Cognition 210 (2021) 104604

Contents lists available at ScienceDirect

Cognition

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

Children integrate speech and gesture across a wider temporal window than speech and action when learning a math concept

Cristina Carrazza [a,*], Elizabeth M. Wakefield [b,1], Naureen Hemani-Lopez [a], Kristin Plath [a], Susan Goldin-Meadow [a]

[a] University of Chicago, United States
[b] Loyola University Chicago, United States

ARTICLE INFO

Keywords:
Gesture
Action
Mathematical equivalence
Learning
Retention
Temporal synchrony

ABSTRACT

It is well established that gesture facilitates learning, but understanding the *best* way to harness gesture and *how* gesture helps learners are still open questions. Here, we consider one of the properties that may make gesture a powerful teaching tool: its temporal alignment with spoken language. Previous work shows that the simultaneity of speech and gesture matters when children receive instruction from a teacher (Congdon et al., 2017). In Study 1, we ask whether simultaneity also matters when children themselves are the ones who produce speech and gesture strategies. Third-graders ($N = 75$) were taught to produce one strategy in speech and one strategy in gesture for correctly solving mathematical equivalence problems; they were told to produce these strategies either simultaneously (S + G) or sequentially (S➔G; G➔S) during a training session. Learning was assessed immediately after training, at a 24-h follow-up, and at a 4-week follow-up. Children showed evidence of learning and retention across all three conditions. Study 2 was conducted to explore whether it was the special relationship between speech and gesture that helped children learn. Third-graders ($N = 87$) were taught an action strategy instead of a gesture strategy; all other aspects of the design were the same. Children again learned across all three conditions. But only children who produced *simultaneous* speech and action retained what they had learned at the follow-up sessions. Results have implications for why gesture is beneficial to learners and, taken in relation to previous literature, reveal differences in the mechanisms by which doing versus seeing gesture facilitates learning.

1. Introduction

When people talk, they convey information not only through spoken language, but also through gesture—movements of the hands that express meaning. Decades of research show that gestures facilitate learning, whether students produce the gestures themselves as they learn a new concept (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Cook, & Mitchell, 2009; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014), or observe the gestures that teachers produce as they explain a new concept (e.g., Congdon et al., 2017; Singer & Goldin-Meadow, 2005; Wakefield, Novack, Congdon, Franconeri, & Goldin-Meadow, 2018). This phenomenon has been well-studied in mathematical equivalence—an important pre-algebraic concept underlying children's understanding that the two sides of an equation must be equal (e.g., $8 + 4 + 3 = \_ + 3$; McNeil, 2014). Researchers have found that adding gesture to spoken instruction in mathematical equivalence leads to more immediate learning gains than presenting spoken instruction alone (e.g., Goldin-Meadow et al., 2009; Wakefield, Novack, et al., 2018), and that gesture aids in establishing long-lasting (e.g., Congdon et al., 2017; Cook et al., 2008) and flexible (e.g., Novack et al., 2014) understanding of this concept.

That gesture facilitates learning is well established, but understanding the *best* way to harness gesture and *why* gesture helps learners are still open questions. Gaining a better grasp on how and when to use gesture to promote learning is important not only for theoretical reasons, but also because it will allow researchers to make specific recommendations to educators about incorporating gesture into lesson plans. Researchers have consequently begun to focus on properties of gesture that have the potential to make it uniquely powerful as a teaching tool, and to systematically study how these properties impact

* Corresponding author at: Department of Psychology, 5848 S. University Avenue, Chicago, IL 60637, United States.
E-mail address: cristinac@uchicago.edu (C. Carrazza).
[1] Co-first author.

https://doi.org/10.1016/j.cognition.2021.104604
Received 12 March 2020; Received in revised form 12 January 2021; Accepted 13 January 2021
Available online 3 February 2021
0010-0277/© 2021 Elsevier B.V. All rights reserved.

learning outcomes.

One property of gesture that has the potential to make it a powerful learning tool is its temporal alignment with spoken language. Gesture is synchronized with the speech it accompanies (Kendon, 1980) and listeners seamlessly integrate gesture into the speech they are processing (McNeill, 1992). Even when gesture conveys different information from the speech it accompanies (cf. Goldin-Meadow, 2003), listeners integrate the two channels to form a single representation (Cassell, McNeill, & McCullough, 1999). Listening to speech, and watching the gestures that go along with it, thus permits learners to be simultaneously exposed to different, but complementary, ideas. In fact, there is evidence that children are more likely to learn from instruction when they are given two different strategies, one in gesture and one in speech, than when they are given the same strategy in both speech and gesture, or a single strategy in speech alone. Importantly, learning is greater when the two different strategies are presented simultaneously in speech and gesture, and not when they are presented sequentially in speech (Singer & Goldin-Meadow, 2005). This finding suggests that simultaneous presentation of speech and gesture may be crucial for learning—an idea that has been proposed more broadly in dual-coding theories of learning (e.g., Baddeley, 1999; Chandler & Sweller, 1991; Mayer, 2002, 2005). According to these theories, learners benefit from input presented simultaneously in two modalities because our ability to process information from one input channel (e.g., hearing speech) has limits and adding a second input channel (e.g., seeing gesture) helps us go beyond those limits. Simultaneous gesture and speech input can thus be understood as a specific case of dual-coding.

Recent work by Congdon et al. (2017) experimentally explores the impact on children's learning of temporally aligning input from two modalities. Congdon et al. (2017) tested whether the temporal relation between speech and gesture produced by an instructor affects child learning outcomes by directly comparing simultaneous versus sequential presentation of speech and gesture strategies during math instruction. In the study, two strategies for solving mathematical equivalence problems were used: an equalizer strategy (the idea that the two sides of an equation need to sum to the same number) and an add-subtract strategy (the idea that problems can be solved by summing the addends on the left side of the equation and subtracting the addend on the right). Children were randomly assigned to one of three groups: (1) the teacher/experimenter produced the two strategies sequentially, equalizer followed by add-subtract, both in speech; (2) she produced the two strategies sequentially, equalizer in speech followed by add-subtract in gesture; (3) she produced the two strategies simultaneously, equalizer in speech along with add-subtract in gesture. Congdon and colleagues found that simultaneity mattered—children who saw simultaneous speech and gesture during the lesson retained what they had learned better than children in the other two conditions.

The present study builds on this work by addressing a question of practical and theoretical importance—whether temporal synchrony between speech and gesture is also crucial for learning when speech and gesture are produced by the student, rather than the teacher; in other words, when the learner *does* gesture rather than *sees* someone else do it. From a practical perspective, we know that teachers *and* students use gesture in the classroom. But we do not know whether the recommendations we make to teachers about how *they* should use gesture also apply to how *students* should use gesture. Is it necessary for children to produce strategies in speech and gesture simultaneously, or can they produce the strategies sequentially, in order to reap learning benefits?

From a theoretical perspective, we have begun to understand the mechanisms by which gesture shapes learning, but do not know how general these mechanisms are. The temporal synchrony between speech and gesture during instruction may be necessary for learning, and it may therefore be the source of gesture's power as a teaching tool, whether the gestures are produced or observed. Alternatively, the temporal synchrony between speech and gesture during instruction may matter only when gesture is observed; when it is produced by the learner, it may not be necessary to temporally align the two modalities. There is evidence that learning outcomes differ when children produce gestures themselves versus observing an experimenter produce gestures. For example, Goldin-Meadow et al. (2012) found that children learned more on a mental rotation task when they were taught to *produce* meaningful gestures about translation and rotation than when they *observed* an experimenter producing these same gestures. In a different paradigm, Wakefield, Hall, et al. (2018) demonstrated that these learning differences extend across time. Children either produced gestures for an action, or observed an experimenter produce the same gestures, during a word learning task. They were better at remembering the newly taught word for the action 24 h later if they themselves had produced the gestures than if they had watched the experimenter produce the gestures. Similarly, in a recent meta-analysis, Dargue, Sweller, and Jones (2019) found that children showed greater learning gains when producing gesture than when observing gesture. These findings suggest that the mechanisms responsible for learning from the gestures one produces oneself, and learning from the gestures others produce, might differ.

Consistent with this view, there are hints in the literature that children can learn when they produce gesture without speech, suggesting that the temporal alignment between speech and gesture may not be that important when learners themselves produce the gesture. Cook et al. (2008) modeled an equalizer strategy for solving mathematical equivalence problems for children to produce in speech alone, gesture alone, or speech and gesture simultaneously during a math lesson. Children in all three groups improved after the lesson, but children who produced the strategy in speech alone did not retain what they learned over a 4-week delay. In contrast, children who produced the strategy in gesture, either on its own or with speech, performed well on retention measures. Children seem to be able to benefit simply from *doing* gesture, with or without speech. Brooks and Goldin-Meadow (2016) modeled an equalizer gesture or a control gesture (which was not interpretable in the context of a mathematical equivalence problem), neither of which was accompanied by speech, for children to produce during a math lesson. Children who produced the equalizer strategy in gesture were more likely to profit from the math lesson than children who produced the control gesture. Gesture can be a powerful force when it is in the hands of the learner, even when it is produced without speech.

In Study 1, we ask whether Congdon et al.'s (2017) findings for seeing gesture extend to doing gesture—we ask whether gesture and speech need to be produced simultaneously in order for learning to occur when learners themselves produce the gestures. We gave children who were unable to solve mathematical equivalence problems on a pretest models for problem-solving strategies in speech and gesture that they were then asked to produce during a math lesson. We then measured gains in knowledge immediately after the lesson, at a one-day follow-up session, and at a four-week follow-up session—the same time points at which Congdon and colleagues measured learning gains after learners saw the experimenter produce gestures during instruction. Children were randomly assigned to one of three groups and taught a lesson on how to solve mathematical equivalence problems (e.g., $3 + 6 + 5 = \_ + 5$). Children in all three groups were taught to produce an equalizer strategy (the idea that the two sides of the equation need to be equal) in speech, and a grouping strategy (the idea that the two unique numbers on the left side of the equation can be grouped and summed to arrive at the number that goes in the blank on the right side of the equation) in gesture. One group was taught to produce the two strategies simultaneously (S + G). Two groups were taught to produce the two strategies sequentially, one in which gesture preceded speech (G➔S) and one in which speech preceded gesture (S➔G). We hypothesized that if gesture facilitates learning through the same mechanisms when it is produced by learners as when it is observed by learners, children should display the best long-term learning after producing speech and gesture strategies simultaneously during the math lesson (S + G). Although we had no a priori hypothesis about whether the sequence in which speech and gesture appeared would affect learning, it is possible that one modality serves as better contextual support for the other (i.e., that gesture serves as better contextual support for speech than speech serves for gesture, or vice versa). We therefore varied the order of speech and gesture, and tested whether learning differed in the G➔S and S➔G conditions. A final possibility is that the timing between speech and gesture does not matter when learners produce the strategies themselves. If so, all three conditions (S + G, G➔S, S➔G) ought to be equally good for learners.

2. Study 1

2.1. Participants

Data from 75 third-grade students ($M = 9.06$ years, $SD = 0.52$; 40 females) were analyzed in Study 1. The study focused on children of this age because third-graders typically do not understand mathematical equivalence and fail to solve problems of this format (e.g., McNeil, 2014). In order to ensure that all participants had the same starting knowledge, children were excluded from the study if they solved any of the pretest problems correctly. An additional 70 children were tested, but excluded because they solved one or more pretest problem correctly, and one additional child was excluded because the child was not proficient in English. Although there was a high no-response rate on our demographic questionnaire (52%), from the data collected, participants were racially, ethnically, and socioeconomically diverse (27% White, 8% Asian, 5% Black, 4% More than one race, 3% Native Pacific Islander, 1% Native American). Overall, the sample came from lower SES households: 52% of parents reported having a high-school degree or less; only 16% reported having a college or graduate degree. Prior to the study, parents provided consent and children gave assent. Children received a small prize and certificate of participation, and teachers of participating classrooms received a gift card to a local learning store.

2.2. Materials

Math problems were written by the experimenter on a white dry-erase magnetic board. To be consistent with similar math equivalence instruction studies (Novack et al., 2014), black magnetic number tiles were placed over each number during the training phase. Children were asked to solve two types of problems at pretest and posttest. In Form A problems, 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 problems, the first addend on the left side of the equals sign was repeated on the right side (e.g., $4 + 7 + 2 = 4 + \_$). Form A problems were used during the training phase. On posttest and follow-up assessments, we also included a third type of problem. In Form C problems, there were no identical addends on the two sides of the equal sign (e.g., $5 + 3 + 4 = 2 + \_$).

2.3. Design and procedure

Children were tested individually at their school across three testing days (see Fig. 1). Testing Day 1 consisted of a pretest, training math lesson, and immediate posttest. Children were randomly assigned to one of three instruction conditions at training: (1) simultaneous speech and gesture (S + G); (2) sequential speech and gesture, with gesture preceding speech (G➔S); (3) sequential speech and gesture, with speech preceding gesture (S➔G). To measure retention of the instruction material, children also completed follow-up tests after a 24-h (Testing Day 2) and four-week (Testing Day 3) delay.

2.4. Testing Day 1

2.4.1. Pretest

All children completed a paper-and-pencil pretest consisting of six problems (3 Form A; 3 Form B). The experimenter then asked children to explain their answers, writing each problem with the child's answer on a white board. Only children who solved all problems incorrectly continued to the training phase.

2.4.2. Training

Children were randomly assigned to one of three conditions. In the simultaneous speech and gesture (S + G; $n = 25$) condition, children were taught to produce the strategy in speech along with the strategy in gesture. In one of the sequential speech and gesture conditions, children were taught to produce the gesture strategy followed by the speech strategy (G➔S; $n = 25$). In the other sequential condition, children were taught to produce the speech strategy followed by the gesture strategy (S➔G; $n = 25$).

In all three conditions, children were taught the equalizer strategy in speech ("I want to make one side equal to the other side") and the grouping strategy in gesture (a V-handshape with the index finger placed under the first two numbers, followed by an index finger placed under the blank on the right side of the problem, Fig. 2). These two strategies express different but complementary ways of solving mathematical equivalence problems. Equalizer highlights the principle underlying mathematical equivalence; grouping highlights a procedure for solving problems of this type (the number in the blank is equal to the sum of the two addends on the left side that do not appear on the right side).

Children were first taught the speech and gesture strategies in a pre-training phase. The experimenter modeled the speech and gesture strategies and asked the child to practice producing the strategy in relation to two Form A problems that were written on the board. During pre-training, children did not solve any problems. In the simultaneous condition (S + G), the two strategies were modeled together, and the child was asked to "repeat the words and hand movements." In the sequential conditions (G➔S, S➔G), the child learned the strategies separately. For example, in the G➔S condition, the child was shown the gesture and asked, "Can you repeat those hand movements?" and then told the speech strategy and asked, "Can you repeat those words?" Once children were able to produce the strategies that had been modeled for them, they moved on to a training phase in which they and the experimenter took turns solving 12 Form A problems on the whiteboard, 6 problems each. When it was the experimenter's turn to solve a problem, she wrote the correct solution in the blank and then produced the equalizer strategy in speech. The experimenter did not perform any hand movements during training. When it was a child's turn to solve a problem, the child was asked, "Can you say your words and do your movements?" both before and after solving the problem. The experimenter told children whether their answer was correct or incorrect, but did not provide the correct answer or any additional feedback. If children produced their strategy incorrectly, the experimenter corrected them; however, this rarely happened, as children were required to produce the speech and gesture strategies without the help of the experimenter in order to begin training.[2]

2.4.3. Posttest

After training, children completed a paper-and-pencil posttest that was identical in format to the pretest, with the addition of Form C problems. Children were not reminded of their strategy before this assessment, and did not actively use their strategy when solving problems. Upon completion, they were asked to explain their solutions to Form A and B problems.

[2] As a measure of compliance, we selected 50% of the data and considered the number of times children had to be corrected during the training session because they did not produce the taught strategy accurately. We found that children were only corrected an average of 0.10 times per problem and the number of corrections did not differ between the three conditions $F(2, 40) = 0.49$, $p = .62$.

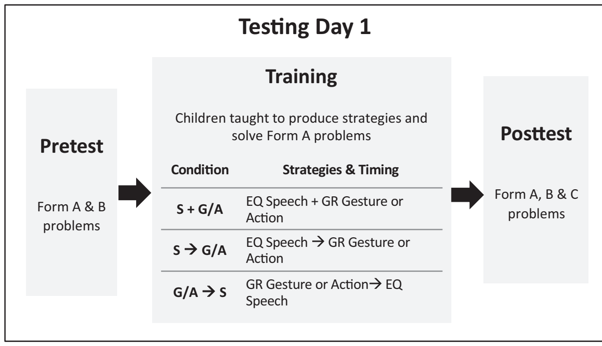

Fig. 1. Experimental design for studies 1 and 2. G = gesture, which was manipulated in Study 1; A = action, which was manipulated in Study 2.

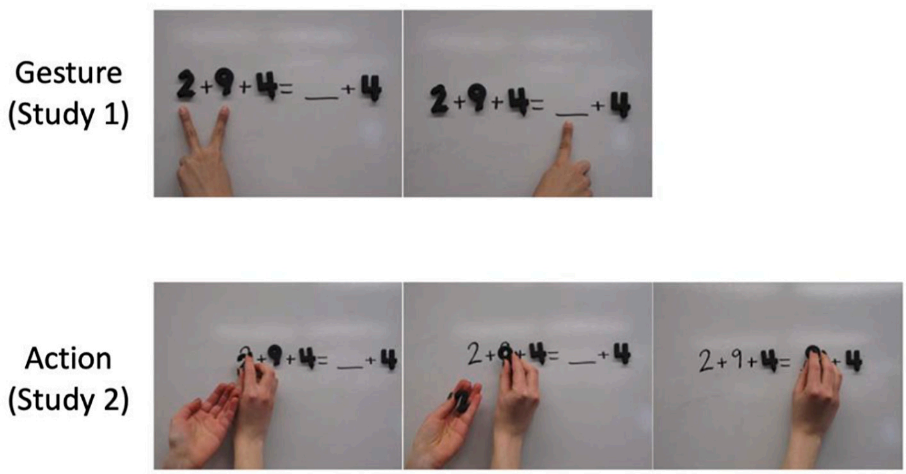

Fig. 2. Example of the grouping strategy children were taught to produce in gesture (study 1) and action (study 2).

2.5. Testing Days 2 and 3

Children completed two follow-up pencil-and-paper tests, one day and four weeks after Testing Day 1. Assessments consisted of 9 problems, and contained all problem types (Forms A, B, and C). Children were not reminded of their strategies nor did they actively use their strategies on these assessments.

3. Results

All analyses were conducted through R Studio (version 1.2.1335), supported by R version 3.6.1 "Action of the Toes" (R Core Team, 2019). Analyses relied on the lme4 package, which allows for mixed effects modeling (Bates, Mächler, Bolker, & Walker, 2015). When running mixed effects models through lme4, dummy coding was used, the default option for coding in this package. Appropriate reference levels for factors were assigned before each model was run: for testing day, the immediate posttest was the reference level; for timing of speech and gesture, sequential use of strategies was the reference level; for problem type, Form A problems on which children received training was the reference level. An alpha level of 0.05 was used when evaluating statistical significance.

Table 1 summarizes the average proportion correct for all problem types and testing days by condition. We first consider children's performance on the immediate posttest. As expected, across conditions children showed the highest proportion correct on Form A problems, the problem type used during the training session, with lower proportion correct on Form B and Form C problems. There were no systematic differences across conditions (see Table 1 for means).

To test for statistically significant effects at the immediate posttest, we constructed a binomial logistic regression model with trial-level accuracy (0,1) as the dependent measure, and condition (S + G, G➔S, S➔G) and problem type (A, B, C) as fixed factors. Participant was included as a random factor, as was random slope for the interaction of problem type and participant. Problem type significantly predicted accuracy ($\chi^2(2) = 6.33$, $p = .04$); this effect was driven by the difference between Form A and Form C problems ($\beta = -8.13$, $SE = 4.12$, $t = -0.78$, $p = .04$). The model did not reveal a significant effect of condition on children's accuracy on the immediate posttest ($\chi^2(2) = 2.38$, $p = .30$). Children who learned from producing the gesture strategy followed by the speech strategy did not significantly differ in accuracy from children who learned from producing speech followed by gesture ($\beta = -7.32$, *SE*

Table 1
Average proportion correct for each problem type and testing day. Standard errors are displayed in parentheses below group means.

Condition	Testing Day 1			Testing Day 2			Testing Day 3		
	Form A	Form B	Form C	Form A	Form B	Form C	Form A	Form B	Form C
G➔S	0.84 (0.07)	0.59 (0.10)	0.60 (0.10)	0.78 (0.08)	0.64 (0.09)	0.70 (0.09)	0.72 (0.09)	0.58 (0.09)	0.64 (0.10)
S➔G	0.73 (0.08)	0.53 (0.10)	0.54 (0.10)	0.70 (0.09)	0.52 (0.10)	0.50 (0.10)	0.70 (0.09)	0.62 (0.10)	0.58 (0.10)
S + G	0.75 (0.09)	0.64 (0.09)	0.62 (0.09)	0.68 (0.10)	0.56 (0.09)	0.58 (0.09)	0.66 (0.09)	0.56 (0.10)	0.48 (0.10)

= 4.91, $t = -1.49$, $p = .14$) or speech and gesture simultaneously ($\beta = -0.95$, $SE = 1.96$, $t = -0.48$, $p = .63$). The lack of a significant effect of condition also held when we collapsed across the two sequential conditions (S➔G; G➔S) and used sequential versus simultaneous production of speech and gesture strategies as a fixed factor in the model ($\chi^2(1) = 0.05$, $p = .82$). Because we had no a priori hypotheses about how order of speech and gesture would affect learning, and found no condition differences at immediate posttest, we collapse across the two sequential conditions for all subsequent analyses.

We next turn to our main question—do children show different rates of long-term learning when they themselves produce problem-solving strategies simultaneously versus sequentially? Congdon et al. (2017) found an advantage for simultaneously produced strategies for long-term learning when children observed teachers produce the strategies. We ask whether timing matters when gesture is in the hands of the learner. To address this question, we constructed a binomial logistic regression model with trial-level accuracy (0, 1) as the dependent measure. Fixed factors included timing (simultaneous vs. sequential, collapsing across S➔G and G➔ S), testing day (posttest, next day follow-up, 4-week follow-up), and problem type (A, B, C), and a timing by testing day interaction (a significant interaction would indicate differences in retention across the three time points). Participant was included as a random factor, as were random slopes for the interaction of problem type and participant, and testing day and participant.

The model revealed no significant effect of testing day($\chi^2(1) = 0.03$, $p = .87$), suggesting that children retained what they had learned. There was also no evidence of a testing day by strategy timing interaction ($\chi^2(1) = 1.68$, $p = .19$), suggesting that children retained what they learned whether they had produced gesture and speech simultaneously or sequentially during instruction. A post-hoc power analysis calculating observed power suggests this null effect is not due to lack of statistical power (power = 0.90).

Although not central to this study, an analysis of variance of the model revealed a significant effect of problem type ($\chi^2(2) = 23.12$, $p < .001$). Post-hoc analyses showed that children performed best on Form A problems (the type of problem used during the training session), compared to either Form B ($\beta = 2.84$, $SE = 0.60$, $t = 4.75$, $p < .001$) or Form C ($\beta = 3.03$, $SE = 0.68$, $t = 4.43$, $p < .001$) problems; there was no difference between performance on Forms B and C ($\beta = 0.18$, $SE = 0.36$, $t = 0.49$, $p = .62$). Finally, to ensure that we had not missed a potential emerging difference between the sequential conditions over time, we conducted a similar analysis comparing children's accuracy across all three time points for these conditions. The model revealed no significant differences between the gesture followed by speech (G➔S) and speech followed by gesture (S➔G) instruction conditions ($\chi^2(1) = 1.99$, $p = .16$).

4. Study 2

In Study 1, we found that, when gesture is in the hands of the learner, it facilitates long-lasting learning, whether it is produced simultaneously or sequentially with speech. This finding underscores the need to consider whether gesture is produced by students or teachers when we recommend how gesture should be used as a teaching tool. Congdon et al.'s (2017) findings suggest that *teachers* should use gesture and spoken instruction simultaneously; our findings suggest that this recommendation does not necessarily hold for *students.* The power that self-produced gesture has to promote learning does *not* seem to rely on whether it is produced simultaneously with speech.

Congdon et al. (2017) argued that simultaneous speech and gesture was a more effective teaching tool than sequential speech and gesture in their study because, in temporally aligned speech-plus-gesture instruction, speech provided context that allowed students to interpret the gesture as meaningful. When gesture does not co-occur with speech, children may struggle to connect the gesture they see to the preceding speech or to the math problem. Without the context provided by speech, children may interpret the movements they see as a meaningless, rather than as gesture representing a useful problem-solving strategy. Indeed Novack, Wakefield, and Goldin-Meadow (2016) have shown that, when speech is added to, and co-occurs with, observed hand movements, those hand movements are particularly likely to be interpreted as meaningful gestures. The advantage that simultaneous speech and gesture had over sequential speech and gesture in the Congdon et al. study also aligns with general findings about learning from multimodal instruction, which suggest that the most effective way to get integration across two channels of information is to present them to a learner simultaneously (e.g., Mayer, 2002, 2005).

Why then were the children in our Study 1 able to benefit equally from simultaneous and sequential speech and gesture when the two were self-produced? One possibility is that learners can integrate speech and any manual movement over a longer temporal window when they produce the speech and the movement than when they see others produce the two. If so, learners might be able to benefit from information in gesture produced along with actions on objects whether they produce the two simultaneously or sequentially. We can test this hypothesis by asking learners to produce speech along with actions on objects.

In Study 2, we use the same basic design, but teach children to produce simultaneous or sequential strategies in speech and *action*, a hand movement that is less closely aligned with spoken language than gesture is. When college students are asked to explain how to use a common object (e.g., a hair brush or water bottle) and are told to either act on the object during the explanation, or show how to use the object without touching it (i.e., to gesture), the onset of the self-produced gestures is more closely timed with speech than the onset of the self-produced actions (Church, Kelly, & Holcombe, 2014). Moreover, adults find it harder to ignore information conveyed in gesture when it is incongruent with information in the accompanying speech than to ignore information conveyed in action when it is incongruent with information in speech (Kelly, Healy, Ozyurek, & Holler, 2014); in other words, we bind information from concurrent speech and gesture more easily than information from concurrent speech and action on objects. Together, these findings suggest that it might be easier to integrate information from speech and gesture during the learning process than information from speech and action on objects.

This difference in how easily speech + gesture versus speech + action are integrated has the potential to help us understand our Study 1 findings. When children in Study 1 produced speech and gesture sequentially, the tight natural connection between speech and gesture may have allowed them to hold both pieces of information in mind and integrate across them, even though they were not simultaneously produced. In other words, it was the natural synchrony and integration of speech and gesture that allowed children to learn from these strategies, even when they produced them sequentially. But action on objects and speech are less well synchronized than gesture and speech. As a result, the timing of self-produced actions might matter—integrating across sequential action + speech may be more difficult than integrating across simultaneous action + speech, leading us to predict that children will learn less well from sequential action + speech than from simultaneous action + speech.

Alternatively, self-produced movements could have a large impact on learning regardless of their relation to speech. We know that self-produced action and gesture both facilitate learning, in part, because they engage the motor system during the learning process. Learning through self-produced action (e.g., James, 2010; James & Swain, 2011; Longcamp, Anton, Roth, & Velay, 2003; Longcamp, Tanskanen, & Hari, 2006) or gesture (e.g., Macedonia, Muller, & Friederici, 2011; Wakefield, Congdon, Novack, Goldin-Meadow, & James, 2019) leaves a rich, lasting sensori-motor representation of the learned information that can be drawn upon when learners subsequently process the information. If the temporal alignment between speech and gesture in Study 1 was irrelevant to the findings and simply producing movement was key to the learning effect, temporal alignment should also be irrelevant for action and speech. Children should then learn equally well from action + speech produced simultaneously or sequentially.

In addition to distinguishing between two possible explanations for our Study 1 findings, Study 2 has the potential to address a shortcoming of Study 1—interpreting the null effect as evidence that children benefit equally from simultaneous versus sequential speech and gesture. If we *do* see a simultaneity advantage emerge when children produce action and speech strategies (i.e., if we see an effect of temporal alignment in Study 2), we can conduct a planned-comparison across studies to determine whether the patterns seen after action versus gesture training statistically differ from one another. This analysis could then lend further support to the conclusion that children benefit equally well from gesture + speech produced simultaneously or sequentially.

Study 2 followed the same experimental design as Study 1, but with action on objects rather than gestures. The experimenter modeled the equalizer strategy in speech and moved plastic numbers that were placed over the numbers in a mathematical equivalence problem (the plastic numbers were present in Study 1 but children did not touch them); the experimenter's movements instantiated the grouping problem-solving strategy used in Study 1 (see Fig. 2). Children were asked to produce the grouping movements along with the equalizer strategy in speech during a training session. As in Study 1, gains in knowledge were measured immediately after training, and at one-day and four-week follow-up sessions. Children were randomly assigned to one of three groups for training. All children were taught to produce an equalizer strategy in speech, and a grouping strategy in action, which they produced either simultaneously with speech (S + A) or sequentially with speech (A➔S; S➔A).

4.1. Participants

Third-grade students (M = 8.79 years, SD = 0.42; 54 females) participated in Study 2.[3] As in Study 1, children were excluded from the study if they solved any of the pretest problems correctly (n = 22), or if there was a language barrier (n = 9), leaving a sample of 87 children. As in Study 1, there was a high no-response rate on the demographic questionnaire (62%), but based on available data, the sample was racially, ethnically and socio-economically diverse (22% White, 3% Black, 7% More than one race, 6% Native American). As Study 1, this sample was mostly low SES: 48% of parents reported having a high-school degree or less; 15% reported having a college or graduate degree. Prior to the study, parents provided consent and children gave assent. Children received a small prize and certificate of participation and teachers of participating classrooms received a gift card to a local learning store.

4.2. Design and procedure

The design and procedure of Study 2 was identical to Study 1 (see Fig. 1), except that children were taught to produce the grouping problem-solving strategy in action rather than gesture. The actions were performed on magnetic number tiles; children picked up the first two number tiles on the left side of the equation, and then placed them together on the blank (see Fig. 2). All of the children were also taught to produce the equalizer strategy in speech. Children were randomly assigned to one of three training conditions, which determined the temporal alignment between the speech and action they were instructed to produce: simultaneous speech and action (S + A; n = 29); action followed by speech (A➔S; n = 29); speech followed by gesture (S➔A; n = 29). As in Study 1, children completed three days of testing, and assessments included a mix of Form A, B, and C problem types. As in Study 1, children were compliant— they rarely needed to be corrected during training on the strategies they had learned during pre-training,[4] and they completed the post-test and follow-up tests without being reminded of their strategies.

5. Results

As in Study 1, we considered how children performed at immediate posttest before addressing our main question (how retention was affected by type of training). Table 2 presents the average proportion correct for all problem types and testing days by condition. Children performed best on Form A problems, the problem type used during the training session, and performed less well on Form B and C problems.

To test for statistically significant differences, we used the same approach as in Study 1. We constructed a binomial logistic regression model with trial-level accuracy (0, 1) as the dependent measure, and included condition (S + A, A➔S, S➔A) and problem type (A, B, C) as fixed factors. Participant was included as a random factor, and we included random slope for the interaction of problem type and participant. Problem type significantly predicted accuracy ($\chi^2(2) = 11.79$, $p < .01$), an effect driven by the difference between Form A and Form B problems ($\beta = -1.93$, $SE = 0.56$, $t = -3.43$, $p < .001$). The model did not reveal a significant effect of condition on children's accuracy on the immediate posttest ($\chi^2(2) = 0.65$, $p = .72$). Children who learned from producing the action strategy followed by the speech strategy did not significantly differ in accuracy from children who produced speech followed by action strategies ($\beta = -0.67$, $SE = 1.10$, $t = -0.61$, $p = .54$) or simultaneous speech and action ($\beta = 0.20$, $SE = 1.11$, $t = -0.18$, $p = .86$). As in Study 1, we found that this effect was stable when the two sequential conditions were collapsed ($\chi^2(1) = 0.43$, $p = .51$).

Our main question was whether retention is the same when children use action and speech strategies during the lesson as when they use gesture and speech strategies (i.e., no effect of temporal alignment between speech and gesture on retention). To address this question, we used a binomial logistic regression model with trial-level accuracy (0, 1) as the dependent measure. Fixed factors included timing (simultaneous vs. sequential, collapsing across S➔A and A➔ S), testing day (posttest, next day follow-up, 4-week follow-up), and problem type (A, B, C), and a timing by testing day interaction.[5] Participant was included as a random factor, and we included random slopes for the interaction of problem type and participant, and for the interaction of testing day and participant.

Unlike Study 1, there was a marginal effect of testing day ($\chi^2(1) = 3.57$, $p = .06$), suggesting that children showed a drop-off in what they retained from training over the 4-week retention period. There was also an interaction that approached significance between testing day and timing ($\chi^2(1) = 3.80$, $p = .05$), with a small to medium effect size (OR = 2.09, 95% CI: 1.00, 4.36). In other words, there was a difference in how children performed over time depending on the temporal alignment of their speech and action. Post-hoc analyses revealed a significant effect of testing day for children who produced sequential speech and action ($\beta = -0.55$, $SE = 0.19$, $t = -2.87$, $p < .01$), but no effect for children who produced simultaneous speech and action ($\beta = 0.16$, $SE = 0.39$, $t = 0.42$, $p = .67$). Children who produced sequential speech and action displayed

[3] Of these 87 students, 82 completed all three days of testing; the remaining 5 students completed testing days 1 and 2, but not day 3. These students were distributed across the experimental conditions.

[4] As a measure of compliance, we selected 50% of the data and considered the number of times children had to be corrected during the training session because they did not produce the taught strategy accurately. We found that children were only corrected an average of 0.06 times per problem, and there were no differences across conditions $F(2, 34) = 2.22$, $p = .13$.

[5] A timing by testing day interaction was expected, as the interaction indicates different retention rates depending on whether children produced action and speech strategies simultaneously versus sequentially. An analysis comparing a model with only main effects, versus the interaction term, revealed that the model with the interaction term was a better fit ($\chi^2 = 3.73$, $p = .05$).

Table 2
Average proportion correct for each problem type and testing day. Standard errors are displayed in parentheses below group means.

Condition	Testing Day 1			Testing Day 2			Testing Day 3		
	Form A	Form B	Form C	Form A	Form B	Form C	Form A	Form B	Form C
A➔S	0.67 (0.08)	0.47 (0.09)	0.39 (0.09)	0.67 (0.08)	0.41 (0.09)	0.51 (0.09)	0.59 (0.09)	0.27 (0.08)	0.40 (0.09)
S➔A	0.61 (0.08)	0.44 (0.09)	0.48 (0.09)	0.66 (0.09)	0.46 (0.08)	0.44 (0.08)	0.46 (0.09)	0.33 (0.08)	0.44 (0.08)
S + A	0.68 (0.08)	0.55 (0.08)	0.49 (0.09)	0.69 (0.08)	0.51 (0.09)	0.59 (0.08)	0.67 (0.09)	0.48 (0.09)	0.58 (0.09)

a drop-off over the 4-week retention period in the knowledge they had gained; children who produced simultaneous speech and action were, in contrast, able to retain the knowledge they had gained.

Although not of central interest in the present study, an analysis of variance of the initial model revealed an effect of problem type ($\chi^2(2) = 18.22$, $p < .001$). As found in Study 1, post-hoc analyses showed that children performed better on Form A problems (the type of problem used during the training session) than on either Form B ($\beta = 1.81$, $SE = 0.35$, $t = 5.22$, $p < .001$) or Form C ($\beta = 2.96$, $SE = 0.57$, $t = 5.21$, $p < .001$) problems. Unlike Study 1, children also performed better on Form B than Form C problems ($\beta = 1.15$, $SE = 0.42$, $t = 2.76$, $p < .01$). Finally, to ensure that we had not missed a potential emerging difference between the sequential conditions over time, we conducted a similar analysis comparing children's accuracy across all three time points for these two conditions. The model revealed no significant differences between the action followed by speech condition and the speech followed by action condition ($\chi^2(1) = 0.15$, $p = .70$).

6. Comparing across studies 1 and 2

Using the same testing procedures in Studies 1 and 2 allows for comparison across action and gesture training. Although students were not randomly assigned to each study, all participants were third graders who were not able to solve any of the problems correctly before instruction. Study 1 and Study 2 were also conducted by the same three experimenters, who were blind to the hypotheses of the study. The students in the studies came from three different public schools within the same city, and Study 1 and Study 2 have two of these schools in common.[6] We therefore combined the datasets from the two studies and directly compared performance over time (see Fig. 3). As in other analyses, a binomial logistic regression model was used with trial-level accuracy (0, 1) as the dependent measure. Fixed factors included movement type (gesture, action), timing (simultaneous, sequential [collapsing across sequential orders]), testing day (posttest, next day follow-up, 4-week follow-up), and problem type (A, B, C). In order to explore whether there were differences in retention over time as a function of timing (simultaneous versus sequential) and movement type (action, gesture), we included a three-way interaction with these predictors, and therefore two-way interactions between each of these predictors as well. Finally, as in other models, participant was included as a random factor, and we included random slopes for the interaction of problem type and participant, and of testing day and participant.

In terms of main effects, we found a marginal effect of testing day (posttest, next day follow-up, 4-week follow-up), with children's performance decreasing over time $\chi^2(1) = 2.72$, $p = .10$). We also found significant main effects of timing (sequential, simultaneous; $\chi^2(1) = 4.17$, $p = .04$) and problem type (A, B, C; $\chi^2(2) = 55.34$, $p < .001$). Children performed better when they produced their movements simultaneously with speech than when they produced them sequentially with speech; and they performed better on Form A problems than Form B ($\beta = -2.17$, $SE = 0.31$, $t = -7.08$, $p < .01$) or Form C ($\beta = -2.84$, $SE = 0.42$, $t = -6.84$, $p < .01$) problems. However, there was no significant main effect of movement type (action, gesture; $\chi^2(1) = 1.90$, $p = .17$), and we found no evidence of any significant 2-way interactions (movement type by testing day, $\chi^2(1) = 0.72$, $p = .40$; strategy timing by movement type, $\chi^2(1) = 1.63$, $p = .20$; testing day by strategy timing, $\chi^2(1) = 0.19$, $p = .67$).

Most importantly, however, the model reveals a significant three-way interaction between movement type, testing day, and timing ($\chi^2(1) = 5.85$, $p = .02$), with a medium to large effect size (OR = 4.76, 95% CI: 1.35, 16.87). Post-hoc analyses based on this interaction underscore the patterns previously reported in Studies 1 and 2—children who learned with the help of gesture retained knowledge over time, whether they learned with simultaneous or sequential gesture and speech ($\beta = -0.78$, $SE = 0.60$, $t = -1.30$, $p = .19$). In contrast, children who learned with the help of action retained knowledge better when they learned with simultaneous action and speech than with sequential action and speech ($\beta = 0.74$, $SE = 0.38$, $t = 1.95$, $p = .05$). In other words, temporal alignment with speech had an impact on learning through self-produced action, but no impact on learning through self-produced gesture.

7. Discussion

Previous work suggests that gesture is a powerful teaching tool, in part, because it can express information simultaneously with spoken instruction and thus promote integration across the two channels. Supporting this idea, Congdon et al. (2017) found that children retained what they had learned from mathematical equivalence instruction significantly better if their teacher produced gesture simultaneously with speech than if she produced gesture sequentially with speech. However, what was not yet known was whether gesture needs to occur simultaneously with speech when it is in the hands of the learner, not the teacher—that is, when it is produced rather than seen by the learner.

Our findings from Study 1 indicate that, when gesture is in the hands of the learner, the temporal alignment between speech and gesture does *not* affect whether gesture benefits the learner—children learn and retain information from gesture whether or not the strategies they produce in gesture are temporally aligned with the strategies they produce in speech. Our findings from Study 2 suggest that this is not a general feature of all types of self-produced hand movements, but rather a special feature of gesture—children learn and retain information from action better if the strategies they produce in action *are* temporally aligned with the strategies they produce in speech. These findings have implications both for understanding the mechanisms underlying gesture's impact on learning, and for developing recommendations to teachers about how gesture can best be used in the classroom.

7.1. How does gesture promote learning?

The literature has moved beyond showing that gesture *can* help learning to focus on *how* it helps. Our results add to this discussion. First, we contribute to the small but growing literature suggesting that gesture can impact learning differently when the learner produces it vs. when the learner observes it (Goldin-Meadow et al., 2012; Wakefield, Hall, et al., 2018). Taken together with Congdon et al. (2017), Study 1 conceptually replicates these findings and, in conjunction with Study 2,

[6] One difference between participants in the two studies was when in the school year they were tested. In Study 1, 55 students were tested in the spring, 20 in the fall. In Study 2, 20 students were tested in the spring, 67 in the fall. However, time of year did not predict student's learning in either study (Gesture $\chi^2(1) = 0.08$, $p = .60$; Action $\chi^2(1) = 0.27$, $p = .60$).

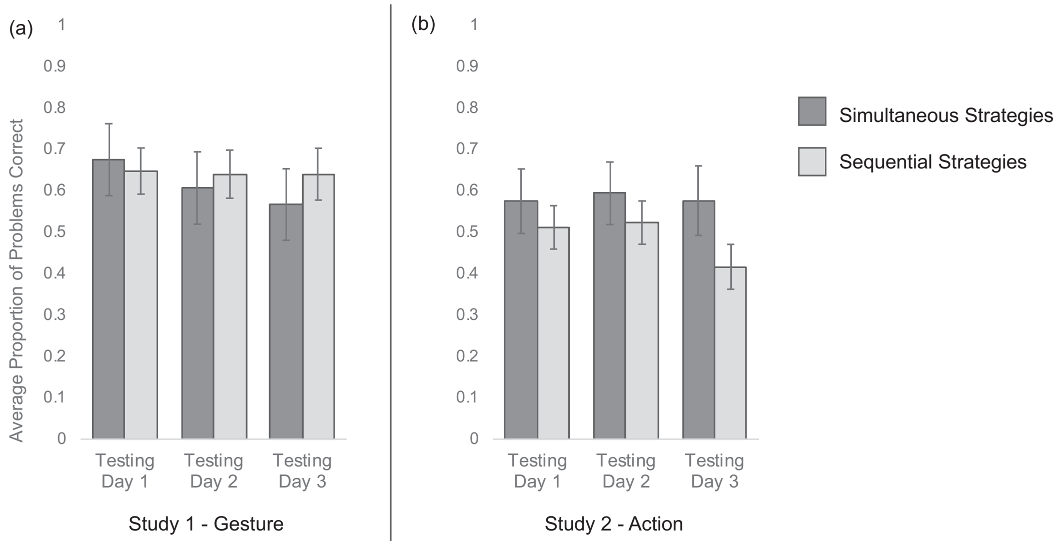

Fig. 3. Average proportion of problems correct at each time point for gesture training in Study 1 (panel a) and action training in Study 2 (panel b). Testing session had no effect on retention following gesture training, but did have a significant effect on retention following action training—children retained what they had learned better when action and speech were produced simultaneously than when the two were produced sequentially.

sheds light on a potential mechanism.

7.1.1. Doing gesture is different from seeing gesture

When children observe a teacher producing gesture, that gesture is a more effective teaching tool if it is produced simultaneously with speech than if it is produced sequentially with speech (Congdon et al., 2017). But when children produce the gestures themselves, gesture is effective in promoting learning whether it is produced simultaneously or sequentially with speech (Study 1). Synchronizing gesture with speech thus appears to matter for seeing gesture, but not for doing gesture. This finding is methodologically important because researchers who try to blindly generalize across studies using self-produced gesture and studies using observed gesture may arrive at a muddled picture of how gesture helps learners.

The finding that gesture and speech do not need to be *produced* simultaneously to promote learning also fleshes out previous findings on self-produced gesture. Both Cook et al. (2008) and Brooks and Goldin-Meadow (2016) found that children improved on a mathematical equivalence task after producing a problem-solving strategy in gesture (without speech) and subsequently listening to an instructor describe how to solve the problem in speech. We suggested in the introduction that this effect might reflect gesture's ability to influence learning when it is produced on its own without speech. However, in light of our Study 1 findings, another possibility is that, when gesture is produced by a learner, there may be a long window during which the gesture can be integrated with speech (much longer than the window when the learner observes gestures). Perhaps children in the previous studies were able to integrate information gleaned from the gestures they themselves produced with information later produced in speech by the experimenter. Manipulating the timing of the experimenter's subsequent spoken instruction could shed light on how long the integration window is for speech and self-produced gesture.

7.1.2. Action is less tightly tied to speech than gesture is

Study 1 shows us that it is not essential to produce gesture simultaneously with speech to promote learning and retention. Study 2 tells us that this effect is particular to gesture and does not extend to hand movements produced on objects. Like gesture, action on objects promotes learning. However, unlike gesture, action is more likely to lead to retention when the actions are produced simultaneously with speech than when they are produced sequentially with speech. We suggest that action on objects may behave differently from gesture because action is less tightly integrated with speech than gesture is (see Church et al., 2014; Kelly et al., 2014). When children produce one problem-solving strategy in speech and a different strategy in action, they may have to expend a significant amount of effort to process and integrate the information in speech and action. As a result, they may not be able to hold the two strategies in mind as easily as they can when they produce these same two strategies in speech and gesture.

According to dual-coding theories of learning (e.g., Baddeley, 1999; Chandler & Sweller, 1991; Mayer, 2002, 2005), learners benefit from information presented in more than one modality because there are limitations on our ability to process information in any one modality. Adding a second modality can help us go beyond those limitations and therefore benefit from information presented in more than one modality. In both Studies 1 and 2, children use two modalities to express strategies for solving the mathematical equivalence problems, and learn to solve the problems after producing these strategies.

The fact that temporal alignment works differently for speech and gesture than for speech and action does not contradict the tenets of dual-coding. However, it does suggest that the temporal window over which information from the two modalities can be integrated depends on the modalities. We hypothesize that the integration window for speech and gesture may be longer than the window for speech and action. Speech and gesture are more tightly aligned than many other dual inputs a child receives (e.g., speech and action on objects; speech and pictures; speech and diagrams). As a result, integrating speech with other non-gesture modalities (including speech and action) may require more cognitive effort than integrating speech with gesture. Expending cognitive effort may, in turn, limit the size of the window over which speech and non-gesture modalities can be integrated (at least when they are produced by the learner).

7.1.3. Speech may not be needed to provide context for self-produced gesture

In their study of teacher-produced speech and gesture, Congdon et al. (2017) hypothesized that, when observed by learners, simultaneous speech and gesture is more effective than sequential speech and gesture because the temporal synchrony between the two modalities allows speech to provide relevant context for gesture. This context may make it easier for learners to see gesture as meaningful. But our data suggest that, when produced by learners, simultaneous speech and gesture is *not* more effective than sequential speech and gesture. When children produce one problem-solving strategy in speech and a different strategy in gesture, they seem to be able to hold both strategies in mind over a longer window than when they observe their teacher producing these same two strategies. This longer window might then render the temporal alignment between gesture and speech less important.

Support for this hypothesis comes from our finding that order in the two sequential gesture conditions did not affect children's learning and retention. Children learned and retained knowledge of mathematical equivalence whether they produced gesture before speech or speech before gesture—it does not seem to matter whether gesture provides context for speech or speech provides context for gesture because, under this hypothesis, when self-produced, the two are held in memory over a relatively long window.

7.1.4. The motor system cannot fully account for the effect gesture has on learning

Our data also speak to the hypothesis that producing gesture facilities learning because it engages the motor system in the learning process (e.g., Macedonia et al., 2011; Wakefield et al., 2019). When we compared levels of retention after learning through gesture vs. through action, we found that, overall, retention was significantly better after children produced gesture than after they produced action. Engaging the motor system may be a factor in explaining how producing gesture helps learning, but it cannot be the whole story—or else we would not have found a difference between learning through gesture and learning through action, both of which engage the motor system.

The natural connection between gesture and speech may have given children who learned through gesture an extra boost over children who learned through action, which aligns less naturally with speech than gesture does (Church et al., 2014; Kelly et al., 2014). Gesture may therefore be a better teaching tool than action, at least in this context. In terms of generalizing our findings, it is worth pointing out that the actions we used in this study resemble what Clark and Gerrig (1990) call *demonstrations*. Unlike many actions, demonstrations have a link to the speech they accompany, although it is less tight than the link between gesture and speech. In this context, it is important to note that even demonstration actions that link to speech are a less powerful teaching tool than gesture.

7.2. Implications for practice

In addition to contributing to our understanding of the mechanisms underlying gesture's impact on learning, our findings highlight two suggestions for teachers.

First, although teachers need to be aware of the temporal alignment between their own speech and gesture (and produce gesture simultaneously with speech), they need not insist on temporal alignment between their students' speech and gesture. It is not essential that children produce their speech and gesture together in order for them to take full advantage of its learning benefits. Simply encouraging students to produce meaningful gesture and speech strategies, in any order, during a math lesson is enough to scaffold learning.

Second, encouraging students to produce meaningful gesture during a math lesson has a bigger effect on retention than encouraging them to produce meaningful actions. Involving the body in instruction does promote learning. But recruiting the body to gesture rather than act directly on objects leads not only to immediate learning, but also to learning that lasts, no matter what the temporal alignment between gesture and speech.

7.3. Remaining questions

Our studies contribute to our understanding of the role that gesture, action, and speech play in learning. However, questions remain.

One question to consider is whether the results would have been the same if the movement strategies (action or gesture) had conveyed the same information as the speech strategy—in other words, do the findings depend on the fact that children were required to integrate two distinctly different strategies (equalizer and grouping) for solving mathematical equivalence problems? Although this question needs to be tested empirically, it is worth noting that children can display similar rates of learning when they produce the same strategy in speech and gesture as when they produce different strategies in speech and gesture (Wakefield & James, 2015). Simultaneity of speech and gesture was not manipulated in that study, but the study does suggest that having two routes to a problem-solving strategy, one in speech and the other in gesture, can benefit a learner no matter how much overlap there is in the information conveyed in the two modalities. Future work is needed to determine whether the temporal alignment between speech and gesture, when the two convey the same strategy, has the same impact on learning from self-produced gesture and learning from observed gesture.

A second question pertains to our finding that, unlike retention following self-produced gesture, retention following self-produced action is better when the action occurs simultaneously with speech than when it occurs sequentially with speech. Future work is needed to determine whether this difference extends to observed actions. Because temporal alignment with speech matters for retention even for gesture when it is observed (Congdon et al., 2017), we suspect that temporal alignment with speech will also matter when action is observed—that is, the difference between gesture and action that we have found here for self-produced movements may disappear when the two types of movements are observed.

8. Conclusion

Previous work has shown that the temporal alignment between speech and gesture matters for learning when the learner *observes* the teacher produce speech and gesture—learning and retention are better when learners see gesture produced simultaneously with speech than when they see it produced sequentially with speech (Congdon et al., 2017). We have found here that temporal alignment does not matter for learning and retention when speech and gesture are *produced* by the learner. Importantly, this effect does not extend to all types of self-produced hand movements. Instead, it appears to be a special feature of gesture—children learn and retain information from action better if the strategies they produce in action *are* temporally aligned with the strategies they produce in speech. Taken together, our findings implicate the tight relation that gesture has to speech in explaining gesture's impact on learning, and point to an important difference in how seeing gesture vs. doing gesture affects learning.

Funding

Funding for this study was provided by the National Science Foundation (EHR 1561405) to Susan Goldin-Meadow (PI) and Elizabeth M. Wakefield (co-PI). We also thank Casey Hall for her help with data coding.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104604.

References

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science*, 1–19. https://doi.org/10.1111/cogs.12292 (online first).

Cassell, J. D., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition, 7*, 1–34. https://doi.org/10.1075/pc.7.1.03cas.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332. https://doi.org/10.1207/s1532690xci0804_2.

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. https://doi.org/10.1080/01690965.2013.857783.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language, 66*, 764–805.

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. https://doi.org/10.1016/j.learninstruc.2017.03.005.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047–1058. https://doi.org/10.1016/j.cognition.2007.04.010.

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin, 145*, 765–784. https://doi.org/10.1037/bul0000202.

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

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. (2009). Gestures gives children new ideas about math. *Psychological Science, 20*, 267–271. https://doi.org/10.1111/j.1467-9280.2009.02297.x.

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science, 15*, 876–884. https://doi.org/10.1111/j.1467-7687.2012.01185.x.

James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental Science, 13*, 279–288. https://doi.org/10.1111/j.1467-7687.2009.00883.x.

James, K. H., & Swain, S. N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental Psychology, 14*, 1–6. https://doi.org/10.1111/j.1467-7687.2010.01011.x.

Kelly, S. D., Healy, M., Ozyurek, A., & Holler, J. (2014). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-014-0681-7.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton and Co.

Longcamp, M., Anton, J.-L., Roth, M., & Velay, J.-L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *NeuroImage, 19*, 1492–1500. https://doi.org/10.1016/s1053-8119(03)00088-0.

Longcamp, M., Tanskanen, T., & Hari, R. (2006). The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *Neuroimage, 33*, 681–688. https://doi.org/10.1016/j.neuroimage.2006.06.042.

Macedonia, M., Muller, 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. https://doi.org/10.1002/hbm.21084.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6.

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.

McNeil, N. M. (2014). A "change-resistance" account of children's difficulties understanding mathematical equivalence. *Child Development Perspectives, 8*, 42–47. https://doi.org/10.1111/cdep.12062.

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

Novack, M., Congdon, E., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*, 903–910. https://doi.org/10.1177/0956797613518351.

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition, 146*, 339–348. https://doi.org/10.1016/j.cognition.2015.10.014.

R Core Team. (2019). *R version 3.6.1 "action of the toes"*. Vienna, Austria: R Foundation for Statistical Computing.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*, 85–89. https://doi.org/10.1111/j.0956-7976.2005.00786.x.

Wakefield, E. M., Congdon, E. L., Novack, M. A., Goldin-Meadow, S., & James, K. H. (2019). Learning math by hand: The neural effects of gesture-based instruction in 8-year-old children. *Attention, Perception, & Psychophysics*, 1–11. https://doi.org/10.3758/s13414-019-01755-y.

Wakefield, E. M., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental Science, 21*, Article e12656. https://doi.org/10.1111/desc.12656.

Wakefield, E. M., & James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Developmental Psychology, 5*, 1105–1114. https://doi.org/10.1037/a0039471.

Wakefield, E. M., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science, 21*, Article e12664. https://doi.org/10.1111/desc.12664.