



Cognitive Science (2018) 1–22

Copyright © 2018 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12603

## Mental Transformation Skill in Young Children: The Role of Concrete and Abstract Motor Training

Susan C. Levine,<sup>a</sup> Susan Goldin-Meadow,<sup>a</sup> Matthew T. Carlson,<sup>b</sup>  
Naureen Hemani-Lopez<sup>c</sup>

<sup>a</sup>*Department of Psychology, Department of Comparative Human Development and Committee on Education, University of Chicago*

<sup>b</sup>*Department of Spanish, Italian, and Portuguese and Center for Language Science, The Pennsylvania State University*

<sup>c</sup>*Department of Psychology, University of Chicago*

Received 11 May 2016; received in revised form 20 January 2018; accepted 24 January 2018

---

### Abstract

We examined the effects of three different training conditions, all of which involve the motor system, on kindergarteners' mental transformation skill. We focused on three main questions. First, we asked whether training that involves making a motor movement that is relevant to the mental transformation—either concretely through action (action training) or more abstractly through gestural movements that represent the action (move-gesture training)—resulted in greater gains than training using motor movements irrelevant to the mental transformation (point-gesture training). We tested children prior to training, immediately after training (posttest), and 1 week after training (retest), and we found greater improvement in mental transformation skill in both the action and move-gesture training conditions than in the point-gesture condition, at both posttest and retest. Second, we asked whether the total gain made by retest differed depending on the abstractness of the movement-relevant training (action vs. move-gesture), and we found that it did not. Finally, we asked whether the time course of improvement differed for the two movement-relevant conditions, and we found that it did—gains in the action condition were realized immediately at posttest, with no further gains at retest; gains in the move-gesture condition were realized throughout, with comparable gains from pretest-to-posttest and from posttest-to-retest. Training that involves movement, whether concrete or abstract, can thus benefit children's mental transformation skill. However, the benefits unfold differently over time—the benefits of concrete training unfold immediately after training (online learning); the benefits of more abstract training unfold in equal steps immediately after training (online learning) and during the intervening week with no additional training (offline learning). These findings have implications for the kinds of instruction that can best support spatial learning.

---

Correspondence should be sent to Susan C. Levine, Department of Psychology, Department of Comparative Human Development and Committee on Education, University of Chicago, 5848 S. University Ave., Chicago, IL 60637. E-mail: s-levine@uchicago.edu

*Keywords:* Mental transformation; Mental rotation; Training study; Gesture; Development; Motor system

---

## 1. Introduction

Spatial thinking, including the ability to mentally transform objects, is related to achievement in the STEM (Science, Technology, Engineering, and Mathematics) disciplines, even after controlling for language and mathematical skills (e.g., Casey, Nuttall, & Pezaris, 2001; Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). Fortunately, spatial skills are malleable and can be improved by engaging in a variety of activities and practice (e.g., Baeninger & Newcombe, 1989; De Lisi & Wolford, 2002; Ehrlich, Levine, & Goldin-Meadow, 2006; Green & Bavelier, 2003; Sorby, 2009; Terlecki, Newcombe, & Little, 2008; Uttal et al., 2013). However, we know little about the relative effectiveness of different kinds of instruction in promoting spatial thinking, a question that has implications for educational efforts to increase the STEM pipeline. In this study, we address this question by comparing the gains young children make on an age-appropriate mental transformation task after they receive different kinds of training.

A substantial body of literature suggests that mental transformation tasks engage the motor system. Notably, mental rotation in adults shares behavioral and neural signatures with actually rotating the objects (e.g., Ganis, Keenan, Kosslyn, & Pascual-Leone, 2000; Gardony, Taylor, & Brunyé, 2014; Kosslyn, Digirolamo, Thompson, & Alpert, 1998; Kosslyn, Thompson, Wraga, & Alpert, 2001; Parsons, 1987; Parsons et al., 1995; Sekiyama, 1983; Shepard & Cooper, 1982; Shepard & Metzler, 1971; Vingerhoets, de Lange, Vandemaële, Deblaere, & Achten, 2002; Wexler, Kosslyn, & Berthoz, 1998; Wiedenbauer, Schmid, & Jansen-Osmann, 2007; Wohlschläger & Wohlschläger, 1998). Moreover, motor system involvement in mental rotation tasks has been found across a broad range of ages including infants in the first year of life (e.g., Frick & Wang, 2014; Mohring & Frick, 2013; Schwarzer, Freitag, Buckel, & Lofruth, 2013; Schwarzer, Freitag, & Schum, 2013) and children (e.g., Frick, Daum, Walser, & Mast, 2009; Wiedenbauer & Jansen-Osmann, 2008).

Consistent with this evidence, encouraging the use of the motor system during mental rotation tasks can improve or interfere with performance, depending on whether the movement made is consistent with, or conflicts with, the mental transformation being performed. For example, training that encourages adults (Chu & Kita, 2011) or children (Goldin-Meadow et al., 2012) to gesture the movements they are carrying out when solving mental rotation problems improved performance, compared to control conditions. In contrast, having participants perform conflicting motor movements during a mental rotation task resulted in decrements in the mental rotation performance of 5- and 8-year-old children but not of 11-year-olds, suggesting that motor processes may play an even larger role in mental rotation in younger than in older children (Frick et al., 2009).

Building on these findings, in this study, we for the first time compare the effects of different kinds of motor training in improving young children's ability to visualize the results of a spatial transformation. Although most studies have focused on a specific kind of mental transformation—mental rotation—we have found that, in young children, the ability to mentally rotate shapes is related to the ability to mentally translate shapes, and our task involves both kinds of transformations (Levine, Huttenlocher, Taylor, & Langrock, 1999). The transformations we use involve mentally combining two pieces to make a whole shape, through rotation, translation, or both. We focus on kindergarten-age children for two reasons, the first practical and the second theoretical. Practically, this is the youngest age at which children perform above chance on our target task, a task that involves mentally transforming pieces to form a whole object (Levine et al., 1999). Theoretically, we focus on young children because of evidence that spatial thinking may be more susceptible to training during early years (e.g., Uttal et al., 2013).

Our study design included three training conditions: (a) *action training*, in which children were asked to move two pieces together to make a target shape; (b) *move-gesture training*, in which children were asked to gesture the movement needed to bring the two pieces together to make the target shape; unlike the action condition, this condition does not involve direct manipulation of objects, but rather represents that manipulation in a more abstract way; and (c) *point-gesture training (our control condition)*, in which children were asked to point to the two pieces that needed to be brought together to make the target shape; this condition engages the motor system, but not in any way that reflects the manipulation and movement of the objects. We use these three training conditions to address three questions.

1. Does movement relevant to the task improve learning better than irrelevant movement?

We first ask whether the action and 916 Run Run Shaw Tower Centennial Campus training are more effective in improving children's mental transformation skill than the control point-gesture training. Consistent with Hostetter and Alibali's (2008) hypothesis that gesture reflects action simulation, we hypothesized that action and move-gesture training would result in significantly greater improvement in mental transformation skill than the control point-gesture condition. Of note, unlike the action and move-gesture training conditions, the control condition does not provide movement information relevant to the mental transformation task, but rather provides deictic information, focusing the child's attention on the relevant pieces. Thus, comparing the action and move-gesture conditions to the control condition allows us to examine whether movement-relevant information is important to children's learning.

2. Does the abstractness of the relevant movement (action vs. move-gesture) affect how much is learned overall?

Second, we ask whether the abstractness of the relevant movement in training matters in terms of learning outcomes. That is, does concrete action training support learning on the mental transformation task differently from the more abstract move-gesture training?

Favoring the hypothesis that action training is particularly effective in improving the mental transformation skill of young children, traditional cognitive development theories posit that children often solve problems by acting on physical objects prior to being able to solve them symbolically (e.g., Bruner, Olver, & Greenfield, 1966; Piaget, 1953). If so, action training, which involves directly manipulating objects, might result in more learning than gesture training, which does not involve direct object manipulation.

In contrast to this traditional view, and favoring the hypothesis that move-gesture training is particularly effective in improving young children's mental transformation skill, recent studies show that concreteness can hurt generalization by focusing attention on perceptual details that are peripheral or irrelevant to the task (e.g., Goldstone, Medin, & Gentner, 1991; Goldstone & Son, 2005; McNeil, Uttal, Jarvin, & Sternberg, 2009; Mix, 2010, 2010; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014; Uttal, Scudder, & Deloache, 1997). In the context of our mental transformation task, acting directly on objects might encourage learners to focus on the outcome of the movement, rather than on the process of the transformation itself. In contrast, relevant move-gestures might help learners focus on the process of transforming the shapes and visualizing the outcome of the transformation, skills that are essential to success in mental transformation tasks (Ehrlich et al., 2006; Goldin-Meadow, 2010). Furthermore, because gesturing does not result in a transformed object, it might create "transfer appropriate learning" by engaging learners in the same kind of processing during training that they engage in at test (e.g., Franks, Bilbrey, Lien, & McNamara, 2000; Morris, Bransford, & Franks, 1977). That is, move-gesture training, like the test trials on the mental transformation task, requires the child to imagine the transformation of pieces and the resultant shape they make. In contrast, action training, because it involves moving pieces together to form the resultant of shape, does not. Thus, the processing required by the move-gesture training, at least on the face of it, seems more similar to the processing required on the test trials (where it is not possible to move pieces), and this similarity might favor learning in the move-gesture condition.

3. Does the abstractness of the relevant movement (action vs. gesture) affect how learning unfolds over time?

The third question we ask is whether the time course of improvement differs, particularly in the action versus move-gesture conditions where improvement is expected. The majority of spatial training studies, particularly those with children, examine learning immediately after training. The few studies that have examined performance after a delay have generally shown durable gains of spatial training (Uttal et al., 2013). But the question we ask here is whether type of training affects how gains unfold over time.

We know from previous studies of motor learning that easier training regimens often result in the greatest immediate performance gains, but do not always lead to the greatest long-term learning gains (e.g., Guadagnoli & Lee, 2004). We might then expect action training to result in more immediate gains than move-gesture training because it provides an easier form of training, one that reveals the answer to the child. In contrast, the more challenging move-gesture training might result in more long-term learning or, at the least, more prolonged learning.

Along these lines, recent evidence raises the possibility that the effects of motor training may actually continue to grow after training has ceased—an effect dubbed “offline” learning (e.g., Censor, Sagi, & Cohen, 2012; Debarnot, Piolino, Baron, & Guillot, 2013; Sami, Robertson, & Miall, 2014). However, to date, no study has examined whether the offline learning following motor training differs as a function of the abstractness of the motor training by comparing training that involves directly moving objects (action) to training that involves gesturing about those movements (move-gesture). There is reason to believe that move-gesture training could improve offline learning more than action training because, unlike action training, it involves generating an answer to a problem, which has been shown to lead to less immediate benefit, but greater long-term benefit in both children and adults, consistent with greater offline learning (e.g., generation effects, Bjork & Bjork, 2014; Vlach, 2014; Vlach, Ankowski, & Sandhofer, 2012). Prior studies have found that desirable difficulties, or the related construct of optimal challenge points, can slow down initial learning, but then lead to enhanced offline learning and greater long-term gains (e.g., Auble & Franks, 1978; Benjamin, Bjork, & Schwartz, 1998; Bjork, 1994; Guadagnoli & Lee, 2004; Halamish & Bjork, 2011; McDaniel & Mason, 1985; Vlach, 2014; Vlach et al., 2012). The results of one previous study are broadly consistent with the hypothesis that gesture training can lead to offline learning. Children were taught how to solve math equivalence problems, and their improvement after the lesson was measured immediately after training, and also 4 weeks later. Children who both saw and produced gesture during training performed no better on the test immediately after training than children who did not see or produce gesture during training. However, the children who experienced move-gesture training performed significantly better at the 4-week test, suggesting that gesture training had led to offline learning (Cook, Mitchell, & Goldin-Meadow, 2008; see also Congdon et al., 2017).

To recap, we extend research on training mental rotation skill in three ways. First, we examine whether training that engages the motor system in relevant movements via action or gesture is more effective than a control point condition, which engages the motor system in a movement that guides attention but is irrelevant to the mental transformation. Second, we directly compare the gains children make when given training that involves the motor system at different levels of abstraction—action on objects versus movement gestures that simulate these actions. Finally, we examine the time course of learning when the motor system is engaged at different levels of abstraction, exploring the possibility that, because it creates a “desirable difficulty,” move-gesture training might lead to fewer gains at immediate posttest than action training, but to greater gains 1 week later even without any intervening training.

## 2. Methods

### 2.1. Participants

A total of 114 5-year-old and 6-year-old children attending kindergarten (62 girls and 52 boys,  $M_{age} = 74.18$  months,  $SD = 4.82$  months; range: 62–84 months) participated in

the study. Thirty-seven additional children were eliminated because they were not native English speakers and their level of English proficiency was not sufficient for them to understand the task instructions, which were given in English ( $n = 33$ ); because they did not follow the experimenter's instructions during training ( $n = 1$ ); or because they performed near ceiling (responding correctly on 11 or 12 of the 12 pretest problems) at pretest ( $n = 3$ ).

Kindergarten classrooms in a large urban area were recruited through phone calls and e-mails to school principals. The study was conducted at schools and all children had a signed parental consent form with prior assent from the parents for the children's participation. The population was predominantly Caucasian—of the 114 children who participated in the study, 102 reported ethnicity (72.5% Caucasian, 7.8% Asian, 2.9% Black, 13.7% mixed; and 2.9% other). Socioeconomic status was predominantly middle to upper-middle class, based on parents' self-reported education level. Of the 103 participants who reported parental education level, 88.35% reported an education level at a bachelor's degree or higher.

## 2.2. Materials and stimuli

During the first session, participants were given a pretest assessment of their mental transformation skill, training, and a posttest assessment of their mental transformation skill. The mental transformation task was adapted from Levine et al. (1999). During the second session, 1 week later, they were administered a second test, which we call the retest. The pretest, posttest, and retest each consisted of 12 test items, followed by six additional explanation items. Children were asked to solve each problem and, on the last six problems, to tell the experimenter how they got their answers immediately following each of these problems.

### 2.2.1. Pretest, posttest, and retest stimuli

Stimuli were presented in a vertically oriented loose-leaf binder where the “pieces card” (containing pictures of the two pieces that needed to be put together) and a  $2 \times 2$  “choice card” array were simultaneously shown to the child. Each of these cards was presented on an  $8.5 \times 11$  inch piece of paper, with the pieces card presented below and closer to the child than the choice array. Participants were asked to choose the target shape that could be formed by the two pieces, which were created by halving the target form along an axis of symmetry. Half of the problems involved pieces that were symmetrical along the horizontal axis; half were symmetrical along the vertical axis. The location of the target shape on the choice card was randomized across trials with the constraint that consecutive trials did not have the same location for the target shape.

There were four types of problems that differed in the spatial transformation needed to create the target shape (see Fig. 1): (a) *Direct Translation*, where pieces had to be moved perpendicular to the line of symmetry of the target shape; (b) *Diagonal Translation*, where pieces had to be moved diagonally to create the target shape; (c) *Direct Rotation*, where pieces had to be rotated 45 degrees and moved perpendicular to the line of

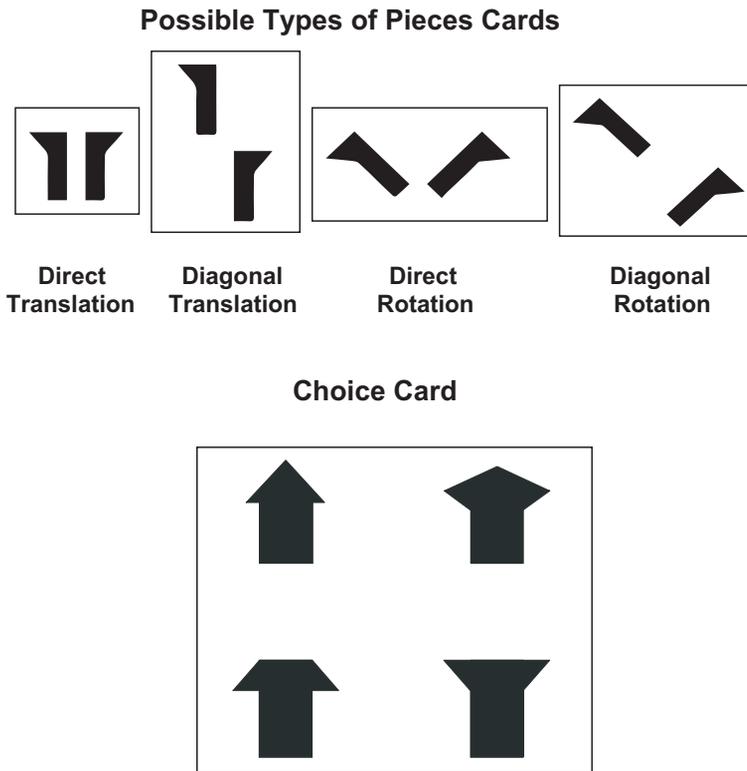


Fig. 1. The pieces cards show the four different types of spatial transformations used on different assessment items. Each choice card is paired with only one type of pieces card. Piece card type varies across the items. Note that both the choice array and the pieces card are displayed on 8"  $\times$  11" pieces of paper.

symmetry to create the target shape; and (d) *Diagonal Rotation*, where pieces had to be rotated 45 degrees and then moved diagonally to create the target shape. In half of the *Diagonal Translation* and *Diagonal Rotation* problems, the piece on the left was higher than the piece on the right, and vice versa for the other half. Previous studies report that translation problems are easier than rotation problems (but performance on these problem types is correlated), and that patterns of performance on these problem types is similar with respect to strategy and sex differences (Ehrlich et al., 2006; Levine et al., 1999). We included these types of problems to make sure that we were able to sample the variability in mental transformation skill likely to be present among 5- to 6-year-olds, not to examine whether the effects of training differed depending on problem type, which would require many more items of each type.

The four problem types were counterbalanced across participants, using two different order sequences. At each time point (pretest, posttest, and retest), 12 problems were presented, three instances of each of the four types (i.e., 3 *Direct Translation*, 3 *Diagonal Translation*, 3 *Direct Rotation*, and 3 *Diagonal Translation*). Following the administration of these problems, six explanation problems were given consisting of the following

problem types: 1 *Direct Translation*, 2 *Diagonal Translations*, 2 *Direct Rotations*, and 1 *Diagonal Rotation*.

Two different forms of the mental transformation test, each involving different shapes, were used for pretest and posttest and were counterbalanced across participants. Half of the participants received Form A for pretest and Form B for posttest; the other half received Form B for pretest and Form A for posttest. In each group, the retest form was the same as the participants received at the immediate posttest. At each testing time point, problems were presented in a different fixed random order.

### 2.2.2. *Pre-training and training stimuli*

There were four pre-training items on which the experimenter taught the child to make the relevant action or gesture, including one item of each transformation type—direct translation, diagonal translation, direct rotation, and diagonal rotation. Following these pre-training items, the child completed eight training problems, two of each problem type. The training problems differed from the pretest, posttest, and retest problems in that the pieces cards were replaced by actual black wooden pieces placed in a covered, clear Plexiglas container, 7.5 inches square. Furthermore, the shapes and pieces presented were not the same as those presented at the various testing time points.

### 2.3. *Design and procedure*

All sessions were videotaped with prior consent of the participants' parents. All children were given a pretest, training, and an immediate posttest during Session 1, and a retest 1 week later during Session 2. On the first pretest trial, the experimenter said, "Look at the pieces" (while pointing at the pieces card). "Now look at the shapes" (while pointing at the choice card). "If you put these pieces (point at pieces card) together they will make one of these shapes (point at choice card). Point to the shape that the pieces make." On subsequent trials, the experimenter only said, "Point to the shape that the pieces make." On the six explanation problems, the child answered each problem and was then asked to explain how he or she arrived at the answer. No feedback was given on any pretest or explanation item.

Children were tested individually and randomly assigned to one of three training conditions: action ( $n = 41$ ), move-gesture ( $n = 38$ ), point-gesture ( $n = 35$ ). At the start of the training portion of the experiment, a second experimenter showed participants the action or gesture they were to produce before choosing the target shape. Instructions and procedures for each of the training conditions are shown in Table 1. In all three conditions, two wooden pieces were shown to the child inside the Plexiglas container described earlier. In the action condition, the lid of the container was removed so the child could move the pieces and, in the other two conditions, the lid remained on so the child could not touch or move the pieces. In the action training condition, after the participant physically moved the pieces together, the experimenter separated the pieces, returned them to their original location, and placed the clear lid back on before asking the child to choose the answer. This procedure prevented the child from choosing the answer by simply matching

Table 1

Instructions given on the eight training problems to children in each of the three conditions. Instructions that varied across condition are in bold italics

Intervention Condition	Instructions and Procedure
Action	<b>Experimenter to child:</b> “If these two pieces are moved together ( <i>point with flat hand above pieces</i> ), they will make one of these shapes ( <i>point with flat hand above shapes</i> ). First, <b><i>show me how to move the pieces together with both hands (child moves pieces, and then experimenter moves them apart)</i></b> . Now point to the shape the pieces make.”
Move-gesture	<b>Experimenter to child:</b> “If these two pieces are moved together ( <i>point with flat hand above pieces</i> ), they will make one of these shapes ( <i>point with flat hand above shapes</i> ). First, <b><i>show me with both hands how to move the pieces together (child moves hands)</i></b> . Now point to the shape the pieces make.”
Point-gesture	<b>Experimenter to child:</b> “If these two pieces are moved together ( <i>point with flat hand above pieces</i> ), they will make one of these shapes ( <i>point with flat hand above shapes</i> ). First, <b><i>point to the pieces (child points to pieces)</i></b> . Now point to the shape the pieces make.”

the shape completed by the action with the target shape. In the move-gesture condition, children used two flat hands to gesture moving the pieces together without touching the pieces before choosing the target shape. In the point-gesture condition, children pointed to the pieces with a flat hand before choosing the target shape.

There were four pre-training trials on which the second experimenter and the child took turns solving the problems (two problems each), with the experimenter going first. On each of these problems, the experimenter/child performed the action or gesture for the training condition the child was randomly assigned to complete (action, move-gesture, point-gesture). Following the pre-training trials, the child completed eight training trials. On each trial, the child was asked to perform the movement that was taught (action, move-gesture, point-gesture) before choosing the answer. After the training was completed, the first experimenter returned to administer the posttest in the same manner as the pretest (the posttest contained different problems than the pretest). One week later, the child was re-administered the same posttest that was given immediately after training (the retest). The activities the child completed at each of the two sessions are summarized in Table 2.

#### 2.4. Coding

We coded the card choices that children made on the 12 individual problems at each test time as correct or incorrect; we did not include the six problems on which children also gave explanations in this score. We used the six explanation problems given at pretest and posttest to code the movement gestures children spontaneously produced along with their speech. The co-speech gestures produced at pretest were entered as a control variable in our main analysis examining the effects of training condition because previous

Table 2  
Description of activities child engages in during each of the two test sessions

Session 1	Pretest	<ul style="list-style-type: none"> <li>• Child solves 12 mental rotation problems presented on paper.</li> <li>• Child then solves an additional six problems and explains how they arrived at their answer on each.</li> </ul>
	Pre-training	<ul style="list-style-type: none"> <li>• A second experimenter and child alternate practicing relevant action/gesture (depending on condition) with the experimenter going first using pieces presented in a Plexiglas container (four trials, two performed by the experimenter and two by the child).</li> </ul>
	Training	<ul style="list-style-type: none"> <li>• On each of eight training problems, the child carries out the relevant action/gesture and then tries to choose the correct shape from among four alternatives.</li> </ul>
	Immediate posttest	<ul style="list-style-type: none"> <li>• Child solves 12 mental rotation problems presented on paper.</li> <li>• Child then solves an additional six problems and explains how they arrived at their answer on each.</li> </ul>
Session 2	Retest	<ul style="list-style-type: none"> <li>• One week later, the child solves 12 mental rotation problems presented on paper.</li> <li>• Child then solves an additional six problems and explains how they arrived at their answer on each.</li> </ul>

work had found that children who spontaneously produce move-gestures on problems of this type tend to be more advanced in mental rotation than children who do not produce these gestures (Ehrlich et al., 2006). Gesture was coded as referring to movement if the child moved his or her hands in a straight or curved line, indicating that the pieces underwent a change in location, or if the child rotated his or her hands, indicating that the pieces underwent a change in orientation. The children typically used pointing hand-shapes, flat hands, or C-hands in their move-gestures. We calculated the total number of problems on which a child produced a move-gesture and used this number to create a covariate in our analyses.

### 3. Results

#### 3.1. Model structure and fitting

We analyzed the test trial-level data (pretest, posttest, retest) using mixed effects logistic regression, fitted using the *glmer* function from the *lme4* package in *R* (Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2014). Post hoc comparisons were carried out using the *glht* function in the *multcomp* package (Hothorn, Bretz, & Westfall, 2008), with *p*-values corrected for multiple comparisons based on Westfall, Tobias, Rom, Wolfinger, & Hochberg (1999). The full model specification is described in the remainder of this section, and full estimates are given in Table 3 (Wald tests of individual coefficients are given, each of which has one degree of freedom). Significance tests of three-level factors and interactions were performed by likelihood ratio tests and are reported in the text as relevant.

Table 3  
Model specifications and estimates

Dependent Variable: accuracy (by trial); 0 = incorrect, 1 = correct					
Fixed Effects	exp( $\beta$ )	$\beta$ (in logits)	SE	Wald $z$	$p$
Intercept	1.36	0.31	0.19	1.58	.11
Test_1 (pretest vs. posttest)	1.08	0.08	0.17	0.47	.64
Test_2 (pretest vs. retest)	1.37	0.32	0.18	1.72	.08
Condition_1 (Point vs. Move-Gesture)	0.78	-0.25	0.18	-1.39	.16
Condition_2 (Point vs. Action)	0.95	-0.05	0.18	-0.29	.77
Gender	0.76	-0.27	0.26	-1.02	.31
Pretest Gesturing	2.23	0.80	0.32	2.44	.01
Test_1: Condition_1	1.36	0.30	0.24	1.29	.20
Test_1: Condition_2	1.89	0.64	0.23	2.73	.006
Test_2: Condition_1	1.63	0.49	0.26	1.92	.06
Test_2: Condition_2	1.96	0.67	0.25	2.67	.008
Test_1: Gender	0.68	-0.39	0.34	-1.14	.26
Test_1: Pretest Gesturing	0.74	-0.31	0.42	-0.73	.47
Test_2: Gender	0.58	-0.54	0.37	-1.46	.14
Test_2: Pretest Gesturing	0.92	-0.09	0.44	-0.20	.84
Condition_1: Gender	0.83	-0.18	0.36	-0.50	.62
Condition_1: Pretest Gesturing	0.75	-0.28	0.48	-0.59	.55
Condition_2: Gender	1.53	0.43	0.37	1.16	.24
Condition_2: Pretest Gesturing	0.30	-1.18	0.46	-2.57	.01
Test_1: Condition_1: Gender	1.68	0.52	0.47	1.10	.27
Test_1: Condition_1: Pretest Gesturing	1.83	0.60	0.61	0.99	.32
Test_1: Condition_2: Gender	1.29	0.25	0.48	0.53	.60
Test_1: Condition_2: Pretest Gesturing	1.57	0.45	0.60	0.74	.46
Test_2: Condition_1: Gender	2.02	0.71	0.51	1.37	.17
Test_2: Condition_1: Pretest Gesturing	0.74	-0.31	0.65	-0.47	.64
Test_2: Condition_2: Gender	1.41	0.35	0.52	0.67	.50
Test_2: Condition_2: Pretest Gesturing	1.22	0.20	0.64	0.31	.76
Random Effects					
Grouping Factor	Effect	SD	Correlations <sup>a</sup>		
Subject	Intercept	0.46			
	Test_1	0.43	0.57		
	Test_2	0.52	0.67	0.99	
Item	Intercept	0.70			

Note. <sup>a</sup>The correlations are present only to capture variance due to the random grouping factors, and their significance is not tested.

The outcome variable was the binary accuracy on each trial. Test time (pretest, posttest, retest) was a within-subjects factor, and training condition (also with three levels, point-gesture, move-gesture, and action) and gender were between-subjects factors. Test time and training condition were treatment-coded with *pretest* and *point-gesture* as the reference levels. The interaction of test time and training condition tests our three central

research questions, and the selection of these reference levels was made with an eye to each question. In Table 3, the contrasts for test time are labeled Test\_1 (comparing *Pretest* to *Posttest*) and Test\_2 (*Pretest* vs. *Retest*). The contrasts for training condition are labeled Condition\_1 (*point-gesture* vs. *move-gesture*) and Condition\_2 (*point-gesture* vs. *action*).

Gender was coded as 0 (male) and 1 (female), and centered at its mean. Test time and training condition were allowed to interact in order to test our central hypotheses about the relative amounts of improvement exhibited by children in the three training conditions.

Based on evidence from Ehrlich et al. (2006) that producing movement gestures on the mental transformation task is correlated with improvement on the task, we also included the number of explanation problems at pretest on which children produced a move-gesture as a control variable (labeled Pretest Gesturing in Table 3). These were the additional six problems completed after the pretest, but before training, on which children were asked to explain their answers. The majority of these movement gestures occurred with speech about movement (79%).<sup>1</sup> Visual inspection revealed that this count was distributed bimodally, with 12 and 11 participants gesturing on 0 or 1 trial, respectively, 2 participants on 2 trials, 9 participants on 3 trials, and the remaining 83 gesturing on 4–6 trials. For this reason, children were divided into low-gesturers (0 or 1 trial, 20% of children overall; 7/36 in the point-gesture condition, 7/39 in the move-gesture condition, and 9/42 in the action condition) and high-gesturers (two or more trials, the remaining 80% of the children).<sup>2</sup> This variable, producing movement gestures on pretest explanation trials, was coded as 0 if the child produced move-gestures on zero or one trial, and 1 if the child produced move-gestures on two or more trials, and it was then centered at its mean. A likelihood ratio test showed that adding this variable, plus its interactions with test time and training condition (including the three-way interaction), did not improve the model ( $p > .23$ ). Similarly, adding the interactions of gender with test time and training condition also failed to show significant improvement of the model ( $p > .65$ ). Nonetheless, based on other studies showing gender effects on mental transformation tasks (e.g., Levine et al., 1999; Linn & Petersen, 1985; Moore & Johnson, 2008; Quinn & Liben, 2008; Voyer, Voyer, & Bryden, 1995) and the Ehrlich et al. (2006) findings showing a relation between movement gestures and learning, we retained these predictors and interactions in the model.

Random intercepts were included by item and subject. By-subject slopes for test time, including the random correlations between the by-subject intercepts and slopes, were also included. Regression coefficients,  $\beta$ , are reported in logit units and in odds ratios,  $\exp(\beta)$ .

### 3.2. *Learning as a function of condition*

We first compare the three groups' performance on pretest, and the effects of the covariates, before turning to our central research questions. The model revealed no group differences at pretest ( $p > .15$ ), as would be expected given that children were randomly

assigned to these conditions. As in previous studies using the mental transformation test, boys' scores were numerically higher than girls' at pretest ( $M_{\text{boys}} = 6.86$  [2.10];  $M_{\text{girls}} = 6.35$  [2.11]), but this difference was not significant, nor was there evidence that boys and girls differed in their improvement at the later tests or in the effects of training condition (all  $p > .2$  for effects and interactions involving gender).

Mean pretest performance was also slightly higher for children who had produced at least two movement gestures on pretest explanation problems ( $M_{\text{two or more gestures}} = 6.72$  [2.02];  $M_{\text{zero or one gesture}} = 6.00$  [2.47]). A significant interaction emerged between spontaneously producing move-gestures on pretest explanations and training condition ( $X^2(2) = 7.06$ ,  $p < .05$ ). Post hoc comparisons showed that children in the point-gesture condition who produced two or more move-gestures on pretest scored higher than those who produced zero or one gesture ( $\beta = 0.80$ ,  $SE = 0.33$ ,  $Z = 2.44$ ,  $p < .05$ ,  $\exp(\beta) = 2.22$ ). However, no significant effects of pretest move-gestures were found in the move-gesture ( $\beta = 0.27$ ,  $SE = 0.38$ ,  $Z = .70$ ,  $p > .85$ ,  $\exp(\beta) = 1.31$ ) or action ( $\beta = -0.44$ ,  $SE = 0.37$ ,  $Z = -1.18$ ,  $p > .55$ ,  $\exp(\beta) = 0.64$ ) conditions. Moreover, making more move-gestures during explanations on pretest did not interact with test time, nor was there a three-way interaction of these variables with training condition (all likelihood ratio tests yielded  $p > .5$ ).

Crucially, model comparison using a likelihood ratio test revealed that the interaction between test time and training condition (see the interaction terms for Test\_1 and Test\_2 with Condition\_1 and Condition\_2 in Table 3), central to our research questions, was significant ( $X^2(4) = 9.64$ ,  $p < .05$ ). To illustrate this interaction, the mean observed number of problems correct (out of 12), by test time and training condition, are given in Table 4, and the corresponding mean proportions of correct responses are plotted in Fig. 2, with standard errors estimated over 1000 nonparametric bootstrap samples (Agresti, 2012). We unpack this interaction by focusing on the trajectory of improvement across the three test times in each condition, examining online improvement (from pretest to posttest immediately following training) and offline improvement (from posttest to retest a week after training, with no further training occurring in the intervening period).

Our first question was whether motor-relevant training (i.e., action and move-gesture) led to greater gains than non-motor-relevant training (i.e., point-gesture). Planned comparisons confirmed that this was the case: the action group improved significantly more ( $\beta = 0.67$ ,  $SE = 0.25$ ,  $Z = 2.67$ ,  $p < .01$ ,  $\exp(\beta) = 1.95$ ), and the move-gesture group improved marginally more ( $\beta = 0.49$ ,  $SE = 0.26$ ,  $Z = 1.92$ ,  $p = .06$ ,  $\exp(\beta) = 1.63$ ), than the control point-gesture group.

Table 4

Mean (*SD*) number of correct responses (out of 12 possible) on pretest, posttest, and retest by condition

	Pretest	Posttest	Retest
Point-Gesture	6.80 (2.10)	6.88 (2.60)	7.38 (3.00)
Move-Gesture	6.26 (2.05)	7.16 (2.60)	8.06 (2.32)
Action	6.71 (2.20)	8.29 (1.98)	8.79 (2.03)

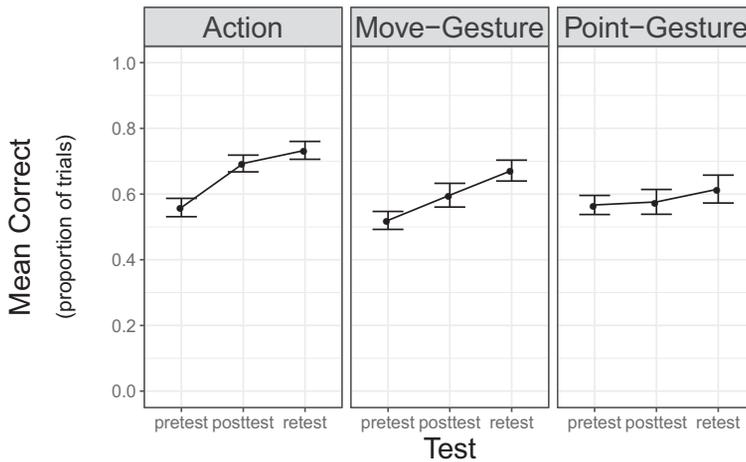


Fig. 2. Mean proportion correct, by test time and training group. Error bars show 95% CIs estimated via nonparametric bootstrap.

Our second question was whether the concreteness of the motor-relevant training (i.e., action vs. move-gesture) influenced total gains. We found that it did *not* in additional planned comparisons: the amount of improvement displayed over the entire period (pretest to retest) did not differ significantly for the move-gesture and action groups ( $\beta = 0.18$ ,  $SE = 0.25$ ,  $Z = .73$ ,  $p > .46$ ,  $\exp(\beta) = 1.20$ ).

Our third question was whether the concreteness of motor-relevant training (action vs. move-gesture) affects how children's gains unfold over time. Planned comparisons showed that children in the action group (Fig. 2, left panel) made significant online gains from pretest to posttest ( $\beta = 0.72$ ,  $SE = 0.16$ ,  $Z = 4.49$ ,  $p < .001$ ,  $\exp(\beta) = 2.05$ ), but showed no significant offline improvement from posttest to retest ( $\beta = 0.27$ ,  $SE = 0.16$ ,  $Z = 1.66$ ,  $p > .09$ ,  $\exp(\beta) = 1.31$ ). Overall, their performance at the retest was significantly better than at pretest; in other words, although they did not gain further after training, they showed no sign of loss of the gains made after training ( $\beta = 0.99$ ,  $SE = 0.17$ ,  $Z = 5.67$ ,  $p < .001$ ,  $\exp(\beta) = 2.69$ ). By comparison, children in the move-gesture group (Fig. 2, center) made approximately the same amount of improvement online and offline (online improvement  $\beta = 0.39$ ,  $SE = 0.16$ ,  $Z = 2.36$ ,  $p = .02$ ,  $\exp(\beta) = 1.48$ ; offline improvement  $\beta = 0.42$ ,  $SE = 0.16$ ,  $Z = 2.58$ ,  $p = .01$ ,  $\exp(\beta) = 1.52$ ). As in the action group, their retest performance was significantly better than their initial performance at pretest ( $\beta = 0.81$ ,  $SE = 0.18$ ,  $Z = 4.47$ ,  $p < .001$ ,  $\exp(\beta) = 2.25$ ) but, unlike the action group, their retest performance (as just noted) was also significantly better than their performance at posttest ( $\beta = 0.42$ ,  $SE = 0.16$ ,  $Z = 2.58$ ,  $p = .01$ ,  $\exp(\beta) = 1.52$ ). For completeness, we note that children in the point-gesture group (Fig. 2, right panel) showed no significant improvement whatsoever, either online ( $\beta = -0.08$ ,  $SE = 0.17$ ,  $Z = -0.47$ ,  $p > .63$ ,  $\exp(\beta) = 0.92$ ), offline ( $\beta = 0.23$ ,  $SE = 0.16$ ,  $Z = 1.43$ ,  $p > .15$ ,  $\exp(\beta) = 1.26$ ), or overall (i.e., from pretest to retest ( $\beta = 0.32$ ,  $SE = 0.18$ ,  $Z = 1.72$ ,  $p > .08$ ,  $\exp(\beta) = 1.38$ )).

Fig. 3 depicts another view of the results by presenting mean improvement for each group during online and offline learning. Note that both online and offline change is minimal for the point-gesture group, whereas improvement in both time frames was significant (see above for tests) for the move-gesture group, and only online improvement was significant for the action group.

To recap, both the action and move-gesture groups made significant gains over the course of the entire experiment, between pretest and retest, whereas the point-gesture group made no significant gains at all. Importantly, although the overall gains made by the action and move-gesture groups were not significantly different, the trajectories these groups followed to make these gains did differ. The action group profited immediately from training, as shown by their improved performance right after they received the training. However, their performance did not change further (in either direction) in the week following training, even though they were not at ceiling at either the time of the posttest or retest (see Fig. 2, left panel). In contrast, the move-gesture group not only showed gains right after training, but they also showed gains of similar magnitude in the week following training.

### 3.3. Training trial performance

We examined whether children's performance on the training trials differed across conditions, and whether their performance improved across the eight training trials. Accuracy for these trials was analyzed using mixed effects logistic regression with training trial order and condition as fixed effects, and covariates as above (a by-subject random slope for trial order was also included). We found better overall training trial performance in the action group than in the other two groups ( $\beta = 0.52$ ,  $SE = 0.20$ ,  $Z = 2.64$ ,  $p = .008$ ,  $\exp(\beta) = 1.69$ ), which is not surprising since the correct answer was visible after the action (but not after the gesture) in each training trial. However, we did not find significant improvement across the training trials in any group (all  $p > .4$ ).

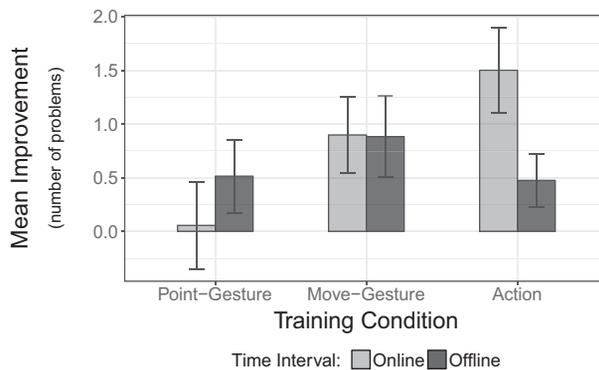


Fig. 3. Mean online (from pretest to posttest) and offline (from posttest to retest) improvement, by training group. Error bars show 95% CIs.

#### 4. Discussion

Our study focused on three questions. First, we asked whether training that involves the motor system—either concretely through action, or more abstractly through gestural movements that represent action—results in more gains in mental transformation skill, relative to a point-gesture control group that does not involve relevant motor movements. Second, we asked whether the overall magnitude of the improvement differs as a result of action versus gesture training. Finally, we asked whether the time course of improvement differs, particularly for the two movement training conditions.

With respect to the first question, we found greater improvement in children's mental transformation skill between pretest and retest (1 week after initial training) for children in both the action and move-gesture training conditions (overall improvement did not differ significantly for these two groups) than in the point-gesture condition, where children did not perform significantly higher than at pretest. This condition difference indicates that improvement in the action and move-gesture conditions, both of which involve movements that are relevant to the mental transformation, do not merely reflect gains that occur from practicing mental transformation problems, as the children in the point-gesture condition did not improve. These findings extend the literature supporting motor involvement in mental transformation by showing that encouraging relevant motor system involvement at different levels of abstraction holds promise for enhancing children's spatial thinking over time.

With respect to the second question, we did *not* find evidence that the move-gesture and action training conditions differed in terms of overall durable gains as assessed 1 week after training. Rather, both of the conditions resulted in significant gains, compared to children's level of performance at pretest, as well as compared to the control condition.

However, with respect to our third question about the time course of gains, we did find evidence of differences in the time course of gains between the move-gesture and action training conditions. Children in the action training condition made significant improvement on the mental transformation task immediately following training, with no significant subsequent improvement 1 week later—that is, they experienced online learning, but not offline learning. In contrast, children in the move-gesture training condition made similar gains from pretest to posttest and from posttest to retest; thus, experiencing both online and offline learning.

A key question is why children continued to show improvement even after training ended in the move-gesture condition. One factor that has been found to lead to offline learning is the similarity between task demands at training and at test (e.g., Franks et al., 2000; Halamish & Bjork, 2011; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006). According to Roediger and Karpicke's (2006) transfer-appropriate-processing hypothesis, training that engages learners in processes that are similar to those that will be needed at test are most effective in supporting performance at test. The desirable difficulty literature posits another, non-mutually exclusive factor that leads to greater offline learning—specifically, retrieval difficulty of learning trials (e.g., Vlach, 2014; Vlach

et al., 2012). According to the desirable difficulty framework (e.g., Bjork, 1994), as well as the optimal challenge point framework (e.g., Guadagnoli & Lee, 2004), training that engages people in deeper processing may slow down learning in the short run, but result in more learning gains over time (Vlach et al., 2012; Szpunar, Khan, & Schacter, 2013). The move-gesture training condition, which resulted in less online and more offline learning than the action training condition, contains both of these factors that have been shown to support robust long-term learning gains—similarity between training and test requirements, and challenging training that requires retrieval processes.

With respect to the first factor, similarity between training and test items, move-gesture training requires generating the result of a spatial transformation as does solving mental transformation problems at test. Because the move-gesture training condition does not yield an outcome, it might encourage learners to focus on the process of the transformation rather than on irrelevant perceptual details of particular transformations (Ehrlich et al., 2006; Goldin-Meadow, 2010), which, in turn, might lead to greater increments in mental transformation skill over time. Move-gesture training may thus involve more “transfer-appropriate processing,” and hence result in more offline learning, than action training.

With respect to the second factor, level of challenge in training, the move-gesture training condition is more difficult than the action training condition as evidenced by children’s worse performance on the training items in the move-gesture condition than in the action condition. In the action training condition, performance on training trials was high, as learners needed only to remember the shape that was formed for a brief period before selecting the correct shape from an array of four alternatives. Unlike in the action training condition, in the move-gesture training condition, children were required to generate the results of the transformation. Our pattern of results—robust online gains with no subsequent offline gains in the action condition and more modest online gains with continued offline gains in the move-gesture condition—is consistent with the desirable difficulty framework (e.g., Bjork, 1994; Vlach, 2014; Vlach et al., 2012) as well as with the transfer-appropriate processing framework (e.g., Franks et al., 2000; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006; Thomas & McDaniel, 2007). One or both of these factors could contribute to the timeline differences in learning in the action versus move-gesture training conditions.

However, the results of our point-gesture training condition make it clear that these factors (retrieval difficulty of training and match between training and test requirements), either together or separately, are not sufficient to account for our results. Like the move-gesture condition, the point-gesture condition requires retrieving the answer to the mental transformation problems (retrieval difficulty), as well as imagining the result of the transformation (matching training and test requirements). Nonetheless, the point-gesture condition does *not* result in either online or offline learning, likely because it does not recruit the motor system in a task-relevant way and thus is too challenging to support learning for kindergarten-age children. In accord with Guadagnoli and Lee’s (2004) optimal challenge point framework, the child’s challenge may need to be just right, not too big and not too small to support long-term learning. Our findings suggest that engaging the motor

system in a relevant way can help children meet the challenge that mental transformation presents.

The current findings raise some interesting questions for future studies. First, might optimal training involve first using concrete action training and then more abstract move-gesture training? Such an approach might help young children realize immediate gains and then build on these gains by increasing both challenge and the similarity of training demands to test demands. Second, might the effectiveness of action versus move-gesture training vary as a function of the level of proficiency the child begins with? It might be more effective to use action training with children who have low levels of skill on the mental transformation task, and to use move-gesture training with children who have higher levels of skill. Both of these hypotheses are motivated by, and are consistent with, the optimal challenge point framework (Guadagnoli & Lee, 2004). A third question asks whether long-term learning gains following move-gesture training will *exceed* gains following action training if more time is allowed to pass between training and retest. Such a finding would be consistent with the desirable difficulty and optimal challenge point frameworks, which suggest that the more difficult the “final test” (e.g., as a result of a longer delay), the more important it is to have training that requires retrieval or generation of information (e.g., Guadagnoli & Lee, 2004; Halamish & Bjork, 2011). Furthermore, previous work examining the role of gesture in learning suggests that the answer to this question may be *yes* because gesture training encourages learners to retain what they have learned (Cook et al., 2008) and even to go beyond this level (Congdon et al., 2017), at least in mathematics problem solving. Future studies are needed to determine whether learning at time points more remote from instruction than 1 week results in a learning advantage for an important spatial skill following move-gesture training, compared to action training. A positive answer to this question would hold important implications for the design of interventions aimed at improving spatial thinking, a cognitive capacity that has been implicated in STEM success.

## Notes

1. In addition to coding movement gestures on pretest explanation problems, we coded it on the immediate posttest explanation problems; we examined whether number of movement gestures produced at pretest correlated with number of movement gestures produced at posttest, and found a strong correlation,  $r = .70$ ,  $p < .001$ . However, no changes in gesturing on these explanation trials were observed between test times in any of the groups, nor were there group differences in gesturing on the explanation problems at either pretest or posttest (point-gesture group:  $M_{\text{pretest}} = 3.71$ ,  $SD = 2.01$ ,  $M_{\text{posttest}} = 3.80$ ,  $SD = 2.25$ ; move-gesture group:  $M_{\text{pretest}} = 4.26$ ,  $SD = 1.95$ ,  $M_{\text{posttest}} = 4.53$ ,  $SD = 1.66$ ; action group:  $M_{\text{pretest}} = 4.15$ ,  $SD = 2.03$ ,  $M_{\text{posttest}} = 4.15$ ,  $SD = 2.10$ ).

2. Changing the cutoff to include the two participants who produced move-gestures on two trials in the low gesture group, or encoding this variable as continuous did not change the results.

## References

- Agresti, A. (2012). *Categorical data analysis*. 3rd edition. Hoboken, NJ: John Wiley & Sons.
- Auble, P. M., & Franks, J. J. (1978). The effects of effort toward comprehension on recall. *Memory & Cognition*, *6*(1), 20–25. <https://doi.org/10.3758/BF03197424>.
- Baeninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*(506), 327–344. <https://doi.org/10.1007/BF00287729>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(7), 1–23.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. <https://doi.org/10.1037/0096-3445.127.1.55>.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bjork, E. L., & Bjork, R. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed.) (pp. 59–68). New York: Worth Publishers.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. New York: Wiley.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, *32*, 28–57. <https://doi.org/10.2307/749620>.
- Censor, N., Sagi, D., & Cohen, L. G. (2012). Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience*, *13*(9), 658–664. <https://doi.org/10.1038/nrn3315>.
- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*, 102–116. <https://doi.org/10.1037/a0021790>.
- 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>.
- De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology*, *163*(3), 272–282. <https://doi.org/10.1080/00221320209598683>.
- Debarnot, U., Piolino, P., Baron, J. C., & Guillot, A. (2013). Mental rotation: Effects of gender, training and sleep consolidation. *PLoS ONE*, *8*(3), e60296. <https://doi.org/10.1371/journal.pone.0060296>.
- Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*(6), 1259–1268. <https://doi.org/10.1037/0012-1649.42.6.1259>.
- Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory and Cognition*, *28*(7), 1140–1151. <https://doi.org/10.3758/BF03211815>.

- Frick, A., Daum, M. M., Walser, S., & Mast, F. W. (2009). Motor processes in children's mental rotation. *Journal of Cognition and Development, 10*(1–2), 18–40. <https://doi.org/10.1080/15248370902966719>.
- Frick, A., & Wang, S. (2014). Mental spatial transformations in 14- and 16-month-old infants: Effects of action and observational experience. *Child Development, 85*(1), 278–293. <https://doi.org/10.1111/cdev.12116>.
- Ganis, G., Keenan, J. P., Kosslyn, S. M., & Pascual-Leone, A. (2000). Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex, 10*(2), 175–180. <https://doi.org/10.1093/cecror/10.2.175>.
- Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2014). What does physical rotation reveal about mental rotation? *Psychological science, 25*(2), 605–612. <https://doi.org/10.1177/0956797613503174>.
- Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1–19. <https://doi.org/10.1515/LANGCOG.2010.001>.
- 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*(6), 876–884. <https://doi.org/10.1111/j.1467-7687.2012.01185>.
- Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non independence of features in similarity judgments. *Cognitive Psychology, 23*, 222–264. [https://doi.org/10.1016/0010-0285\(91\)90010-L](https://doi.org/10.1016/0010-0285(91)90010-L).
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*, 69–110. [https://doi.org/10.1207/s15327809jls1401\\_4](https://doi.org/10.1207/s15327809jls1401_4).
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423* (6939), 534–537. <https://doi.org/10.1038/nature01647>.
- Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*(2), 212–224. <https://doi.org/10.3200/JMBR.36.2.212-224>.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 801–812. <https://doi.org/10.1037/a0023219>.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review, 15*, 495–514. <https://doi.org/10.3758/PBR.15.3.495>.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. <https://doi.org/10.1002/bimj.200810425>.
- Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology, 35*, 151–161. <https://doi.org/10.1017/S0048577298001516>.
- Kosslyn, S. M., Thompson, W. L., Wraga, M., & Alpert, N. M. (2001). Imagining rotation by endogenous versus exogenous forces: Distinct neural mechanisms. *NeuroReport, 12*, 2519–2525. <https://doi.org/10.1097/0000175620010808000046>.
- Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial ability. *Developmental Psychology, 35*(4), 940–949. <https://doi.org/10.1037/0012-1649.35.4.940>.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498. <https://doi.org/10.2307/1130467>.
- McDaniel, M. A., & Mason, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385. <https://doi.org/10.1037/0278-7393.11.2.371>.
- McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction, 19*, 171–184. <https://doi.org/10.1016/j.learninstruc.2008.03.005>.
- Mix, K. S. (2010). Spatial tools for mathematical thought. In K. S. Mix, L. B. Smith, & M. Gasser (Eds.), *The spatial foundations of cognition and language*. New York: Oxford University Press.

- Mohring, W., & Frick, A. (2013). Touching up mental rotation: Effects of manual experience on 6-month-old infants' mental object rotation. *Child Development, 84*(5), 1554–1565. <https://doi.org/10.1111/cdeb.12065>.
- Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science, 19*, 1063–1066. <https://doi.org/10.1111/j.1467-9280.2008.02200.x>.
- Morris, D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*, 519–533. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9).
- Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*(4), 903–910. <https://doi.org/10.1177/0956797613518351>.
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*(1), 18–22. <https://doi.org/10.1037/0022-0663.74.1.18>.
- Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology, 19*(2), 178–241. [https://doi.org/10.1016/0010-0285\(87\)90011-9](https://doi.org/10.1016/0010-0285(87)90011-9).
- Parsons, L. M., Fox, P. T., Down, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., & Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature, 375*, 54–58. <https://doi.org/10.1038/375054a0>.
- Piaget, J. (1953). How children form mathematical concepts. *Scientific American, 189*(5), 74–79. <https://doi.org/10.1038/scientificamerican1153-74>.
- Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19*, 1067–1070. <https://doi.org/10.1111/j.1467-9280.2008.02201.x>.
- R Core Team (2014). *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>.
- Sami, S., Robertson, E. M., & Miall, R. C. (2014). The time course of task-specific memory consolidation effects in resting state networks. *The Journal of Neuroscience, 34*(11), 3982–3992. <https://doi.org/10.1523/JNEUROSCI.4341-13.2014>.
- Schwarzer, G., Freitag, C., Buckel, R., & Lofruthe, A. (2013). Crawling is associated with mental rotation ability by 9-month-old infants. *Infancy, 18*(3), 431–442. <https://doi.org/10.1111/j.1532-7078.2012.00132.x>.
- Schwarzer, G., Freitag, C., & Schum, N. (2013). How crawling and manual object exploration are related to the mental rotation abilities of 9-month-old infants. *Frontiers in Psychology, 4*, 97. <https://doi.org/10.3389/fpsyg.2013.00097>.
- Sekiyama, K. (1983). Mental and physical movement of hands: Kinesthetic information preserved in representational systems. *Japanese Psychological Research, 25*, 95–102. <https://doi.org/10.4992/psycholres.1954.25.95>.
- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Education Psychology, 93*(3), 604–614. <https://doi.org/10.1037/0022-0663.93.3.604>.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science, 171*, 701–703. <https://doi.org/10.1126/science.171.3972.701>.
- Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education, 31*(3), 459–480. <https://doi.org/10.1080/09500690802595839>.
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110*(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>.

- Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, *22*(7), 996–1013. <https://doi.org/10.1002/acp.1420>.
- Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, *33*(5), 668–678. <https://doi.org/10.3758/BF03193305>.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. <https://doi.org/10.1037/a0028446>.
- Uttal, D. H., Scudder, K. V., & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, *18*, 37–54. [https://doi.org/10.1016/S0193-3973\(97\)90013-7](https://doi.org/10.1016/S0193-3973(97)90013-7).
- Vingerhoets, G., de Lange, F. P., Vandemaele, P., Deblaere, K., & Achten, E. (2002). Motor imagery in mental rotation: An fMRI study. *NeuroImage*, *17*, 1623–1633. <https://doi.org/10.1006/nimg.2002.1290>.
- Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. <https://doi.org/10.1111/cdep.12079>.
- Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieved dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 246–254. <https://doi.org/10.1047/a0025260>.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial ability: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270. <https://doi.org/10.1037/0033-2909.117.2.250>.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Education Psychology*, *101*, 817–835. <https://doi.org/10.1037/a0016127>.
- Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (1999). *Multiple comparisons and multiple tests using the SAS system*. Cary, NC: SAS Institute Inc.
- Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*, 77–94. [https://doi.org/10.1016/S0010-0277\(98\)00032-8](https://doi.org/10.1016/S0010-0277(98)00032-8).
- Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction*, *18*, 30–41. <https://doi.org/10.1016/j.learninstruc.2006.09.009>.
- Wiedenbauer, G., Schmid, J., & Jansen-Osmann, P. (2007). Manual training of mental rotation. *European Journal of Cognitive Psychology*, *19*, 17–36. <https://doi.org/10.1080/09541440600709906>.
- Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 397–412. <https://doi.org/10.1037/00961523.24.2.397>.