Learning and Instruction 50 (2017) 65–74

Contents lists available at ScienceDirect

Learning and Instruction

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

Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention

Eliza L. Congdon[*], Miriam A. Novack, Neon Brooks, Naureen Hemani-Lopez, Lucy O'Keefe, Susan Goldin-Meadow

University of Chicago, United States

A R T I C L E I N F O

Article history:
Received 31 May 2016
Received in revised form
3 March 2017
Accepted 26 March 2017
Available online 7 April 2017

Keywords:
Gesture
Mathematics
Learning
Transfer
Multimodality

A B S T R A C T

When teachers gesture during instruction, children retain and generalize what they are taught (Goldin-Meadow, 2014). But why does gesture have such a powerful effect on learning? Previous research shows that children learn most from a math lesson when teachers present one problem-solving strategy in speech while simultaneously presenting a different, but complementary, strategy in gesture (Singer & Goldin-Meadow, 2005). One possibility is that gesture is powerful in this context because it presents information simultaneously with speech. Alternatively, gesture may be effective simply because it involves the body, in which case the timing of information presented in speech and gesture may be less important for learning. Here we find evidence for the importance of simultaneity: 3rd grade children retain and generalize what they learn from a math lesson better when given instruction containing simultaneous speech and gesture than when given instruction containing sequential speech and gesture. Interpreting these results in the context of theories of multimodal learning, we find that gesture capitalizes on its synchrony with speech to promote learning that lasts and can be generalized.

© 2017 Elsevier Ltd. All rights reserved.

Gestures are hand movements, often accompanying spoken language, that are meaningful and convey information to listeners (see Goldin-Meadow, 2003; Goldin-Meadow & Brentari, 2015; Kendon, 2004; McNeill, 1992). For example, a gesture can represent the approximate size of an object, point to or reference an important part of the visual environment, or demonstrate a mimed action, such as how to rotate an object in space. Across a wide range of academic domains and age groups, students learn better from spoken instruction that includes gesture than from spoken instruction that does not include gesture (e.g., Goldin-Meadow & Singer, 2003; Macedonia, Müller, & Friederici, 2011; Ping & Goldin-Meadow, 2008; Singer & Goldin-Meadow, 2005; Valenzeno, Alibali, & Klatzky, 2003). Instruction that combines speech with gesture is particularly beneficial in helping learners generalize what they learn to new problems and retain that understanding over time (e.g., Cook, Duffy, & Fenn, 2013; Goldin-Meadow, 2014). Most research on gesture and learning to date has focused on *whether* gesture-based instruction is beneficial to learners. Here, we move beyond this question to ask *how* incorporating gesture into instruction improves learning.

The first possibility is that gesture is a powerful learning tool because it is an action performed by the hands and is thus an instance of embodied cognition (e.g., Glenberg, 2008; Niedenthal, 2007; Raymond & Gibbs, 2006; Smith, 2005; Wilson, 2002; Clark, 2007). For example, when individuals move their hands in a problem-solving scenario, they do well with respect to learning (Brooks & Goldin-Meadow, 2016), retention (Cook, Mitchell, & Goldin-Meadow, 2008), and insight (Thomas & Lleras, 2009). In these cases, gesture may be helping learners use their own bodies to create an enriched representation of a problem grounded in physical metaphors (see Alibali & Nathan, 2012; Cook et al., 2008; Hostetter & Alibali, 2008; Nathan, 2008). Importantly, learners can also learn from other peoples' gestures, perhaps because observing gesture engages the observer's motor system (e.g., Macedonia et al., 2011; Ping, Goldin-Meadow, & Beilock, 2014; Wakefield, James, & James, 2013).

But gesture's impact on learning may result not only from the fact that it is produced by the body, but also from the fact that it is produced simultaneously with speech. Gesture's ability to co-occur with speech allows instructors to convey two separate, yet complementary, messages *at the same time* (e.g., Goldin-Meadow &

---

[*] Corresponding author. Department of Psychology, University of Chicago, Chicago, IL 60637, United States.

E-mail address: econgdon08@gmail.com (E.L. Congdon).

http://dx.doi.org/10.1016/j.learninstruc.2017.03.005
0959-4752/© 2017 Elsevier Ltd. All rights reserved.

Singer, 2003; Goldin-Meadow, Cook, & Mitchell, 2009; Singer & Goldin-Meadow, 2005). Under this hypothesis, temporal synchrony between speech and gesture instruction during a learning episode is necessary to promote the best learning outcomes.

There are several pieces of evidence that provide support for this hypothesis. First, speech and gesture have been argued to share a common cognitive origin that forms an integrated system in communicative contexts (e.g., Loehr, 2007; McNeill, 1992). In natural discourse, gesture tends to be initiated just prior to speech production, and statistical analysis of naturalistic conversations reveals tight rhythmic synchrony between gesture and speech (e.g., McNeill, 1992). This tight relationship seems to be specific to gesture——speech is *more* closely synchronized with gesture than with other types of movement (e.g., action on objects, Church, Kelly, & Holcombe, 2014). And if the synchrony between speech and gesture is artificially disrupted by a mere 360 ms, overall comprehension is impaired (Habets, Kita, Shao, Özyurek, & Hagoort, 2011).

In addition, Mayer's (2005) work on multimedia learning suggests that any instruction that takes advantage of dual channels of information processed simultaneously (visual/pictorial *and* auditory/verbal) can be beneficial for learning. This and other dual-coding theories of processing are based on the assumption that humans have limited processing capacity in any single input channel, but can increase their overall cognitive processing load by taking in information in two channels at once (Baddeley, 1986, 1999; Chandler & Sweller, 1991). Importantly for the current paper, students learn most effectively when verbal and visual information are temporally and spatially integrated rather than presented separately (Mayer, 2002). This idea, known as the Contiguity Principle, has been used to suggest that learners can make more meaningful connections and better integrate verbal and pictorial input when the two are presented at the same time, as long as neither channel is subject to processing overload (Mayer & Moreno, 2003). Gesture, while not pictorial in the same way as other traditional instructional images, relies on the visual channel and thus may be effective because it can present information in a non-verbal format simultaneously with verbal information.

To date, all studies exploring the effects of instructor gesture on student learning have presented gesture simultaneously with speech and thus do not put the simultaneity hypothesis to the test. Here we directly explore whether simultaneity with speech is necessary for an instructor's gesture to have an effect on learning by manipulating the timing of an instructor's speech and gesture.

Our study design builds on the work of Singer and Goldin-Meadow (2005). In their study, all participants were given instruction containing either one or two correct strategies for solving difficult missing-addend mathematical equivalence problems (e.g., $4 + 3 + 7 = \_\_ + 7$). These types of math problems are an important precursor to algebra, and predict math fluency when controlling for IQ, race, socioeconomic status, and gender (McNeil, Fyfe, & Dunwiddie, 2015). Children in the United States as old as 4th grade have a fragile or non-existent understanding of the equals sign and consistently solve this type of problem incorrectly (e.g., Falkner, Levi, & Carpenter, 1999; Ginsburg, 1989; Saenz-Ludlow & Walgamuth, 1998). In Singer and Goldin-Meadow's most successful training condition, children were simultaneously presented with one strategy in speech and a different, but complementary, strategy in gesture. In speech, they heard the Equalizer (EQ) strategy, a principle that focuses on the idea that the two sides of an equation must be equal. In gesture, children saw a series of hand movements representing the Add-Subtract (AS) strategy, an algorithm in which all the addends on the left side of the equation are added, and then the number on the right side is subtracted from that total to get the answer. Children learned significantly more from this "mismatching" instruction containing EQ in speech and AS in gesture, compared to instruction containing these same two strategies, EQ and AS, presented entirely in speech without any gesture (and also compared to instruction containing only one of the strategies, EQ, presented in speech and gesture).

Learning a conceptual strategy (like EQ) along with a procedural strategy (like AS) in instruction leads to more flexible and generalizable learning than focusing on only one type of strategy (Baroody, 2003; Rittle-Johnson & Siegler, 1998). Gesture may be particularly good at promoting an understanding of EQ in relation to AS because the simultaneous presentation of two messages (which is not possible within the spoken modality, but often happens across speech and gesture, Goldin-Meadow, 2003) facilitates integration and comprehension of the two messages, leading to better learning outcomes than sequential presentation of the same two messages. Alternatively, gesture's ability to co-occur and be seamlessly integrated with speech may *not* heighten its power as a teaching tool. If so, presenting two complementary ideas across two modalities without integrating them temporally may be just as beneficial for learning as presenting the ideas simultaneously. In fact, sequential presentation of speech and gesture could be even more effective for learning since it might allow children time to independently focus on, and process, incoming information from each of the two modalities.

We tested these alternatives by presenting children with two complementary strategies for solving a mathematical equivalence task: EQ and AS. Following Singer and Goldin-Meadow (2005), we gave one group both strategies in speech, which were necessarily presented sequentially (S➔S), and we gave a second group the strategies across modalities, EQ in speech and AS in gesture, presented simultaneously (S✚G). To determine whether gesture's power as a teaching tool comes from its ability to be simultaneously produced with speech, we added a third group who also received EQ in speech and AS in gesture, but the two strategies were presented sequentially (S➔G). Given previous research on gesture's power to promote generalization and retention over time (Cook et al., 2013), we were specifically interested in whether the benefits of gesture would emerge (1) after a delay of either one day or one month, and (2) on generalization problems, either immediately or after a delay.

The instruction procedure was based on Singer and Goldin-Meadow (2005), but in order to ensure that the conditions were equal in terms of clarity of instruction, we made several modifications to the instruction procedure. In the original study, children in the S➔S condition heard one speech strategy on the first presentation of a problem, and then heard the second speech strategy on a second presentation of the same problem. Separating the two presentations in this way may have discouraged the children from integrating the two strategies. To avoid this potential confound, in both sequential conditions (S➔S and S➔G), we gave the two strategies within a single presentation of the math problem, with no break in between the two strategies, just as we did in the simultaneous condition (S✚G). Singer and Goldin-Meadow (2005) found a difference in learning between the two groups in their study, our S✚G and S➔S groups, immediately after instruction. Because we were primarily interested in the impact of gesture on generalization and retention, and because we were concerned that presenting speech and gesture sequentially (S➔G) might be off-putting (since it is an unusual way to present the strategies), we decided to try to increase initial learning across the groups by providing experimenter feedback on all problems. We reasoned that if children in all conditions were to perform similarly on the problems solved immediately after the lesson, we could then be confident that children can learn from an unusual lesson (i.e., from S➔G) and that any subsequent differences found across the groups in their performance on generalization problems or retention over a delay do not stem from any one group's failure to attend to and comprehend the lesson initially, but rather from differences across the groups in how the lesson was processed and internalized.

1. Method

1.1. Participants

One hundred and three third-grade children were tested in schools in the Chicagoland area in a pretest, training, posttest design. To ensure that all participants were naïve to the material, children were excluded if they solved any of six pretest problems correctly ($n = 30$). In addition, one child was excluded because he was not fluent in English. The remaining 72 children (45 girls and 27 boys) ranged in age from 8.38 to 10.29 years ($M = 9.10$ years, $SD = 0.41$ years). The population was identified by a parent questionnaire as primarily Hispanic (75%, with 4% non-Hispanic and 21% unreported) and was socioeconomically diverse based on reports of parental education (14% of parents had less than a high school degree; 43% had a high school degree or GED; 21% had some college; 5% had a college degree or higher; 17% of parents did not report education levels). The sample size (24 participants per condition) was chosen to match the sample size range reported in Singer and Goldin-Meadow ($n = 24–31$ per condition). Schools were recruited through phone calls and e-mails to principals. Children who participated had submitted written consent from their parents prior to the study. Children received small prizes for participating, and the teachers of participating classrooms received gift certificates to a learning store.

1.2. Design and procedure

Children were tested individually at school across three separate days. The initial testing day included a pretest, one of three between-subjects instruction conditions, and immediate tests of basic and generalization problems (described in more detail below). Retention was measured with follow-up visits the following day and four weeks after the initial testing day. Follow-up tests included basic and generalization assessments (see Fig. 1). Experimenter A administered the pretest, immediate test, and both follow-up tests. Experimenter B administered the instruction. Both experimenters were blind to the study hypotheses.

1.2.1. Pretest

Children first completed a paper and pencil pretest containing six problems in a basic format that assessed knowledge of mathematical equivalence. Basic format problems included two different forms presented in a mixed order. On three problems (ABC form), the last number on the left side of the equation was repeated on the right side (e.g., $a+b+c=\_\_+c$). On the remaining three problems (PQR form), the first number on the left side was repeated on the right side (e.g., $p+q+r=p+\_\_$). After children solved all six basic problems, Experimenter A turned on a video camera to begin recording the session. She then wrote each problem, together with the child's answer, on a dry-erase board one at a time and asked children to explain how they got each answer. Children were not given feedback about whether or not their answer was correct.

1.2.2. Instruction

For the instruction task, children were randomly assigned to one of three conditions: Speech-then-Speech (S➔S) ($n = 24$; $M_{age} = 9.03$ years; eight boys), Speech-then-Gesture (S➔G) ($n = 24$; $M_{age} = 9.02$ years; 12 boys), or Speech-with-Gesture (S✚G) ($n = 24$; $M_{age} = 9.24$ years; seven boys). In all conditions, the experimenter and child solved four new basic problems on the dry-erase board. The experimenter wrote a problem on the board leaving the answer blank, and then explained how to correctly solve the problem using the gesture and speech that was specific to the condition (described in further detail below). The child was then asked to solve the same problem and to explain his or her answer. The child was told whether the answer was right or wrong, but was not given the correct answer if the answer was incorrect. The child's solution was then erased and the experimenter repeated the lesson on the same problem a second time. The child was then asked to solve the problem again, but this time was not asked to explain the answer and was not given feedback. Thus, on each of the four instructional trials, children observed the instruction twice and had the opportunity to try solving the same problem twice themselves. Children rarely gestured during training aside from occasional finger-counting.

1.2.3. Training conditions

In the S➔S condition, the experimenter provided two strategies in speech with no accompanying hand movements. The first strategy, the Equalizer (EQ) strategy, explains the conceptual principle that the two sides of the equation must be equal. The second strategy, the Add-Subtract (AS) strategy, describes a procedural algorithm for adding up numbers on the left side and subtracting the number on the right side. For example, for the problem $7+8+5=\_\_+5$, the experimenter said, “I want to make one side equal to the other side. You see, seven plus eight plus five is twenty, and fifteen plus five is twenty. Okay? Or we can also solve this problem by adding seven plus eight plus five which equals twenty, and then subtracting the other five from twenty to get fifteen as the answer.”

As in the S➔S condition, in the S➔G condition, the experimenter

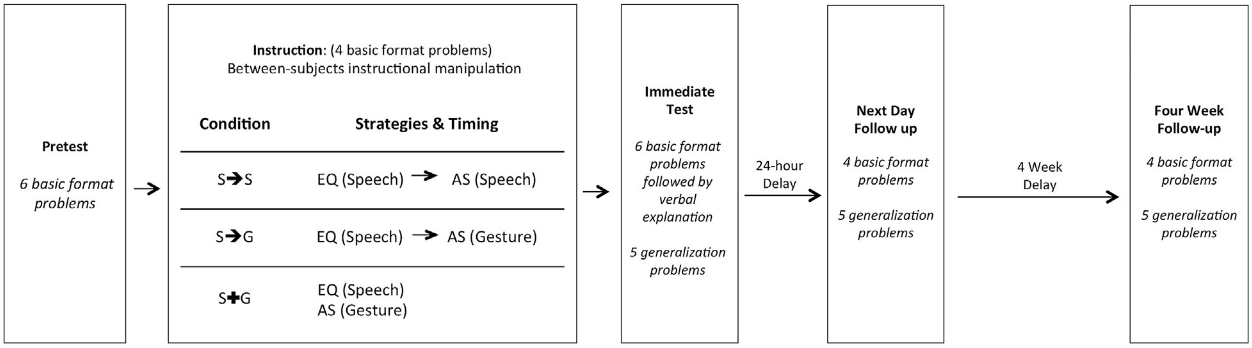

Fig. 1. Diagram of the experimental design.

first explained how to solve the problem using the EQ strategy in speech. The experimenter then produced the AS strategy in gesture with no accompanying speech (see Fig. 2).

In the S✚G condition, the experimenter explained how to solve the problem using the EQ strategy in speech while simultaneously performing the AS strategy in gesture. All instruction sessions were video recorded.

1.2.4. Immediate test: basic problems and explanations

After instruction, children were seated and given a six-question paper and pencil test that was identical in format to the pretest. During this portion of the session, the video recorder was paused as children were sitting quietly and not generating audio or visual data (children did not spontaneously gesture during any of the paper-and-pencil tests). This immediate test included three ABC form and three PQR form problems to measure how well children had learned to solve the types of problems that they saw during training. The specific problems on the pretest and immediate test were counterbalanced across participants. As in the pretest, children first completed the paper and pencil problems, and then were asked to explain how they got their answers on the dry-erase board with Experimenter A. Children's explanations of these basic problems contained unscripted speech and gestures. These explanations were videotaped and the speech was later coded according to a previously established system (Perry, Church, & Goldin-Meadow, 1988).

Coders classified children's explanations by strategy. For example, a child was given the code EQ if her explanation indicated an understanding that both sides of the equation must equal the same amount (e.g., “I put down ten because three plus seven is ten, plus five is fifteen, and [*indicating the other side*] ten plus five is fifteen”). A child was given the code AS if her explanation described adding one side of the equation, and subtracting a number from the other side (e.g., “I added six, five, three and took away six and it equaled seven”). Other correct strategies, such as noticing that two of the addends can be grouped and added (e.g., “I added four plus five to get nine) were also coded, and were combined for our analyses. Children also provided incorrect strategies, such as adding up one side of the equation and putting that sum in the blank, or adding together all of the numbers in the problem. Three independent coders coded each strategy. Inter-rater reliability was high (Kappas between pairs of coders ranged from 0.81 to 0.89) and disagreements were settled by discussions between the first two authors.

1.2.5. Immediate test: generalization problems

Following the immediate test with basic problems, each child was given a paper-and-pencil generalization test, which was not video or audio recorded. This test contained five math problems, which differed in format from the problems children saw during training. Problems differed from the training problems in up to two ways – some had a blank space on the left side of the equation rather than on the right (e.g., $\_\_ + 7 = 3 + 9 + 7$), others did not contain a pair of equal addends; (e.g., $5 + 3 + 7 = 4 + \_\_$), and some had both of these characteristics (e.g., $4 + \_\_ = 6 + 5 + 2$).

1.2.6. Follow-up tests

All children were given follow up assessments one day after instruction, and again four weeks after instruction. The assessment at each time point was a pencil and paper worksheet that contained 4 basic problems and 5 generalization problems.

2. Results

2.1. Analytical approach

We analyzed performance separately for each problem type at each time point. Each child's response to each problem on the written tests was categorized as correct or incorrect. Because individual children often got either all or none of a given problem type correct, the data were strongly bimodal and thus non-normally distributed. Accordingly, we conducted all of our analyses as binomial regression models testing the probability that a given response would be correct or incorrect, with participant as a random effect to account for common variance between problems answered by the same participant. Due to the highly bimodal nature of the data, this random effect of participant accounted for a large proportion of the variance in the data in all reported models. For ease of visual interpretation, both Table 1 and Fig. 3 report the raw average proportion of correctly answered problems aggregated

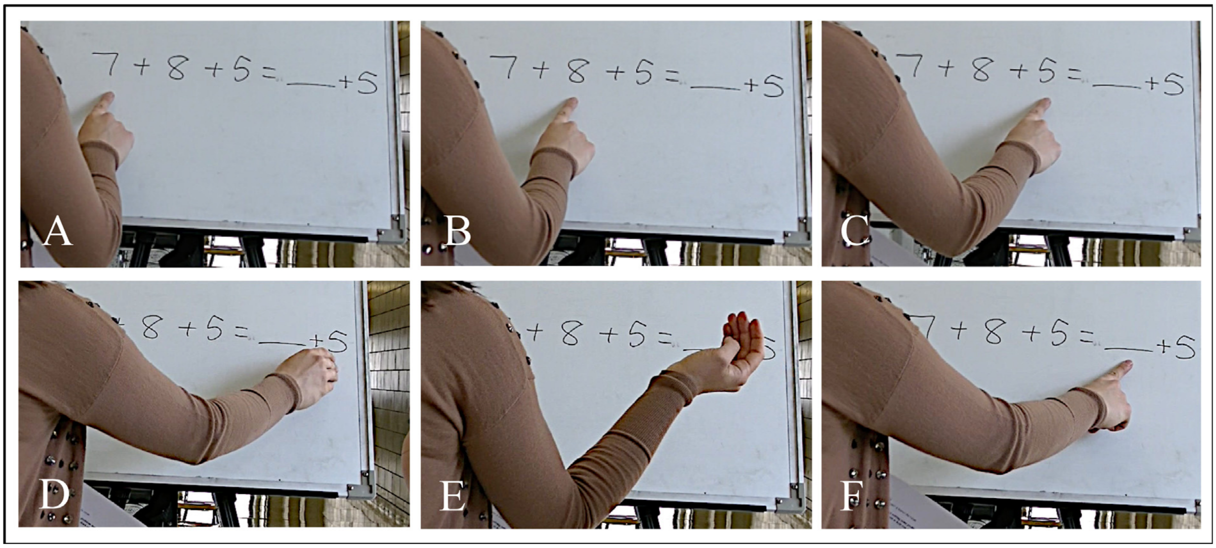

Fig. 2. Still frames taken from video of the AS gesture. The experimenter points to each addend on the left (A, B, C), then pulls away from the addend on the right (D, E), and finally points to the blank space (F).

Table 1
Average proportion correct for each problem type and time point.

Condition	Immediate Test		Next Day Follow-Up		Four Week Follow-Up	
	Basic	Generalization	Basic*	Generalization*	Basic*	Generalization*
S➔S	0.41 (0.49)	0.41 (0.49)	0.33 (0.47)	0.35 (0.47)	0.35 (0.47)	0.28 (0.45)
S➔G	0.42 (49)	0.44 (0.49)	0.39 (0.49)	0.39 (0.49)	0.40 (0.49)	0.35 (0.47)
S✚G	0.50 (0.50)	0.55 (0.49)	0.57 (0.49)	0.59 (0.49)	0.60 (0.49)	0.64 (0.48)

Note. Standard deviations are displayed in parentheses below the group means. Asterisks indicate tasks on which there was a significant main effect of condition.
$^*p < 0.05$.

across individual participants, and do not account for shared variance within individuals to the same extent as the binomial models. The overall patterns of the data with respect to the outcomes of interest do not change with or without the random effect of participant. For each model we report, we present chi-square values that represent the overall effect of condition in each analysis, as well as beta values that reflect relative differences between conditions.

Table 1 displays the raw average proportion correct for each training condition at each time point, separated by problem type. Recall that the data were non-normally distributed (bimodal), which is reflected in the large standard deviation of each average score.

2.2. Initial differences in learning

We first asked whether the lessons in each condition were equally clear and comprehensible. To make this assessment, we looked specifically at performance on the basic problems at immediate test as a measure of improvement. Recall that all children included in the sample got zero problems correct before instruction. There was improvement on the basic problems after instruction in all three conditions (mean proportions correct reported in Table 1). A mixed-effects binomial regression with condition as a fixed effect and participant as a random effect showed no significant differences in performance at immediate test among any of the three conditions (Main effect of condition: χ^2 (2) $= 0.95$, $p = 0.62$; Relative beta values: S➔S vs. S➔G: $\beta = 0.40$, $z = 0.26$ $p = 0.80$; S➔S vs. S✚G: $\beta = 1.43$, $z = 0.93$ $p = 0.35$; S✚G vs. S➔G: $\beta = 1.0$, $z = 0.70$, $p = 0.49$). This lack of difference on basic problems immediately after instruction suggests that any differences between conditions found on generalization and retention problems are unlikely to stem from baseline differences in surface-level comprehension of the lesson.

Next, we compared children's performance on the generalization problems at the immediate test. A regression model predicting the probability of a correct answer by condition with participant as a random effect showed no significant differences in performance on novel problem types at immediate test among the three conditions (Main effect of condition: χ^2 (2) $= 2.03$, $p = 0.36$; Beta values: S➔S vs. S➔G: $\beta = 0.90$, $z = 0.59$, $p = 0.56$; S➔S vs. S✚G: $\beta = 2.26$, $z = 1.41$ $p = 0.16$; S✚G vs. S➔G: $\beta = 1.36$, $z = 0.90$, $p = 0.37$).

2.3. Next day follow-up

We then examined children's performance on the basic problems on the next day follow-up (see Table 1 for means). Here, unlike performance at the immediate test, there was a significant effect of condition (χ^2 (2) $= 42.29$, $p < 0.001$), with children in the S✚G condition performing better than children in both the S➔S condition ($\beta = 17.76$, $z = 6.17$, $p < 0.001$) and the S➔G condition ($\beta = 17.26$, $z = 5.98$, $p < 0.001$). There was no significant difference in performance between participants in the S➔S and S➔G conditions ($\beta = 0.50$, $z = 0.25$, $p = 0.80$).

Performance on the generalization problems on the next day follow-up showed a similar pattern as the basic problems. We again found a significant effect of condition (χ^2 (2) $= 31.72$, $p < 0.001$), with children in the S✚G condition performing better than children in both the S➔S condition ($\beta = 16.27$, $z = 5.56$, $p < 0.001$) and the S➔G condition ($\beta = 15.54$, $z = 4.87$, $p < 0.001$). There was no significant difference in performance between participants in the S➔S and S➔G conditions ($\beta = 0.72$, $z = 0.35$, $p = 0.73$).

Thus, although there were no differences in performance immediately after training, after a one-day delay we found a significant advantage for children in the S+G condition relative to the two sequential training conditions on both the basic and generalization problem types, but no advantage for children in the S➔G condition compared to the S➔S condition.

2.4. Four-week follow-up

Finally, we considered children's performance on the basic and generalization problems four weeks after instruction (see Table 1 for means). On basic problems, there was a main effect of condition (χ^2 (2) $= 32.09$, $p < 0.001$): Children in the S✚G condition performed better on the four-week test than children in both the S➔S condition ($\beta = 16.28$, $z = 5.49$, $p < 0.001$), and the S➔G condition ($\beta = 15.75$, $z = 5.15$, $p < 0.01$). Children in the S➔S and S➔G conditions did not differ significantly in performance on basic problems ($\beta = 0.53$, $z = 0.26$, $p = 0.79$).

Performance on the generalization problems also showed a main effect of condition (χ^2 (2) $= 10.76$, $p < 0.01$). Children in the S✚G condition continued to outperform children in both the S➔S condition ($\beta = 4.19$, $z = 3.18$, $p < 0.001$) and the S➔G condition ($\beta = 4.85$, $z = 2.51$, $p < 0.01$). Children in the S➔S and S➔G conditions did not differ significantly in performance ($\beta = 1.14$, $z = 0.54$, $p = 0.58$). Thus, after a four-week delay, children performed significantly better on both problem types after receiving simultaneous S✚G instruction than after receiving either of the sequential instructions.

2.5. Exploring emerging differences across time

These results suggest that the differences in training conditions emerge only after a delay. However, it is unclear whether this emerging difference is driven by an increase in performance in the S✚G group over time, a decrease in performance in the S➔S and S➔G groups over time, or both. To address this question, we collapsed across problem type to examine how children's performance changed across the three sessions. Fig. 3 presents the proportion of correct responses on all problems for each condition (S➔S; S➔G; S✚G) at each time point (immediate test; next day follow-up; four week follow-up). An initial model with condition, time point, and their interaction as fixed effects, and participant as a random effect, revealed a significant interaction between condition and time point (χ^2 (4) = 19.99, $p < 0.001$). Neither the main effect of time point, nor the main effect of condition was significant on its own (time point: χ^2 (2) = 4.68, $p = 0.10$; condition: χ^2 (2) = 2.71, $p = 0.26$).

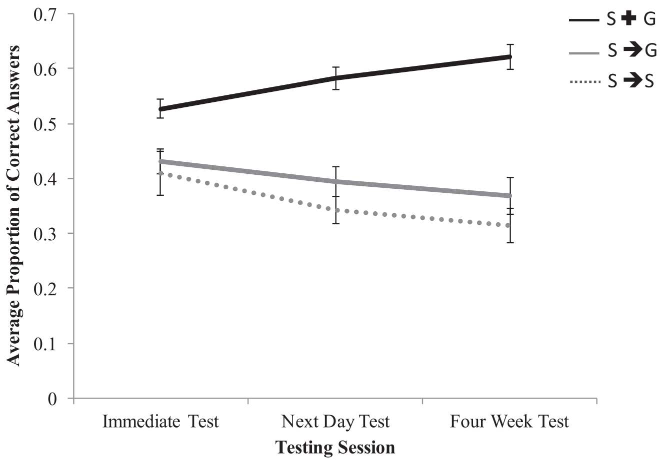

Fig. 3. Change in performance over time by condition. Error bars indicate within-subject standard errors of the mean.

To explore this interaction, we fitted a series of mixed-effects logistic regression models to data from each training condition to predict children's performance as a function of time point with participant as a random effect. Within the S✚G condition, children's performance improved over time: Children in this condition performed significantly better at next day follow-up than at immediate test ($\beta = 0.82$, $z = 2.46$, $p = 0.01$), and marginally better at four week follow-up than at immediate test ($\beta = 0.64$, $z = 1.92$ $p = 0.055$). In contrast, children's performance in the two sequential conditions, S➔G and S➔S, did not improve, and, at some timepoints, worsened. For children in the S➔G condition, performance did not differ between immediate test and next day follow-up ($\beta = 0.42$, $z = -1.43$, $p = 0.15$), but worsened significantly between immediate test and four week follow-up ($\beta = -0.67$, $z = -2.14$, $p = 0.032$). For children in the S➔S condition, performance worsened significantly between immediate test and next day follow-up ($\beta = -0.78$, $z = -2.44$, $p = 0.01$) and between immediate test and four week follow-up ($\beta = -1.10$, $z = -3.2$, $p = 0.001$). For all three conditions, there were no significant differences between the next-day and the four-week follow-ups.

Taken together, our results suggest that gesture's ability to be presented simultaneously with speech is an important aspect of its effect on learning. Of the three instruction groups, children in the S✚G condition were best able to retain and generalize the knowledge they had gained during instruction. This effect emerged most strongly after the longest delay. The lack of significant differences between the S➔S and the S➔G conditions suggests that using multiple modalities *per se* is not likely to be driving the benefits observed in the S✚G training condition.

2.6. The effect of instruction on spoken explanations: immediate test speech strategies

Although there were no significant differences in immediate performance (i.e., number of correct solutions) on basic problems across conditions, there were differences in the types of problem-solving strategies that children in the three training groups spontaneously produced in speech when asked to explain these solutions (see Fig. 4). Most of the children in the S➔S condition produced the AS strategy (68% of total explanations in this condition) to explain their correct answers to the basic problems on the immediate posttest. This preference may reflect a recency effect, as all of the children in this condition heard the AS strategy after the EQ strategy during training. A plurality of children in the S➔G condition produced the EQ strategy (41%), which they heard first in speech during training, but many also produced other correct strategies (20%) that were not explicitly taught in either speech or gesture (e.g., the grouping strategy——group and sum the addends on the left side of the equation that do not appear on the right and put the sum in the blank). Note that none of the children in this condition produced the AS strategy (which was presented in gesture), even though they experienced it second (and thus most recently) during instruction. In contrast, a number of children in the S✚G condition did produce instances of the AS strategy (14%), which they, too, saw in gesture. Children in this condition were most likely to use the EQ strategy (70%). Indeed, in a regression model predicting EQ strategy use by condition, we found that children in the S✚G condition produced the EQ strategy significantly more than children in either the S➔S condition ($\beta = 2.54$, $t = 4.19$, $p < 0.001$) or the S➔G condition ($\beta = 1.54$, $t = 2.54$, $p = 0.01$).

Differences in how often the EQ strategy was used on the immediate posttest are particularly intriguing simply because children in all three conditions heard the EQ strategy equally often during training. Nevertheless, the children reproduced the EQ strategy at different rates after training. We took the rate at which children used the EQ strategy at immediate posttest as a measure of how well they had learned the principle of equivalence (i.e., that the two sides of an equation must be equal) during training. Given the differences in strategy usage by condition, and the condition differences in generalization and retention, we hypothesized that the relation between training condition and success on both the generalization problems and basic problems at later time points may have been mediated by how often children incorporated the EQ strategy into their problem-solving repertoires. We conducted a post-hoc mediation analysis to explore this possibility.

We began by establishing that our data met the conditions needed for a mediation analysis. We found, first, that condition was a significant predictor of a composite score of performance on all of the generalization problems and the two sets of basic problems administered at the follow-up sessions ($\beta = 3.13$, $t = 2.49$, $p = 0.01$, Fig. 5A), controlling for performance on the basic problems at posttest, which was also a significant predictor of the composite score ($\beta = 3.11$, $t = 13.46$, $p < 0.001$). We next found that condition was a significant predictor of how often the EQ strategy was produced in speech when explaining answers to basic problems on the immediate posttest ($\beta = 2.04$, $t = 3.84$, $p < 0.001$, Fig. 5B, arrow on the left). Finally, we found that the rate at which the EQ strategy was produced on the immediate posttest was a significant predictor of the composite performance score ($\beta = 1.15$, $t = 3.95$, $p < 0.001$, Fig. 5B, arrow on the right), controlling for performance on basic problems at posttest ($\beta = 2.65$, $t = 11.06$, $p < 0.001$).

Given these three necessary pieces of information, we then performed a mediation analysis to test whether use of the EQ strategy mediated the relation between training condition and performance. We found that the effect of condition became non-significant when EQ strategy production was included in the analysis ($\beta = 1.02$, $t = 0.813$, $p = 0.41$, Fig. 5B, arrow on the bottom), suggesting that use of EQ in the explanations following training

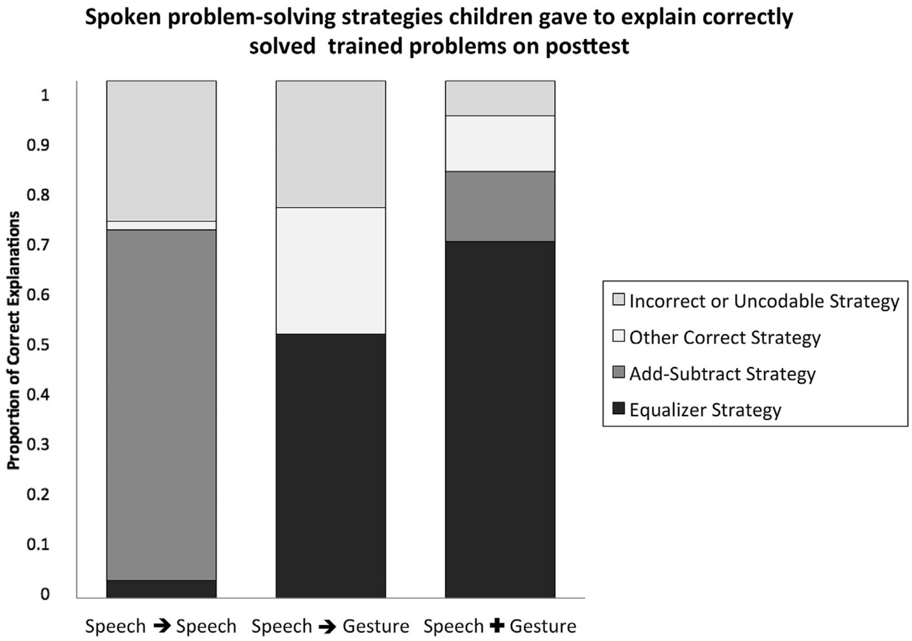

Fig. 4. Proportion of explanations of correctly solved basic problems on the immediate posttest that contained each of the following problem-solving strategies in speech: Add-Subtract, Equalizer, Other Correct, Incorrect/Uncodable.

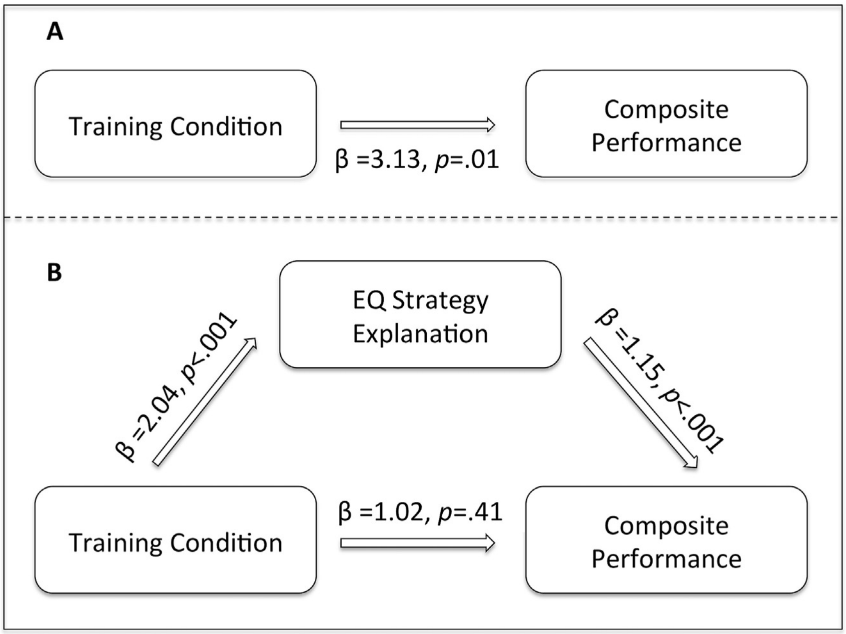

Fig. 5. Depiction of the mediation analysis. Part A shows that training condition predicts the composite performance score (generalization tests at all three time points and basic problems at the next day and 4-week follow-up time points). Part B shows the significant relationship (1) between training condition and EQ strategy use in speech immediately after training, and (2) between EQ strategy use and the composite performance score, and confirms that adding EQ strategy use into the model mediates the effect of training condition on performance.

may have had a mediating impact on children's lasting learning effects. A bootstrapping estimation approach with 1000 samples (Preacher & Hayes, 2004) confirmed that there was a significant indirect effect of condition on the composite performance score through use of the EQ strategy produced prior to all of the problems included in creating the composite score (point estimate = 2.35, 95% confidence interval = 0.740 to 4.784). These findings suggest that presenting the EQ strategy simultaneously with the AS strategy (which happened only in the S✚G condition) may have encouraged children to incorporate the broad, conceptual EQ strategy into their procedural understanding of how to solve the problems, which, in turn, may have led to a deep, generalizable, and long-lasting understanding of mathematical equivalence.

Although these mediation results are based on a post-hoc analysis of children's speech, they suggest that presenting speech and gesture simultaneously may have been powerful in this study because it encouraged children to develop a deeper understanding of the conceptual equalizer principle and to integrate that principle with the algorithmic add-subtract problem solving strategy.

3. Discussion

Speech and gesture are often assumed to have a common cognitive origin and to form an integrated system in communicative contexts (e.g., Loehr, 2007; McNeill, 1992). Consistent with this framework, our data demonstrate that, in a learning context, input received simultaneously in speech and gesture had a bigger effect on learning than the same input received sequentially in the two modalities. When presented at the same time, speech and gesture appeared to encourage learners to simultaneously attend to and integrate ideas conveyed in the two modalities and thus create long-lasting and flexible new concepts. These findings emphasize the importance of timing in dual-channel processing, and suggest that one key to the efficacy of gesture instruction lies in gesture's ability to be temporally synchronized with speech.

One tenet of successful dual-channel processing is that learners make connections between the information encoded in the verbal and non-verbal systems (Paivio, 1986). Children's spoken explanations of basic problems on the immediate test provide suggestive evidence for this type of conceptual integration during simultaneous speech and gesture instruction. We found that children in the S✚G condition were more likely to use the EQ strategy to explain their correct solutions on the immediate test than children in the other two conditions, even though all children heard the EQ strategy during instruction. A post-hoc mediation analysis suggested that condition effects were mediated by this use of EQ at immediate test. That is, increased use of the EQ strategy during posttest was associated with optimal learning outcomes, and this flexible understanding was best promoted by the simultaneous presentation of AS (in gesture) and EQ (in speech) during training.

In light of these findings, one might guess that children would learn most effectively if presented with *only* the EQ strategy in speech (or with a matching EQ strategy in gesture) during instruction. However, Singer and Goldin-Meadow (2005) found that when children heard EQ in speech alone or accompanied by EQ in gesture, they did *not* learn as much as when they heard EQ in speech while seeing AS in gesture. Thus, the most parsimonious explanation of the current data is that learners in our S✚G condition are, on some level, integrating representations of the two complementary strategies across modalities to arrive at a more robust understanding of the equalizer principle than would be obtained from exposure to the EQ strategy alone.

Simultaneous speech and gesture may have been more effective than sequential speech and gesture because the speech made it easier for children to glean meaning from gesture by providing context. Even adults have a difficult time interpreting another person's gestures as meaningful if there is too little context to scaffold the movements (Novack, Wakefield, & Goldin-Meadow, 2016). Children in the S➔G training condition may not have spontaneously mapped the AS gesture onto the context provided by the preceding speech. That is, they may not have connected the AS gesture to either the preceding spoken EQ strategy or the mathematical equation written on the board. Instead, they may have seen the AS gesture as a meaningless movement rather than as movement produced to convey information. This hypothesis is consistent with the fact that not a single child in the S➔G condition translated the gesture-based AS strategy into speech in their immediate posttest explanations, whereas 14% of children in the S✚G condition produced AS in speech to explain their solutions.

In general, it may be that sequential presentation of information in our paradigm pushes children to see one of the two streams of information as irrelevant or unimportant to the task. We have some evidence that children in the sequential conditions may simply have discounted one of the two strategies——most children in the S➔S condition produced only AS in their spoken immediate posttest explanations (the strategy they heard last), and most children in the S➔G condition produced only EQ (the strategy they heard in speech). In contrast, children in the S✚G condition produced *both* AS and EQ in their spoken posttest explanations, suggesting that they took both modalities to be providing relevant information.

There are a few limitations that stem from the design of the current study. One limitation is that in both the S✚G condition and the S➔G condition, the EQ strategy was presented in speech and the AS strategy was presented in gesture. It is possible that different rates of learning may have occurred if children received the AS strategy (i.e., the algorithm) in speech, either concurrently or followed by the EQ strategy (the principle) in gesture. Future work comparing additional strategy/modality combinations can provide insight into the ways in which strategy type, modality and order may interact to influence learning. In addition, we did not have a G➔S training condition in the current design. While the idea was initially rejected for fear that presenting an initial gesture with no spoken context would seem unnatural, there is some research suggesting that such a design could be a more effective form of sequential information presentation. In comprehension of a written text, reading a passage that has low cohesion before a passage that has high cohesion is better for learning because readers are forced to use their preexisting knowledge to fill in the missing information (McNamara, 2001). If we conceptualize gesture as having lower cohesion than speech, it is possible that a G➔S condition may be more beneficial than a S➔G condition, at least for some learners (see McNamara, 2001 for a discussion of how learning from low cohesion then higher cohesion is primarily helpful for readers with high prior knowledge). Future studies may consider manipulating the order of sequential instruction to test whether this phenomenon from reading carries over to spoken and gestural communication.

Ultimately, the fact that children in the S➔G condition performed significantly worse than children in the S✚G condition, combined with the lack of significant differences between the S➔G and the S➔S conditions, suggests that the embodied nature of gesture, on its own, does not account for gesture's powerful role on this learning task. However, other research has demonstrated that gesture can have an effect on learning even when it occurs without any speech at all (e.g., Brooks & Goldin-Meadow, 2016; Chu & Kita, 2008, 2011; Cook et al., 2008). How can our current findings be reconciled with previous work? One key difference between the current study and this previous work is that, in our study, the instructor, not the learner, is producing the gesture. Gesture might need to be tightly integrated with speech when it is produced by others, but not when it is self-produced by the learner, in order for it to affect learning. It will be important in future work to ask whether temporal integration with speech is as important to learning when learners produce their own gestures as it appears to be when learners watch someone else produce the gestures.

Our data raise the question of whether any type of instruction that takes advantage of dual channels of information (see Mayer, 2005), particularly those that can be simultaneously presented (e.g., visual/pictorial information combined with auditory/verbal information), will benefit learning as much as simultaneous speech and gesture did in our study. It may be important not only that the second channel be simultaneously produced with the first, but also that it be produced in the manual modality——note that our findings do not rule out this possibility. If the manual modality does indeed play a special role in learning, as suggested by embodied cognition theory, other channels may fall short relative to gesture, even though they can be presented along with speech. Whatever the answers to these questions, it is clear from our findings that, on its own, the manual modality is not sufficient to promote generalization and retention.

4. Conclusions

Children who received simultaneous speech and gesture instruction were better able to retain and generalize a mathematics lesson than children who received the same speech and gesture presented sequentially. Indeed, when information was presented sequentially, we found no evidence that receiving spoken and gestured instruction was superior to hearing two strategies in speech. Although these findings do not rule out the possibility that gesture's impact on learning may stem, at least in part, from the fact that it is produced by the body, they do make it clear that the manual nature of gesture is not sufficient to give it its power as a teaching tool——for an instructor's gesture to facilitate deep learning, generalization, and retention over time, it must be presented simultaneously with speech.

Funding

This work was supported by the National Institutes of Health Grant R01-HD047450 and by the National Science Foundation Grant BCS-0925595 to S. Goldin-Meadow; the National Science Foundation Grant SBE-0541957 (Spatial Intelligence and Learning Center, S. Goldin-Meadow is a co-principal investigator); and a grant from the Institute of Education Sciences, U.S. Department of Education (R305 B090025) to the University of Chicago in support of E. Congdon and M. Novack. The opinions expressed are those of the authors and do not represent views of the NIH, NSF, IES or the US Department of Education.

Author contributions

E. Congdon, M. Novack, N. Brooks and S. Goldin-Meadow contributed to the study concept and design. Testing and data collection were performed by N. Hemani-Lopez and L. O'Keefe. E. Congdon, M. Novack and N. Brooks performed the data analysis and interpretation under the supervision of S. Goldin-Meadow. E. Congdon drafted the manuscript, and all authors provided critical revisions and approved the final version of the manuscript.

Acknowledgements

We wish to thank the students, teachers and principals from Chicago Public Schools who agreed to participate in this study.

References

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

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–34). Mahwah, NJ: Erlbaum.

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science, 40*(7), 1831–1849.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*, 706–723.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology. General, 140*, 102–116.

Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience, 29*(3), 345–354.

Clark, A. (2007). Re-inventing ourselves: The plasticity of embodiment, sensing, and mind. *Journal of Medicine and Philosophy, 32*(3), 263–282.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development, 84*, 1863–1871.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047–1058.

Falkner, K. P., Levi, L., & Carpenter, T. (1999). Children's understanding of equality: A foundation for early algebra. *Teaching Children Mathematics (Early Childhood Corner), 6*(4), 232–236.

Ginsburg, H. P. (1989). *Children's arithmetic- how they learn it and how you teach it*. Austin, Texas: Pro·ed.

Glenberg, A. M. (2008). Embodiment for education. *Handbook of Cognitive Science: An Embodied Approach*, 355–372.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S. (2014). How gesture works to change our minds. *Trends in Neuroscience and Education, 30*, 23.

Goldin-Meadow, S., & Brentari, D. (2015). Gesture, sign and language: The coming of age of sign language and gesture studies. *The Behavioral and Brain Sciences*, 1–82.

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

Goldin-Meadow, S., & Singer, M. A. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology, 39*, 509–520.

Habets, B., Kita, S., Shao, S., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845–1854.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review, 15*, 495–514.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Chicago, IL: The University of Chicago Press.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture, 7*, 179–214.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32*, 982–998.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85–139.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51–62.

McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of math equivalence. *Journal of Educational Psychology, 107*(2), 423–436.

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

Nathan, M. J. (2008). An embodied cognition perspective on symbols, grounding, and instructional gesture. In M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment and meaning: A debate* (pp. 375–396). Oxford, England: Oxford University Press.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005.

Novack, M. A., Wakefield, E., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition, 146*, 339–348.

Paivio, A. (1986). *Mental representations*. New York: Oxford University Press.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3*, 359–400.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44*, 1277–1287.

Ping, R., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General, 43*, 195–204.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717–731.

Raymond, G., & Gibbs, J. (2006). *Embodiment and cognitive science*. New York: Cambridge University.

Rittle-Johnson, B., & Siegler, R. S. (1998). The relationship between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75–110). Hove, UK: Psychology Press.

Saenz-Ludlow, A., & Walgamuth, C. (1998). Third Graders interpretations of equality and the equal symbol. *Educational Studies in Mathematics, 35*, 153–187.

Singer, M., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85–89.
Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25*(3), 278–298.
Thomas, L. E., & Lleras, A. (2009). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin & Review, 16*, 719–723.
Valenzeno, L., Alibali, M. A., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28*(2), 187–204.
Wakefield, E. M., James, T. W., & James, K. H. (2013). Neural correlates of gesture processing across human development. *Cognitive Neuropsychology, 30*, 58–76.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.