American Psychological Association

Developmental Psychology

ISSN: 0012-1649

2020, Vol. 56, No. 4, 686–698
http://dx.doi.org/10.1037/dev0000899

Piecing Together the Role of a Spatial Assembly Intervention in Preschoolers' Spatial and Mathematics Learning: Influences of Gesture, Spatial Language, and Socioeconomic Status

Corinne Bower
Temple University

Laura Zimmermann and Brian Verdine
University of Delaware

Tamara Spiewak Toub and Siffat Islam
Temple University

Lindsey Foster
University of Delaware

Natalie Evans
Temple University

Rosalie Odean
University of Delaware

Amanda Cibischino
Temple University

Calla Pritulsky
University of Delaware

Kathy Hirsh-Pasek
Temple University

Roberta Michnick Golinkoff
University of Delaware

Spatial skills are associated with mathematics skills, but it is unclear if spatial training transfers to mathematics skills for preschoolers, especially from underserved communities. The current study tested (a) whether spatial training benefited preschoolers' spatial and mathematics skills, (b) if the type of feedback provided during spatial training differentially influenced children's spatial and mathematics skills, and (c) if the spatial training's effects varied by socioeconomic status (SES). Preschoolers ($N = 187$) were randomly assigned to either a 'business-as-usual' control or one of three spatial training groups (modeling and feedback [MF]; gesture feedback [GF]; spatial language feedback [SLF]). Three-year-olds were trained to construct puzzles to match a model composed of various geometric shapes. New models were created similar to the 2-dimensional trials of the Test of Spatial Assembly (TOSA). Training was given once per week for 5 weeks. Preschoolers were pretested and posttested on 2D and 3D TOSA trials, spatial vocabulary, shape identification, and 2 mathematics assessments. Results indicate that first, any spatial training improved preschoolers' 2D TOSA performance, although a significant interaction with SES indicated improvement was driven by low-SES children. Furthermore, low-SES children showed greatest gains on the 2D TOSA with MF and GF. Second, MF and GF improved low-SES children's performance on the 3D TOSA. Third, only low-SES children with MF saw improvements in far-transfer to mathematics (Woodcock-Johnson: Applied Problems, but not the Test of Early Mathematical Ability). Results indicate that, especially for low-income learners, spatial training can improve children's early spatial and mathematics skills.

Keywords: spatial development, spatial training, mathematics learning, spatial language, gesture

Supplemental materials: http://dx.doi.org/10.1037/dev0000899.supp

---

Corinne Bower, Department of Psychology, Temple University; Laura Zimmermann and Brian Verdine, School of Education, University of Delaware; Tamara Spiewak Toub and Siffat Islam, Department of Psychology, Temple University; Lindsey Foster, School of Education, University of Delaware; Natalie Evans, Department of Psychology, Temple University; Rosalie Odean, School of Education, University of Delaware; Amanda Cibischino, Department of Psychology, Temple University; Calla Pritulsky, School of Education, University of Delaware; Kathy Hirsh-Pasek, Department of Psychology, Temple University; Roberta Michnick Golinkoff, School of Education, University of Delaware.

This research was supported by an Institute of Education Sciences Grant R305A140385 awarded to Roberta Michnick Golinkoff and Kathy Hirsh-Pasek. We want to thank Maya Marzouk and Jelani Medford for their help with the set-up and early data collection of this project. We would also like to thank all the children who participated, their parents, and our preschool partners in northern Delaware and the Philadelphia area.

Correspondence concerning this article should be addressed to Corinne Bower, Department of Psychology, Temple University, Weiss Hall, 1701 North 13th Street, Philadelphia, PA 19122. E-mail: corinne.bower@temple.edu

The growth rate of employment in STEM (science, technology, engineering, and mathematics) occupations is faster than non-STEM occupations (Fayer, Lacey, & Watson, 2017). Despite its global importance, however, American students have fallen far behind their international peers in the mastery of STEM competencies; they score well below the international average in mathematics achievement (OECD, 2016), behind countries such as China, Canada, Sweden, the Netherlands, and the United Kingdom. Although the findings are mixed, one potential route worth exploring in trying to promote STEM achievement is to train children's spatial skills, which are malleable skills known to be associated with STEM achievement. The current study takes this approach by providing preschoolers with experiences in spatial skills to see if these experiences can boost spatial and mathematics performance.

We recruit spatial skills in everyday activities when we pack a car trunk with luggage, navigate our way in an unfamiliar place, or imagine what something looks like from another perspective. Specifically, spatial skills involve our ability to mentally visualize, transform, and manipulate objects or scenes. Spatial skills are strongly associated with achievement in STEM fields (e.g., Wai, Lubinski, Benbow, & Steiger, 2010), and mathematics specifically (e.g., Burnett, Lane, & Dratt, 1979; M. B. Casey, Nuttall, & Pezaris, 2001; Delgado & Prieto, 2004; Gunderson, Ramirez, Beilock, & Levine, 2012; Hegarty & Kozhevnikov, 1999; Mix & Cheng, 2012; Mix et al., 2016; Rittle-Johnson, Zippert, & Boice, 2019; Sortor & Kulp, 2003; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017). Findings strongly suggest that placing numbers on a number line, an inherently spatial task, reflects an understanding of magnitude (e.g., de Hevia & Spelke, 2009; Lourenco & Longo, 2009; Moyer & Landauer, 1967; Newcombe, 2017; Siegler & Opfer, 2003; van Galen & Reitsma, 2008). Furthermore, Lourenco, Cheung, and Aulet (2018), Mix (2019), and Mix and Cheng (2012) provided comprehensive reviews on the relation between spatial skills (e.g., mental rotation, spatial visualization, and spatial scaling) and mathematics achievement. For example, visualizing a scaled representation across space (e.g., a number line from 0 to 10) and mentally transforming this representation (e.g., estimating the location of the number 4) may be bolstered by dynamic spatial imagery (e.g., Möhring, Frick, & Newcombe, 2018). Specifically, children might solve simple calculation problems by finding the location on a number line of the first addend and then counting up the number of spaces for the second addend to determine the final solution. Or, children might visualize pushing two sets of objects together to help with addition problems, for example. Having strong dynamic spatial imagery skills may facilitate these processes. Given the pervasiveness of these tasks that call on spatial skills, attention to spatial learning has become an educational priority over the last 20 years (e.g., National Council of Teachers of Mathematics, 2000; National Research Council, 2006).

Because spatial learning is implicated in later spatial and mathematical outcomes, it should follow that training in spatial reasoning should impact not only spatial outcomes, but also mathematical outcomes. Uttal and colleagues (2013) noted spatial training is effective in improving children, adolescents, and adults' performance on spatial tasks as well as transferring to performance on spatial tasks that were not directly trained. Moreover, Uttal and colleagues (2013) found that the benefits of spatial skill training remain even up to 4 months after the initial intervention. This malleability, durability, and transferability of spatial skill training provides a foundation for the possible transfer of spatial training to more generalized performance improvements in STEM disciplines. Although the idea of training spatial skills to transfer to STEM-related tasks is not new (e.g., Mix & Cheng, 2012), prior work is limited. One study conducted by Cheng and Mix (2014) provided one group of 6- to 8-year-olds with a single session of training on a mental transformation task (based on Levine, Huttenlocher, Taylor, & Langrock's, 1999, mental transformation task) while the other group completed crossword puzzles. Children in the spatial training group significantly improved their performance on calculation problems while children in the control group did not. The authors attribute this significant transfer effect to the potential mechanism of increased mental rotation skills or visuospatial working memory capacity as prior work provides evidence of the link between visuospatial working memory and counting task performance (Kyttälä, Aunio, Lehto, Van Luit, & Hautamaki, 2003). If the spatial training increased the children's spatial skills, then this increased capacity may also facilitate the spatial skills needed for mental calculations.

In another study, Hawes, Moss, Caswell, and Poliszczuk (2015) provided one group of 6- to 8-year-olds with computerized mental rotation training for 6 weeks, whereas the other group received literacy training. Children in the spatial training group performed better than control children on two measures of mental rotation as well as an untrained mental transformation task. Spatial training, however, did not transfer to mathematics tasks, including nonverbal exact arithmetic and missing term problems. The authors attribute this null effect to the fact that they waited 3 to 6 days to posttest the children—unlike Cheng and Mix (2014) who gave an immediate posttest. Thus, both immediate and delayed posttesting are crucial in determining the effectiveness of spatial training.

Another study extended spatial training to examine its effects on other mathematics skills, including number line estimation. Three- to 5-year-olds who received one 15–20 min session of spatial training on Levine and colleagues' (1999) mental transformation task showed improvements on the task but not on number ordering or number line estimation tasks (Xu & LeFevre, 2016). Also, Cornu, Schiltz, Pazouki, and Martin (2019) conducted a twice per week 20-min tablet-based visuospatial training with kindergarteners spread over the course of 10 weeks. However, the training only facilitated spatial skills and did not transfer to mathematics skills (e.g., magnitude and number comparison). There are several possible rationales for the conflicting results among these past four spatial training studies. First, the interventions vary in length and are relatively short. Second, the literature approaches spatial training as a singular construct. In other words, training children's spatial skills may require a more comprehensive approach that incorporates a variety of spatial skills (e.g., mental rotation, spatial visualization, scaling) and strategies to teach these skills—especially for younger children. Third, the transfer between the spatial training and particular types of mathematics assessments may be too far. Thus, the type of mathematics skills assessed needs to be comprehensive to understand the transferability of spatial training to mathematics skills.

Unfortunately, there is wide variability among American students, with many low-socioeconomic status (SES) preschoolers lagging behind their middle-SES peers in both spatial (e.g., Verdine, Golinkoff, Hirsh-Pasek, Newcombe, et al., 2014) and mathematical school-readiness scores (Child Trends Databank, 2015). Thus, it is important to consider who might benefit more or less from the various kinds of spatial training. Converging findings suggest that low-SES children might profit more from spatial training than middle- to high-SES children, because low-SES 4- to 7-year-olds' spatial skills are already lagging behind those of their high-SES peers (Jirout & Newcombe, 2015; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Verdine, Golinkoff, Hirsh-Pasek, Newcombe, et al., 2014) and more of them likely have more room for growth. One possible reason for these SES differences in responses to spatial training could be caused by environmental differences. Lower-SES 3-year-olds hear significantly less language than their high-SES peers (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Hart & Risley, 2003). Thus, it could be that lower-SES children also hear less spatial language. Another possibility is that low-SES children may have fewer toys that are beneficial for their spatial development (e.g., puzzles, blocks, and board games; Jirout & Newcombe, 2015; Levine et al., 2012). However, one study found no significant difference in the amount of spatial play between SES groups (Jirout & Newcombe, 2015). Even if frequency of spatial play does not directly contribute to SES differences in spatial performance, possible differences in the *quality* of spatial play (i.e., if a child received corrective feedback or not) could be a major contributing factor. Thus, high-quality spatial training may be especially beneficial for low-SES children. Indeed, Uttal and colleagues' (2013) meta-analysis examining the malleability of spatial skills found that SES significantly negatively correlated with the effect size of spatial training. Specifically, lower SES was associated with larger responses to spatial skill training. In the current study, the spatial training provides frequent, high-quality spatial play along with manipulations of condition-specific feedback (gesture, spatial language) offered by a nurturing adult. Offering spatial play to low-SES children who may lack such experiences may aid them relative to their middle- to high-SES peers who likely already have such experiences.

In short, there is research to suggest a strong and reliable link between early spatial skills and later spatial and mathematical outcomes. Yet, training studies have not been comprehensive in their study of this link nor have the results that have emerged been consistent. This study tests a spatial training intervention that fills these research gaps for even younger children—3-year-olds. We used a modified version of the original two-dimensional Test of Spatial Assembly (2D TOSA; Verdine, Golinkoff, Hirsh-Pasek, et al., 2017; Verdine et al., 2014) as both a spatial assessment and training tool for the current study with different trials for each. The original TOSA has two- and three-dimensional trials and asks children to build a set of target constructions from models using a set of pieces that are identical to the model components (see Figure 1B and 1C). In prior research, the performance of 3-year-olds predicted performance on other spatial measures at age 4 and at age 5, as well as mathematics performance at age 5 (Verdine et al., 2017). In the current study, new trials modeled after the 2D TOSA were administered as a pre- and posttest assessment to see if spatial assembly skills improved due to spatial assembly training. Because the transferability of spatial training to other spatial tasks is supported in prior work (Uttal et al., 2013), the current study explored the transfer of the 2D spatial assembly training to a three-dimensional spatial assembly task (3D Test of Spatial Assembly; 3D TOSA) using LEGO-like blocks.

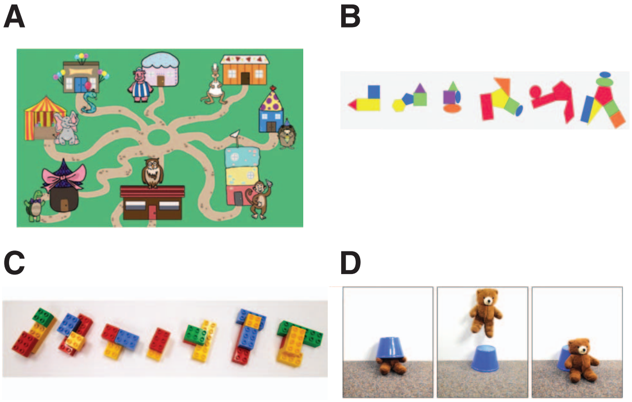

Figure 1. A: Map scene for telling Raffo the Giraffe's Birthday Story and allowing children to select animal friends to visit B. Examples of test trials in the modified 2D Test of Spatial Assembly (2D TOSA; adapted from Verdine et al., 2017). C: Practice and testing trials in the 3D Test of Spatial Assembly (3D TOSA; adapted from Verdine et al., 2017). D: Sample item from the spatial vocabulary assessment ("Point to, 'the bear is under the bucket'"). See the online article for the color version of this figure.

One of the main goals of the current study was to examine gains on spatial and mathematics skills from pre- to posttest based on any spatial assembly training (vs. a business-as-usual control group). Another main goal of the study was to isolate and explore the benefits of different training types for supporting children's spatial learning. The most basic strategy was modeling and feedback (MF) in which the trainer responded to children's answers by saying which shapes were incorrectly assembled and also by helping to place the incorrect pieces in the correct locations if the child continued to be incorrect on a second attempt. We also explored whether layering additional supports on top of this MF would be more effective. Specifically, gesture or spatial language were selected as additional strategies the trainer used to deliver the corrective feedback. Thus, the current study's spatial training implemented three types of feedback when correcting children's errors: (a) simple corrective feedback (MF); (b) with the experimenter performing informative gestures (GF); and (c) with the experimenter providing corrective spatial language (SLF). If one of these training conditions were to have a larger effect over the others in facilitating spatial and/or mathematics skills, then we could deduce important theoretical implications for both formal and informal educational contexts in early childhood experiences.

Gesture was selected as a feedback strategy in the spatial training because research suggests that gesture can support spatial reasoning. Children who were instructed to gesture while trying to solve a spatial task improved significantly more on the task at posttest compared to those who did not gesture (e.g., Levine, Goldin-Meadow, Carlson, & Hemani-Lopez, 2018; Ping, Ratliff, Hickey, & Levine, 2011). Importantly, having children produce strategy-relevant gestures while solving a spatial task helped their performance more than just watching someone else produce the same gestures (e.g., Goldin-Meadow et al., 2012). Moreover, gestures can also assist with mathematical reasoning (e.g., Goldin-Meadow, Cook, & Mitchell, 2009).

It is likely that gestures support children's learning by alleviating an individual's cognitive load or freeing-up working memory and allowing the allocation of more resources to solving the task at hand (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wesp, Hesse, Keutmann, & Wheaton, 2001). In the current study, if the child incorrectly assembled the shapes on their second attempt, the experimenter gestured where the incorrect pieces should go to help the child visualize the correct spatial orientations between the puzzle pieces. This nonverbal spatial strategy may be particularly helpful if children have limited vocabularies for the spatial relations they are trying to reproduce or if they are struggling to visualize a different arrangement of the pieces; gesturing may "ground" the abstract nature of spatial problem solving with concrete, bodily actions (e.g., Beilock & Goldin-Meadow, 2010).

Spatial language was selected as another important feedback strategy because the use of spatial language, such as talking about spatial properties and the location of objects like *big, next to,* and *behind,* may help children attend to and encode spatial information. Pruden, Levine, and Huttenlocher (2011) found that children's production of spatial language during daily tasks between the ages of 14 and 46 months predicted their performance on spatial tasks at age 54 months. In addition, 3-year-olds solved a spatial analogy task successfully if they heard language describing spatial relations (e.g., *in, on,* or *under*; Loewenstein & Gentner, 2005). Hearing spatial relations expressed in language may help children encode and recall important spatial information. The current study not only focuses on providing children with this type of spatial language feedback (SLF), but also with shape names. By 3 years of age, children know many shape names even though it takes years for them to appreciate shapes' critical features (e.g., Satlow & Newcombe, 1998; Verdine, Lucca, Golinkoff, Hirsh-Pasek, & Newcombe, 2016). Thus, shape names and spatial relational language may have an additive effect on spatial assembly skills by increasing awareness of the individual elements of the designs (shape names) and of the relationship between those elements (e.g., *above*).

In addition to facilitating spatial problem-solving, spatial language may also facilitate mathematical outcomes. Work by Purpura and Reid (2016) suggests that 3- to 5-year-olds' spatial-mathematical language (e.g., *a lot*, *more*, *nearest*) is associated with their numeracy performance. Moreover, children who were in a dialogic reading intervention that included quantitative and spatial mathematical language performed better on mathematical outcomes compared to children not in the intervention (Purpura, Napoli, Wehrspann, & Gold, 2017). Even though these prior correlational and experimental studies did not isolate spatial language per se, spatial-mathematical language was effective in boosting mathematical knowledge. The current study extends this prior work by isolating the transfer effects of spatial-only language on mathematical outcomes.

Low- and high-SES preschoolers were randomly assigned to either a MF spatial training, gesture training, spatial language training, or a "business-as-usual" control group. All children were pre- and posttested on a battery of spatial and mathematics assessments. We hypothesized that (a) children in any of the spatial training groups would profit by demonstrating stronger gains in spatial learning and mathematics than the control group; (b) children who received richer feedback (spatial language and gesture conditions) would excel to a greater degree than those whose feedback was limited (MF condition); and (c) low-SES children would profit more than their high-SES peers.

Method

Participants

A total of 187 children participated (M_{Age} = 42.65 months; *SD* = 3.37; range = 36.19 to 47.93; 96 females; 91 males). Based on parent report, the sample was 51% Caucasian, 27% Black, 17% other, and 5% unreported. Of all the children, 14% were Hispanic or Latino. All children were native English speakers. Children were recruited from Head Start facilities as well as private preschools in two U.S. northeastern states. Inclusion criteria required the children to be 3 years of age, proficient in English, and not have any apparent developmental delays.

The project title is "Spatial Instruction in Preschool: Identifying the Malleable Factors" and was approved by the University of Delaware and Temple University's institutional review boards (IRB protocol numbers: 632397–14; 22370, respectively). Parents returned the signed consent form with a background questionnaire that requested the primary caregiver's education level, which was argued by Hoff (2013) to be the most critical SES component for development. Thus, caregivers with a bachelor's degree or higher were categorized as high SES (49%; n = 91) and those with an associate's degree or less were categorized as low SES (47%; n = 88). Because eight caregivers declined to answer this question, the type of school the children attended was used to categorize their SES: Children from Head Start centers were categorized in the low-SES group (n = 5) and children from private preschools in the high SES group (N = 3). Overall, there were 94 children categorized as low SES (50.3%) and 93 children categorized as high SES (49.7%).

An a priori power analysis using a medium effect size (f^2 = .15) and setting power to .80 indicated that 143 cases would be necessary. Assuming a 20% attrition rate over the 7-week period, the minimum sample size required was 172 children (or 176 children to be evenly distributed across the two sites). There were 46 children in the control group (24 girls; 23 low SES); 46 children in the MF training condition (22 girls; 23 low SES); 48 children in the gesture training condition (25 girls; 25 low SES); and 47 children in the spatial language training condition (25 girls; 23 low SES). Boys and girls were evenly distributed within SES within each condition. Children received stickers as thanks for their participation.

Procedure

Children were pretested (Week 1), trained (if in an experimental condition; Weeks 2–6), and posttested (Week 7) individually in a private room outside of the preschool classroom. See Supplemental Table S1 in the online supplemental material for a table of the procedure and measures used by skill area.

Pretest. All participants were pretested on spatial (modified 2D TOSA, 3D TOSA, shape identification test, and spatial vocabulary assessment), mathematics (Woodcock-Johnson–Applied Problems [WJ-AP] and a subtest of the Test of Early Mathematical Ability, 3rd ed. [TEMA-3]), and general vocabulary (Woodcock-Johnson–Picture Vocabulary; WJ-PV) assessments. The order of tasks was randomized before the pretest. The average length of time needed to administer all of the assessments during the pretest was 40 min.

Spatial assessments.

2D TOSA. The 2D TOSA was modified (e.g., Verdine et al., 2017) by including a different set of six trials, including more difficult trials that included five- and seven-piece puzzles. These modifications were made so we could administer the modified 2D TOSA as a pre- and posttest assessment as well as for use during the training without having children experience the same trial more than once. To keep children engaged during the training and testing, this modified 2D TOSA was transformed into a fun game guided by a narrative in which the child was helping to plan Raffo the Giraffe's birthday party and visit Raffo's friends, portrayed on a map (see Figure 1A), to collect items for the party. Each of the six friends had a spatial assembly trial for the participant to complete.

The modified 2D TOSA required participants to recreate a picture of a design using foam cutouts of geometric shapes (see Figure 1B). There were two practice trials and six test trials. For each trial, a picture of the stimulus design (3 cm × 3 cm) was placed in the upper left-hand corner of a whiteboard (22 × 28 cm) and the same exact pieces needed to complete the design were randomly placed on the right-hand side of the whiteboard—the "well." For the first practice trial, the experimenter pointed to the two shape pieces on the side of the board and explained, "I'm going to try to make my pieces look just like this [pointing to the picture of the stimulus design]." The experimenter then placed the two pieces together to match the stimulus design and said, "Now my pieces look just like the picture [while pointing to the picture of the stimulus design]." The experimenter then reset the pieces and placed them in the well and told the participant, "Now it's your turn. Can you make these pieces look just like the picture?". After the participant completed the design, if the design was correct (i.e., both pieces were in the correct location), the experimenter would say that the design was correct and then move on to the next trial puzzle. If the design was incorrect (i.e., at least one piece was in the incorrect location), the experimenter would say "That's not quite right. Let's try it again" and then place both of the pieces back into the well. If the participant got the design correct on the second attempt, the experimenter would say that the design was correct and then move onto the next trial puzzle. If the design was incorrect, then the experimenter would explain "That's not quite right. This piece[s] goes here [while placing the incorrect piece(s) into the correct configuration]." This process was repeated for the second practice trial, which used another stimulus design with two different shape pieces.

Once the two practice trials were completed, the participant moved on to the six test trials—all using different puzzle designs of increasing levels of difficulty. For the test trials, the participant did not receive any corrective feedback. After participants completed their design, they moved on to the next test trial, regardless of accuracy.

Each of the test trials was coded for accuracy. Each trial included one larger "base" piece and between 1 and 6 smaller, other "component" pieces. Scoring was determined based on three dimensions for each component piece. For the first dimension, *adjacent pieces*, a single point was awarded if children placed a component piece beside its correct neighboring piece (i.e., within 1 cm). For component pieces with multiple adjacent pieces, each piece was scored. Thus, the total adjacent piece score for each puzzle was the composite of scores of each component piece to its neighboring piece(s). The second dimension, *horizontal and vertical direction*, assessed whether the child correctly placed the component pieces either above or below or to the left or to the right of the base piece. To assess this, a set of perpendicular x- and y-axes was drawn on a transparency, placed over the center of the base piece, and aligned with the sides of the whiteboard. If at least 50% of a component piece's volume was within the same quadrant as its correct location in the target model, then it received a score of 1. Thus, the total horizontal and vertical direction score for each puzzle was the composite of scores of each component piece. For the third dimension, *relative position*, a transparency of the correct configuration was placed over the child's configuration by matching up the base pieces to maximize the child's component shape pieces that could match the correct configuration's component shape pieces. Each component piece of the child's configuration that was within 1 cm of the same component piece in the correct configuration was awarded 1 point. Thus, the relative position score for each puzzle was the composite of scores of each component piece. All points from these three coding dimensions across all completed test trials were summed for each child (total possible for Set A = 73; total possible for Set B = 72) and a proportion score of this composite was used in analyses. For more details about the coding system and trial procedure, see Verdine, Golinkoff, Hirsh-Pasek, et al., 2017. Combined with the 3D TOSA items, prior studies with 3-year-olds indicate Cronbach's alpha to be .747. To ensure interrater reliability, 20% of participants were scored by two coders with an intraclass correlation (ICC) of .990.

3D TOSA. The 3D TOSA (Verdine et al., 2017; see Figure 1C) required participants to recreate a model made of colored plastic LEGO DUPLO blocks using a matching set of blocks. There were two practice trials and seven test trials. For each trial, the model was placed in front of the participant as well as a duplicate set of detached blocks needed to complete the design. The practice and test trials were administered the same way as the 2D TOSA trials. Overall, the block designs all included one larger base piece and at least one other component block. The six test trials increased in difficulty (i.e., the number of blocks increased in the later test trials). The accuracy of each of the six test trials was a composite sum of each dyad of blocks' score on three dimensions. In other words, if one test trial model included a total of three blocks, then there would be three dyads of blocks to assess (i.e., base block and component block #1; base block and component block #2; both component blocks). Each dyad of blocks was assessed on three dimensions, with a few exceptions (for more details, see Verdine, Golinkoff, Hirsh-Pasek, et al., 2017). For the first dimension of *vertical location*, one point was awarded to each dyad that had the two blocks on the correct level. For the second dimension of *rotation*, one point was awarded to each dyad if a block's axis was oriented correctly to the other block's axis. For the third dimension of *translation*, one point was awarded to each dyad if a block was placed over the correct studs of the other block. A composite score was created for each of the three dimensions (vertical location, range from 0 to 25; rotation, range from 0 to 11; and translation, range from 0 to 25) as well as an overall composite dimension score (sum of all dimension scores; range was from 0 to 61). For more details about the coding system and trial procedure, see Verdine et al. (2017). To ensure interrater reliability, 20% of participants were scored by two coders with an ICC of .971.

Spatial vocabulary assessment. This spatial vocabulary assessment (Bower et al., 2020) was given to assess children's knowledge of spatial relations, such as above, behind, and middle (Park & Casasola's, 2017). The experimenter asked the child to point to the one photograph out of three options that matched a spatial configuration (e.g., "the bear is under the bucket"). There were 16 test items. Each test item consisted of three photographs that displayed a toy bear in various locations and orientations either in relation to another object (e.g., "inside the bucket") or by itself (e.g., "the bear is upside down"). All three photographs of one test item included the same objects in all 3 pictures, just in varying locations. For example, for the "bear is under the bucket" test item, all three photographs showed the bear and bucket in different locations relative to one another (see Figure 1D). After every four test items, a nontest item (e.g., "point to the fish") would appear to ensure children were attending to and understanding the task directions. The total test items correct (out of 16) was the number used in the analyses. Cronbach's alpha was .711.

Mathematics.

Shape identification test. A shape identification test was given to assess each child's knowledge of shape names. The 12 shapes were a square, hexagon, two rectangles of varying size, three types of triangles (isosceles, equilateral, and right), a parallelogram, an oval, a pentagon, a kite, and a circle. The shape outlines were printed on paper (one shape per sheet) and the child was asked to identify the name of each shape. For the three types of triangle, the child only needed to respond with "triangle" to get a correct response. The score used in analyses was the total number correct (out of 12).

WJ-AP. The WJ-AP (Schrank, McGrew, & Mather, 2014) assessed children's counting skills (e.g., "how many apples are in this picture?") and addition/subtraction skills (e.g., "three birds were sitting on the park bench. One flew away. How many birds were left?"). This task was administered according to the testing manual. Testing began with Item 1 and ceased when five consecutive test items were answered incorrectly. This mathematics assessment has an internal reliability of .91 for 5- to 19-year-olds (Schrank et al., 2014), and .82 with 3-year-olds in the current study.

TEMA-3. The TEMA-3 (Ginsburg & Baroody, 2003), designed for children of ages 3 years to 8 years, tests a range of early mathematics skills. The current study used a subset of TEMA-3 items, including four trials of *nonsymbolic number knowledge* that asked the child to point to the side (out of two) that has more dots. Second, three trials of *nonverbal production* items were administered. The experimenter showed a certain number of chips and put them under a mat, requiring the child to reproduce the same number of hidden chips. Third, five trials of *nonverbal addition and subtraction* items (e.g., Levine, Jordan, & Huttenlocher, 1992) involving hidden objects that demonstrated mental computation skills were administered. The experimenter showed a certain number of chips and put them under a mat. Then, the experimenter either showed additional chips and slid them under the mat, or took some chips away and asked the child to declare the number of chips currently under the mat (requiring addition and subtraction of the hidden chips). Fourth, three trials of *number constancy* items were administered. The experimenter placed a certain number of chips in a line and counted them. Then, the experimenter moved the same chips into another configuration and asked the child how many chips there were. Fifth, two trials of a 2-item *Give-N task* (e.g., Wynn, 1990) that assessed counting principles were administered. The experimenter asked the child to give a specific number of chips (three and five, respectively) from a larger set (10) of chips. Sixth, a *number order* task with three trials was administered. These trials asked children to indicate sequential knowledge of numerals (e.g., "What number comes next? 3 and then comes . . .?"). Finally, 10 trials assessing *number knowledge* were administered that asked the child which number is more (e.g., 4 or 5).

A child received a score of 1 for each correct response: A total of 29 points were possible. The number used in analyses was the total number correct. The established reliability for this measure is relatively high with alphas ranging between .92 and .6 for children between the ages of 3 to 8 years (Bliss, 2006). The alpha for the pretest data was .85.

Vocabulary.

WJ-PV. The WJ-PV (Schrank et al., 2014) is a flipbook-style standardized test of vocabulary knowledge in which children name target objects illustrated in pictures. Standardized testing procedures were followed as indicated in the manual. The WJ-PV was included to control for the extent to which language ability and general intelligence contributed to the relations between the spatial training and spatial and mathematical abilities.

Spatial skill training. Children were randomly assigned to one of three training conditions and a business-as-usual control group. Children in the three training conditions received five spatial training sessions (10 min each) using a modified 2D TOSA over the course of an average of 5 weeks. In sum, children in the training groups received a total of 50 min of spatial training over 5 weeks. There were eight adult trainers throughout the course of the study, but children in the training conditions had the same trainer for each of their five training sessions. Children in the business-as-usual control group were pre- and posttested just like the other three training conditions, but instead of participating in the training, the children stayed in their classrooms and participated in their usual classroom activities with their classmates.

Each training session consisted of two parts: (a) a shape parade and (b) spatial assembly training. The shape parade was included as part of the training to introduce and practice elements of condition-specific strategies that would be used during the spatial assembly training. During the shape parade, the trainer sequentially displayed the nine shapes for three seconds each in the following order to the child: circle, oval, triangle, square, rectangle, kite, parallelogram, pentagon, hexagon. After each shape was displayed, the trainer placed it back in line with the rest of the shapes.

The second part of the training—spatial assembly training—consisted of a total of seven different designs including one practice trial, per session. The child had a maximum of two attempts to correctly place the puzzle pieces together to match the model design, for each of the six different designs. If the puzzle pieces were placed correctly on the child's first attempt, the trainer would indicate the puzzle was correct and move onto the next puzzle. If the puzzle pieces were placed incorrectly on the child's first attempt, then the trainer would indicate the pieces that were incorrect by removing the incorrect pieces and placing them to the side of the board (without any additional feedback) so the child could attempt the puzzle a second time. This was the same across all three training conditions. Each of the three training conditions presented the shape parade differently and provided different feedback on the child's incorrect second attempt as discussed next.

Training conditions. The MF condition used a bare-bones corrective procedure. For the shape parade, the trainer showed the child the nine shapes sequentially while saying "Let's look at this piece" for each piece. For the puzzle training, if the puzzle was incorrect a second time, the trainer would explain "Mr./Mrs. [*animal name*] says this piece goes here" and then assemble the piece correctly in front of the child until all pieces were correctly placed. If the puzzle was correct on the second attempt, the trainer would say so and move onto the next puzzle.

The gesture feedback (GF) condition allowed for the use of gesture to function as a way to plan where to place pieces. For the shape parade, the trainer explained, "I'm going to trace [the pieces] with my finger, and then you can trace the piece with your finger! Let's trace this piece. I'll try first, and then you can try!" and then continued to trace with a fingertip along the edge of each shape and then prompted the child to do the same. (No labels were offered.) This was repeated until the last shape was traced. For the gesture group's puzzle training, the process was the same as MF training, but with the addition of the use of gesture. If the puzzle was incorrect on the second attempt, the trainer verbally indicated which pieces were incorrect (e.g., "Mr./Mrs. [*animal name*] says this piece goes here") and would then use hand gestures to show the child where the incorrect pieces were meant to go and then put the pieces in the correct places. For example, a horizontal turning hand motion was used to indicate that a piece needed to be rotated. The trainer also traced the shape of a piece in the location where it was to be placed.

The SLF condition used spatial words that described spatial relations. For the shape parade, the shape was named and the child was given a description of its properties (e.g., "This is a hexagon. It has six straight sides and six corners"). For the puzzle training, the process was the same as MF training, but with the addition of the use of spatial language. If the puzzle was incorrect on the second attempt, the trainer would tell the child where the incorrect pieces should go using specific spatial language for shape and location (e.g., "Mr./Mrs. [*animal name*] says the *circle* goes *on top of the triangle*"). The trainer would then move the piece to its correct location.

The training protocol was designed so that gestures were never used in the MF or SLF training and spatial terms were never used in the MF or GF training.

Posttest. All participants were tested on the same set of assessments at both pre- and posttest, with an average of 7 weeks in between. The average length of time in the study was 43 days ($SD = 11$ days; range: 19–82 days) as the testing team and trainers needed to work around preschool holidays, teacher schedules, and university schedules. For each child, the same randomized order of tasks was used for pretest and posttest.

Results

Overview

For a summary of significant training effects on spatial and mathematics outcomes see Table 1. For descriptive statistics for the pre- and posttests see Table S2 in online supplemental material. The results are organized into three sections based on our hypotheses: (a) overall training effects, (b) specific feedback effects, and (c) the role of SES as a moderator of the training on spatial and mathematics outcomes.

Linear regressions were conducted to examine the spatial training condition effects on each spatial and mathematics outcome. The outcome's pretest score, child characteristics (gender, age in months, SES), and prevocabulary score were covariates. For those linear regressions examining the additional interaction between SES and condition, interaction terms were added as predictor variables.

Hypothesis 1: Any Training (i.e., Collapsing Across the Three Training Conditions) Will Increase Performance More Than the Control Group

Children who received any training increased their modified 2D TOSA skills (ß = .15, $p = .005$) more than the business-as-usual control group (see Table 2). Although this main effect is moderated by SES (discussed later in the Hypothesis 3 section), there

Table 1

Summary of Significant Training Effects on Spatial and Mathematics Performance at Posttest Controlling for Child Sex, Age in Months, WJ: Picture Vocabulary Pretest Score, and Assessment of Interest Pretest Score

	Spatial assessments			Mathematical assessments		
Condition vs. control	2D TOSA	3D TOSA	Spatial vocabulary	Shape ID	WJ: Applied Problems	TEMA
Any training	✓					
By SES	✓-low	✓-low			✓-low	
Training type	✓-MF,SLF			✓-SLF		
By SES	✓-MF,GF-low	✓-MF,GF-low			✓-MF-low	

Note. TOSA = Test of Spatial Assembly; WJ = Woodcock-Johnson; TEMA = Test of Early Mathematical Ability; SES = socioeconomic status; MF = modeling and feedback; SLF = spatial language feedback; GF = gesture feedback. Children were trained on two-dimensional puzzles, but both the 2D and 3D TOSA were given as pre- and posttest assessments. Spatial Vocabulary assesses children's comprehension of spatial terms such as *above* and *next to*. Shape ID assesses children's learned names of geometric shapes. WJ: Applied Problems tested for addition and subtraction skills using word problems. TEMA tested for nonverbal addition/subtraction, number knowledge, number constancy, and number order. An "✓" indicates a significant training effect. For "training type," only the significant training condition is indicated: MF, GF, and SLF. For the "By SES" rows, "-low" indicates a training effect occurred only in the low-SES group. The dimension score for the reported 3D TOSA results is the vertical location composite score as the total dimension score did not yield significant results.

Table 2

Significant Condition Main Effects and Interaction Effects With SES on Immediate Post-Test Spatial and Math Outcomes

		2D TOSA				3D TOSA				Shape ID				WJ: Applied Problems			
Model	Condition variable	F	β	p	R^2	F	β	p	R^2	F	β	p	R^2	F	β	p	R^2
1a: Collapsed training condition	Any training	28.91	.15	.005	.49	—	—	—	—	—	—	—	—	—	—	—	—
1b: With SES interactions	Any Training × SES	25.00	.24	.054	.50	18.46	.34	.012	.41	—	—	—	—	47.93	.24	.020	.66
2a: Individual condition	MF	21.72	.18	.009	.50	—	—	—	—	121.13	—	—	.69	—	—	—	—
	GF		—	—							—	—					
	SLF		.17	.010							.14	.006					
2b: With SES interactions	MF × SES	16.25	.20	.052	.51	12.04	.22	.052	.41	—	—	—	—	30.41	.24	.008	.66
	GF × SES		.21	.043			.32	.005							—	—	
	SLF × SES		—	—			—	—							—	—	

Note. TOSA = Test of Spatial Assembly; WJ = Woodcock-Johnson; TEMA = Test of Early Mathematical Ability; SES = socioeconomic status; MF = modeling and feedback; SLF = spatial language feedback; GF = gesture feedback. All regressions control for age in months, child sex, pre-WJ: Picture Vocabulary score, and pre-test score of outcome variable. All reported models have an omnibus test that is $p < .01$ and β is the standardized coefficient. The referent group of all regression models is the control group (not listed in the table). For significant Condition × SES interactions, see the Results section for analyses split by SES. The reported 3D TOSA statistics are with the vertical location dimension score as the dependent variable (total dimension score statistics are not reported here because they were not significant).

was no main effect of overall training on other spatial or mathematics outcomes.

Hypothesis 2: GF and SLF Training Conditions Will Increase Performance More Than the MF and Control Conditions

Children in the MF ($\beta = .18$, $p = .009$) and SLF (ß = .17, $p = .010$) conditions increased their modified 2D TOSA scores more than the control group (see Model 2a in Table 2; Figure 2A). Although this main effect is again moderated by SES, no other condition group comparisons were significant. In addition, children in the SLF—the only condition that labeled shape names—significantly improved their performance on shape identification more than children in the control group, $\beta = .14$, $p = .006$; MF, $\beta = .14$, $p = .007$; and GF, $\beta = .17$, $p = .001$ (see Model 2a in Table 2; Figure 2B). There were no effects of specific types of feedback on children's 3D TOSA, spatial vocabulary, WJ-AP, and TEMA-3 performance.

Hypothesis 3: Low-SES Children in the Training Conditions Will Increase Performance More Than Their High-SES Peers

There was a marginally significant interaction with SES ($p = .054$) such that low-SES children who received any training increased their modified 2D TOSA scores (ß = .28, $p < .001$) more than those in the control group (see Model 1b in Table 2). When exploring the individual training conditions, the MF ($p = .052$) and GF ($p = .043$) conditions yielded significant interactions with SES, such that low-SES children in the MF (ß = .31, $p = .001$) and GF conditions (ß = .25, $p = .007$) increased their 2D TOSA scores with spatial training compared to the control group (see Model 2b in Table 2; Figure 3A). There were no significant condition effects for high-SES children, including analyses looking at performance on only the difficult puzzles (i.e., puzzles with five or more pieces).

When examining SES and training interactions on spatial transfer assessments, there were no significant interactions on children's total 3D TOSA performance. However, when we performed an exploratory analysis to examine differences on the subscores (e.g., vertical placement, rotation, and translation), there was a significant interaction on the vertical location dimension composite score only (i.e., correct level placement of blocks; $p = .012$). Low-SES children who received any training significantly improved their vertical location scores ($\beta = .19$, $p = .027$) compared to the control group (see Model 1b in Table 2). Further examination indicated that only low-SES children who were in the MF (ß = .22, $p = .038$) and GF (ß = .22, $p = .036$) conditions significantly improved their vertical location scores compared to the control group (see Model 2b in Table 2; Figure 3B).

When examining SES and training interactions on mathematics transfer assessments, there was a significant interaction on children's WJ-AP performance (ß = .24, $p = .020$; see Model 1b in Table 2); however, neither SES group was significant by itself (low SES, $p = .078$; high SES, $p = .362$). When exploring the individual training conditions, only the MF group significantly interacted with SES, $\beta = 0.24$, $p = .008$, such that low-SES children in the MF condition improved significantly more on WJ-AP than children in the control condition, ß = .18, $p = .035$ (see Model 2b in Table 2; Figure 3C). There were no significant SES interactions with condition on children's, spatial vocabulary, shape identification, and TEMA-3 performance.

Discussion

The current study was motivated by the fact that preschoolers' mathematics readiness in the United States, especially those from low-SES environments, is lagging (Child Trends Databank, 2015; Ginsburg, Lee, & Boyd, 2008). One factor associated with mathematical learning is spatial skills. Numerous studies support the significant association between spatial and mathematics skills (see Mix & Cheng, 2012, for a review), but prior studies provide inconsistent evidence about the causal role of spatial skills in mathematical learning (Cheng & Mix, 2014; Cornu, Schiltz, Pazouki, & Martin, 2019; Hawes et al., 2015; Xu & LeFevre, 2016).

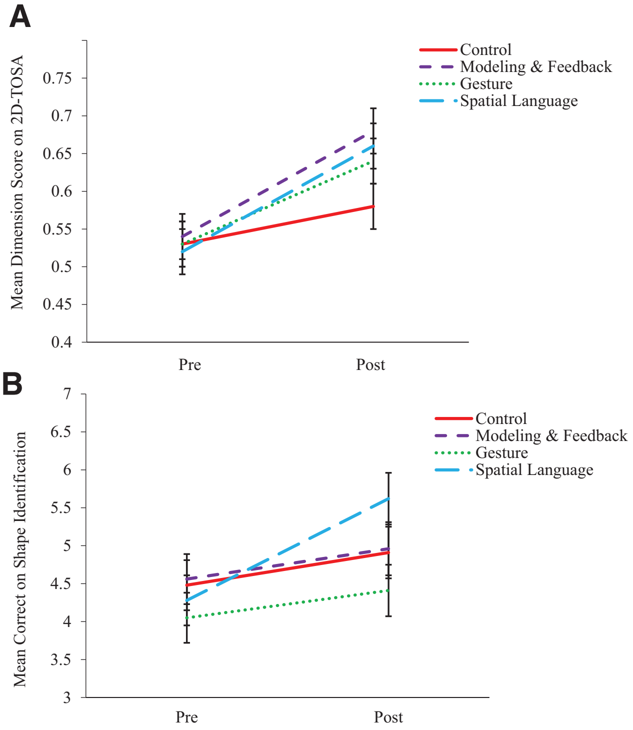

Figure 2. A: The modeling and feedback ($p = .009$) and spatial language ($p = .010$) conditions increased their 2D Test of Spatial Assembly (2D TOSA) performance compared to the control group. B: The spatial language condition increased their Shape Identification scores more than the control group ($p = .006$), modeling and feedback condition ($p = .007$), and gesture condition ($p = .001$). Both graphs display estimated marginal means when controlling for age in months, sex, socioeconomic status, and prevocabulary score. The errors bars represent standard errors. See the online article for the color version of this figure.

The current study is the first to use a playful, spatial assembly training task over multiple sessions with the aim of increasing young children's spatial skills and mathematics achievement. Moreover, we examined the effects of different spatial feedback strategies. Specifically, preschoolers in a spatial assembly intervention across five training sessions moved individual shapes to construct a larger geometric puzzle from a model. The feedback provided to the children based on their performance during this spatial assembly task was manipulated such that some children received just simple corrective feedback (MF), some received corrective feedback accompanied by spatial language (SLF), and some received corrective feedback accompanied by gesture (GF). Because the current study's findings did not perfectly align with our original predictions, the following paragraphs discuss, first, spatial training effects on children's spatial skills as moderated by SES; and second, any transfer of training effects to mathematic skills as moderated by SES.

Spatial Skills: Training Effects

Given that puzzle play and block building are engaging activities that facilitate spatial skills (e.g., Casey et al., 2008; Jirout & Newcombe, 2015; Verdine et al., 2014), we hypothesized that children's spatial skills—regardless of SES—would profit from spatial training, with all types of feedback. The current findings suggest that children who were given five training sessions on a 2D spatial assembly task over the course of five weeks increased their performance on similar trials from pre- to posttest compared to children who did not receive training. However, when examining the interaction that emerged between spatial training and SES, this significant main effect was moderated by SES: only low-SES children benefited. The training effects were not evident among high-SES children, even on the performance of more complex puzzles with five or more pieces. Perhaps this can be attributed to the fact that low-SES children's pretest 2D TOSA performance was significantly lower than high-SES children's performance; low-SES children had more room for improvement. This is congruent with past work that found similar moderation of spatial training outcomes by SES. For example, Uttal and colleagues' (2013) meta-analysis examining the malleability of spatial skills found that SES significantly correlated with the effect size that resulted from spatial training such that lower SES was associated with larger responses to spatial skill training than higher SES.

We also hypothesized that children who received feedback via training would benefit from the inclusion of spatial language or gesture and would excel to a greater degree than those who received MF without these additional supports. For 2D spatial assembly skills, children in the MF and SLF training groups increased their performance more than children who did not receive any training. The GF condition also showed improvement but did not reach traditional levels of significance ($p < .10$). Thus, basic MF and SLF was beneficial for children's spatial skills. The trainer in the MF condition indicated which shapes were placed incorrectly (e.g., "This piece is not quite right, try it again" [*trainer puts the shape placed incorrectly back into the side well*]). Bringing attention to the incorrectly placed pieces allowed the child to learn from their mistake and attempt to place the shape again in its correct spatial orientation and location. For the SLF training, the trainer used this same basic feedback procedure, but added spatial language to help scaffold attention to the correct spatial relations between the individual shapes. For example, the trainer might say "Actually, the blue *square* should go *on top* of the yellow *rectangle*."

Hearing spatial language while solving a spatial problem improves performance (e.g., Loewenstein & Gentner, 2005), although there are two possible explanations for why this occurs. One possibility is that spatial words themselves (such as *on* or *above*) prompt children to encode spatial distinctions needed for task solution (e.g., Feist & Gentner, 2007; Miller, Patterson, & Simmering, 2016). An alternative possibility is that children's selective attention to spatial information is prompted by hearing spatial words (Casasola, 2005; Gentner & Goldin-Meadow, 2003; Miller & Simmering, 2018). In addition, previous work conducted by Pruden and colleagues (2011) finds that 14- to 46-month-old children who were exposed to more parental spatial language had better performance on a spatial task at 54 months. This longitudinal finding, along with the results of the current study, supports the directionality and causal influence spatial language apparently has on children's spatial skills.

These training effects are further moderated by SES with only low-SES children showing change. Low-SES children who received MF and GF training increased their 2D TOSA performance

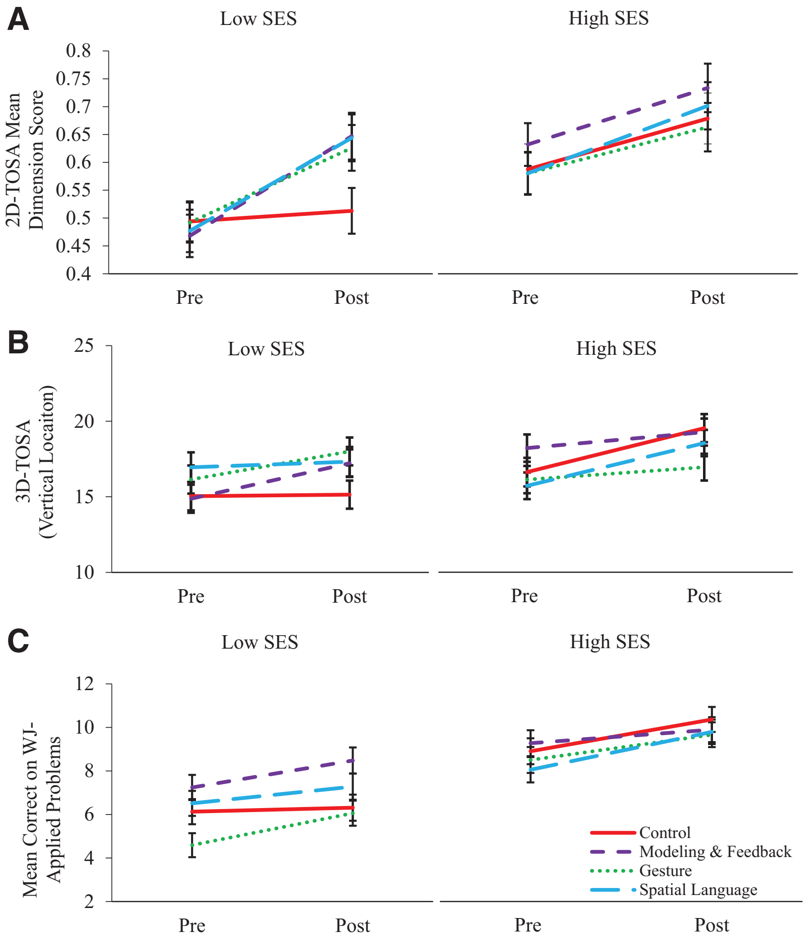

Figure 3. A: 2D Test of Spatial Assembly (TOSA): Low socioeconomic status (SES) children in the modeling and feedback (p = .001) and gesture (p = .007) conditions increased their scores with spatial training compared to the control group. B: 3D TOSA: Low-SES children in the modeling and feedback (p = .038) and gesture (p = .036) conditions increased their scores more than children in the control condition. C: WJ-Applied Problems: Low-SES children in the modeling and feedback training condition increased their score more than their high-SES peers, p = .035. The graphs display estimated marginal means when controlling for age in months, sex, and prevocabulary score. The errors bars represent standard errors. See the online article for the color version of this figure.

more than those who did not receive training. Notably, low-SES children improved their 2D TOSA performance at posttest when they had received nonverbal MF or GF, but not SLF. At pretest, low-SES children performed significantly lower on the WJ-PV and spatial vocabulary assessment compared to their high-SES peers. It is possible that low-SES preschoolers were not familiar enough with spatial language to profit from hearing it. In a way, the SLF may have provided too much information that was not useful, but the simple, nonverbal corrective feedback provided in the MF and GF conditions was beneficial.

Transfer to 3D spatial assembly. Even though there was no overall transfer effect of the 2D spatial assembly training to performance on a 3D spatial assembly task for all children, there was a significant moderation with SES. Low-SES children's correct vertical placement of blocks on the 3D TOSA benefited more from any training, but in particular the MF and GF training, compared to their high-SES peers. Perhaps these training conditions provided low-SES children with a tool, encouraging them to imagine where to put pieces prior to placing them. When the children placed pieces incorrectly, the trainer in these training conditions modeled such planning for them. But the SLF training may have provided too much information with the use of spatial terms. Furthermore, the trainings only transferred to increases in children's correct vertical location placement of blocks and not to the other dimensions (rotation and translation) or the composite of all dimension scores. This is not surprising given that *rotation* on the 3D TOSA meant aligning a piece perpendicular to another piece and *translation* meant putting the piece onto the correct studs. Neither of these skills were modeled in the 2D puzzles. However, even though the spatial assembly training was two-dimensional, children needed to attend to the correct orientation of pieces as they lined up vertically, analogous to vertical levels with the blocks. For example, the third item in Figure 1B has four pieces: The triangle is on the bottom level; square and small oval in the middle level; and large oval on the top level. Thus, attending to and encoding the spatial orientation of pieces according to their vertical placement in the 2D spatial assembly training may be similar to placing the blocks in the correct vertical location in the 3D spatial assembly task.

Transfer to spatial vocabulary. Performance on the spatial vocabulary assessment was expected to improve with SLF training that included not only relational words and phrases (e.g., *next to, above, to the side*), but shape names (e.g., *square, pentagon, triangle*). However, there was no significant training transfer effect on the spatial vocabulary assessment. Children in the SLF condition clearly paid attention to the spatial language because they performed better on the shape identification assessment compared to other children. Then why did the SLF training not transfer to the spatial vocabulary assessment, when it did for shape identification? One possible reason for the lack of SLF transfer to a spatial vocabulary comprehension task is that the children interpreted the relational language more as labels rather than describing interpiece relations. Another possible reason for the lack of SLF transfer could be that the transfer stimuli were too different (e.g., a teddy bear's spatial relation to a bucket) from the geometric puzzle task. On this view, anything that significantly deviated from the trained target was not impacted. Thus, future work should promote spatial language transfer effects by using spatial terms in a range of situations.

Mathematics Skills: Transfer of Training Effects

Even though there was no significant main effect of spatial training on children's mathematics performance, there was a significant SES moderation. Low-SES children's performance on the WJ-AP, but not on the TEMA, demonstrated far-transfer. In other words, low-SES children who received any training, and in particular MF, increased their WJ-AP performance more than those who did not receive training. Arguably, the novel and complex word problems given in the WJ-AP assessment required spatial visualization (e.g., Kaufmann, 1990; Shepard, 1978). One problem, for example, included a drawing of five balloons and asked, "If you had these balloons and someone gave you two more, how many balloons would you have?" Children might solve this problem by visualizing the addition of two balloons to a bunch of five balloons. The spatial assembly training provided similar visualization opportunities: children could imagine placing the individual shapes in their correct orientations before (or as) they actively moved the shapes. Thus, visualization experiences during the spatial assembly training may have supported low-SES children's skill in this domain, a skill that was also then used when solving these word problems. This process-level explanation is supported by research from Lourenco and colleagues (2018) and Mix (2019). Relatedly, this may help explain why simple MF provided during spatial training appeared to affect low-SES children's performance on the WJ-AP, but not on the TEMA-3. The latter assessed nonsymbolic number knowledge and number order, and then used tokens to assess nonverbal addition and subtraction, number constancy, and counting, not seeming to elicit spatial visualization skills as much as the WJ-AP test.

To review, simple MF during 2D spatial assembly training facilitated low-SES children's 2D and 3D spatial assembly performances and performance on one mathematics assessment. Moreover, GF also improved low-SES children's 2D spatial assembly performance and vertical placement of blocks on the 3D TOSA. Perhaps high-SES children are already receiving these types of spatial experiences in their homes and informal and formal educational contexts; however, low-SES children may not be. Providing low-SES children with corrective feedback on a 2D spatial assembly task can benefit their 2D and 3D spatial and mathematics skills.

Limitations and Future Directions

Despite the fact that 3-year-olds enjoyed the training and low-income children showed transfer to mathematics and elements of the 3D TOSA, the training consisted of only five 10-min training sessions, one session per week. Given the significant results in the current study with this relatively short training, future training studies might extend the training in an attempt to generate better transfer to spatial and mathematics tasks. Another limitation is that the children who received GF training did not gesture themselves, but only observed the trainer's gestures. Goldin-Meadow and colleagues (2012) provided evidence that children who produced their own task-relevant, iconic gestures while solving a spatial mental transformation task improved their performance more than those who only observed others perform similar gestures. The authors attribute this advantage of the 'learning-by-doing' approach to four factors: grounding children's cognition by recruiting the motor system; constructing new long-lasting action representations; elaborating and revising perceptual information; and guiding general attention to the task.

The current findings provide evidence that a playful, five-session spatial training can improve 3-year-olds' performance on the trained spatial task with some transfer to other spatial tasks and a mathematics task, especially for low-SES children. These findings highlight the need to extend these findings and design effective spatial training for young children who might be lacking such experiences in a way that facilitates both their spatial and mathematics skills.

References

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21,* 1605–1610. http://dx.doi.org/10.1177/0956797610385353

Bliss, S. (2006). Test of Early Mathematics Ability (3rd ed.). *Journal of Psychoeducational Assessment, 24,* 85–98. http://dx.doi.org/10.1177/0734282905282839

Bower, C., Foster, L., Zimmermann, L., Verdine, B., Marzouk, M., Islam, S., . . . Hirsh-Pasek, K. (2020). *Three-year-olds' spatial language comprehension and links with mathematics and spatial performance*. Manuscript submitted for publication.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence, 3,* 345–354. http://dx.doi.org/10.1016/0160-2896(79)90003-5

Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology, 41,* 183–192. http://dx.doi.org/10.1037/0012-1649.41.1.183

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction, 26,* 269–309. http://dx.doi.org/10.1080/07370000802177177

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. http://dx.doi.org/10.2307/749620

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15,* 2–11. http://dx.doi.org/10.1080/15248372.2012.725186

Child Trends Databank. (2015). *Early school readiness*. Retrieved from https://www.childtrends.org/wp-content/uploads/2015/01/indicator_1422334019.296.html

Cornu, V., Schiltz, C., Pazouki, T., & Martin, R. (2019). Training early visuo-spatial abilities: A controlled classroom-based intervention study. *Applied Developmental Science, 23,* 1–21. http://dx.doi.org/10.1080/10888691.2016.1276835

de Hevia, M. D., & Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition, 110,* 198–207. http://dx.doi.org/10.1016/j.cognition.2008.11.003

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32,* 25–32. http://dx.doi.org/10.1016/S0160-2896(03)00061-8

Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, And future* (pp. 1–35). Washington, DC: U.S. Bureau of Labor Statistics.

Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition, 35,* 283–296. http://dx.doi.org/10.3758/BF03193449

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/4117.001.0001

Ginsburg, H., & Baroody, A. (2003). *TEMA-3 examiners manual*. Austin, TX: Pro-Ed.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report, 22,* 1–24. https://eric.ed.gov/?id=ED521700. http://dx.doi.org/10.1002/j.2379-3988.2008.tb00054.x

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20,* 267–272. http://dx.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. http://dx.doi.org/10.1111/j.1467-7687.2012.01185.x

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12,* 516–522. http://dx.doi.org/10.1111/1467-9280.00395

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development, 90,* 985–992. http://dx.doi.org/10.1111/cdev.13128

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48,* 1229–1241. http://dx.doi.org/10.1037/a0027433

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27,* 4–9.

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4,* 60–68. http://dx.doi.org/10.1016/j.tine.2015.05.001

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91,* 684–689. http://dx.doi.org/10.1037/0022-0663.91.4.684

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49,* 4–14. http://dx.doi.org/10.1037/a0027238

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26,* 302–310. http://dx.doi.org/10.1177/0956797614563338

Kaufmann, G. (1990). International library of psychology. Imagery: Current developments. In P. J. Hampson, D. F. Marks, & J. T. E. Richardson (Eds.), *Imagery effects on problem solving* (pp. 169–196). Florence, KY: Taylor & Frances/Routledge.

Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J., & Hautamaki, J. (2003). Visuospatial working memory and early numeracy. In S. J. Pickering (Ed.), *Memory and education* (pp. 65–76). Leicester, UK: British Psychological Soc.

Levine, S. C., Goldin-Meadow, S., Carlson, M. T., & Hemani-Lopez, N. (2018). Mental transformation skill in young children: The role of concrete and abstract motor training. *Cognitive Science, 42,* 1207–1228. http://dx.doi.org/10.1111/cogs.12603

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35,* 940–949. http://dx.doi.org/10.1037/0012-1649.35.4.940

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53,* 72–103. http://dx.doi.org/10.1016/S0022-0965(05)80005-0

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48,* 530–542. http://dx.doi.org/10.1037/a0025913

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50,* 315–353. http://dx.doi.org/10.1016/j.cogpsych.2004.09.004

Lourenco, S. F., Cheung, C.-N., & Aulet, L. S. (2018). Is visuospatial reasoning related to early mathematical development? A critical review. In A. Henik & W. Fias (Eds.), *Heterogeneity of function in numerical cognition* (pp. 177–210). Cambridge, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-811529-9.00010-8

Lourenco, S. F., & Longo, M. R. (2009). Multiple spatial representations of number: Evidence for co-existing compressive and linear scales. *Experimental Brain Research; Heidelberg, 193,* 151–156. http://dx.doi.org/10.1007/s00221-008-1698-9

Miller, H. E., Patterson, R., & Simmering, V. R. (2016). Language supports young children's use of spatial relations to remember locations. *Cognition, 150,* 170–180. http://dx.doi.org/10.1016/j.cognition.2016.02.006

Miller, H. E., & Simmering, V. R. (2018). Children's attention to task-relevant information accounts for relations between language and spatial cognition. *Journal of Experimental Child Psychology, 172,* 107–129. http://dx.doi.org/10.1016/j.jecp.2018.02.006

Mix, K. S. (2019). Why are spatial skill and mathematics related? *Child Development Perspectives, 13,* 121–126. http://dx.doi.org/10.1111/cdep.12323

Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: Developmental and educational implications. In J. Benson (Ed.), *Advances in Child Development and Behavior* (Vol. 42, pp. 197–243). Waltham, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145,* 1206–1227. http://dx.doi.org/10.1037/xge0000182

Möhring, W., Frick, A., & Newcombe, N. S. (2018). Spatial scaling, proportional thinking, and numerical understanding in 5- to 7-year-old children. *Cognitive Development, 45,* 57–67. http://dx.doi.org/10.1016/j.cogdev.2017.12.001

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215,* 1519–1520. http://dx.doi.org/10.1038/2151519a0

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2006). *Learning to think spatially: GIS as a support in the K-12 curriculum*. Washington, DC: The National Academies Press.

Newcombe, N. (2017). *Harnessing spatial thinking to support stem learning* (OECD Education Working Papers No. 161). http://dx.doi.org/10.1787/7d5dcae6-en

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and equity in education*. Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2015-results-volume-i_9789264266490-en

Park, Y., & Casasola, M. (2017). The impact of object type on the spatial analogies in Korean preschoolers. *Cognitive Psychology, 94,* 53–66. http://dx.doi.org/10.1016/j.cogpsych.2017.02.001

Ping, R., Ratliff, K., Hickey, E., & Levine, S. C. (2011). Using manual rotation and gesture to improve mental rotation in preschoolers. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 459–464). Boston, MA: Curran Associates, Inc.

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science, 14,* 1417–1430. http://dx.doi.org/10.1111/j.1467-7687.2011.01088.x

Purpura, D. J., Napoli, A., Wehrspann, E., & Gold, Z. (2017). Causal connections between mathematical language and mathematical knowledge: A dialogic reading intervention. *Journal of Research on Educational Effectiveness, 10,* 116–137. http://dx.doi.org/10.1080/19345747.2016.1204639

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly, 36,* 259–268. http://dx.doi.org/10.1016/j.ecresq.2015.12.020

Rittle-Johnson, B., Zippert, E. L., & Boice, K. L. (2019). The roles of patterning and spatial skills in early mathematics development. *Early Childhood Research Quarterly, 46,* 166–178. http://dx.doi.org/10.1016/j.ecresq.2018.03.006

Satlow, E., & Newcombe, N. (1998). When is a triangle not a triangle? Young children's developing concepts of geometric shape. *Cognitive Development, 13,* 547–559. http://dx.doi.org/10.1016/S0885-2014(98)90006-5

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside Publishing.

Shepard, R. N. (1978). The mental image. *American Psychologist, 33,* 125–137. http://dx.doi.org/10.1037/0003-066X.33.2.125

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237–243.

Sortor, J. M., & Kulp, M. T. (2003). Are the results of the Beery-Buktenica Developmental Test of Visual-Motor Integration and its subtests related to achievement test scores? *Optometry and Vision Science, 80,* 758–763. http://dx.doi.org/10.1097/00006324-200311000-00013

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,* 352–402. http://dx.doi.org/10.1037/a0028446

van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology, 101,* 99–113. http://dx.doi.org/10.1016/j.jecp.2008.05.001

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2014). Finding the missing piece: Blocks, puzzles, and shapes fuel school readiness. *Trends in Neuroscience and Education, 3,* 7–13. http://dx.doi.org/10.1016/j.tine.2014.02.005

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and mathematical skills across the preschool years. *Monographs of the Society for Research in Child Development, 82,* 1–109.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85,* 1062–1076. http://dx.doi.org/10.1111/cdev.12165

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology, 126,* 37–51. http://dx.doi.org/10.1016/j.jecp.2014.02.012

Verdine, B. N., Lucca, K. R., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2016). The shape of things: The origin of young children's knowledge of the names and properties of geometric forms. *Journal of Cognition and Development, 17,* 142–161. http://dx.doi.org/10.1080/15248372.2015.1016610

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102,* 860–871. http://dx.doi.org/10.1037/a0019454

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology, 114,* 591–600. http://dx.doi.org/10.2307/1423612

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36,* 155–193. http://dx.doi.org/10.1016/0010-0277(90)90003-3

Xu, C., & LeFevre, J.-A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology, 52,* 854–866. http://dx.doi.org/10.1037/dev0000124

Received April 15, 2019
Revision received October 11, 2019
Accepted December 23, 2019 ■