dreyfus, 1999; Knuth, Choppin, & Bieda, 2009); therefore, there is a need to explore how to better identify and support valid proof practices. We follow Harel and Sowder's (2005) definition of proof as the process of reasoning about a conjecture and communicating the believed truth or falseness of the conjecture, and consequently we use the terms *proof* and *justification* interchangeably. According to Harel and Sowder (2005), mathematically valid proofs must be (1) general in nature, such that they hold for all cases, (2) based on logical inferences in which conclusions are drawn from valid premises, and (3) based on operational thought, in which there is progression through goals and subgoals. Harel and Sowder's (2005) taxonomy focuses on traditional methods of communicating proof with written and spoken language. Although mathematical proofs often are verbalized instead of formally written in K-12 classrooms, the content and structure of students' speech alone may not provide a complete picture of students' mathematical knowledge. In particular, studies have shown that people often express information in gestures that they do not express in speech (Alibali & Goldin-Meadow, 1993; Alibali & Kita, 2010; Nathan & Johnson, 2010; Williams et al., 2012). We therefore argue that the body is an important modality in which complementary or even novel mathematical information can be communicated. In this work, we extend Harel and Sowder's (2005) framework by invoking views of embodied cognition to include gestures as an additional means of supporting both mathematical cognition and mathematical communication.

### Gestures and Embodiment in Proof Practices
Theories of embodied cognition reject traditional views of cognition as based on manipulating amodal symbol systems in the brain based on purely syntactic rules (Shapiro, 2011). In contrast, within an embodied framework, cognition comprises a mutual feedback loop between the brain and the body. For example, gestures may reveal cognitive processes, and the process of gesturing in turn affects those cognitive processes (Alibali & Nathan, 2012; Goldin-Meadow, Cook, & Mitchell, 2009; Hostetter & Alibali, 2008; Shapiro, 2011). Several embodied cognition scholars (Barsalou, 2008; Hostetter & Alibali, 2008) acknowledge the role of gesture as simulated action that reenacts perceptual, motor, and mental states that arise from prior experience. In this work, we investigate whether certain types of gesture are associated with valid proof construction. We draw upon work by Göksun, Goldin-Meadow, Newcombe, and Shipley (2013), who categorized gestures used in a mental rotation task as "static" if they referred to individual objects, and "dynamic" if they indicated movement such as rotation or direction. Newcombe and Shipley (2012) also differentiate static gestures, those that *indicate* spatial features or locations, versus dynamic gestures that *transform* spatial features or locations. We build upon these definitions by operationalizing simulated actions as *static depictive gestures* if the hands display a fixed, unchanging mathematical object and *dynamic depictive gestures* if the hands are used to enact transformations on mathematical entities and relations. We refer to these in the rest of the paper simply as "dynamic" or "static" gestures, but we are referring to the dynamic and static nature of the *objects*, and not whether or not the gesture itself contains movement. We make this distinction because we believe that dynamic gestures are closely related to learners' use of simulated action to represent multiple cases and fluid transformations between depicted entities, which can promote sophisticated mathematical reasoning. Furthermore, our prior work has found a significant correlation between using dynamic gestures and generating valid mathematical justifications across six different mathematical tasks (Walkington et al., this volume).

### Use of Speech in Proof Practices
Although theories of embodied cognition underscore the importance of the body when engaging in cognitive tasks such as proof production, proofs themselves are linguistic in nature. Mathematical proofs can be thought of as a specific kind of disciplinary discourse practice (e.g., Gee, 2007; Gresalfi & Cobb, 2006), and in K-12 classrooms, proofs often take spoken—rather than formal, written—forms (Healy & Hoyles, 2000). Here we focus on mathematical proofs that are communicated through spoken language and gesture, hypothesizing that specific features of participants' speech patterns may be important for constructing valid proofs. Speech patterns can be analyzed at many levels for a variety of features. For example, we could quantitatively assess how often people use pronouns, adjectives, or words relating to different topics, such as work or socializing, when they are speaking. We could also quantify how much overlap there is between different words or sentences in a block of continuous speech, or what tense a person tends to use. Indeed, the recent emergence of software-based text-mining tools has allowed us to better understand characteristics of human speech that are associated with

learners' making gains in conceptual understanding (see Jeon & Azevedo, 2007; Williams & D'Mello, 2010). When people are asked to verbally construct mathematical justifications, some of these quantitative measures of their speech may be indicative of a well-constructed, logically sound mathematical argument. By understanding the verbal properties of valid proofs, we can better grasp how to scaffold learners to adopt successful proof practices.

## Hypotheses and Predictions

This study investigates how gestures, speech patterns, and the production of valid proofs are associated with one another. In doing so, this study seeks to deepen our understanding of whether—and how—gestures provide additional information, beyond speech, about the validity of students' verbally communicated proofs. This leads us to formulate two hypotheses: **Hypothesis 1** states that speech and gesture may constitute equivalent but distinct manifestations of the same underlying cognitive processes that result in participants' generating valid proofs. In other words, while both speech and gesture may each predict valid proofs separately, together they provide redundant measures of the same underlying construct relating to valid mathematical reasoning. Alternatively, **Hypothesis 2** claims that speech and gesture may indicate distinct underlying cognitive processes involved in proof production. For instance, gestures may reveal distinct aspects of the learner's mental action-based simulations that are not conveyed in their speech patterns. In order to test these competing hypotheses, we formulated the following research questions. **Research Question 1** asks, Is the use of dynamic gesture correlated with the validity of one's proof? This enables us to investigate whether gesture does, in fact, provide information about proof validity. **Research Question 2** asks, Are particular speech patterns correlated with the validity of one's proof? This allows us to capture what characteristics of speech tend to indicate logical and valid mathematical reasoning. **Research Question 3** asks, Are dynamic gestures correlated with the use of particular speech patterns during the proof process? If speech and gesture are highly correlated with one another, but accounting for both does not improve the model's prediction of the validity of participants' proofs, then this would provide evidence for Hypothesis 1. In order to allow for the falsification of Hypothesis 1 in favor of Hypothesis 2, **Research Question 4** asks, Do speech and dynamic gestures contribute *uniquely* to models for predicting proof validity? An affirmative response would indicate support for Hypothesis 2, which stipulates that gesture and speech reveal distinct underlying processes of proof practices.

## Methods

Undergraduates ($N$ = 120) ($M$ age = 19.2 years, 51% female) enrolled in a psychology course at a large Midwestern university were prompted to verbally justify two tasks from distinct mathematical domains: one relating to planar geometry, the other to an inference about parity in the number system. Participants read each conjecture (order was counterbalanced) out loud and were prompted to think aloud as they verbalized their justifications. The first task was, "Mary came up with the following conjecture: 'For any triangle, the sum of the length of any two sides must be greater than the length of the remaining side.' Provide a justification as to *why* Mary's conjecture is true or false." The second task was, "An unknown number of gears are connected together in a chain. If you know what direction the first gear turns, how could you figure out what direction the last gear turns? Provide a justification as to *why* your answer is true."

## Coding of Proof Accuracy

Participants were asked to think aloud as they provided their justifications, and videotapes of the sessions were uploaded into Transana, a software program that allows for transcribing and analyzing video data (Woods & Fassnacht, 2012). After transcription, the videos were split into 240 separate transcripts (one for each task for each participant). Based on the verbal transcript and video, each justification was each coded as *Valid* (1) or *Invalid* (0) based on Harel and Sowder's (2005) definition of valid proofs. Examples of spoken valid and invalid proofs for each of the two tasks are shown in Table 1. Overall, 40.8% and 50% of participants provided valid proofs for the gear and triangle tasks, respectively. Inter-rater reliability for coding the validity of participants' justifications was high, with a Kappa value of 0.84 for a 20% subset.

## Coding of Speech

The 240 transcripts were analyzed using the LiWC (Pennebaker, Booth, & Francis, 2007) and Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) text analysis software programs. LiWC comprises various dictionaries of words assigned to different topic categories, such as social process words (e.g., words relating to family or friends) and cognitive process words (e.g., words describing causation or certainty). LiWC's output consists of the percentage of words the participants used from each dictionary. LiWC provides a measure of the content of a text, whereas the second text-mining program, Coh-Metrix, analyzes the quality of a text. Coh-Metrix contains 108 different indicators of text readability (for a full list, see http://cohmetrix.memphis.edu), which are broadly organized into categories such as syntactic simplicity, word concreteness, and deep cohesion. Coh-Metrix's output provides continuous quantitative measures of the degree to which these characteristics are

present in a text. Therefore, using both software programs in conjunction allows for analysis of the technical aspects of the language and readability gathered from Coh-Metrix, as well as the content and topic of the language that participants use from LiWC, providing a more holistic picture of the nature of participants' speech. We included all speech categories from LiWC and Coh-Metrix in our analyses, with the exception of categories specific to the number of paragraphs or number of sentences, as we were using natural speech instead of written text, and these demarcations vary greatly by transcription norms.
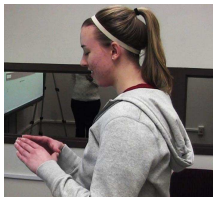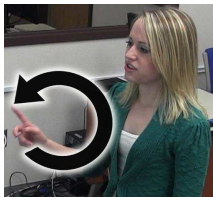
Table 1: Examples of spoken valid and invalid participant proofs from each task

| Proof Type | Example |
|---|---|
| Triangle, Valid | "Mary's conjecture is true, because if the one side is long–is longer than the sum of the other two sides then the other two sides won't be able to touch at the top and it won't be a triangle." |
| Triangle, Invalid | "That isn't true. Uh it's false because you could have a triangle where one side is very long and the other two sides are shorter, um very short, and so they add up to a length that is shorter than the longest side." |
| Gear, Valid | "Um obviously the gear after the first one turns in the opposite direction and the next one turns in the opposite direction and so on and so on, so I guess if there's an odd number of gears it will turn in the same direction as the first gear and if there's an even number of gears it'll turn in the opposite direction." |
| Gear, Invalid | "Um I feel that all the gears should turn the same way because it's a chain reaction so it should turn in the same direction as the first gear." |

## Coding of Gestures

The 240 video clips were also coded for whether participants generated dynamic or static gestures while engaging in proof activities. For both the triangle and gear tasks, we coded a gesture as *Static* if participants indicated a single entity with their hands, but coded it as *Dynamic* if they indicated multiple entities that were connected together via a depicted relationship. For the triangle task, an example of a *Static* gesture would be indicating a side of a triangle or a full triangle, whereas a *Dynamic* gesture would be fluidly depicting several different sized triangles in a row to show a transformation, or showing a single triangle that grew in size. For the gear task, an example of a *Static* gesture would be portraying a single gear moving in one direction, whereas a *Dynamic* gesture would be two gears moving in the same or in opposite directions, or a chain of gears turning in alternate directions. Note that the distinction is not between gestures that "move" versus those that "don't move," since gestures that don't move can still simulate action or perceptual change. Instead, the distinction lies in whether participants used gestures to represent multiple mathematical objects that were related to one another. Table 2 provides images of participants creating each of these gesture types. A complete clip was coded as *Dynamic* (1) if it contained any instances of dynamic gestures. A clip was coded as *Not Dynamic* (0) if it contained only static gestures or no gestures. Our Kappa reliability value for a subset of 30 video clips (12.5% subset from a total of 240 videos; 15 gear, 15 triangle) was 0.85. This was for the categories of "No Gestures," "Static Only" and "Dynamic," which were assigned holistically to the entire clip (both attempts).

Table 2: Examples of dynamic and static gestures

| Task | Example of Dynamic Gesture | | Example of Static Gesture | |
|---|---|---|---|---|
| Triangle |  | *Participant uses both hands to make two sides of a triangle and fluently moves from making a flattened triangle to a normal triangle.* |  | *Participant uses both hands to create a full triangle that does not move or change.* |
| Gear |  | *Participant uses both hands to show two gears moving in opposite directions.* |  | *Participant uses right index finger to show a single gear turning in one direction.* |

## Results and Discussion

To examine relations among gesture, speech categories, and proof validity, we calculated Pearson correlation coefficients and examined those correlations that were both in the same direction (positive or negative) and significant for both the gear and triangle tasks. The following three sections detail the results of those analyses.

### Dynamic Gestures and Valid Proofs

In order to answer Research Question 1, we investigated the correlation between dynamic gesture and proof validity. Results showed that participants who used dynamic gestures were more likely to produce valid proofs than those who did not, both on the triangle task ($r = 0.454$, $p < .001$) and on the gear task ($r = 0.255$, $p = .005$)—that is, across two distinct mathematical domains. We also looked at the correlations between static gesture and valid proofs; static gestures were significantly negatively correlated with valid proofs for the triangle task ($r = -0.183$, $p = .045$), but were not significantly related to the validity of proofs for the gear task. However, a prior exploratory study showed that the relationship between dynamic gestures and valid proofs holds across six different mathematical tasks (Walkington et al., this volume). By answering Research Question 1 affirmatively, we do not directly falsify either of our competing hypotheses, but it was important to establish the relationship between dynamic gesture and proof validity prior to answering our other research questions.

### Speech Categories and Valid Proofs

To answer Research Question 2, we investigated the correlation between proof validity and the speech categories derived from LiWC and Coh-Metrix. In prior analyses of this data set (Pier et al., 2014), we found that four speech categories were predictive in regression models of valid proofs across the two tasks: *Present tense* verbs, lexical diversity (as measured by the *type-token ratio* for content words), *discrepancy* words, and *temporal connectives*. Of those four categories, use of the *present tense* and *type-token ratio* were each negatively correlated with valid proofs, while use of *discrepancy* words and use of *temporal connectives* were each positively correlated with valid proofs. Table 3 describes each of the speech categories that were significantly correlated with valid proofs.

Table 3: Speech categories significantly correlated with valid proofs

| Speech Category | Description of Speech Category | Triangle Task | Gear Task |
|---|---|---|---|
| *Present tense*[L] | Verbs in the present tense | $r = -0.280$** | $r = -0.328$*** |
| Lexical diversity (*type-token ratio*)[C] | Ratio of the number of unique words (*types*) divided by the number of times that word occurs (*tokens*); higher ratios indicate more unique words | $r = -0.280$** | $r = -0.500$*** |
| *Discrepancy* words[L] | Words in LiWC *Discrepancy* dictionary, e.g., "could," "would," "should" | $r = 0.264$** | $r = 0.202$* |
| *Temporal Connectives*[C] | Words connecting clauses that indicate time order, e.g., "before," "then," "after" | $r = 0.265$** | $r = 0.240$** |
| *Note.* *$p < .05$, **$p < .01$, and ***$p < .001$. [L] is a measure from LiWC, and [C] a measure from Coh-Metrix. | | | |

Following the methodology used by Pier et al. (2014), we systematically investigated the 20 transcripts that scored highest and the 20 that scored lowest on each category in order to determine which aspects of the transcripts seemed to be associated with high or low scores on each of these linguistic categories. These analyses revealed that the negative correlation between proof validity and present tense was due to participants' use of self-conscious statements such as, "I don't know," or "I don't understand." The negative correlation with type-token ratio indicated that participants with high type-token ratios verbalized more varied, unrelated words, lacking continuity of ideas. In contrast, participants with a low type-token ratio repeated a consistent set of words related to reasoning through the conjecture. Finally, using both discrepancy words and temporal connectives stemmed from more conditional (i.e., "if…then") statements indicating participants were making logical inferences—one of the features of valid proofs per the proof scheme from Harel and Sowder (2005).

### Dynamic Gestures and Speech Categories

In order to answer Research Question 3, whether gesture use is correlated with speech patterns during the proof process, we investigated the correlations between dynamic gesture and speech categories from LiWC and Coh-Metrix. Although we also calculated correlations between static gesture and speech categories, none of these were both statistically significant and in the same direction (positive or negative) for both tasks. Table 4 describes each of the speech categories that were significantly correlated with dynamic gestures. Four speech categories were negatively correlated with dynamic gestures (*present tense*, *quantifiers*, *insight* words, and *cognitive processes* words), whereas three speech categories were positively correlated with dynamic gestures (*deep cohesion*, *all connectives*, and *temporal connectives*). Thus, not only are dynamic gesture and speech each

individually correlated with valid proofs, but they are correlated with one another; furthermore, many of the same categories that were found in prior work to be significantly predictive of proof validity in speech are also significantly correlated with dynamic gesture (i.e., *present tense* and *temporal connectives*) (Pier et al., 2014). However, in order to evaluate our competing hypotheses, we need to address Research Question 4: whether speech and gesture explain *unique* variance in determining proof validity, or whether they instead explain overlapping, redundant variance in proof validity.

We again systematically investigated the transcripts with the 20 highest and 20 lowest scores for each speech category. We found that when participants frequently mentioned their own mental state or cognitive processing, they were not likely to produce dynamic gestures relevant to the proofs, since they seemed to focus more on their thinking than on the content of the proof. Such patterns resulted in negative correlations between dynamic gestures and the categories of *present tense*, *insight* words, and *cognitive processes* words.

Table 4: Speech categories significantly correlated with dynamic gesture

| Speech Category | Description of Speech Category | Triangle Task | Gear Task |
|---|---|---|---|
| *Present tense*[L] | Verbs in the present tense | $r = -0.188$* | $r = -0.207$* |
| *Quantifiers* [L] | Words referring to general quantities, e.g., "all," "more," "greater" | $r = -0.184$* | $r = -0.200$* |
| *Insight words*[L] | Words in LiWC *Insight* dictionary, e.g., "think," "know," "understand" | $r = -0.214$* | $r = -0.337$*** |
| *Cognitive processes words*[L] | Words in LiWC *Cognitive Processes* dictionary, e.g., "ought," "know," "cause" | $r = -0.228$* | $r = -0.321$*** |
| *Deep cohesion*[C] | Incidence of causal and intentional connectives, indicating higher coherence | $r = 0.257$** | $r = 0.217$* |
| *All connectives*[C] | Incidence of all connective words, i.e., conjunctions | $r = 0.267$** | $r = 0.225$* |
| *Temporal connectives*[C] | Words connecting clauses that indicate time order, e.g., "before," "then," "after" | $r = 0.248$** | $r = 0.214$* |
| *Note.* *$p < .05$, **$p < .01$, and ***$p < .001$. [L] is a measure from LiWC, and [C] a measure from Coh-Metrix. | | | |

Additionally, when participants used temporal connectives and connectives more generally to cohesively link together different statements, they tended to produce dynamic gestures. Both dynamic gestures and verbal statements with connective words typically express relationships between entities; thus, reasoning about relationships relevant to the proof content was manifested in the production both of cohesive speech and dynamic gestures. This drove the positive correlations between dynamic gesture and the speech categories of *deep cohesion, all connectives*, and *temporal connectives*.

## Stepwise Regression Analysis

To answer Research Question 4, we assessed whether speech and gesture each explained unique variance in the models' predicting proof validity. We ran two stepwise logistic regression analyses using the *lmer*() function (Bates & Maechler, 2010) in the *R* software environment. In each model, participant was a random effect and task (i.e., gear, triangle) was a fixed effect. The outcome was whether the participant generated a valid proof for the task, coded as a 0 or 1. Interactions were tested for significance, but none were present in the final model. In the first analysis, we added predictors in the following order: Dynamic gesture, speech categories significantly associated with dynamic gestures, and speech categories significantly associated with valid proofs. Predictors were tested for inclusion in the model using the *anova*() function, which uses a chi-square reference distribution to test for significant reductions in deviance. In the second analysis, we added predictors in the following order: Speech categories significantly associated with valid proofs, speech categories significantly associated with dynamic gestures, and dynamic gestures. Thus, in one analysis we added terms for dynamic gestures into the model first, and in the other analysis we added dynamic gestures last into the model. Both analyses resulted in the same final model, so we only discuss the results in terms of one stepwise regression analysis. Furthermore, we included controls for the word count of each proof, but this factor was not a significant covariate.

Table 5 provides the results of the stepwise regression. Three of the speech categories were significant predictors of valid proofs: *discrepancy words* ($z = 4.21$, $p < .001$), *type-token ratio* of content words ($z = -3.08$, $p = .0020$), and *cognitive processes* words ($z = -3.29$, $p = .001$). Also, use of dynamic gestures was a significant, positive predictor of valid proofs ($z = 3.00$, $p = .003$). Adding in the speech category predictors reduced the overall deviance of the model by 24.1% ($\chi^2(3) = 74.75$, $p < .001$), and adding the dynamic gesture predictor reduced the deviance by an additional 7.0% ($\chi^2(1) = 16.44$, $p < .001$). Overall, language and gesture predictors combined accounted for approximately 30% of the variance in whether participants generated valid proofs.

The regression analyses thus answer Research Question 4 by demonstrating that dynamic gestures uniquely predict the validity of proofs, which provides support for Hypothesis 2. This means that while both

speech and gesture independently predict whether or not a participant verbalized a valid proof, there is distinct information in each modality. In other words, gesture and speech are not redundant, but instead appear to provide unique but related insight into the cognitive processes involved in successful proof construction. Thus, consideration of *both* speech and gesture is important to understanding the depths of students' understandings—and misunderstandings.

Table 5: Stepwise regression analysis results

| | Estimate | Std. Error | *z* value | *p* value | Sig. |
|---|---|---|---|---|---|
| (Intercept) | 18.28 | 8.42 | 2.17 | 0.0300 | * |
| *Discrepancy* words | 3.17 | 0.75 | 4.21 | <.001 | *** |
| *Type-token Ratio* (LDTTRc) | -49.25 | 15.97 | -3.08 | 0.0020 | ** |
| *Cognitive processes* words | -0.94 | 0.29 | -3.29 | 0.0010 | ** |
| Dynamic Gestures | 8.92 | 2.98 | 3.00 | 0.0027 | ** |
| *Note.* \*$p < .05$, \*\*$p < .01$, and \*\*\*$p < .001$. | | | | | |

## Conclusion

The findings in this study suggest that gesture and speech are not merely interchangeable windows into cognitive processes. Instead, gesture and speech each provide *unique* insight into student proof generation. From an embodied cognition perspective, these findings underscore the idea that since gesture and speech convey some distinct information, both are important components of cognitive processing to consider when evaluating student performance on a task, such as valid proof construction. In other words, although a student may not be able to specifically articulate a valid mathematical proof, his or her gestures can shed light on potential key mathematical insights that he or she possesses but is not yet able to verbalize. However, as with any study, our conclusions are limited by our measurement techniques; there may be other components of language that are important to proof construction that our tools were not able to capture, or other ways of classifying gesture that would allow for even more explanatory power. In addition, examining these patterns across more mathematical tasks would certainly be useful. Nonetheless, these findings suggest that attending to student gestures could assist teachers in diagnosing student misunderstandings when formulating mathematical proofs.

In addition, given that learners typically struggle to grasp the discursive, logical structure of explicit proof statements even with continued instruction, interventions seeking to change students' gestures may offer a feasible alternative channel for learning. Teachers could model and encourage students to use dynamic gestures in the classroom, while concurrently accentuating the importance of language-based aspects of proof. Therefore, understanding how gesture can uniquely explain proof validity can inform potential educational interventions.

Overall, this work indicates that gestures have explanatory power about students' mathematical knowledge and skills above and beyond that available from speech alone. Understanding the complex interplay between gesture and speech as reasoning unfolds is critical to supporting students in engaging in mathematics learning. By creating environments in which students engage in justification practices using both gesture and speech, we can help students to "learn and become" as they take part in mathematical practices.

## References

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.

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture, 10*(1), 3-28.

Alibali, M., & Nathan, M. (2012). Embodiment in mathematics teaching and learning: Evidence from students' and teachers' gestures. *Journal of the Learning Sciences, 21*(2), 247-286.

Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Johnson, C. V., Jacobs, S. A., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition and Instruction.*, *32(1)*, 65 – 100.

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

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

Boncoddo, R., Williams, C., Pier, E., Walkington, C., Alibali, M., Nathan, M., & Dogan, F. (2013). Gesture as a window to justification and proof. In Martinez, M., & Castro Superfine, A. (Eds.), *Proceedings of the 35th annual meeting of PME-NA* (pp. 229-236). Chicago, IL: University of Illinois at Chicago.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics, 38*(1-3), 85-109.

Gee, J. P. (2007). *Social linguistics and literacies: Ideology in discourses*. New York: Routledge.

Göksun, T., Goldin-Meadow, S., Newcombe, N.,& Shipley, T. (2013). Individual differences in mental rotation. *Cognitive Processing (Special Issue on Spatial Learning and Reasoning)*, 14, 153-162.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267-72. doi:10.1111/j.1467-9280.2009.02297.x

Graesser, A., McNamara, D., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, (36), 193–202.

Gresalfi, M. S., & Cobb, P. (2006). Cultivating students' discipline-specific dispositions as a critical goal for pedagogy and equity. *Pedagogies*, *1*(1), 49-57.

Harel, G., & Sowder, L. (1998). Students' proof schemes. *Issues in Mathematics Education, 7*, 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*, NCTM.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, *31*(4), 396–428.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514. doi:10.3758/PBR.15.3.495

Jeon, M., & Azevedo, R. (2007). Analyzing human tutorial dialogues for cohesion and coherence during hypermedia learning of a complex science topic. *In D.S. McNamara & J.G. Trafton (Eds.), Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society.

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.

Nathan, M. J. & Johnson, C. V. (2010, August). The relation of situation models to gesture production when learning from a scientific text. *Poster presented at the Twentieth Anniversary Meeting of the Society for Text & Discourse,* Chicago, IL.

Nathan, M.J., Srisurichan, R., Walkington, C., Wolfgram, M., Williams, C., & Alibali, M. (2013). Building cohesion across representations. *Journal of Engineering Education*, 77-116.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics.* Reston, VA: NCTM. Retrieved from http://www.nctm.org/standards/content.aspx?id=26792.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common Core State Standards: Mathematics. Washington D.C..

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *Educational Studies in Mathematics*, *70*(2), 159-174.

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. New York, NY: Springer.

Pennebaker, J. W., Booth, R. J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX: LIWC.net.

Pier, E. L., Walkington, C. A., Clinton, V., Boncoddo, R., & Nathan, M. J. (2014). Exploring computerized text analysis to predict the validity of students' proof construction. *Paper presented at the 2014 American Educational Research Association Annual Meeting and Exhibition.* Philadelphia, PA.

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

Schoenfeld, A. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior*, *13*(1), 55–80.

Shapiro, L. (2011). *Embodied Cognition*. New York: Routledge.

Walkington, C., Nathan, M., Alibali, M., Pier, E.L., Boncoddo, R., & Williams, C.C. (2013). *Projection as a mechanism for grounding mathematical justification in embodied action.* Paper presented at the 2013 American Educational Research Association Annual Meeting and Exhibition. San Francisco, CA.

Williams, C., & D'Mello, S. (2010, January). Predicting student knowledge level from domain-independent function and content words. In *Intelligent Tutoring Systems* (pp. 62-71). Springer Berlin Heidelberg.

Williams, C., Walkington, C., Boncoddo, R., Srisurichan, R., Pier, E., Nathan, M., & Alibali, M. (2012). Invisible proof: The role of gestures and action in proof. In L.R. Van Zoest, J.-J. Lo, & J.L. Kratky (Eds.), *Proceedings of the 34th annual meeting of PME-NA* (pp. 182-189). Kalamazoo, MI: WMU.

Woods, D., & Fassnacht, C., (2012). Transana v2.52. http://transana.org. Madison, WI: The Board of Regents of The University of Wisconsin System.

## Acknowledgements