Received: 3 July 2017 | Accepted: 13 February 2018

DOI: 10.1111/desc.12664

PAPER

WILEY Developmental Science

Gesture helps learners learn, but not merely by guiding their visual attention

Elizabeth Wakefield[1,2] | Miriam A. Novack[1,3] | Eliza L. Congdon[1,4] | Steven Franconeri[3] | Susan Goldin-Meadow[1]

[1]Department of Psychology, University of Chicago, Chicago, IL, USA

[2]Department of Psychology, Loyola University, Chicago, IL, USA

[3]Department of Psychology, Northwestern University, Evanston, IL, USA

[4]Department of Psychology, Bucknell University, Lewisburg, PA, USA

Correspondence
Elizabeth Wakefield, Department of Psychology, Loyola University Chicago, Chicago, IL, USA.
Email: ewakefield1@luc.edu

Funding information
NICHD (R01-HD47450), NSF BCS 1056730, NSF 1561405, and the Spatial Intelligence and Learning Center (SBE 0541957) through the National Science Foundation

Abstract

Teaching a new concept through gestures—hand movements that accompany speech—facilitates learning above-and-beyond instruction through speech alone (e.g., Singer & Goldin-Meadow, 2005). However, the mechanisms underlying this phenomenon are still under investigation. Here, we use eye tracking to explore one often proposed mechanism—gesture's ability to direct visual attention. Behaviorally, we replicate previous findings: Children perform significantly better on a posttest after learning through Speech+Gesture instruction than through Speech Alone instruction. Using eye tracking measures, we show that children who watch a math lesson with gesture *do* allocate their visual attention differently from children who watch a math lesson without gesture—they look more to the problem being explained, less to the instructor, and are more likely to synchronize their visual attention with information presented in the instructor's speech (i.e., *follow along with speech*) than children who watch the no-gesture lesson. The striking finding is that, even though these looking patterns positively predict learning outcomes, the patterns do not *mediate* the effects of training condition (Speech Alone vs. Speech+Gesture) on posttest success. We find instead a complex relation between gesture and visual attention in which gesture *moderates* the impact of visual looking patterns on learning—*following along with speech* predicts learning for children in the Speech+Gesture condition, but not for children in the Speech Alone condition. Gesture's beneficial effects on learning thus come not merely from its ability to guide visual attention, but also from its ability to synchronize with speech and affect what learners glean from that speech.

RESEARCH HIGHLIGHTS

- Instruction through gesture facilitates learning, above-and-beyond instruction through speech alone. We replicate this finding and investigate one possible mechanism: gesture's ability to guide visual attention.
- Seeing gesture during math instruction changes children's visual attention: they look more to the problem, less to the instructor, and synchronize their attention with speech.
- Synchronizing attention with speech positively predicts learning outcomes but only within the gesture condition; thus, gesture *moderates* the impact of visual looking patterns on learning.
- Gesture's learning effects come not merely from guiding visual attention, but also from synchronizing with speech and affecting what learners glean from that speech.

1 | INTRODUCTION

Teachers use more than words to explain new ideas. They often accompany their speech with gestures—hand movements that express information through both handshape and movement patterns. Gesture is used spontaneously in instructional settings (Alibali et al.,

https://doi.org/10.1111/desc.12664

2014) and controlled experimental studies have found that children are more likely to learn novel ideas from instruction that includes speech and gesture than from instruction that includes only speech (e.g., Ping & Goldin-Meadow, 2008; Singer & Goldin-Meadow, 2005; Valenzeno, Alibali, & Klatzky, 2003). In the current study, we move beyond asking *whether* gesturing towards a novel mathematical equation improves children's learning outcomes to ask *how* gesturing improves learning. Specifically, we investigate the ways in which adding gesture to spoken instruction changes visual attention as children are learning the concept of mathematical equivalence (that the two sides of an equation need to be equivalent). We then ask whether these patterns of visual attention help explain differences in learning outcomes.

One understudied potential benefit of gesture is that it directs visual attention towards important parts of instruction. But gesture could affect visual attention in one of two ways. First, gesture may help learners by boosting effective looking patterns—patterns that are already elicited by verbal instruction, but may be *heightened* or *encouraged* by adding gesture. For example, if time spent looking at a key component of a math problem during instruction boosts the likelihood of insight into the problem, gesture might facilitate learning by encouraging children to attend to that key component more than they would have if instruction contained only speech. In other words, the positive effects of gesture on learning may be mediated by the heightened use of specific looking patterns that children already use. Alternatively, gesture may impact learning by working synergistically with speech. If so, gesture may facilitate learning, not by guiding visual attention to the problem per se, but by encouraging children to combine and integrate the information conveyed in speech with the information conveyed in gesture.

There is, in fact, some support for the idea that gesture does more for learners than (literally) point them in the "right" direction for learning. In a study conducted by Goldin-Meadow, Cook, and Mitchell (2009), children were taught to produce a strategy in speech to help them learn to solve a math equivalence problem (e.g., 5+3+2 = _+2), *I want to make one side, equal to the other side*. One group produced this spoken strategy without gestures. The other two groups were told to produce the spoken strategy while performing one of two gesture strategies. In the "correct" gesture condition, a V-handshape indicated the two numbers on the left side of the equation that could be *grouped* and added together to arrive at the correct answer (the 5 and the 3 in the above example), followed by a point at the blank. In the "partially-correct" gesture condition, the V-handshape indicated two numbers that did not result in the correct answer when added together (the 3 and the 2 in the example), again followed by a point at the blank; the gesture was partially correct in that the V-hand represented *grouping* two numbers whose sum could then be placed in the blank. If gesture aids learning solely by directing visual attention to important components of the problem, the "partially-correct" gesture should lead to poor learning outcomes. However, children performed *better* at posttest if they learned with the partially-correct gesture than if they learned with the spoken strategy alone. Still, children learned *best* with the "correct" gesture strategy, suggesting that guiding children's visual attention to important components of a problem may underlie part of gesture's impact on learning.

Previous eye tracking research shows that, in general, listeners will visually attend to the parts of the environment that are referenced in a speaker's words (e.g., Altmann & Kamide, 1999; Huettig, Rommers, & Meyer, 2011). This phenomenon has been extensively documented in the visual world paradigm where participants hear sentences while viewing scenes that contain a variety of objects. Even when they are not told to direct their attention toward any particular objects, individuals tend to visually fixate on objects mentioned in speech. For example, given a scene depicting a boy seated on the floor with a cake, ball, truck, and train track surrounding him, listeners who hear "The boy will move the cake" fixate on the boy and the cake rather than the other items in the scene (example drawn from Altmann & Kamide, 1999; for review of the visual word paradigm, see Huettig et al., 2011). Not surprisingly, similar looking patterns arise when spoken language is instructional. For example, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) monitored participants' visual attention to objects placed in front of them when they heard instructions like, "Put the apple that's on the towel in the box". In this example, participants were likely to look first at the apple that was on top of a towel, and then at the box, ignoring other objects in the visual array. Together, these findings suggest a tight alignment between verbal instruction and visual attention.

But when learning a new concept like mathematical equivalence, the connection between an instructor's speech and aspects of the instructional scene may not be as apparent to children as the connection between words and their referents in the visual world paradigm. To master the concept of mathematical equivalence, children must understand what it means for two sides of an equation (e.g., 3+5+4 = _+4) to be equivalent, and the mathematical operations that can be used to arrive at a balanced equation. Many children reveal a deep misunderstanding when they incorrectly solve missing addend problems like this. Children either add all of the numbers in the problem (and put 16 in the blank) or they add the numbers up to the equal sign (and put 12 in the blank) (Perry, Church, & Goldin-Meadow, 1988). During a math lesson, children may hear spoken instruction like, "You need to make one side equal the other side", and not be able to connect the words to the appropriate referents. If so, it may be difficult for children to learn from verbal instruction alone, raising the question of exactly how gesture facilitates learning in instances with potentially ambiguous speech.

In the current study, we ask how gesture directs visual attention for 8- to 10-year-old children who are learning how to solve missing addend equivalence problems (e.g., 2+5+8 = _+8). We use eye tracking to compare children's visual attention during instructional videos with either speech alone or speech with accompanying gesture. Previous work on mathematical equivalence has found that giving children relatively brief instruction on example problems and allowing them to solve problems themselves results in an increased understanding of mathematical equivalence. Importantly, incorporating gesture into the instruction boosts this understanding relative to instruction with speech alone (e.g., Congdon et al., 2017; Singer & Goldin-Meadow, 2005). In the present study, we use the *grouping* gesture, described earlier, during instruction. This gesture involves pointing a V-handshape at the first two numbers in a missing addend equivalence problem, followed by a point to the blank space. As described earlier, the V-handshape represents the idea that the equation can be solved by adding the two numbers indicated by the gesture, and putting that total in the blank. The V-handshape gesture is produced spontaneously by children who already understand how to solve mathematical equivalence problems (e.g., Perry et al., 1988) and has also been shown to facilitate learning when taught to children (Goldin-Meadow et al., 2009). Note that this gesture contains both *deictic* properties (pointing to specific numbers) and *iconic* properties (representing the idea of grouping). The benefits of learning through this type of gesture could thus arise from looking to the gesture itself, from looking to the numbers that the gesture is referencing, or from some combination of the two.

2 | METHOD

2.1 | Participants

Data from 50 participants were analyzed for the present study. Children between the ages of 8 and 10 ($M_{\text{speech alone}}$ = 8.53 years, *SD* = 0.53, $M_{\text{speech+gesture}}$ = 9.02 years, *SD* = 0.56) were recruited through a database maintained by the University of Chicago Psychology Department, and tested in the laboratory. The sample was racially diverse (42% White, 24% Black, 16% More than one race, 4% Asian, 14% Unreported) and included 26 children in the Speech+Gesture condition (14 females) and 24 children in the Speech Alone condition (14 females). Overall, the sample came from moderately high SES households: on average, at least one parent had earned a college degree, although the sample ranged from families where the highest parental education level was less than a high school degree, to households in which at least one parent had earned a graduate degree. Although not matched for subject variables across conditions, through random assignment, children were relatively equally distributed in terms of ethnic background, gender, and SES.

All children in the current sample scored a 0 (out of 6) on a pretest,[1] indicating that they did not know how to correctly solve mathematical equivalence problems at the start of the study. Data from all 50 children were included in the behavioral analyses. Data from five of the 50 children (two in Speech+Gesture, three in Speech Alone) were excluded from the eye tracking analyses because calibration had noticeably shifted from the target stimuli for these children (see details in the Results section). Prior to the study, parents provided consent and children gave assent. Children received a small prize and $10 in compensation for their participation.

2.2 | Materials

2.2.1 | Pretest/posttest

The pretest and posttest each contained six missing addend equivalence problems, presented in one of two forms. In Form A, the last addend on the left side of the equals sign was repeated on the right side (e.g., 5+6+3=_+3) and in Form B, the first addend on the left side of the equals sign was repeated on the right side (e.g., 4+7+2=4+_). Both pretest and posttest consisted of three of each problem type.

2.2.2 | Eye tracker

Eye tracking data were collected via corneal reflection using a Tobii 1750 eye tracker with a 17-inch monitor and a native sampling frequency of 50 Hz. Tobii software was used to perform a 5-point calibration procedure using standard animation blue dots. This step was followed by the collection and integration of gaze data with the presented videos using Tobii Studio (Tobii Technology, Sweden). Data were extracted on the level of individual fixations as defined by the Tobii Studio software—an algorithm determines if two points of gaze data are within a preset minimum distance from one another for a minimum of 100 msec, allowing for the exclusion of eye position information during saccades. After extraction, fixation location was queried at 20 msec intervals, to align with the native sampling frequency of the eye tracker.

2.2.3 | Instructional videos

Two sets of six instructional videos were created to teach children how to solve Form A missing addend math problems (e.g., 5+6+3=_+3)—one set for children in the Speech Alone condition, and one set for children in the Speech+Gesture condition. All videos showed a woman standing next to a Form A missing addend math problem, written in black marker on a white board. At the beginning of each video, the woman said, "Pay attention to how I solve this problem", and then proceeded to write the correct answer in the blank (she wrote 11 in the above example). She then described how to solve the problem, explaining the idea of *equivalence*: "I want to make one side equal to the other side. 5 plus 6 plus 3 is 14, and 11 plus 3 is 14, so one side is equal to the other side." During this spoken instruction, the woman kept her gaze on the problem. In the Speech+Gesture videos, the woman accompanied her speech with a gesture strategy. When she said "I want to make **one side**…", she simultaneously pointed a V-handshape (using her index and middle fingers) to the first two numbers in the problem, then, as she said "…the **other side**" she moved her hand across the problem, bringing her fingers together to point to the answer with her index finger (see Figure 1). The gesture was selected to complement and clarify the spoken strategy. The woman produced no gestures in the Speech Alone videos. To ensure that the speech was identical across the two training conditions, prior to taping, the actress recorded a single audio track that was used in both the Speech Alone and Speech+Gesture videos. Each of the 12 videos was approximately 25 seconds long.

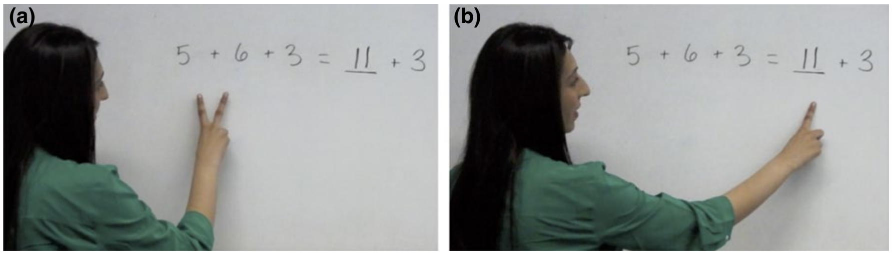

FIGURE 1 Panel (a) shows the experimenter's gesture when she is saying "one side" and panel (b) shows the gesture when she is saying "other side"

2.3 | Procedure

Children participated individually in a quiet laboratory space, and were randomly assigned to the Speech Alone or Speech+Gesture training condition. Figure 2 shows the study procedure. Children first completed a written pretest containing six missing addend math problems. None of the children solved any of the problems correctly. The experimenter then wrote children's (incorrect) answers on a white board and asked them to explain how they got each answer. Children were not given any feedback about their answers or explanations.

Next, children sat in front of the eye tracking monitor, approximately 60 centimeters from the screen, and were told they would watch instructional videos that would help them understand the type of math problems they had just solved. After their position was calibrated and adjusted if necessary, they began watching the first of the six instructional videos (either Speech Alone or Speech+Gesture, depending on the assigned training condition). At the conclusion of each of the six videos, children were asked to solve a new missing addend problem on a small, hand-held whiteboard, and were given feedback on whether or not their answer was correct (e.g., "that's right, 10 is the correct answer" or "no, actually 10 is the correct answer"). All problems shown in the instructional videos were Form A, and all problems that children had the opportunity to solve were Form A.

After watching all six instructional videos and having six chances to solve their own problems during training, children completed a new six-question paper-and-pencil posttest. The posttest, like the pretest, included three Form A problems and three Form B problems. As children saw only Form A problems during training, we refer to these as "Trained" problems and Form B as "Transfer" problems.

3 | RESULTS

3.1 | Behavioral results

3.1.1 | Training

Figure 3 shows the proportion of participants in each condition who answered problems correctly during training. Although none of the children included in our sample knew how to solve the problems before the study, after watching the first instructional video, 10 of 24 children (41.7%) in the Speech Alone condition and 10 of 26 children (38.5%) in the Speech+Gesture condition answered their own practice problem correctly, indicating rapid learning in both training conditions. Learning then continued to increase across the lesson and, by the final training problem, over 90% of participants in each group were answering the training problems correctly. A mixed-effects logistic regression predicting the log-odds of success on a given training problem with problem number (1 through 6) and condition (Speech Alone, Speech+Gesture)

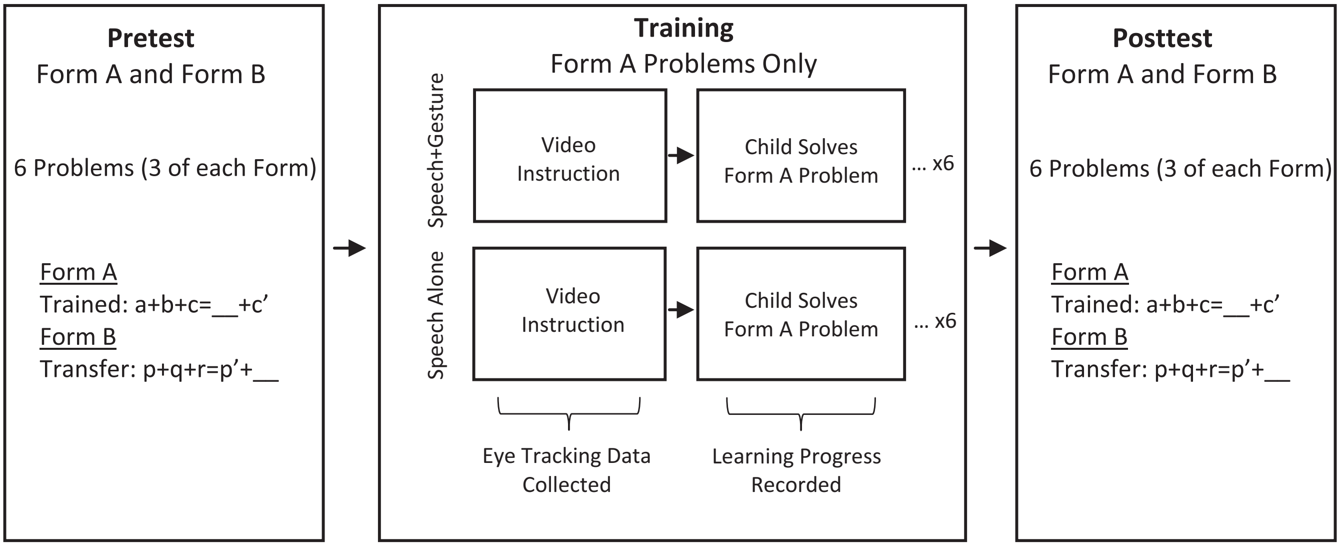

FIGURE 2 Diagram of procedure

as fixed factors, and participant as a random factor, revealed a positive effect of training problem (β = 0.91, *SE* = 0.15, *z* = 6.21, *p* < .001), indicating that children were more likely to correctly answer problems as training progressed. There was no effect of condition during training (β= 0.03, *SE* = 0.72, *z* = 0.04, *p* = .96), indicating that learning rates during training did not differ for children who did or did not receive gesture in the instruction. Together, these findings indicate that the two types of instruction were equally comprehensible, and did not differ in their effect on performance during training.

3.1.2 | Posttest

Although the groups did not differ in performance during training, their scores on an immediate posttest revealed an advantage for having learned through Speech+Gesture instruction (see Figure 4). Participants in the Speech+Gesture condition answered significantly more problems correctly at posttest (*M* = 4.11, *SD* = 2.04) than participants in the Speech Alone condition (*M* = 2.64, *SD* = 2.08). A mixed-effects logistic regression with problem type (Form A: trained, Form B: transfer) and condition (Speech+Gesture, Speech Alone) as fixed factors, and participant as a random factor, showed a significant effect of condition (β = 2.60, *SE* = 0.99, *z* = 2.59, *p* < .01), indicating that posttest performance in the Speech+Gesture condition was better than performance in the Speech Alone condition. There was also a significant effect of problem type (β= 2.27, *SE* = 0.43, *z* = 5.31, *p* < .001), demonstrating that performance on Form A (trained problems) was better than performance on Form B (transfer problems). This main effect was expected, as children received instruction on the trained problem form, but not the transfer problem form. There was no significant interaction between condition and problem type (β= 0.29, *SE* = 0.79, *z* = −0.37, *p* = .71).[2]

3.2 | Eye tracking results

3.2.1 | Data selection for eye tracking analyses

Next, we address the question of how gesture influenced visual attention by analyzing the eye tracking data from the instructional videos. To begin, we characterized broad differences in allocation of visual attention for children in the Speech Alone vs. Speech+Gesture conditions, asking how including gesture in instruction changes looking patterns. We considered whether gesture changed (1) the proportional amount of fixation time to the three major instructional elements (instructor, problem, and gesture), and (2) the degree to which children followed along with spoken instruction. Second, we asked whether differences found in visual attention predicted learning outcomes, as measured by children's posttest scores. To consider the relation between eye tracking measures and posttest scores we use linear regression models, reporting beta value coefficients and their corresponding *t* statistics. As will be described next, we averaged looking measures across all eligible trials *before* a "learning moment" (see below for definition), which varied by child. In other words, we predicted a child's posttest score using the average proportion of fixation time to each AOI, and the average degree of following across learning trials.

We reasoned that the way children attended to instruction *before* learning how to correctly solve mathematical equivalence problems is likely to differ from how they attended *after* they started solving problems correctly. Because we were ultimately interested in whether visual attention patterns would predict learning outcomes, we focused our analyses on data collected from each child before his or her personal "learning moment".[3] Recall that, on each

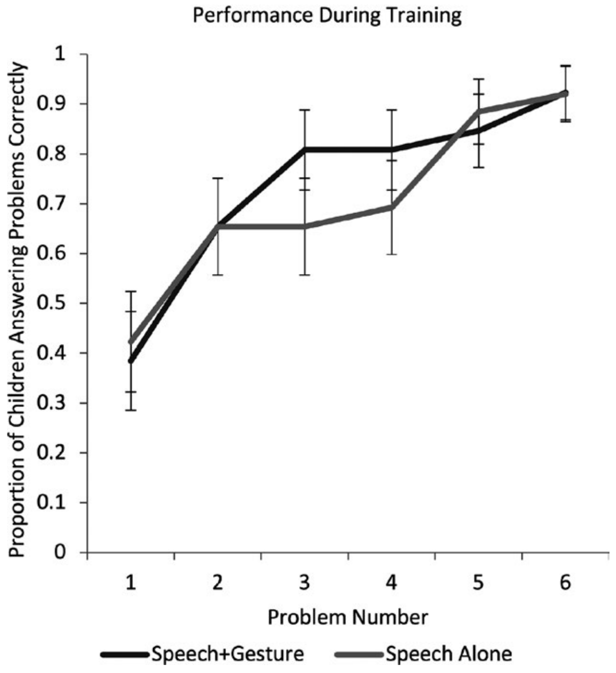

FIGURE 3 Performance during training on practice problems. Bars represent ± 1 *SE* of the mean

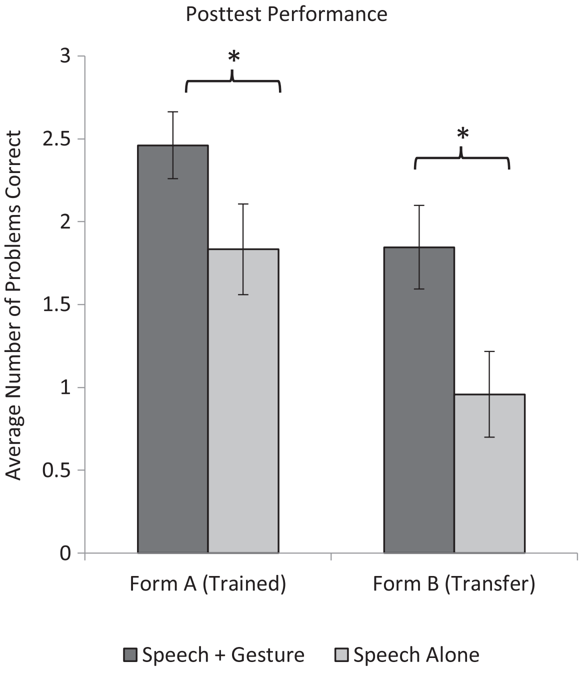

FIGURE 4 Posttest performance by condition and problem type. Error bars represent ±1 standard error of the mean

training trial, children first watched a training video with one problem, and then had an opportunity to solve another problem on a white board. The *learning moment* was defined as the point at which the child started answering his or her own white board problems correctly, and continued to provide correct answers on all subsequent problems. We included eye tracking data from all instructional videos prior to a child's learning moment, including the video that directly preceded a child's first correct answer on a white board problem. For example, if a child correctly answered problem 2, and also correctly answered the remaining problems, eye tracking data from instructional problems 1 and 2 were analyzed. If a child correctly answered problem 2, incorrectly answered problem 3, and then correctly answered problems 4 through 6, eye tracking data collected during instructional problems 1 through 4 were analyzed. Based on these criteria, children from the Speech+Gesture group contributed data from an average of 2.58 of 6 problems (*SD* = 1.90) and those from the Speech Alone group contributed data from an average of 2.71 of 6 problems (*SD* = 1.84). Given our behavioral finding that children in the two conditions followed a similar learning trajectory across training (see Figure 3), we were not surprised to find that condition did not significantly predict learning moment ($t(43) = 0.23$, $p = .82$).

Eye tracking data were excluded from the analyses if visual inspection of the eye tracking playback video of a given trial indicated unreliable tracking. For example, if the playback showed tracking consistently in the space above the math problem, and above the head of the experimenter, it was assumed that the child was not actually looking at the blank space, but rather that the child was looking at the problem, and that the tracking was inaccurate. This inspection was performed by a research assistant who was blind to the hypotheses of the study. This stipulation resulted in the exclusion of five participants (Speech+Gesture: n = 2; Speech Alone: n = 3), and the exclusion of at least one additional trial from seven other children. Within the remaining sample (Speech+Gesture: n = 24; Speech Alone: n = 21), eye tracking analyses were performed on clean trials that occurred before each child's learning moment. On average, after exclusions, children in the Speech+Gesture condition contributed data from 2.38 (*SD* = 1.56) trials, and children in the Speech Alone condition contributed data from 2.10 trials (*SD* = 1.45). A similar number of trials were thus considered for analysis across conditions ($t(43) = 0.63$, $p = .54$).

A multistep process was used to extract data and prepare it for analysis: (1) Areas of interest (AOIs) were generated for the instructor, problem, and gesture space[4] (see Figure 5) using Tobii Studio. The problem space was further separated by addend to calculate *Following Scores*, described later in this section. The remaining spaces outside of these AOIs were collapsed into an "Other" AOI. (2) Data were extracted and processed so that the AOI a participant fixated could be determined at 20 msec intervals across the entire length of each problem (see Materials section for further details regarding processing). (3) Time segments of interest, during which a particular event was happening in the videos, were defined. Certain time segments captured large amounts of data (e.g., the instructor stating the equalizer strategy, "*I want to make one side equal to the other side*"), whereas other time segments captured smaller amounts of data (e.g., the instructor referring to one of the addends in the problem, for example, "*five*"). (4) Within the defined time segments, the total gaze duration during a given time segment in each AOI was computed (e.g., 1000 msec), as well as whether there was a "hit" in each AOI (i.e., a score of "1" was assigned if a child looked to the AOI during the time segment; "0" was assigned if a child did not look to the AOI during the time segment).

3.2.2 | Fixation to instructional elements

To determine whether patterns of visual attention differed when children were instructed through speech alone or speech with gesture, we calculated the proportion of time children spent in each AOI for two time segments of interest (see Figure 6). The *strategy* segment encompassed time when the instructor stated the equalizer strategy: "*I want to make one side equal to the other side*." During this segment, spoken instruction was identical across conditions, but children in the Speech+Gesture condition also saw co-speech instructional gestures. As the strategy was explained twice per problem, data from these epochs were combined into one segment of interest. The *explanation* segment encompassed time when the instructor elaborated on the strategy, highlighting the particular addends in the problems (e.g., "*5 plus 6 plus 3 is 14, and 11 plus 3 is 14*"). No gestures were produced during this segment. As a result, the segment was visually identical across the two conditions, allowing us to ask whether the presence of gesture during the preceding *strategy* segment influenced the way children in the Speech+Gesture condition deployed their visual attention in the subsequent *explanation* segment. If so, eye tracking during this segment should differ for children in the Speech+Gesture vs. Speech Alone conditions.

Within the strategy and explanation time segments, we calculated the proportion of time a participant spent in each AOI, collapsing over the participant's included problems.

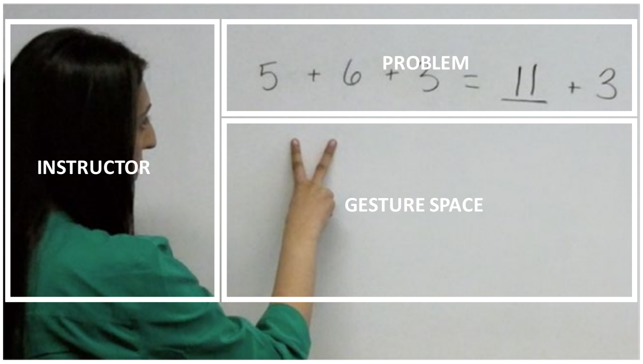

FIGURE 5 Example of areas of interest (AOIs). Depending on the specific analysis, the problem AOI was further subdivided into left side vs. right side, and individual addends

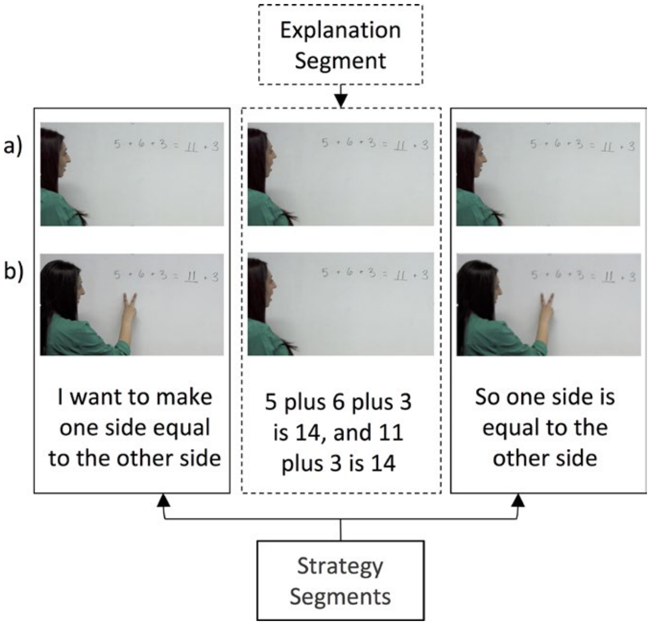
Explanation Segment
a)
b)
I want to make one side equal to the other side
5 plus 6 plus 3 is 14, and 11 plus 3 is 14
So one side is equal to the other side
Strategy Segments

FIGURE 6 Time segments of interest for (a) Speech Alone training, and (b) Speech+Gesture training. Instruction differs for Speech Alone vs. Speech+Gesture training during the strategy segments, and is identical during the explanation segment

3.2.2.1 | Strategy segment

Figure 7 shows the proportion of time children spent looking in each of the AOIs during the strategy segment in each condition, excluding the "Other" AOI which children rarely fixated during either segment. On average, children in the Speech+Gesture condition spent a greater proportion of time looking to the problem itself than children in the Speech Alone condition (64.6% versus 50.1%) (β = 0.15, *SE* = 0.05, *t* = 2.65, *p* < .05). In contrast, children in the Speech Alone condition allocated more visual attention to the instructor than children in the Speech+Gesture condition (45.2% vs. 14.7%) (β = 0.31, *SE* = 0.05, *t* = 6.39, *p* < .0001). Finally, children in the Speech+Gesture condition spent 18.4% of the time looking to the gesture space. Not surprisingly, children in the Speech Alone condition spent significantly less time (2.7%) looking to this AOI (β = 0.16, *SE* = 0.03, *t* = 5.17, *p* < .0001) as there was nothing there to draw their visual attention. Taken together, these results suggest that adding gesture to verbal instruction leads participants to look more at the objects mentioned in speech, and less at the instructor herself.

3.2.2.2 | Explanation segment

Figure 7 also shows the proportion of time spent in the problem, instructor, and gesture space AOIs during the explanation segment. No significant differences between conditions were found in the proportion of time children spent in the instructional AOIs during this time. All children predominately looked to the problem (Speech Alone: 61.4% Speech+Gesture: 68.4%; β = 0.07, *SE* = 0.06, *t* = 1.20, *p* = .24), although they also allocated a substantial proportion of their attention to the instructor (Speech Alone: 36.5% Speech+Gesture: 28.4%; β = 0.08, SE = 0.06, *t* = 1.44, *p* = .16). Children looked very little towards gesture space (Speech Alone: 0.6% Speech+Gesture: 1.0%; β = 0.004, *SE* = 0.005, *t* = 0.80, *p* = .43) or "Other" space (Speech Alone: 1.4% Speech+Gesture: 2.2%; β = 0.008, *SE* = 0.01, *t* = 0.81, *p* = .42), suggesting that during the explanation segment, most children were on task. Together with the analysis from the strategy segment, these findings suggest that adding gesture to instruction influences children's visual attention in the moments when gesture is used, but that this effect does *not* extend to subsequent spoken instruction without gesture.

3.2.3 | Following along with spoken instruction

To determine how well children were following the spoken instruction, we once again considered the strategy segment and the explanation segment separately. During the strategy segment, when the instructor stated the equalizer strategy, "*I want to make **one side** equal to the **other side***", we defined "following" as visually attending to one side of the problem when the instructor said "one side" and then switching to focus on the other side of the problem when the instructor said "other side". That is, we created shorter time segments, capturing the time during which the instructor said "one side" and "other side" as the two time segments of interest, and determined whether children had "hits" in AOIs encompassing the left side of the problem versus the right side of the problem during these time segments. For each instance of the

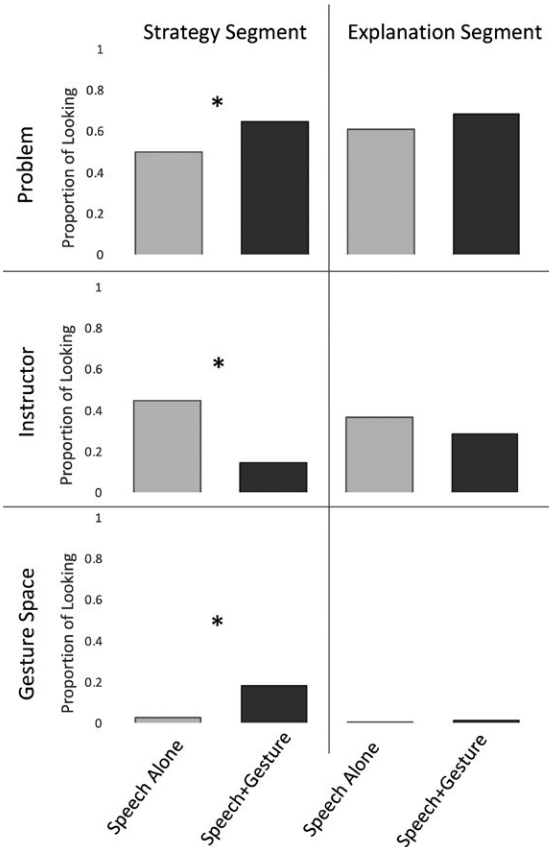

FIGURE 7 Proportion of time children spent looking to each AOI, separated by condition and problem segment

equalizer strategy, children were given a score of “1” if, across the “one side” and “other side” time segments, they had hit both of the AOIs. This pattern of attention indicated that they had followed along with the instructor’s speech, moving their visual attention between the two sides of the problem as each was indicated in speech.

Under this approach, a child received a following score of “1” for a problem on which she followed during both instances of the equalizer strategy in a given trial; “0.5” if she followed during one, but not both, instances of the equalizer strategy in a given trial; and “0” if she followed during neither instance of the equalizer strategy in a given trial. Following scores were averaged across all included problems for a given child to obtain one overall following score per participant.

We found that children in the Speech+Gesture condition had a higher following score than children in the Speech Alone condition (Speech+Gesture: $M = 0.75$, $SD = 0.27$; Speech Alone: $M = 0.55$, $SD = 0.27$). A model predicting following score by condition showed a significant effect of condition ($\beta = 0.17$, $SE = 0.09$, $t = 2.02$, $p < .05$), suggesting that the addition of gesture in instruction significantly increased children’s ability to follow along with spoken instruction.

We took a similar approach to assess how well children followed along with spoken instruction that was *not* accompanied by gesture in either condition, that is, during the explanation of the problem (e.g., “**5** plus **6** plus **3** is 14, and **11** plus **3** is 14”). We isolated the time segments during which each addend was mentioned in spoken instruction and created AOIs around each individual addend. Children were assigned a “hit” during a time segment if they visually attended to the corresponding AOI during that segment (e.g., based on the problem above, looking at the “5” when the instructor said “five”). As there were five time segments of interest per problem, we calculated a following score between 0 and 1 for each problem by dividing the number of hits by five. Thus, if a child visually attended to two of the addends when the instructor mentioned them in her speech, he received a following score of 0.4 for that problem. Following scores were then averaged across all included problems for a given child to obtain one overall following score per participant.

In line with our finding that overall visual attention to instructional elements did not differ during the instructor’s explanation, we found no condition difference in following score during this time segment (Speech+Gesture: $M = 0.44$, $SD = 0.22$; Speech Alone: $M = 0.38$, $SD = 0.18$, $\beta = 0.06$, $SE = 0.06$, $t = 0.92$, $p = .36$).

3.2.4 | How do looking patterns relate to posttest score?

Our results show that including gesture in spoken instruction *does* change how children visually attend to that instruction: when gesture is present, children look significantly more to the problem and the gesture space, significantly less to the instructor, and follow along better with spoken instruction, than when gesture is absent. Having established these differences, we ask whether there is a significant relation between looking patterns and posttest score and, if so, how gesture influences that relation.

First, we find that three of the four looking patterns that significantly differ by condition also independently predict posttest score (Table 1 shows a breakdown of all analyses). Looking to the problem,

TABLE 1 Summary of analyses showing how looking patterns predict posttest performance

Looking pattern	Does looking pattern *predict* posttest performance?	Does looking pattern *mediate* the effect of condition on posttest performance?	Is gesture a *moderator* between looking pattern and posttest performance?
Fixation to Problem	**Yes** $\beta = 3.86$ $SE = 1.66$ $t = 2.33$ $p < .05$	No ACME: 0.40 CI: −0.24 to .97 $p = .14$	No $\beta = 4.27$ $SE = 3.57$ $t = 1.20$ $p = .24$
Fixation to Gesture Space	No $\beta = 2.23$ $SE = 2.67$ $t = 0.84$ $p = .41$		
Fixation to Instructor	**Yes** $\beta = 3.68$ $SE = 1.46$ $t = 2.52$ $p < .05$	No ACME: 0.73 CI: −0.71 to 1.80 $p = .23$	No $\beta = 6.82$ $SE = 4.31$ $t = 1.58$ $p = .12$
Following Speech	**Yes** $\beta = 2.45$ $SE = 1.15$ $t = 2.13$ $p < .05$	No ACME: 0.34 CI: -0.11 to 0.99 $p = .17$	**Yes** $\beta = 5.76$ $SE = 2.26$ $t = 2.55$ $p < .05$

Note. *SE* refers to one standard error of the mean, CI stands for confidence interval, and ACME stands for Average Causal Mediation Effect.

looking to the instructor, and following along with speech each predicts posttest scores—the *more* children look to the problem, the *less* they look to the instructor, and the *more* they follow along with speech, the better they do on the posttest. In contrast, looking to gesture space does not predict posttest score. Thus, not only does gesture *change* looking in these key ways, but these changes in looking patterns have a relationship with subsequent learning outcomes.

Next, we probe this relation further to ask whether the predictive looking patterns were specifically mediated by the presence of gesture. In other words, we ask whether the relationship between condition and posttest seen in Figure 4 (posttest performance is better after Speech+Gesture instruction than Speech Alone instruction) can be explained by a simple increase or decrease in effective looking patterns. To consider this question, we conducted separate mediation analyses, using the bootstrapping method described by Preacher and Hayes (2004). For each analysis, 1000 simulations were used, and we asked whether the Average Causal Mediation Effect (ACME) was significant. Table 1 shows that none of the looking patterns is a significant mediator for the effect of gesture in instruction on posttest score.

Finding no conclusive evidence that visual looking patterns mediate the relation between condition and posttest, we conclude that gesture's power as a teaching tool stems from more than its ability to guide overt visual attention. We next asked whether gesture *moderates* the relation between visual looking pattern and posttest score. In other words, we asked whether the changes in looking patterns, which can be attributed to including gesture in instruction, were subsequently beneficial to learning *if and only if* there was also a gesture present.

To demonstrate that gesture moderates the relation between looking pattern and posttest score, we need evidence that there is a significant interaction between a given looking pattern and posttest score (Hayes, 2013). We looked separately at the three measures of looking that had a significant effect on posttest—(1) looking to the instructor, (2) looking to the problem, and (3) following along with speech. For the first two measures, we found no significant interaction with condition (see Table 1). However, we *did* find a significant interaction between following along with spoken instruction and condition (Table 1). To explore the interaction effect, we asked whether following score predicted posttest score for children in each of the two conditions. Following score was a significant predictor of posttest score for children in the Speech+Gesture condition (β = 4.39, *SE* = 1.38, *t* = 3.18, *p* < .01), but not for children in the Speech Alone condition (β = 1.37, *SE* = 1.83, *t* = 0.75, *p* = .46) (Figure 8).[5] This finding suggests that, during the strategy segment, including gesture in instruction fundamentally changes *how* following along with speech facilitates learning. Following along with speech is not, on its own, beneficial to learning (otherwise, we would have seen this effect in the Speech Alone condition as well). Rather, following along with speech supports learning outcomes when it is accompanied by a representational gesture that clarifies the speech.

4 | DISCUSSION

Our study builds on decades of research that have established a beneficial connection between including gesture in instruction and learning outcomes (for review see Goldin-Meadow, 2011). As in previous work, we find that children who were shown instructional videos that included spoken and gestured instruction performed significantly better on a posttest than children who learned through spoken instruction alone. Moving beyond previous work, our study reveals two important findings. First, we find that watching an instructor gesture changes how children allocate their visual attention—children look more to the problem and gesture space, less to the instructor, and are better able to follow along with ambiguous spoken instruction (Figure 7). Second, our results indicate a complex relationship between gesture and visual attention in which gesture *moderates* the effect of visual looking patterns on learning. Following along with speech predicted learning for children in the Speech+Gesture condition, but not for children in the Speech Alone condition (Figure 8). This finding suggests that following along with speech not only increases in *frequency* when gesture is included in instruction, but also in *efficacy*. Note that we found no *mediation* effects for any of our eye tracking measures. Despite the fact that looking to the problem, looking away from the instructor, and following along with spoken language were all more common in the Speech+Gesture condition

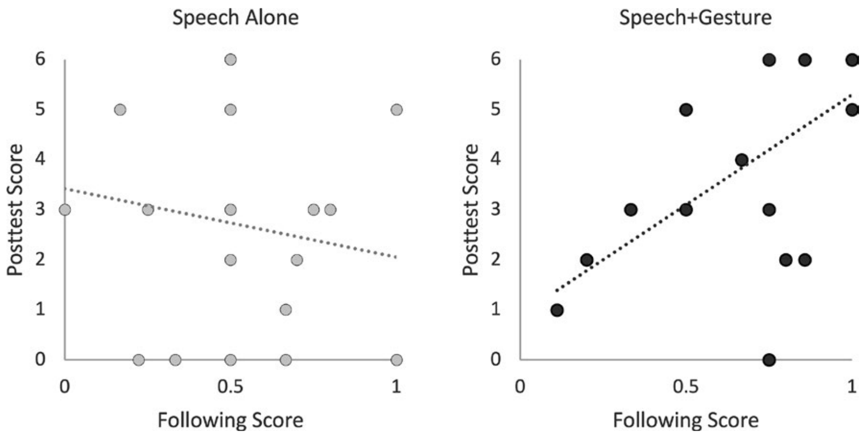

FIGURE 8 Relation between strategy segment Following Score and Posttest Score for the Speech Alone condition (left graph) and the Speech+Gesture condition (right graph)

than in the Speech Alone condition, and also predicted posttest performance, none of the measures was a significant mediator of the relation between condition and posttest performance (Table 1, middle column). These findings have implications for understanding how gesture functions to direct attention in an instructional context, and for understanding the mechanisms underlying gesture's effect on children's learning outcomes.

Our eye tracking results demonstrate that, at a global level, gesture directs visual attention towards referents mentioned in speech in an instructional context—we found a significant difference between the amount of visual attention children allocated to the mathematical problem in the Speech+Gesture condition, compared to the Speech Alone condition. This result might strike some as surprising, as the Speech+Gesture videos contained moving hands that could have drawn children's attention away from the problem. However, the finding makes sense in terms of what we know about gesture—gesture is a spatial, dynamic, social cue, and even young infants will shift their visual attention in response to gesture to look at the gesture's referent (Rohlfing, Longo, & Bertenthal, 2012). Interestingly, we also show that the effect gesture has on visual attention is transient—it is only during the times when gesture is produced that children in the Speech+Gesture condition attend differently from their peers in the Speech Alone condition. During the explanation segment, when instruction contained only speech and thus was identical across the two conditions, children's attention did not differ. This finding suggests that gesture has the potential to help children in the moment when it is produced, perhaps to integrate information conveyed across the two modalities.

Although children in the Speech+Gesture condition allocated the majority of their attention to the numbers in the problem that were indicated by gesture, they also spent close to 20% of the strategy segment looking at the gesture itself. This result deviates from previous eye tracking gesture research, which has focused on how observers process naturally occurring gesture during face-to-face communication (e.g., watching a person tell a story). These earlier findings suggest that looking directly toward a speaker's hands is quite rare (Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009). Listeners prefer instead to look primarily at the speaker's face, and spend little time overtly attending to the speaker's hands. For example, in one study of gesture in discourse, only 9% of the gestures that were produced received focal attention (Gullberg & Holmqvist, 2006). On the rare occasions when interlocutors do look directly at a speaker's gesture, it is typically because the speaker himself is looking towards his own hands, or is holding a gesture in space for an extended period of time (Gullberg & Kita, 2009).

These previous findings suggest that observers tend to watch gestures primarily when they expect to receive important information from those gestures, as indicated by the speaker's attention to his own gestures. In our videos, gestures were front-and-center—they were in the middle of the screen, providing a cue to their importance, and the instructor was oriented away from the child and not making eye contact with the child. As a result, it is not surprising that the learners in our instructional experiment, who are seeking useful problem-solving information, spent a sizeable proportion of their time attending to the gesture itself. In addition, the iconic form of the gesture in our videos was informative—the V-handshape represented the fact that the two numbers indicated by the gesture could be *grouped* and added together to arrive at the correct solution. Children may have spent time focusing on the V-handshape in order to glean meaning from its form. Whatever the reasons, our findings make it clear that observers watch gesture during instruction differently from how they watch gesture in other non-learning communicative contexts.

Understanding how gesture shapes visual attention during instruction is important, but the main goal of our study was to provide insight into how gesture interacts with visual attention to support learning. Here, our results suggest that gesture does not merely boost looking patterns that lead to improved learning outcomes. We found no conclusive evidence that the looking patterns that predict learning (looking to the problem, looking away from the instructor, and following along with speech) acted as statistical mediators for the positive effects that gesture had on learning. This finding suggests that the gesture in our instruction helped learners learn through other mechanisms. If this hypothesis is correct, we might expect that including visual highlighting that draws attention to relevant information in the problem during instruction (e.g., through the use of computerized underlining) would *not* facilitate learning as efficiently as gesture. Visual highlighting may draw attention to important components of the problem, but gestures have the potential to bring added value to the learning experience. This hypothesis should be directly tested in future work.

If gesture is *not* simply drawing attention to important components in instruction, what is it doing for learners? We have evidence that two distinct features of gesture need to be working simultaneously to promote optimal learning. First, gesture needs to help learners allocate their attention in ways that can help them interpret ambiguous speech. Our results even suggest that the dynamic, temporal relation between gesture and speech may be centrally important for learning—we found that our "following along with speech" measure moderated gesture's effect on learning outcomes, whereas the total overall time spent looking at various components of the visual scene did not. Researchers have previously highlighted the strong connection between speech and gesture, showing that gesture and speech are more tightly integrated than other forms of action and speech (Church, Kelly, & Holcombe, 2014; Kelly, Healy, Özyurek, & Holler, 2015) and that the simultaneity between speech and gesture is important for learning outcomes (Congdon et al., 2017). But, crucially, looking patterns alone are not enough—if they were, children who followed along with speech during speech alone instruction should have improved just as much as children who followed along with speech during speech and gesture instruction. In other words, following along with speech *and gesture* promotes learning; merely following along with speech does not.

Processing gesture and speech simultaneously thus appears to qualitatively change how children learn from the components of instruction to which they attend. Helpful looking patterns during ambiguous speech were only beneficial for children who were also exposed to an iconic, representational structure that provided additional content about the relational structure of the problem. We therefore hypothesize that it is these two pieces coming together—gesture's ability to direct visual attention, and its ability to simultaneously add content to speech through its iconic representational form—that explain the benefits gesture confers on learning.

Another intriguing finding from our study is that learning rates and performance *during* instruction did not differ across the two training conditions—differences emerged only in the posttest where intermittent reminders of the strategy were not present. This finding suggests that our two instructional conditions were equally comprehensible during the initial learning process, but that children in the Speech Alone condition formed a more fragile representation of correct problem-solving strategies than children in the Speech+Gesture condition. We suggest that in instructional situations where children are overcoming misconceptions for the first time, gesture in instruction can play a particularly powerful role in helping learners solidify brand new knowledge that otherwise deteriorates very quickly. Previous work has shown that gestures are particularly good at promoting long-lasting learning over and above speech alone (e.g., Congdon et al., 2017; Cook, Duffy, & Fenn, 2013; Cook, Mitchell, & Goldin-Meadow, 2008). But ours is the first study to isolate this consolidation and retention effect on a relatively short time scale; that is, between training and an immediate posttest. This result further supports the overarching hypothesis that gesture affects cognitive change in ways that cannot be fully captured by overt behavioral measures taken during the learning process itself (see also Brooks & Goldin-Meadow, 2016). Although beyond the scope of the current paper, other findings hint at mechanisms that may be playing a role in solidifying knowledge. For example, we know that gesture provides learners with a second, complementary problem-solving strategy that can be integrated with spoken language and lead to better understanding of the principles of mathematical equivalence (e.g. Congdon et al., 2017). And we know that not only can watching gesture create a robust motor representation in listeners (Ping, Goldin-Meadow, & Beilock, 2014), but the motor representation created in learners who produce gesture is later reactivated when the learners are exposed again to the math problems (e.g., Wakefield et al., under revision).

In conclusion, our study builds on the existing literature in a number of ways. We show, for the first time, that children visually attend to instruction differently when it includes gesture than when it does not include gesture. We also show that, even though looking patterns heightened through gesture instruction predict learning, gesture's contribution to learning goes above-and-beyond merely directing visual attention.

ACKNOWLEDGEMENTS

Funding for this study was provided by NICHD (R01-HD47450, to Goldin-Meadow), NSF BCS 1056730, NSF 1561405, and the Spatial Intelligence and Learning Center (SBE 0541957, Goldin-Meadow is a co-PI) through the National Science Foundation. We also thank Kristin Plath, William Loftus, Lorenzo Grego, and Aileen Campanaro for their help with data collection, and Amanda Woodward for the use of the eye tracker.

ENDNOTES

[1] An additional 49 children completed the study, but answered at least one pretest problem correctly and are excluded from the current analyses.

[2] We also tested for an effect of age, by including age as a predictor in the same model, and found that age did not predict posttest performance (β= −0.40, *SE* = 0.87, *z* = −0.46, *p* = .64).

[3] For a description of looking patterns across the entire session, ignoring learning moment, see Novack et al. (2016), which describes looking patterns by condition across the entire six-problem set. The proportion of time spent looking to each AOI in each training condition is similar to results presented here, but did not predict any of our posttest measures, indicating important differences in processing before and after a child's "learning moment".

[4] To make comparison between the Speech Alone and Speech+Gesture conditions possible, we identified a "gesture space" in the Speech Alone video (the area where gesture was produced in the Speech+Gesture condition) despite the fact that no gestures were actually produced in the Speech Alone videos.

[5] This finding also held when we considered only the problem *directly preceding* a child's learning moment, rather than averaging across all problems preceding the child's learning moment: following score significantly predicts posttest score for children in the speech+gesture condition (β = 3.75, *SE* = 0.94, *t* = 3.98, *p* < .001), but not for children in the speech alone condition (β = 0.13, *SE* = 1.87, *t* = 0.07, *p* = .94).

REFERENCES

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

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

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

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

Congdon, E.L., Novack, M.A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74.

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.

Goldin-Meadow, S. (2011). Learning through gesture. *WIREs: Cognitive Science*, *2*, 595–607.

Goldin-Meadow, S., Cook, S.W., & Mitchell, Z. (2009). Gestures gives children new ideas about math. *Psychological Science*, *20*, 267–271.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, *14*, 53–82.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*, 251–277.

Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Huettig, F., Rommers, J., & Meyer, A.S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151–171.

Kelly, S.D., Healy, M., Özyurek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin and Review*, *22*, 517–523.

Novack, M.A., Wakefield, E.M., Congdon, E.L., Franconeri, S., & Goldin-Meadow, S. (2016). There is more to gesture than meets the eye: Visual attention to gesture's referents cannot account for its facilitative effects during math instruction. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2141–2146). Austin, TX: Cognitive Science Society.

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. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, *143*, 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.

Rohlfing, K.J., Longo, M.R., & Bertenthal, B.I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science*, *15*, 426–435.

Singer, M.A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*, 85–89.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Valenzeno, L., Alibali, M.W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*, 187–204.

How to cite this article: Wakefield E, Novack MA, Congdon EL, Franconeri S, Goldin-Meadow S. Gesture helps learners learn, but not merely by guiding their visual attention. *Dev Sci*. 2018;e12664. https://doi.org/10.1111/desc.12664