AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

2022, Vol. 58, No. 1, 112–137
https://doi.org/10.1037/dev0001281

Effects of Spatial Training on Mathematics Performance: A Meta-Analysis

Zachary C. K. Hawes[1], Katie A. Gilligan-Lee[2], and Kelly S. Mix[3]

[1] Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto
[2] School of Psychology, University of Surrey
[3] Department of Human Development and Quantitative Methodology, University of Maryland

Prior research has revealed robust and consistent relations between spatial and mathematical skills. Yet, establishing a causal relation has been met with mixed effects. To better understand whether, to what extent, and under what conditions mathematics performance can be improved through spatial training, we conducted a systematic meta-analysis of the extant literature. Our analysis included 29 studies that used controlled pre-post study designs to test the effects of spatial training on mathematics (N = 3,765; k = 89). The average effect size (Hedges's g) of training relative to control conditions was .28 (SE = .07). Critically, there was also evidence that spatial training improved individuals' spatial thinking (g = .49, SE = .09). Follow-up analyses revealed that age, use of concrete manipulatives, and type of transfer ("near" vs. "far") moderated the effects of spatial training on mathematics. As the age of participants increased from 3 to 20 years, the effects of spatial training also increased in size. Spatial training paradigms that used concrete materials (e.g., manipulatives) were more effective than those that did not (e.g., computerized training). Larger transfer effects were observed for mathematics outcomes more closely aligned to the spatial training delivered compared to outcomes more distally related. None of the other variables examined (training dosage, spatial gains, posttest timing, type of control group, experimental design, publication status) moderated the effects. Additionally, analyses of publication bias and selective outcome reporting were nonsignificant. Overall, our results support prior research and theoretical claims that spatial training is an effective means for enhancing mathematical understanding and performance. However, our meta-analysis also highlights a poor understanding of the mechanisms that support transfer. To fully realize the potential benefits of spatial training on mathematics achievement, more theoretically guided studies are needed.

Keywords: spatial training, spatial visualization, mathematics performance, meta-analysis, STEM

Broadly defined as the ability to generate, manipulate, and reason about spatial relations between and within objects, spatial thinking is widely regarded as a key contributor to mathematics performance. Indeed, over a century of research has revealed robust and consistent relations between spatial and mathematical abilities (Galton, 1880; Smith, 1964; Xie et al., 2020; see reviews by Hawes & Ansari, 2020; Lourenco et al., 2018; Mix & Cheng, 2012). Moreover, evidence of this relation is found in people of all ages, from infancy through adulthood, and across a diverse range of populations.

Zachary C. K. Hawes https://orcid.org/0000-0003-4055-7607

We have no conflict of interest to disclose. As per journal requirements we acknowledge that this study was not preregistered. All data, supplementary material, and annotated analyses with the accompanying R code can be accessed on our OSF page (https://osf.io/8yn7m/).

We thank and acknowledge Elizabeth Tipton and James Pustejovsky for generously offering us their time and expertise in RVE analysis. We thank Alice Hamilton for her assistance in the data search and extraction process.

Correspondence concerning this article should be addressed to Zachary C. K. Hawes, Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, Dr. Eric Jackman Institute of Child Study, University of Toronto, 45 Walmer Road, Toronto, ON M5R 2X2, Canada. Email: zack.hawes@utoronto.ca

This evidence, coupled with theoretical accounts of the space-mathematics link, has led to speculation that spatial training may be an effective means of improving mathematical abilities. Researchers have claimed that "spatial instruction can be expected to have a 'two-for-one' effect that yields benefits in mathematics and the spatial domain" (Verdine et al., 2014; p. 12). Others have suggested that spatial training could pay substantial dividends in increasing participation and success in not only mathematics, but the other STEM (Science, Technology, Engineering, and Mathematics) disciplines as well (Newcombe, 2010; Uttal et al., 2013). Given the central role that mathematics plays in academic achievement (Duncan et al., 2007), as well as other important outcomes, such as socio-economic status (SES), health, and personal well-being (Parsons & Bynner, 2005; Ritchie & Bates, 2013), evidence-based approaches to improving mathematics are increasingly in demand. Thus, establishing whether spatial training transfers to mathematics is of both theoretical and practical importance.

Currently, such a prediction rests almost entirely on theoretical claims and correlational evidence (Hawes & Ansari, 2020). Only recently have researchers begun to test whether spatial training generalizes beyond the spatial domain to support improvements in other domains, such as mathematics. Our aim is to provide a systematic meta-analysis of this emerging literature, to determine whether, to what extent, and under what conditions mathematics performance can be improved through spatial training.

Why Spatial Training Should Work

One might question the potential of spatial training to show transfer to mathematics, given that training in other domain-general skills, such as working memory has not shown such transfer (e.g., see Melby-Lervåg et al., 2016; Sala & Gobet, 2019; 2020). Certainly, simply demonstrating that performance in two domains is correlated does not explain why they are connected or guarantee that learning in one will transfer to the other (Hawes & Ansari, 2020; Mix & Cheng, 2012). However, spatial training may have greater promise than other domain-general skills training, given growing evidence that the relations between spatial thought and mathematical thought are based on shared cognitive processes.

First, there is substantial evidence that spatial skills are malleable. A meta-analysis of 206 training studies across a 25-year period (1984–2009) demonstrated that spatial thinking can be improved in people of all ages and through a variety of training approaches (i.e., video games, course training, spatial task training) (Uttal et al., 2013). The average effect size (Hedges's *g*) for training relative to control in these studies was approximately one half a standard deviation (.47). To put this finding into context, the authors suggested that an improvement of this magnitude would approximately double the number of U.S. citizens with the spatial skills associated with receiving a degree in engineering.

Importantly, studies also indicate that spatial processes are activated automatically when people solve mathematics problems (see Hawes & Ansari, 2020, for a review). For example, results from a recent fMRI meta-analysis indicated that spatial (mental rotation) skills and numerical skills activate highly overlapping regions of the intraparietal cortex, suggesting common neural mechanisms (Hawes, Sokolowski, et al., 2019; see also Hubbard et al., 2005). There is also strong evidence that people represent numerical magnitudes in a spatial format (Dehaene et al., 1993; Fias & Fischer 2005; Toomarian & Hubbard, 2018), and rely on spatial processing to decode mathematical symbols, for example, algebra equations (Landy & Goldstone, 2010). Moreover, multiple factor analyses have revealed remarkably strong, sometimes even indistinguishable, relations between spatial and mathematical task performance across development (Frick, 2019; Mix et al., 2016; 2017). Strong evidence of shared processing distinguishes the relation between mathematics and space from relations between mathematics and other domain-general processes (e.g., see Hawes & Ansari, 2020).

Another reason to predict strong transfer from spatial training to mathematics is that people strategically recruit spatial skills and representations when solving mathematics problems (Casey & Fell, 2018; Lowrie & Kay, 2001; Mix, 2019). For example, spatial visualization can be used as a "mental blackboard" on which to model, simulate, and manipulate mathematical relations (e.g., Seron et al., 1992), or to ground the meaning of mathematical symbols (Mix, 2019). People also use spatial tools, such as number lines, base-10 blocks, or diagrams as scaffolding while problem solving (Hegarty & Kozhevnikov, 1999; Mix, 2010)—an approach that is widely used in mathematics education (Carbonneau et al., 2013; Sowell, 1989). The widespread use of spatial strategies in mathematics indicates a natural connection that may be enhanced through training. In contrast, there are fewer examples of similar strategy use for other domain-general skills for which training has failed to transfer.

To summarize, there are a variety of reasons why spatial training may be an effective means of improving mathematics performance. These reasons include shared domain-general processes (e.g., mental manipulation of objects), as well as shared domain-specific processes (e.g., use and knowledge of spatial transformations to solve geometry and measurement problems). Spatial skills have also been linked to more flexible and efficient mathematics problem-solving strategies (Casey et al., 2017; Hegarty & Kozhevnikov, 1999; Laski et al., 2013; Lowrie & Kay, 2001). Thus, spatial training may have the added benefit of encouraging more effective spatially based problem-solving strategies and solutions (e.g., see Casey & Fell, 2018). A similar argument is that spatial instruction provides learners with additional entry points into mathematics—new ways of seeing, understanding, and appreciating mathematics that without explicit spatial instruction will continue to go underrecognized, undervalued, and underdeveloped (Moss et al., 2016; National Research Council, 2006). Together, these reasons not only indicate why spatial training may transfer to mathematics, but further suggest the added value of spatial training that extends beyond standard mathematics instruction alone.

What Spatial Skills to Train?

To date, investigations of space-math relations have focused almost exclusively on spatial visualization (i.e., small-scale reasoning that involves dynamic representations of spatial relations, e.g., Hegarty et al., 2006). Whether large-scale spatial skills, such as navigation, also relate to mathematics remains an open question. Indeed, nearly all studies included in the current analysis focused their training on the development of spatial visualization skills (e.g., mental rotation).

This focus on spatial visualization may be due to the extensive body of research demonstrating moderate to strong correlations between spatial visualization skills and mathematics (Delgado & Prieto, 2004; Hawes, Moss, et al., 2019; Hegarty & Kozhevnikov, 1999; Kyttälä & Lehto, 2008; Tam et al., 2019; Tolar et al., 2009; Wei et al., 2012), as well as the current theoretical accounts described above, which offer cogent explanations for this specific relation (e.g., Seron et al., 1992). Taken together, it makes sense that spatial visualization would be seen as the most promising candidate for spatial training studies. However, this is not to say other spatial skills might not also facilitate mathematics performance. Factor analyses have indicated that spatial skills form a single factor in elementary aged children (Mix et al., 2016, 2017)—a factor that is highly correlated with mathematics—which implies that training in any of these skills should be equally effective at improving mathematics performance. There is also growing evidence that spatial skills besides spatial visualization, such as spatial scaling, figure copying, and proportional reasoning, represent important spatial processes for the learning, doing, and understanding of mathematics (Frick, 2019; Gilligan et al., 2019; Mix et al., 2020; Möhring et al., 2016). Presently, however, the dominant approach to testing whether spatial training generalizes to mathematics involves the training of spatial visualization skills. Our results should be interpreted with this limitation in mind.

The Evidence: Transfer From Spatial Training to Mathematics

As noted above, direct tests of transfer from spatial training to mathematics are relatively recent and limited. Of the published studies in this area, 95% have appeared in or since 2014. Admittedly, the literature that has amassed so far is mixed, but there is emerging evidence to support the promise of spatial training.

Some studies published on this topic have reported significant positive transfer (e.g., Cheng & Mix, 2014; Hawes et al., 2017; Lowrie et al., 2017; Mix et al., 2020). For example, Cheng and Mix (2014) gave 7-year-olds an age-appropriate calculation test, and then, in a later session, had them complete a series of spatial transformation exercises with feedback. Immediately following completion of the spatial exercises, children were given a second calculation test. There was a significant increase in children's calculation scores from pre- to posttest for the spatial training group, but no improvement for a control group that had practiced crossword puzzles instead. Similar findings have been reported for older children (e.g., 12-year-olds, Lowrie et al., 2017; Mix et al., 2020) and following longer periods of training (e.g., 32 weeks, Hawes et al., 2017). Others have reported positive effects on mathematics scores in children following mixed training that included, but was not limited to, spatial skills such as mental rotation (Nelwan & Kroesbergen, 2016). Studies have shown similar improvement among undergraduates taking science and engineering course work following spatial visualization training (Miller & Halpern, 2013; Sorby, 2009; Sorby et al., 2013). Taken together, these studies provide intriguing preliminary evidence that spatial training can improve student outcomes in mathematics.

However, not all attempts to obtain transfer from spatial training to mathematics have been successful. In some studies, spatial training led to improvement in spatial skill, but there was no transfer to numeracy or mathematics (Cornu et al., 2019; Hawes et al., 2015; Rodán et al., 2019; Xu & LeFevre, 2016). For example, Hawes et al. (2015) completed a randomized controlled study in which 6- to 8-year-olds received six weeks of training in either mental rotation or literacy. Children who had practiced mental rotation went on to show significant gains on a test of spatial transfer that included spatial transformation and spatial puzzles, but they showed no improvement on tests of calculation (either symbolic or nonsymbolic). Children in the literacy group had no improvements in spatial or mathematics performance at posttest. Thus, spatial training was effective at improving general spatial skill but did not transfer to mathematics.

It is important to know whether the positive effects of spatial training that have been reported are reliable and whether they reflect a valid causal mechanism for improving mathematics. To address this question, we carried out an extensive literature search to identify studies that tested whether spatial training improves mathematics. We then used meta-analysis techniques to estimate the size of these effects when the entire body of evidence is considered. We also tested a range of potential moderator variables, to determine whether spatial training improves mathematics skills but only under certain conditions. Our rationale for selecting specific moderator variables follows.

Potential Moderators of Spatial Training Effects

We identified and tested nine moderator variables. Eight of these were methodological dimensions that may have contributed to the variability in transfer between studies, each of which is briefly described below. As one of several efforts to address publication bias, we also tested whether results differed for published versus unpublished studies. Whether or not there is a significant overall effect of spatial training, these differences in study design, features of spatial training, and publication status might explain why transfer is reported in some studies and not others.

Participant Age

Spatial processing may be more tightly tied to mathematics at different developmental stages and if so, these age-linked effects could moderate the effectiveness of spatial training. For example, it could be argued that spatial processes are more tightly tied to mathematics in early childhood because young children are preoccupied with grounding symbols and all mathematics content is relatively novel to them. On some accounts, spatial processes may be engaged more frequently when learners are grounding symbols or interpreting novel problems (Mix et al., 2016; Uttal & Cohen, 2012). Alternatively, spatial processes might become more tightly tied to mathematics as children get older and become more skilled in recruiting spatial processes strategically. Although direct comparisons among elementary age groups do not suggest a change in the strength of the correlation between spatial skill and mathematics (e.g., Mix et al., 2016), such a difference might emerge if a wider range of age groups were considered. For example, we know that two of the studies for which spatial training effects were not found (Cornu et al., 2019; Xu & LeFevre, 2016) focused on 3- to 5-year-olds —a younger age group than had been tested in studies showing positive training effects (i.e., 7- to-12-year-olds).

Training Dosage

The number and length of spatial training sessions have varied across studies in this literature. Although it stands to reason that more spatial training would lead to greater effects, the evidence does not seem to indicate a linear relation. For example, one study reported significant transfer following only a single training session lasting 40 minutes (Cheng & Mix, 2014). Similar effects were obtained after 6 weeks of training (Mix et al., 2020). Thus, despite considerable differences in training dosage, comparable levels of transfer to mathematics occurred. One explanation for these nonlinearities may be threshold effects, in which spatial training improves mathematics up to a point, after which further training has less impact (Freer, 2017). However, it is important to first establish whether the patterns related to dosage are linear when the entire body of evidence is considered. The present meta-analysis provides such an opportunity and may either confirm that the relation is nonlinear or reveal stable patterns that might otherwise have been missed.

Spatial Gains

Studies vary in how effective the spatial training delivered is in improving spatial outcomes, that is, the size of spatial gains. One might expect that greater spatial gains following training may be correlated with greater transfer to mathematics outcomes.

Transfer Distance Within Mathematics (Near Versus Far Math Transfer)

Spatial training may have greater effects for mathematics topics that have direct overlap with the spatial training delivered. For example, it could be argued that spatial training using origami has direct overlap with the skills tested in geometry and measurement tasks because both the spatial training and the mathematics tasks require interpreting and manipulating forms. Weak support for this notion comes from cross-domain factor loadings which seemed to indicate that certain mathematics skills were more highly related to spatial skill than others (Mix et al., 2016; 2017). However, the existing literature provides several reasons to question predictions based on mathematical specificity. First, there is little evidence, to date, to suggest that spatial skills are differentially related to different aspects of mathematics (e.g., Mix & Cheng, 2012; Xie et al., 2020). Second, spatial skills are strongly correlated with basic numerical reasoning tasks that—on the surface—do not appear to share conceptual overlap with spatial skills (e.g., comparing the larger of two numbers) (Verdine et al., 2014; Viarouge et al., 2014). Finally, as noted above, the spatial and mathematical domains appear to be unitary but highly correlated (Mix et al., 2016) and the evidence that certain kinds of mathematics are particularly sensitive to certain kinds of spatial training is weak (Mix et al., 2020). That said, attempts to classify mathematical strands (e.g., numeration, algebra, data science, etc.) according to spatial processing demands has scarcely been examined, and it remains to be demonstrated whether the space-math relation depends on the specific spatial paradigm delivered and the characteristics of the mathematics outcomes under investigation.

Training Delivery (Concrete Versus Nonconcrete)

Studies vary in whether spatial training is delivered with objects or through an electronic application. Mix et al. (2020) used an object-based delivery in which children first chose two of four pictures that matched a standard after rotation, and then used three-dimensional block constructions, that they could physically turn, to check their choices. Other studies have programmed similar spatial exercises into software applications that children can use on an iPad or laptop computer (e.g., Hawes et al., 2015; Rodán et al., 2019). Given that spatial processes support and constrain movement through physical space and the manipulation of objects, one might expect spatial training with physical objects to be more effective. However, some studies have shown no difference between implicit video-based training and training with explicit practice and feedback (Gilligan et al., 2019), suggesting that training delivery may not be a meaningful moderator. The present meta-analysis compares studies that use concrete materials (physical manipulatives) in the delivery of spatial training to those that do not, and thus will provide a broader base from which to assess the potential effects of training delivery.

Posttest Timing

In some studies, the mathematics posttest was given immediately following training, while in others, it was given several days, or sometimes more than a week, after training. It is important to examine these differences because the timing of the mathematics posttest could affect the nature of the underlying spatial changes being measured. Posttests administered after longer delays are more likely to reflect stable changes in spatial (and mathematics) skill, whereas posttests administered immediately after training may reflect priming effects rather than stable skill changes.

Control Group Type (Business-as-Usual Versus Active)

Part of our inclusionary criteria was that studies had to include a control group. Studies that used within-group pre-post studies designs were excluded from the present study. Control groups differed according to whether they included a business-as-usual (BAU) control group versus an active control group. For our analyses, we compared the outcomes of studies that delivered BAU mathematics curriculum to children in the control group against the outcomes of studies that provided direct instruction in either mathematics or an unrelated skill to children in the control group. One could argue that active controls provide a more rigorous test of the effects of spatial training given that all children (experimental and control) receive some form of intervention. Indeed, prior meta-analyses have revealed that the type of control group can greatly moderate, even eliminate, the effects and conclusions one can make about the effectiveness of training (e.g., see Green et al., 2019; Melby-Lervåg et al., 2016; Uttal et al., 2013). Thus, it is important to confirm whether there are significant effects of spatial training for both control types. Note, in the present study, there were not enough studies to compare the effects of training against different types of active control groups (e.g., those that received mathematics instruction vs. an unrelated skill, such as literacy training). Thus, our comparison was between BAU and active controls only.

Experimental Design (RCT Versus Quasi-Experimental)

Some of the studies in this corpus were randomized controlled trials (RCTs) with a randomly assigned control group. Others took advantage of existing classroom groupings and delivered the spatial training in a nonrandom, quasi-experimental design. Because RCTs provide a more stringent test, one might expect the effects of spatial training to be greater in, and perhaps even exclusive to, the quasi-experimental designs. To know whether spatial training has robust and consistent effects, it is important to determine whether significant positive effects have been observed in studies of both designs.

Publication Status

As one of several safeguards against publication bias, we included a moderator analysis for published versus unpublished studies. This comparison could indicate whether any significant effects in the overall dataset are attributable to published studies, in whole or in part, as an index of possible file drawer effects.

Other Potential Moderators

Several other factors could plausibly influence the effectiveness of spatial training and the degree to which spatial training transfers to mathematics, but we were unable to test them given the studies that met our inclusion criteria. Specifically, we were unable to explore the effects of different types of spatial skill training because the majority of studies delivered some form of spatial visualization training, and many spatial skills were not represented at all (e.g., large-scale spatial skills such as navigation). In the end, there was not enough variability in the types of spatial skills trained to include it as a moderator. Similarly, although motivational factors may also influence the efficacy of spatial training, only one study measured motivational factors. Finally, it was not possible to investigate the durability of gains over time because only one study in our meta-analysis included a follow-up test after posttest.

Current Study

This study addresses three overarching research questions. First, is there a causal effect of spatial training on spatial skills? Given the assumption that transfer from the spatial to mathematical domain is based on the malleability of spatial thinking, it makes sense to first establish whether the independent variable of interest (spatial skill) is indeed mutable. Moreover, many spatial training studies have been conducted since the Uttal et al. (2013) analysis. The findings from the present study will provide additional insights into the malleability of spatial thinking. Second, our main question asks whether there is a causal effect of spatial training on mathematics. We will determine this by investigating whether spatial training leads to transfer of gains to mathematics outcomes when the entire literature is considered. Based on Pearl's (2009) Interventionalist Theory of Causation, if an effect of spatial training on mathematics is obtained, (i.e., a manipulation in one variable has led to change in another), then a causal relationship can be inferred. Third, how might specific conditions and features of spatial training influence the extent to which spatial training transfers to mathematics? We address this question by examining a range of candidate moderators.

Method

The meta-analysis search was conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (www.prisma-statement.org), (Moher et al., 2009).

Data Collection and Inclusion Criteria

As outlined in Figure 1, the search strategy included several stages. Suitable articles for this meta-analysis were identified using electronic searches of the databases PUBMED, PsycINFO, ERIC, ProQuest Dissertations & Theses, and Google Scholar (November 2019). For each database, we used combinations of the search terms "training," "practice," "experience" and "instruction" with the search terms "spatial," "mental rotation," "scaling," "visualisation (visualization)" and "visuospatial skills," and the search terms "mathematics," "maths," "mathematical," "numeracy," "arithmetic," "calculation" and "geometry." This search yielded 190 results. To avoid publication bias (i.e., the file-drawer effect), we took several additional steps to acquire unpublished data. We searched the tables of contents for recent conference proceedings for any relevant studies. We sent out a call for data that explicitly asked for unpublished work (see Appendix A or visit our Open Science Framework (OSF) page: https://osf.io/8yn7m/). The call was distributed through several international mailing lists, including the Spatial Intelligence Learning Center (SILC; Sep 24th 2019), DEV-Europe (Sep 23rd 2019), DP-Net (October 12th 2019) and the Cognitive Development Society (Cogdevsoc; Oct 13th 2019), as well as the professional networks of the research team. These steps led to the discovery of 28 additional studies.

A three-stage process was used to ensure that all of the studies identified in our search were suitable for inclusion in the meta-analysis (see Table 1). Each stage was completed independently by the first and second authors and then the results were compared to determine interrater agreement.

Stage I

For each identified article, the title was reviewed, and articles were removed if they (i) used nonhuman subjects, (ii) were published in a language other than English, or (iii) did not include behavioral evidence (e.g., used cellular or physiology-based outcomes instead, or were solely review articles). Complete interrater agreement was achieved on all but two studies (99.98% total agreement), and these were both ultimately included following discussion. In total, 111 articles progressed to Stage II.

Stage II

The reviewers read the abstracts of each article and determined whether the studies included (a) a pretest-training-posttest design, (b) at least one mathematics outcome measure, (c) at least one spatial training group and (d) a control group. Interrater reliability was 100% and 47 studies progressed to Stage III.

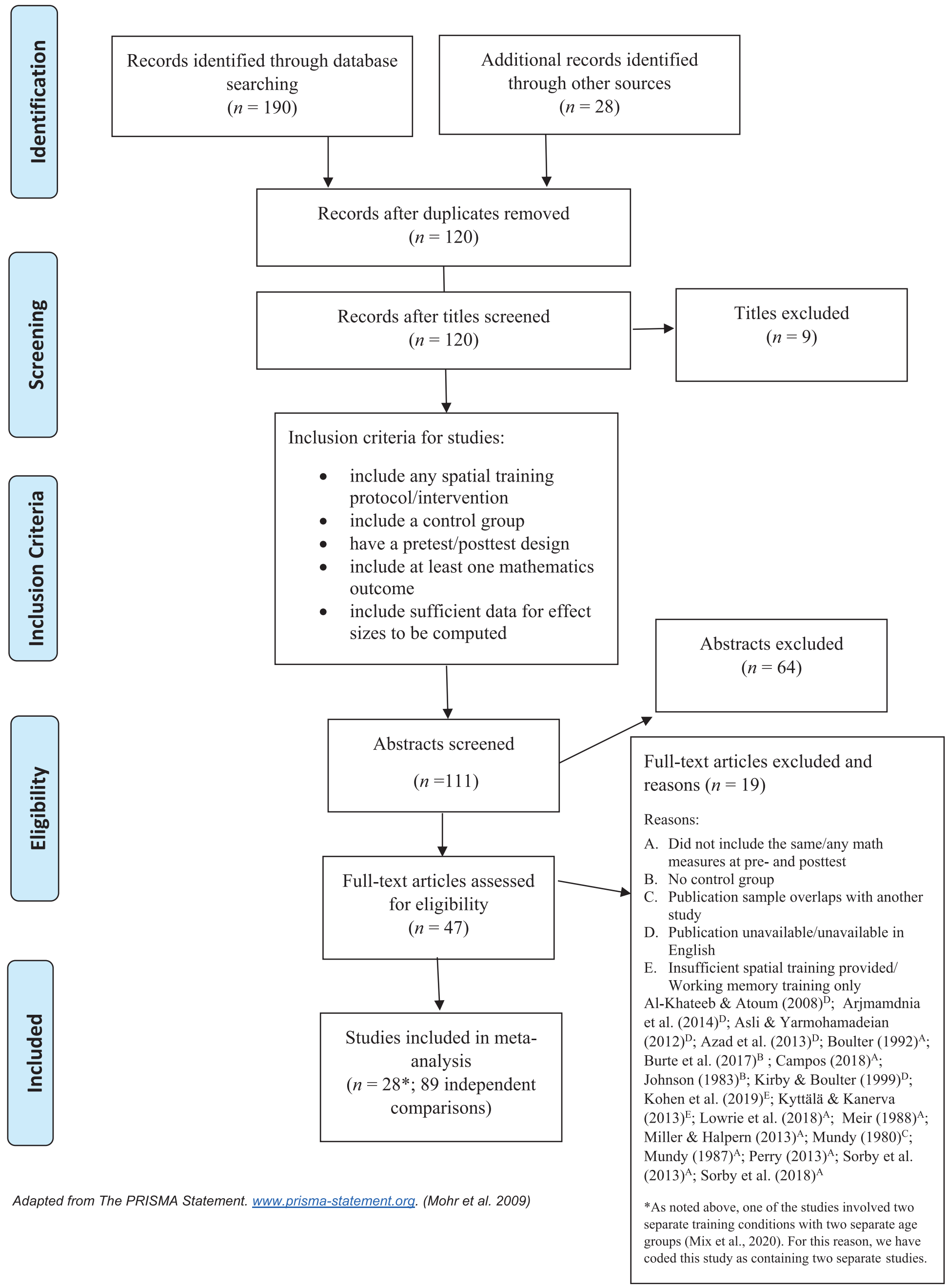
Stage III

Each reviewer read each article in full to confirm its suitability and 19 articles were excluded (reasons for this are given in Figure 1). The interrater reliability at Stage III was 100%. Twenty-eight studies were included in the final meta-analysis. This number increased to 29 studies when we noticed that one of the studies included two distinct age groups (1st and 6th grade students), (Mix et al., 2020), leading us to analyze the two age groups as separate studies (See Table 2). We contacted authors directly if they did not provide enough information in their papers (e.g., pre- and posttraining data), to calculate effect sizes. All authors who were contacted provided this information on request. As a final attempt to locate any missing studies, we also searched the reference lists of all suitable articles identified through the literature search. No additional studies were identified. Interrater reliability was 100%.

Moderator Variables

The operational definitions for our eight moderator variables are presented below. Twenty percent of studies were double coded by

Figure 1
PRISMA Diagram Outlining the Search Strategy and Reasons for Excluded Studies



Note. See the online article for the color version of this figure.

two reviewers (i.e., the first and second authors) and their interrater reliability was 98%. Any discrepancies were resolved through discussion and consulting the original papers. The remaining studies were divided between the two reviewers and coded separately. When information about a particular moderator was missing, we requested the information from the authors directly. For the moderator "Control Group," two studies compared the same experimental group to two different control groups. In both cases, only one comparison was used in the meta-analysis. In one case, we chose the control group that most closely resembled the experimental group. Specifically, the experimental group included participants with math learning difficulties (MLD), so we selected the control group of children with MLD and excluded the control group of children who were typically developing (Krisztián et al., 2015). In the second case, both control groups had similar demographic features, so we took the most conservative approach and selected the control

Table 1
Summary of Inclusion Criteria for the Meta-Analysis

Inclusion criteria	Description
1	Includes at least one mathematical/numerical outcome measure
2	Includes at least one spatial training group and one control group
3	Has a causal, pretest-posttest design
4	Has either a random or quasi-experimental study design
5	Includes effect sizes for intervention effects, or sufficient data so that effect sizes can be generated by the reviewers
6	Based on a human population of any age
7	Written in English (or a suitable translation is available)
8	Data/article available before March 1st, 2020
9	Outlines new data, that is, is not a review article

group that generated the smaller effect size when compared to the experimental group (Freer, 2016). Lastly, it should be noted that one of the selected studies included two posttests (i.e., an immediate posttest and a follow-up posttest; Honour, 2020). However, as this was the only study to include a follow-up posttest, these data were not included in the meta-analysis.

Participant Age

Participant age was coded continuously based on the mean age of the participants measured in months.

Training Dosage

Training dosage was coded continuously based on the overall amount (duration) of spatial training measured in minutes.

Spatial Gains

Spatial gains was coded continuously as the average effect size for gains in spatial outcomes following spatial training compared to the control group.

Transfer Distance Within Mathematics (Near Versus Far Math Transfer)

Transfer distance within mathematics (near versus far math transfer) was coded categorically at the outcome level. If the raters agreed there was clear overlap between the skills practiced during training and the skills measured on the posttest, the effect was categorized as near math transfer (e.g., origami-based spatial training and outcome measures of geometry). If there was not clear overlap between the skills practiced during training and the skills measured on the posttest, the effect was categorized as far math transfer (e.g., origami-based spatial training and arithmetic). For full details on how each math outcome was categorized, see Table 2.

Training Delivery (Concrete Versus Nonconcrete)

Training delivery (concrete versus nonconcrete) was coded categorically at the study level. If the spatial training included the use of concrete materials (e.g., blocks, tangrams, paper folding) the study was coded as concrete. This category included studies where training had both concrete and nonconcrete elements. If a study did not use any concrete materials in the delivery of spatial training (e.g., training was digital practice, paper and pencil worksheets), it was classified as nonconcrete.

Posttest Timing

Posttest timing was coded categorically at the study level based on when posttesting was completed. The categories were (a) immediate posttesting; (b) posttesting between 1 and 7 days after the completion of spatial training; (c) posttesting 8 days or longer following the completion of spatial training. These categories were based on Uttal et al. (2013) and provided a roughly equal number of studies per category.

Control Group

Control group was coded categorically at the study level based on the type of control group used. Studies were classified as having a business-as-usual control if the children in the control group did not complete any additional training sessions beyond their usual classroom lessons. Studies were classified as having an active control if children in the control group completed the same number of training sessions as the experimental group, but either practiced skills in a nonspatial domain (e.g., crossword puzzles), or in two cases, practiced a mathematics skill, such as counting.

Experimental Design

Experimental design was coded categorically for each study as either a randomized controlled trial if participants were randomly assigned to conditions, or a quasi-experimental study if the design was based on preexisting groups.

Publication Status

Publication status was coded categorically for each study as either published if they were published in a peer-reviewed journal, or unpublished if they were (i) an unpublished dissertation, (ii) a study that had yet to be submitted for review, or (iii) a study that the authors had not and did not plan to submit for review.

Analytical Approach

Effect sizes were calculated using the Comprehensive Meta-Analysis Program (Borenstein et al., 2005). Effect sizes were estimated using Hedges's *g*, a standardized mean difference statistic (experimental group vs. control) that corrects for small sample bias (Hedges, 1981). Hedges's *g* values were calculated using mean differences scores and reflect the estimated mean difference in pre-post performance by the experimental group

Table 2

Characteristics of Studies Included in the Meta-Analysis on the Effect of Spatial Training on Math

Study author and year	*N*	Age (years)	Country	Description of spatial training	Duration (min)	Design	Control group	Math outcomes	Transfer	Use of concrete materials
Akayuure et al., 2016	94	20	Ghana	Origami integrated into geometry lessons. Students constructed and discussed the geometrical properties of origami models.	480	QED	Business as usual	Geometric Knowledge for Teaching (GKT) Test score	Near transfer	Yes
Arici & Aslan-Tutak, 2015	184	15	Turkey	Origami integrated into geometry lessons	540	QED	Business as usual	Geometry achievement test and geometric reasoning test	Near transfer	Yes
Baldwin, 1984	88	10–12	USA	Spatial orientation and spatial visualization worksheets	500	QED	Business as usual	Mathematics subtests (concepts and problem solving) of Iowa Tests of Basic Skills	Far transfer	No
Boakes, 2009	56	12–13	USA	Simple and more challenging origami lessons emphasizing math language and geometry principles	300	QED	Business as usual	National Assessment of Educational Progress (NAEP) Geometry Test score	Near transfer	Yes
Bower et al., 2021	144	3.5	USA	Three training groups. For all groups, training included a shape parade followed by spatial assembly training. This was presented digitally on a tablet using an app. The games and feedback were presented by a digital character. 1. Limited scaffolding condition where minimal scaffolding was provided to participants and only "bare bones" feedback was given. 2. Gesture feedback condition where the digital character used gesture to outline shapes in the shape parade, and during feedback showing where to place pieces. 3. Spatial language condition where the digital character used spatial language to name shapes and their properties in the shape parade, and to describe where pieces should go during spatial assembly feedback.	50	RCT	Business as usual	Shape Identification, Applied Problems subtest of Woodcock Johnson Tests of Achievement, and subtest of TEMA-3	Near transfer (Shape Identification) and far transfer (all other measures)	Group 1: No Group 2: No Group 3: No
Bower et al., 2020	187	3.5	USA	Three training groups. For all groups, training included a shape parade followed by spatial assembly training using physical materials. 1. Limited scaffolding condition where minimal scaffolding was provided to participants and only "bare bones" feedback was given. 2. Gesture feedback condition where the experimenter used gesture to outline shapes in the shape parade, and during feedback to show where to place pieces. 3. Spatial language condition where the experimenter used spatial language to name shapes and their properties in the shape parade, and to describe where pieces should go during spatial assembly feedback.	50	RCT	Business as usual	Shape identification, Applied Problems subtest of Woodcock Johnson Tests of Achievement, and subtest of TEMA-3	Near transfer (Shape Identification) and far transfer (all other measures)	Group 1: Yes Group 2: Yes Group 3: Yes
Cheng & Mix, 2014	58	6–8	USA	Mental transformation (mental rotation) where participants are given two shapes and asked what shape would be formed by joining them. They use physical shapes to check their answers.	40	RCT	Active Control: Children completed crossword puzzles.	Arithmetic test assessing number facts, multidigit questions and missing term problems	Far transfer	Yes
Cheung et al., 2020	62	6–7	USA	Computer-based mental rotation training: Children completed trials of the Children's Mental Transformation Task (CMTT). Feedback was given.	52.5	QED	Active Control: Literacy training. Participants choose which of four written words matched a verbally presented word.	Arithmetic test assessing missing term problems and Calculations subtest of Woodcock Johnson Tests of Achievement	Far transfer (both measures)	No

(*table continues*)

Table 2 (*continued*)

Study author and year	*N*	Age (years)	Country	Description of spatial training	Duration (min)	Design	Control group	Math outcomes	Transfer	Use of concrete materials
Cornu et al., 2019	125	4–7	Luxembourg	Computer (tablet)-based intervention targeting a range of visuospatial skills including disembedding, symmetry, rotation, copying, tangrams, figure completion, identifying matching/differing shapes, shape closing, shape and line bisection, row completion, dot connection.	400	QED	Business as usual	Nonsymbolic comparison, free counting, higher counting further counting, number naming, Number comparison, arithmetic, and number line estimation	Far transfer (all measures)	No
Freer, 2016	34	6–8	USA	Mental transformation (mental rotation) where participants are given two shapes and asked what shape would be formed by joining them. They use physical shapes to check their answers.	40	RCT	Active Control: Children completed crossword puzzles.	Mathematic test including word problems, missing term problems, number facts, and calculation items	Far transfer	Yes
Freina et al., 2017	79	9–11	Italy	Diverse visuo-spatial training delivered both digitally and with concrete materials, for example, visual memory, movement and mental rotation, spatial perspective taking, mazes, symmetry, origami, understanding 2D projections of 3D spaces, Minecraft, games using room plans and photographs.	900	QED	Business as usual	Standardized math test (AC-MT 6 −11) including written number operations, identifying biggest decimal, deducing a number from words, ordering numbers, word problems	Far transfer (all measures)	Yes
Gilligan et al., 2019	250	7–9	UK	Two computer-based training groups. 1. Mental rotation training: Participants watched an instructional video on mental rotation or completed trials of a mental rotation task with feedback, that is, selecting which image matched a rotated image. 2. Spatial scaling training: Participants watched an instructional video on spatial scaling or completed trials of a spatial scaling task with feedback, that is, determining whether the position of a target was the same in two pictures at different scaling factors.	5	RCT	Active Control: Participants either watched an instructional video matching written words to pictures or actively selected which of two pictures matched a written word. Feedback was given.	Missing term problems, number line estimation, geometry symmetry test, and geometry shape test	Far transfer (all measures)	Group 1: No Group 2: No
Hawes et al., 2015	58	6–8	Canada	Computer (tablet)-based mental rotation intervention: Three different tasks were completed including rotating shapes to complete a picture, matching shapes that are rotated, and distinguishing rotated and mirror images.	315	RCT	Active Control: Participants completed literacy training (identified correctly spelled words, separated verbs from nouns, recalled object names, completed words, sentences and crosswords, book reading, matched images to sentences).	Nonverbal exact arithmetic task and arithmetic test assessing missing term problems	Far transfer (both measures)	No
Hawes et al., 2017	59	5–7	Canada	Classroom-based intervention with quick challenge activities and full lessons. These targeted mental visualization, mental transformation, symmetry, 2D/3D, mental rotation, visual memory, spatial language comprehension.	2,820	QED	Active Control: Intervention was based on using an enquiry-based approach to environmental science.	Symbolic magnitude comparison test, nonsymbolic magnitude comparison test, and number knowledge test	Far transfer (all measures)	Yes
Honor, 2020	77	10–11	UK	Computer-based model construction requiring mental rotation and perspective taking. Participants were required to build a digital block construction given a 2D top and side view of a model.	150	RCT	Active Control: Participants completed an activity booklet of word tasks (i.e., brain teasers).	Number line estimation (bounded) and number line estimation (unbounded)	Far transfer (both measures)	No
Hung et al., 2012	99	10–11	Taiwan	Two training groups. 1. Participants completed digital games including online tangrams, rotation and manipulation of 3D objects, and a treasure hunt task requiring spatial memory. 2. Participants completed activities with physical materials including completing tangrams, manipulating blocks and clay, and observing geometric figures from different perspectives.	NA	RCT	Business as usual	Math Computer Adaptive Test (MCAT), an adaptive test of math achievement	Far transfer	Group 1: No Group 2: Yes

(*table continues*)

Table 2 (*continued*)

Study author and year	*N*	Age (years)	Country	Description of spatial training	Duration (min)	Design	Control group	Math outcomes	Transfer	Use of concrete materials
Krisztián et al., 2015	25	12	Hungary	Origami: cut out, folded and decorated origami shapes.	600	RCT	Business as usual: Two control groups were included in this study. The meta-analysis only included the group with math learning difficulties.	Numerical ability test including items assessing addition, subtraction, multiplication, and division	Far transfer	Yes
Lowrie et al., 2017	186	11–12	Australia	Training included sessions on mental rotation, spatial orientation, spatial visualization, and integration of each of the aforementioned skills. Physical materials were used.	1,200	QED	Business as usual	MathT test which includes both geometry-measurement and number items. The items were taken from Australia's National Assessment Program (NAPLAN).	Far transfer	Yes
Lowrie et al., 2019	641	12–14	Australia	Training focused on mental rotation, spatial orientation and spatial visualization. Tasks included completing 2D and 3D rotations and using rotational language, representing and mapping different perspectives, constructing and deconstructing shapes (e.g., nets and geometrical shapes), origami and translations.	675	QED	Business as usual	MathT test including geometry and measurement (12 item test), number and algebra (6 item test), statistics and probability (4 item test)	Near transfer (Geometry) and far transfer (all other measures)	Yes
Lowrie et al., 2019	327	10–12	Australia	Training was both digital and physical. It focused on mental rotation, spatial orientation and spatial visualization using topics including reflection, symmetry, paper folding and cutting, nets of solids, hidden blocks, and cross-sectioning of 3D objects.	360	QED	Business as usual	Geometry items, word problems, nongeometry based graphic problems	Near transfer (Geometry) and far transfer (other measures)	Yes
Mix et al., 2020a	134	7	USA	Two training groups. 1. Spatial visualization training including a part–whole object completion task (where participants are given two shapes and asked what shape would be formed by joining them), a rotation task with 2D images and tangram puzzles. Physical materials were provided for checking answers. 2. Form perception/visuospatial working memory (VSWM) training including trials of a VSWM task (participants were shown a display of objects on a grid for a limit time and were then asked to mark on the grid they had been displayed), a Corsi block tapping task, and a figure copying task (where participants copied a set of line drawings).	180	RCT	Active Control: Participants completed language activities using iPad devices. These games included crossword puzzles, rhyming words games, and word search puzzles. Feedback was given.	Composite score of symbol grounding including place value (comparing, ordering, and interpreting multidigit numerals), word problems from the test of Early Mathematical Ability–third edition [TEMA-3]), and number line estimation (0–100/1,000 lines). Composite score of symbol decoding including calculation (multistep arithmetic problems), missing term problems, and notational spacing (vertical calculation problems from the TEMA-3).	Far transfer (both measures)	Group 1: Yes Group 2: No
Mix et al., 2020b	124	12	USA	Two training groups. 1. Spatial visualization training including a part whole object completion task (where participants are given two shapes and asked what shape would be formed by joining them), a rotation task with 3D images and models, and tangram puzzles. Physical materials were provided for checking answers. 2. Form perception/visuospatial working memory (VSWM) training including trials of a VSWM task (participants were shown a display of objects on a grid for a limit time and were then asked to mark on the grid they had been displayed), a Corsi block	180	RCT	Active Control: Participants completed language activities using iPad devices. These games included crossword puzzles, rhyming words games, and word search puzzles. Feedback was given.	Composite score of tasks measuring symbol grounding including place value (rational numbers subtest from the Comprehensive Mathematics Abilities test [CMAT]), word problems (problem-solving subtest from the CMAT), and number line estimation (placing fractions on a 0–1 number line). Composite score of tasks measuring symbol decoding	Far transfer (both measures)	Group 1: Yes Group 2: No

(*table continues*)

Table 2 (*continued*)

Study author and year	*N*	Age (years)	Country	Description of spatial training	Duration (min)	Design	Control group	Math outcomes	Transfer	Use of concrete materials
				tapping task, and a figure copying task (where participants copied a set of line drawings).				including calculation (multi-step arithmetic problems), algebra (algebra subtest from the CMAT), and notational spacing (algebra problems).		
Rodán et al., 2019	81	7–8	Spain	Computer-based mental rotation training: Participants selected which of two shapes could be rotated to fit into an outline. Feedback was given.	90	QED	Business as usual	Numerical aptitude test "N" from the Evaluación Factorial de las Aptitudes Intelectuales (EFAI-1). Participants manipulated numerical symbols and reasoned procedurally with information and quantitative relationships.	Far transfer	No
Sala et al., 2017	159	7.6	Italy	Two training groups using physical materials. 1. Mental translation and rotation: Participants completed trials of the Children's Mental Transformation Task by identifying which two physical shapes could be joined together to make a third target shape. 2. Mental translation: As described for group 1 but without rotation.	60	QED	Business as usual	Arithmetic test including math equations and missing term problems	Far transfer	Group 1: Yes Group 2: Yes
Schmitt et al., 2018	59	4.6	USA	Physical block play intervention where children were given varying levels of prompts across sessions, that is, simple prompts, components prompts.	245	RCT	Business as usual	Preschool Early Numeracy Skills Screener–Brief Version, shape recognition task, and mathematical language task	Far transfer (all measures)	Yes
Sundberg, 1994	32	11–14	USA	Spatial visualization training with concrete materials, for example, tangrams, cube construction, drawing 3D objects, geoblocks, geoboards, puzzles, drawing.	1,950	RCT	Active Control: Geometry lessons	3-Rs test of Mathematics Achievement and Quantitative Abilities. Level 14	Far transfer	Yes
Thompson, 2012	157	6–7	USA	Three spatial visualization training groups. 1. Physical manipulatives training: compare plane and solid figures; make complex objects; determine if objects can stack, slide, or roll; conclude how many faces, edges, and vertices are on a shape; compare figures to everyday items; recognize geometric figures in objects; cut out, assemble, label and count the faces, edges, and vertices of shapes. 2. Multimedia training: videos on combining shapes, solids, finding shapes, symmetry, plane-figures. 3. Both manipulatives and multimedia.	Missing	QED	Business as usual	North Carolina standard course of study mathematics competency goal three test	Far transfer	Group 1: Yes Group 2: No Group 3: Yes
Tillotson, 1984	102	11–12	USA	Mental manipulation of 2D/3D shapes and applying these skills to problem solving. This included mapping between 2D and 3D shapes, identifying properties of 3D shapes, rotations, mental paper folding, and translating verbal problems into pictures.	360	QED	Business as usual	Problem-solving inventory including analytical and spatial problem solving	Far transfer	Yes
Xu & LeFevre, 2016	84	3–5	Canada	Mental transformation: Participants identified which two physical shapes could be joined together to make a third target shape.	17.5	RCT	Active Control: Sequential training where participants identified the successor and predecessor of several target numbers	Rote counting, number identification, number ordering, number line, and number line with midpoint	Far transfer (all measures)	Yes

Note. RCT = randomized controlled trial; QED = quasi-experimental design.

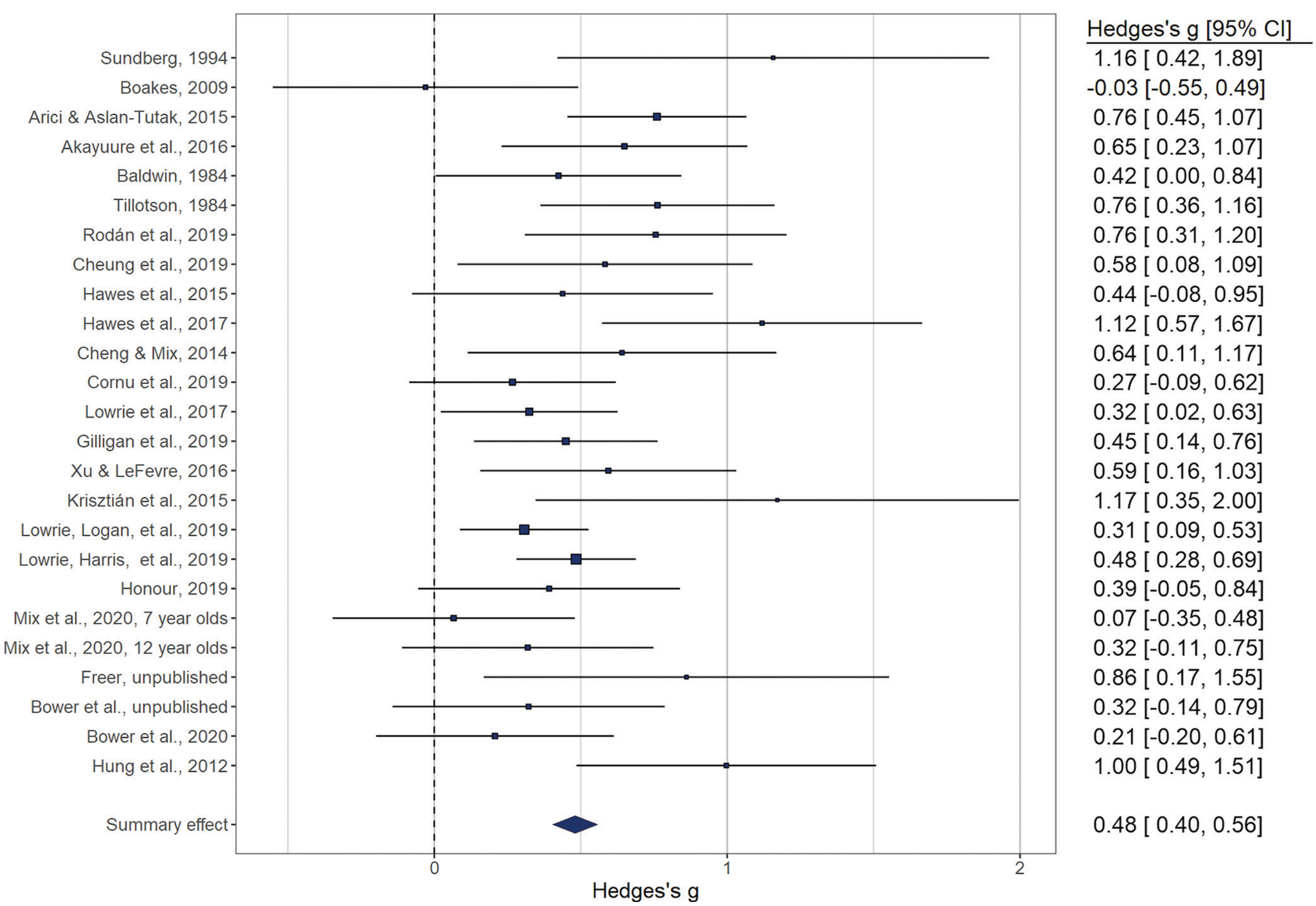
compared to the control group. All effects were coded so that positive values reflect positive effects of the experimental group over the control group.

Meta-analyses and metaregressions were completed in RStudio (Version 1.3.1056) using the robumeta package (Fisher & Tipton, 2015). As outlined in Table 2, several studies included more than one mathematics outcome measure (i.e., contributed more than one effect size to the meta-analysis) which resulted in dependence in the data. Previous studies have dealt with this dependence by creating an unweighted average effect size for any study that includes multiple outcome measures. However, this practice leads to reduced statistical power and a loss of information (Fisher & Tipton, 2015). Instead, we used robust variance estimation (RVE) which allows for the inclusion of multiple effects from a single study, and controls for the dependencies between these effects (Fisher & Tipton, 2015; Tipton & Pustejovsky, 2015). RVE calculates weights based on the correlated effects method and applies small sample corrections (Tipton, 2015). Following Tipton (2015), we set this correlation coefficient (Spearman's rho) at .8. We also completed sensitivity analyses to determine the effect of varying the value of the correlation coefficient from $\rho = .1$ to $\rho = 1.0$ (Tanner-Smith & Tipton, 2014). Sensitivity analysis is a form of quality control. It is used in meta-analysis to determine whether summary effects are robust to assumptions made during analysis, that is, to determine that the choice of correlation coefficient does not lead to substantial differences in the effect sizes reported. Finally, we reviewed the degrees of freedom for all analyses and flagged any results with fewer than four degrees of freedom as these should be interpreted cautiously due to increased risk of Type I error (Tanner-Smith et al., 2016).

Between-study heterogeneity (T^2) and between-study variance (I^2) were reported for all analyses. T^2 provides an estimate of the variance in true effect size and is measured using the same metric as the effect. For each main effect we also reported Prediction Intervals (PI) to give a range of the true study effects. These are calculated as mean effect $\pm 2(T)$ where T is Tau (Borenstein et al., 2017). In contrast, I^2 is a metric of how much between-study variation is attributable to systematic (true) variance as opposed to

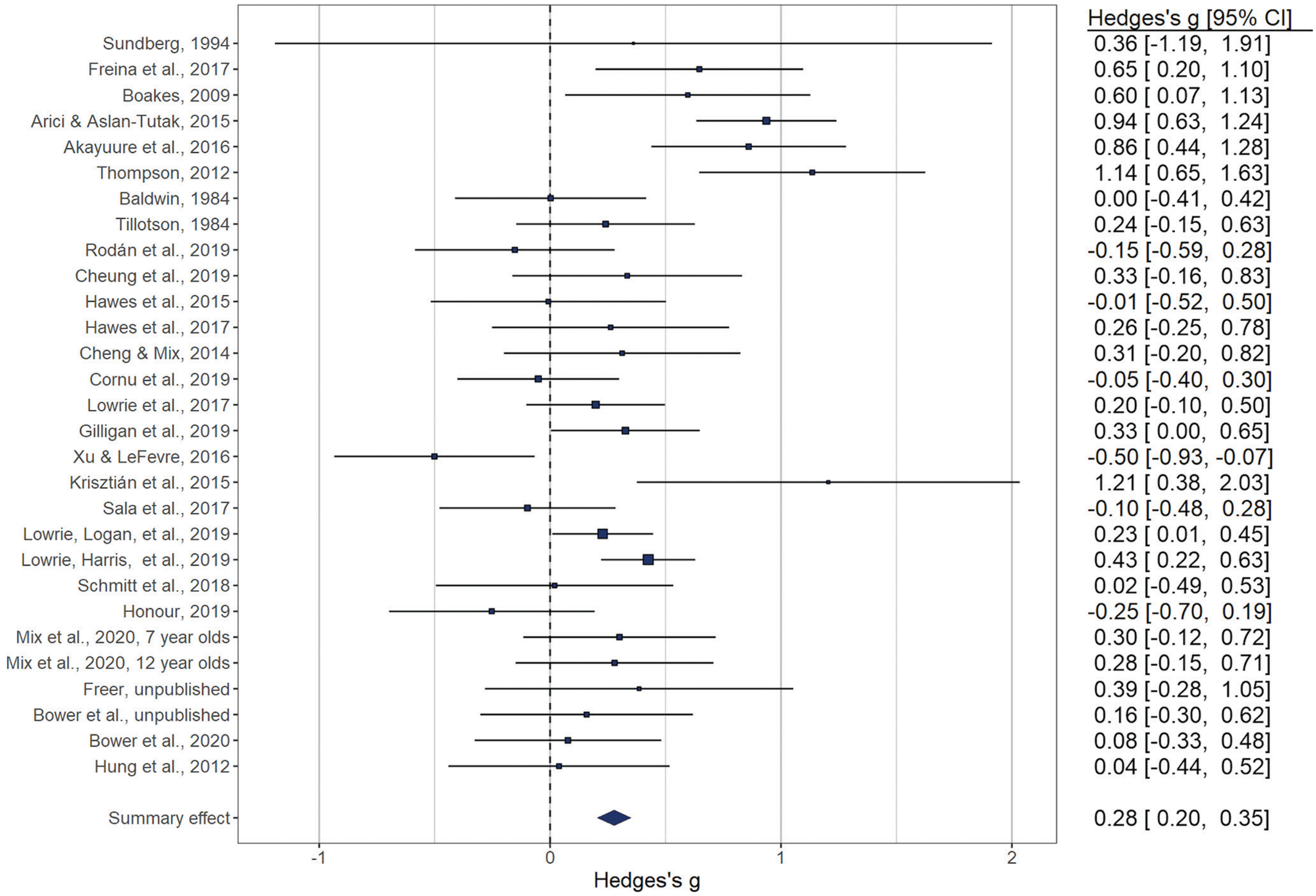
Figure 2

Forest Plot of the Effects of Spatial Training on Spatial Measures



Note. The diamond represents the overall average effect size and confidence interval (Hedges's *g*) and each square represents the effect size for each study (horizontal lines represent confidence intervals). Note that the effect sizes shown here are for illustrative purposes only and vary to a small degree from the RVE analysis. The average generated from the values shown in this plot does not take into account that some studies contributed multiple effect sizes. CI; Confidence Intervals. See the online article for the color version of this figure.

Figure 3
Forest Plot Showing the Effect Sizes of Spatial Training on Mathematics Outcomes



Note. The diamond represents the overall average effect size and confidence interval (Hedges's *g*) and each square represents the effect size for each study (horizontal lines represent confidence intervals). Note that the effect sizes shown here are for illustrative purposes only and vary (to a small degree) from the RVE analysis. The average generated from the values shown in this plot does not take into account that some studies contributed multiple effect sizes. CI; Confidence Intervals. See the online article for the color version of this figure.

random sampling error (Borenstein et al., 2005). Benchmarks have been set for values of I^2, such that values of 25%, 50% and 75% have been defined as low, medium and high ratios of variance, respectively (Higgins et al., 2003). Low I^2 values are favourable as they indicate that observed heterogeneity is predominantly due to random sampling and not due to study differences.

We first calculated the main effects of spatial training on a) spatial performance and b) mathematics performance using RVE. We then used regression models within the robumeta package to investigate moderator effects. Each moderator was entered into the model individually, that is, all moderators were not entered at the same time. For categorical moderators, we completed an omnibus significance test—the Hotelling's T^2 test for multiple-contrasts (Tipton & Pustejovsky, 2015)—executed using the Wald-test function in the clubSandwich package in R (Pustejovsky, 2017). To follow up on any significant effects, we compared effect sizes between categories. As all moderator analyses included fewer than 40 studies, the significance level of the *F*-test was adjusted to .01 (Tanner-Smith et al., 2016). Studies missing information for one of the moderators were excluded from the analysis for that moderator.

For reporting purposes, forest plots were generated for the main effect of spatial training on both spatial and mathematics performance. The effects reported in these plots are unweighted average effect sizes for each study and are shown for illustrative purposes only. These averages differ from the values generated through the main analyses because the plots do not take dependencies into account.

Publication Bias and Small Study Effects

In addition to including publication status as a moderator variable, we took several other precautions to guard against publication bias and small study effects. To ensure that the assumption of independence was met, we reviewed all studies by the same author to ensure that no two studies were based on the same sample. As

Table 3
Results of Moderator Analyses for the Effect of Spatial Training on Mathematics Outcomes

			Heterogeneity					Treatment effects			
Moderator	*m*	*k*	I^2	T^2	*F*	β (Hedges's *g*)	*SE*	95% CI	*t*	*df*	*p*
Continuous variable moderators											
Age (years)	29	89	63.54	.07		.051	.01	[0.02, 0.08]	3.80	7.36	.006
Training dosage (minutes)	27	84	69.24	.09		.000	.00	[−.00, 0.00]	1.12	2.52	.36
Spatial gains (Hedges's *g*)	25	76	69.21	.09		.303	.24	[−0.22, 0.83]	1.28	10.8	.227
Categorical variable moderators											
Transfer distance within math	29	89	64.75	.08	11.1					6.48	.01
Far transfer	26	77				.203	.07	[0.06, 0.34]	3.01	24.02	.006
Near transfer	7	12				.653	.12	[0.34, 0.97]	5.49	4.65	.003
Training delivery	29	89	67.25	.08	10.9					21.5	.003
Concrete	20	45				.416	.08	[0.25, 0.58]	5.29	17.10	< .001
Nonconcrete	13	44				.052	.09	[−0.14, 0.24]	.60	10.70	.562
Type of control group	29	89	72.07	.11	1.26					19.9	.275
Business as usual	18	56				.334	.09	[0.14, 0.53]	3.59	16.50	.002
Active control	11	33				.180	.10	[−0.05, 0.41]	1.78	9.76	.107
Posttest timing	26	82	63.19	.07	.87					10.5	.446
Immediate	5	16				.200	.20	[−0.36,0.76]	1.01	3.80	.374
Between 1 and 7 days	9	34				.126	.09	[−0.08, 0.33]	1.42	7.82	.195
8 days or longer	12	32				.298	.09	[0.11, 0.49]	3.43	10.30	.006
Experimental design	29	89	71.14	.10	1.93					25.6	.176
RCT	14	52				.174	.09	[−0.03, 0.37]	1.90	12.6	.081
Quasi-experimental	15	37				.362	.10	[0.15, 0.58]	3.66	13.7	.002
Publication type	29	89	73.13	.11	.06					9.4	.813
Published	22	71				.269	.08	[0.11, 0.43]	3.50	20.45	.002
Unpublished	7	18				.317	.18	[−0.13, 0.76]	1.75	5.83	.131

Note. These results are based on RVE moderator analyses run separately for each moderator of interest. k = number of effect sizes; m = number of studies for each moderator; I^2 = true heterogeneity; T^2 = variation in effect sizes between studies; 95% CI = 95% confidence interval; df = degrees of freedom. For the analyses of the continuous moderators (top of table), β values represent the slope coefficient estimates and are a measure of the effect (Hedges's g) of each moderator against the intercept; p = p-value of the t-tests that compare each moderator against the intercept. For the analyses of categorical moderators, β values indicate the effect size (Hedges's g) for each level of the moderator compared against zero; F = F-value associated with an omnibus Wald (HTZ) test, used to test the null hypothesis that the average effect size is the same across all levels of the moderator; p = p-value of HTZ Wald tests for overall moderator effects and p-value of t-tests that compare each level of each moderator against zero. Where the degrees of freedom are less than 4, there is a high chance of Type 1 error and results should be viewed as unreliable. This is the case for training dosage and immediate posttest timing.

described in Gunnerud et al. (2020), we investigated publication bias for our main effects in three ways. First, as noted above, we used moderator analysis to compare published to unpublished studies. Next, as recommended in Rodgers and Pustejovsky (2021), we examined funnel plot asymmetry using both effect-size level funnel plots to diagnose selective outcome reporting (disaggregated effect sizes), and study-level funnel plots to diagnose selective publication (aggregated effect sizes). Funnel plots are a graphical representation that show an effect size (x-axis) plotted against its standard error (y-axis). Cases with higher standard errors are plotted at the bottom of a funnel plot and the vertical line down the middle of the plot shows the average effect size. An asymmetrical funnel plot with more studies on the right of the vertical line would suggest bias, that is, selective reporting bias for funnel plots with disaggregated effect sizes, and publication bias for funnel plots with aggregated effect sizes. We used contour-enhanced funnel plots that additionally show bias for different levels of two-tailed p-values (Peters et al., 2008). We visually inspected funnel plots for asymmetry and tested this asymmetry statistically using robust Egger's Regression, that is, the precision-effect estimator with standard error (PEESE) and precision-effect test (PET) for disaggregated dependent effects (Rodgers & Pustejovsky, 2021). As a final measure of publication bias, we completed trim and fill analysis to estimate the number of missing studies in our main analyses and the potential effect that these studies would have on the reported effect sizes (Duval & Tweedie, 2001). The outcome of these tests for publication bias are presented in the next section.

Data Availability and Acknowledgment of Preregistration

We acknowledge that this study was not preregistered. All data, supplementary material, and annotated analyses with the accompanying R code can be accessed on our OSF page: https://osf.io/8yn7m/

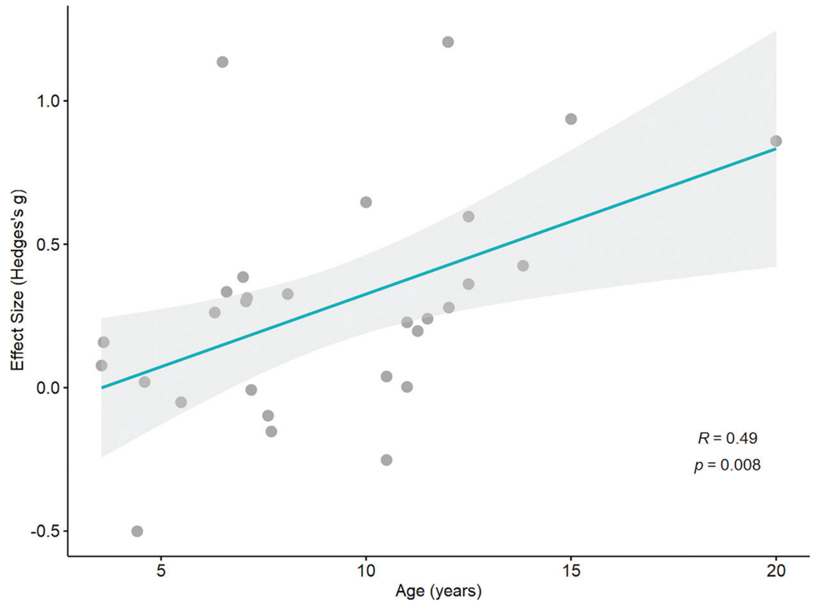
Results

Descriptive Information

The main meta-analysis on the effects of spatial training on mathematics included 89 mathematics outcomes taken from 29 studies. In total, 3,765 participants (Experimental N = 2,403; Control N = 1,362) were included in the meta-analysis. Across studies, the mean number of mathematics outcomes reported per study was 3.07 (min = 1, max = 9). Characteristics of each of the studies are presented in Table 2. The effect size for the meta-analysis

Figure 4

Correlation Between Age (Years) and the Effects of Spatial Training on Mathematics Outcomes



Note. This is a zero-order correlation between the average effect size of spatial training on mathematics outcomes per study. This correlation is not based on the RVE moderator analysis; instead if an individual study included three math outcomes, the average effect across all three outcomes was calculated. See the online article for the color version of this figure.

investigating the effect of spatial training on spatial outcomes was based on 25 studies with 72 effects, and a total sample size of 3,311 (Experimental $N = 2{,}109$; Control $N = 1{,}202$). The mean number of spatial outcomes reported per study was 2.88 (min = 1, max = 12). Characteristics of the studies are available as supplementary material on our OSF page: https://osf.io/8yn7m/

Overall Effect of Spatial Training on Spatial Outcomes

The overall effect of spatial training on spatial reasoning was $g = .521$, $SE = .06$, 95% CI [.41, .60], PI [−.03, 1.07], $p < .001$. Between-study variability, T^2, was estimated to be .08, with approximately 65% of the variance ($I^2 = 65.49$) attributable to systematic/true variance as opposed to random error. Varying the assumed correlation between the within-study effect sizes from .1 to 1, had no impact on the overall effect, including standard error, and minimal impact on estimates of heterogeneity, that is, correlation of .1; $I^2 = 64.96$, $T^2 = .08$; correlation of 1; $I^2 = 65.64$, $T^2 = .08$. Figure 2 shows a forest plot of the average effect size for each study as well as the overall effect. As noted earlier, this plot was generated for illustration purposes and is based on an average effect for each study. Note there was some evidence of publication bias (see Appendix B for details or visit https://osf.io/8yn7m/). An adjusted effect size of .49 was calculated and reported hereafter as the most accurate estimate of the effects of spatial training on spatial outcomes.

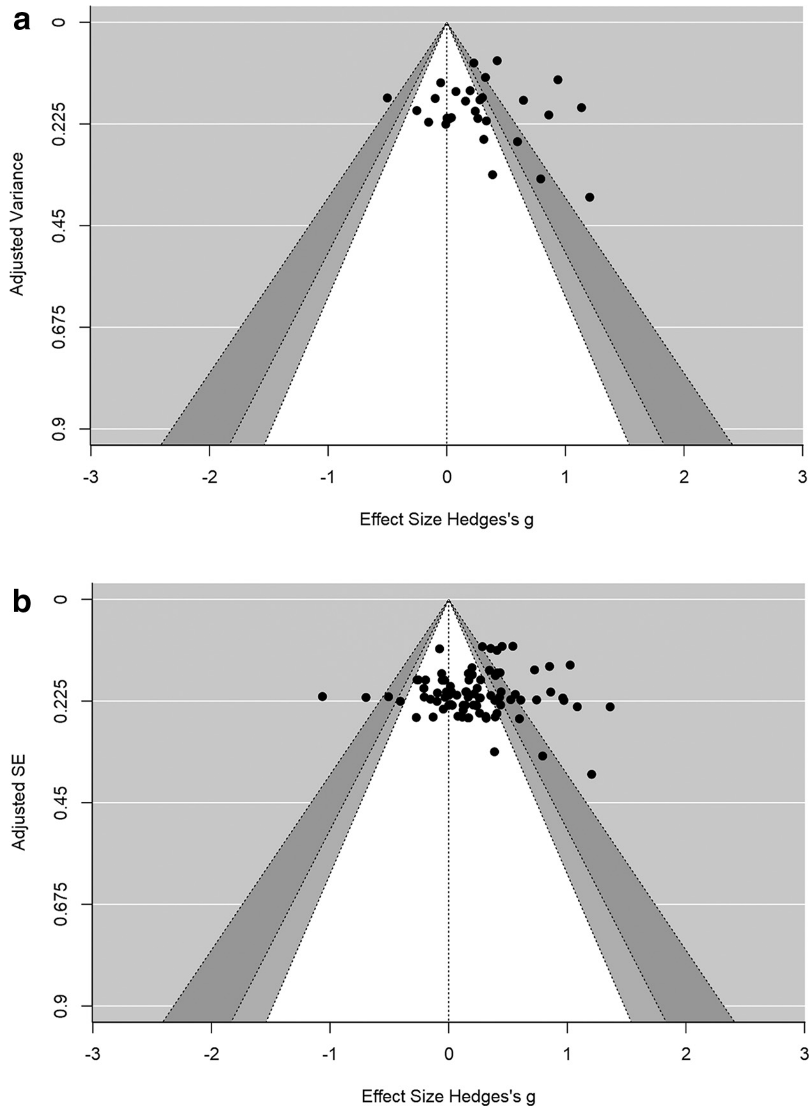
These results replicate the meta-analytic findings of Uttal et al. (2013), who found a comparable effect of spatial training on spatial reasoning ($g = .47$). Note that the present study includes only four studies reported in the Uttal et al. (2013) meta-analysis. Moreover, while most studies in the Uttal et al. (2013) meta-analysis were conducted with adolescents and adults, all but one of the studies in our analysis were carried out with children. Thus, our findings replicate and extend Uttal et al.'s findings that showed spatial thinking is a highly malleable construct. This finding satisfies a critical prerequisite in the effort to improve mathematical performance through spatial training.

Overall Effect of Spatial Training on Mathematics Outcomes

The overall effect of spatial intervention on mathematics outcomes was $g = .279$, $SE = .07$, 95% CI [.14, .42], PI [−.37, .93], $p < .001$. Between-study variability, T^2, was estimated to be .11, with approximately 72% of the variance ($I^2 = 72.22$) attributable to systematic/true variance as opposed to random error. Varying the correlation between the within-study effects from .1 to 1 had no impact on the overall effect of *SE*. This adjustment led to small differences on estimates of heterogeneity, that is, correlation of .1; $I^2 = 71.86$, $T^2 = .11$; correlation of 1; $I^2 = 72.32$, $T^2 = .11$. A forest plot summarizing the effects of spatial training on mathematics outcomes across individual studies is shown in Figure 3. As noted earlier, this plot was generated for illustration purposes and is based on an average effect for each study. This is why there is a slight discrepancy between the effect size reported above and throughout the article ($g = .28$) and the overall effect size reported in Figure 3 ($g = .27$).

Figure 5

(a) Study Level, Contour-Enhanced Funnel Plot of Aggregated Effect Sizes for Studies Investigating the Effects of Spatial Training on Mathematics Outcomes, and (b) Effect Size Level, Contour-Enhanced Funnel Plot of Disaggregated Effects Sizes for Studies Investigating the Effects of Spatial Training on Mathematics Outcomes



Note. Key for color codes: White indicates $.10 < p < .05$; light gray indicates $.05 < p < .01$; Dark gray indicates $p < .01$.

Moderator Effects

The results of the moderator analyses are reported in Table 3. Of the three continuous variables examined, only age was a statistically significant moderator. Figure 4 presents a scatterplot of the relation between age (years) and the average effect of spatial training on mathematics outcomes per study. Contrary to the sensitive period hypothesis, there was a positive association between age and the effects of spatial training on mathematics, $r(27) = .49$, $p = .008$. This finding suggests that transfer of gains from spatial training to mathematics increases as children develop.

Of the six categorical variables examined, two emerged as statistically significant moderators. The first was training delivery (Concrete vs. Nonconcrete). Training that included concrete materials led to larger gains in mathematics (mean effect (g) = .416) compared to training that had no concrete component (mean effect (g) = .052) ($\beta = .36$; $p = .003$). The second significant moderator was transfer distance within mathematics. On average, transfer for mathematics measures that were highly overlapping with the spatial training exercises (near math transfer) was associated with larger effect sizes (mean effect (g) = .653) than for mathematics measures with less overlap (far math transfer; mean effect (g) = .204) ($\beta = .45$; $p = .014$).

Publication Bias

The effects of spatial training on mathematics performance were similar for published studies (mean effect (g) = .269) compared to unpublished studies (mean effect (g) = .367) (β = .05; p = .810) (see Table 3). A Wald test confirmed no significant effect of publication type as a moderator, F (9.40) = .06, p = .813, I^2 = 73.13, T^2 = .11. To further assess publication bias, we first generated a contour-enhanced funnel plot of aggregated effect sizes, that is, one effect size per study. Visual inspection of the plot suggested funnel plot symmetry (Figure 5a) and the PEESE test was not statistically significant (p = .253). To diagnose selective outcome reporting, we also generated a contour-enhanced funnel plot of disaggregated effect sizes. Visual inspection of the plot suggested funnel plot symmetry (Figure 5b) and the PET test was not statistically significant (p = .433). Finally, using the trim and fill method we also found no indications of missing studies given the funnel plot asymmetry. Thus, our results do not appear attributable to significant publication bias or small-study effects.

Discussion

In this study, we addressed the question of whether and to what extent spatial training transfers to mathematics performance. Although decades of evidence from behavioral and brain science have demonstrated close connections between spatial and mathematical thinking, establishing a causal relation has been met with mixed effects. For the first time, the present study reports on a comprehensive and detailed meta-analysis of the effects of spatial training on mathematics. Our results indicate that spatial training provides a highly effective means for improving spatial skills (g = .49) as well as showing significant transfer to mathematics performance (g = .28). Thus, spatial training appears to benefit performance in both spatial and mathematical domains. However, our results also indicated moderate to large between-study heterogeneity in effects, suggesting the need to better understand the specific conditions under which transfer occurs.

Age, use of concrete manipulatives, and type of transfer (i.e., how closely the mathematical outcome measures aligned with the training) all significantly moderated the effects of spatial training on mathematics. In general, as the age of participants increased from 3–20 years, the effects of spatial training also increased in size. Studies that used concrete materials (e.g., blocks, puzzle pieces, paper folding) to train spatial thinking were more effective than studies that did not use concrete materials (e.g., computerized training). In terms of transfer, larger transfer effects were observed on mathematics outcomes more closely aligned to the training program (i.e., near math transfer) when compared to outcomes more distally related to the spatial training provided (i.e., far math transfer). On average, the effect size of spatial training on far mathematics transfer outcomes was .20 compared to a notably larger effect on near mathematics transfer outcomes, .65. None of the other moderator variables examined (training dosage, spatial gains, posttest timing, type of control group, experimental design, publication status) were significant. Moreover, analyses of publication bias and selective outcome reporting were nonsignificant. Overall, our results support prior research and theoretical claims that spatial training may be an effective means for enhancing mathematical understanding and performance. As discussed further below, the implications of these findings have potentially significant and far-reaching consequences. However, this study also makes it clear that this area of research is in its infancy and many questions remain; perhaps most notably, the need to better understand the mechanisms that underlie transfer.

Interpreting Effect Sizes

Our main effect of .28 indicates the average number of standard deviation units by which the intervention group outperformed the control group. Unfortunately, this and other effect size estimates are not inherently meaningful and do not readily translate into practice and application (Funder & Ozer, 2019; Hill et al., 2008). This raises the question of whether an effect size of .20 for far math transfer, or even a more liberal estimate of .28 that included near math transfer as well, represents a large enough effect to be practically meaningful. To address questions such as these, researchers have argued that effect sizes should be judged in comparison to empirically established benchmarks (Hill et al., 2008). In adherence to this guideline, we interpret the present findings in terms of how they compare to (i) normative student growth trajectories in mathematics, and (ii) effect size results from other mathematics and cognitive training interventions.

The overall effect of .28 reported in this study is comparable to the annual gains that occur in Grades 6–10 on U.S.-based nationally normed tests of mathematics (see Bloom et al., 2008; Hill et al., 2008). For younger grades (Kindergarten-Grade 5), where annual gains are much larger, our effect of .28 is comparable to about 25–50% of the annual gains that occur in mathematics. Against this benchmark, the effects observed in the present study appear quite large, especially considering how short the training durations were (ranging from 5 mins to 44 hrs). However, caution is warranted as these benchmark estimates are based solely on U.S. data and from standardized achievement tests. Our sample included studies from around the world, though predominantly from White, educated, industrial, rich, democratic (or WEIRD) samples, and included mathematics measures that were a mix of researcher-developed and normed achievement tests. Whether the annual gains demonstrated by American students are typical and generalize to other populations remains unknown. Furthermore, larger effect sizes are expected to occur on researcher-developed math outcomes, where there is often a more explicit link between the intervention and outcome measures (Lipsey et al., 2012). Thus, the comparisons noted above likely represent an overestimation of the actual effect size. Nonetheless, against this benchmark, the gains observed in the present study appear to have practical significance.

Another way of contextualizing our findings is to compare the effect sizes we observed to those of other educational intervention studies. Historically, it has been reported that successful educational interventions tend to have effects between .25 and .50 standard deviations (e.g., Hattie, 2009; Hill et al., 2008; Lipsey & Wilson, 1993). Against this benchmark, the effect sizes we observed fall toward the lower end of the spectrum. However, more recent research suggests much more conservative effects of educational interventions on academic achievement (Cheung & Slavin, 2016; Lortie-Forgues & Inglis, 2019). For example, a meta-analysis by Cheung and Slavin (2016) found an effect size of .16 for randomized educational interventions and an effect size of .23 for nonrandomized quasi-experimental studies. Similarly, research from Lortie-Forgues and Inglis (2019) found substantially lower effect sizes for large-scale educational RCTs (i.e., .06).

Lortie-Forgues and Inglis also looked at the effects associated with math-specific interventions and found a similarly small overall effect (.04). The effects reported for this study compare favorably with these estimates. However, the wide range of effect sizes for educational interventions and differences in methodologies (e.g., large-scale RCTs vs. quasi-experimental designs) makes it difficult to situate the current effects with any level of precision.

Arguably, a more parsimonious approach involves comparing our results to other cognitive training studies. For this purpose, we turn to the working memory training literature; a large and mixed literature that, in terms of study design (including typical sample sizes), aims, and questions of interest, is highly similar to the spatial training literature. Although exceptions exist, the overall consensus is that working memory training, including visual-spatial working memory training, does not transfer to mathematics (e.g., see Melby-Lervåg et al., 2016; Sala & Gobet, 2020; Schwaighofer et al., 2015). For example, the results of two recent meta-analyses estimate the effect of working memory training on mathematics to lie somewhere between .06 and .12 standard deviations (Melby-Lervåg et al., 2016; Schwaighofer et al., 2015). Moreover, the meta-analysis by Schwaighofer et al. (2015) indicated the effect of working memory training on mathematics was not moderated by the type of working memory training, verbal versus visual-spatial working memory. Against this benchmark, our far mathematics transfer effect of .20 is roughly two to three times as strong. Interestingly, these findings also suggest that spatial visualization training may be a more effective means for improving mathematics than visual-spatial working memory training. This raises the important question, and one that we further expand on below, of why spatial training may be more optimally suited to transfer to mathematics than other forms of cognitive training.

Theoretical Implications

The present study helps advance our understanding of space-math relations in several regards. At a minimum, the present study provides 'proof of concept' that spatial training can transfer to mathematics. This finding has theoretical importance because it demonstrates a causal relation, shifting the focus from *whether* to *how* and *why* spatial training/instruction might facilitate math learning performance. Indeed, moving from broad implications to the specific, the results of the moderator analyses provide new insights into the conditions under which space-math transfer is most (un)likely to occur.

Age Effects

Our results indicated a positive association between age and the effects of spatial training on mathematics. In general, as participants' age increased, so too did the effects of training on math. This result runs contrary to the sensitive period hypothesis, in which earlier interventions are expected to have larger effects than later interventions (e.g., see Heckman, 2007). Why did we find evidence in the opposite direction?

There are several explanations, none of which are mutually exclusive. One possibility is that mathematical content becomes increasingly more spatial as one moves from basic to advanced levels of mathematics. Accordingly, the mathematics outcome measures used in older aged samples may include mathematics that is inherently spatial. Another possibility is that higher-level mathematics affords more opportunities to use and apply one's spatial visualization skills. This is consistent with research suggesting that spatial reasoning is more important for conceptualizing higher-level mathematics (e.g., see Mix & Cheng, 2012). Thus, as individuals age, they may encounter both a broader range of mathematical tasks, but also situations in which to use and apply their spatial visualization skills. Lastly, it is also plausible that as children age and their metacognition matures, they become increasingly aware of their spatial thinking and its potential use in solving mathematical problems. Taken together, these explanations are consistent with evidence that the association between spatial skills and mathematics is stronger in the higher grades (Burnett et al., 1979; Mix & Cheng, 2012; Vernon, 1950; Voyer et al., 1995). Indeed, the role and importance of spatial thinking for mathematics appears to strengthen over development, not weaken (Mix & Cheng, 2012). Moving forward, it is clear that much is to be learned about the role of age and development in uncovering the mechanisms that underly the space-math link.

Concrete Materials

The results of our moderator analyses revealed significantly better outcomes for training with concrete materials versus training that involved worksheets or digital practice only. Our corpus included various approaches to incorporating concrete materials. As noted above, Mix et al. (2020) provided three-dimensional block constructions that children could rotate and compare to the choice drawings. Other studies used paper folding exercises as their spatial training (e.g., Arici & Aslan-Tutak, 2015; Boakes, 2009). What these concrete approaches had in common was training based on perceiving and acting upon physical objects. It makes sense that acting on objects would support improvement in spatial skills, given that spatial relations are instantiated in physical space —whether by navigating one's body through space or moving objects relative to each other or to the viewer. This finding is also consistent with perception-action and embodiment theories (Barsalou, 2008; Glenberg, 2015; Pecher & Zwaan, 2005; Thelen & Smith, 1996), which hold that cognition is rooted in perceptual information generated by bodily movement and actions on objects. From an embodied cognition perspective, the spatial skill of mental rotation, for example, may develop from massive exposure to manipulating objects and observing the objects as they move through various orientations. Similarly, the spatial skill of remembering locations may develop from massive exposure to placing and locating objects in space. Our finding that spatial training was more successful when it included such exercises lends support to these accounts.

The use of concrete materials offers a potential explanation for previous inconsistencies in the literature. For example, Hawes et al.'s (2015) failure to replicate the findings of Cheng and Mix (2014) may be due, in part, to differences in training delivery. Although both studies trained children's spatial transformation skills, Hawes et al.'s (2015) training used a computerized approach while Cheng and Mix (2014) used concrete materials. Cheng and Mix (2014) found evidence of transfer to missing-term problems ($3 + \_\_ = 4$), while Hawes et al. (2015) did not. Taken together, our findings suggest that the use of concrete materials may be an important variable to consider when designing and trying to understand the effects of spatial training on mathematics. Moving forward, research is needed that directly compares the effects of using or not using concrete materials in the delivery of spatial interventions.

Transfer Effects

Larger transfer effects were observed for near mathematics transfer outcomes than far. This finding suggests that transfer depends on the degree to which the training and outcome measures are aligned and recruit similar cognitive processes. This finding may seem rather obvious, but as we noted earlier, attempts to show that spatial skills are differentially related to different aspects of mathematics have not been successful (Mix et al., 2016; Xie et al., 2020). For example, across a wide variety of studies and age-groups, prior research indicates roughly equal associations between overtly spatial aspects of mathematics, such as geometry, and seemingly less spatial aspects of mathematics, such as basic numerical skills (Hawes & Ansari, 2020; Mix & Cheng, 2012; Xie et al., 2020). However, these results are based on correlational studies, which may mask the time-sensitive nature of these associations. By relying on causal research designs, the present study provides new evidence that space-math relations might be more dependent on task-specific shared processes and strategies than previously indicated.

There are several ways in which training-outcome alignment may have impacted the extent of transfer observed. One possibility is that gains in spatial visualization performance reflect authentic changes to one's spatial cognition (specific or general as these changes may be). These changes, in turn, may prove useful when performing mathematical tasks that recruit the same processes. For example, several studies found large gains in students' geometry performance following spatial visualization training (e.g., see Lowrie, Logan, et al., 2019; Lowrie, Harris, et al., 2019). Presumably, both domains shared the need to perform similar mental operations (e.g., visualizing and transforming objects).

However, another possibility is that transfer occurs—not necessarily due to changes in spatial cognition—but due to changes in strategy use. The closer the training resembles the types of mathematics problems under investigation, the more likely one is to recognize strategies and reasoning employed in the spatial context as useful to the mathematical context. Indeed, such priming has previously been hypothesized as one reason why transfer may be observed from space to mathematics, despite limited evidence of change to one's spatial reasoning abilities (e.g., see Cheng & Mix, 2014). This difference might help explain why we failed to obtain dosage effects. Even extremely brief training interventions may be effective if they afford the learner new spatial insights and strategies to be used during mathematics problem solving. For any given mathematics problem, there are often myriad ways of both structuring and solving the problem; with strategies and solutions varying from primarily verbal and formulaic to primarily visual and spatial (e.g., the Pythagorean theorem can be solved formulaically but also through pure visual-spatial proofs). In short, spatial interventions may "prime" individuals to recruit spatial strategies in mathematics problem solving rather than (or in addition to) improving the spatial processes themselves.

One such strategy is spatial visualization. On the surface, there are subdomains of mathematics for which spatial visualization strategies are overtly useful, for example, geometry. However, the same spatial visualization processes that serve certain geometry problems, may also serve arithmetic or other less obviously spatial mathematical tasks (including other aspects of geometry). There is evidence that spatial visualization plays a role in how children mentally organize and model arithmetic problems (Hegarty & Kozhevnikov, 1999; Huttenlocher et al., 1994; Laski et al., 2013). For example, addition and subtraction may be conceptualized and visualized as two sets coming together, taken apart, or recombined in a variety of different ways. This example highlights the possibility that the mental operations serving basic arithmetic are functionally equivalent to those serving more overtly spatial mathematical tasks, such as geometry and measurement. Said differently, at a very basic level, the mental operations and associated neural mechanisms that support the composition/decomposition and transformation of numbers (in the case of arithmetic) or objects (in the case of geometry), might be more similar than they first appear (e.g., see Hawes, Sokolowski, et al., 2019).

Moreover, the mental operations and strategies individuals use to solve mathematics problems change over time and with experience. Somebody with experience and fluency performing basic addition may no longer need to engage in spatial visualization processes to arrive at a solution. This might explain why some studies have shown evidence of transfer to arithmetic while others have not. If spatial abilities play a more important role in the learning of new and unfamiliar content, as many have proposed (e.g., Hawes, Moss, et al., 2019; Lowrie & Kay, 2001; Mix & Cheng, 2012; Uttal & Cohen, 2012), then we should expect to see larger training effects for novel versus familiar math outcomes, and in this regard, the shared processes that link space and math may present a moving target. This hypothesis has yet to be tested directly but suggests an important step moving forward.

The evidence above reveals a critical gap in the training literature; that is, the need to study and better understand the strategies and specific processes used by individuals as they solve mathematical tasks. No studies in our review were designed to address these gaps. Moreover, there was little evidence of theoretically guided decisions as to why transfer should occur on the selected math measures. Without explicit attention to these issues, we are left with little insight into the specific mechanisms that underly the space-math association.

Practical Implications

The present findings provide reason to be cautiously optimistic about the efficacy of spatial training to enhance mathematics learning and performance. In terms of recommendations for practice or policy, our findings suggest that spatial training may be most effective when embedded in practice. This approach resembles what others have referred to as "spatializing" the mathematics curriculum (e.g., Bruce et al., 2015; Casey & Fell, 2018; Newcombe, 2013). Rather than isolating and training spatial skills as something separate from mathematics, it is possible that development of spatial skills and their transfer to mathematics is best achieved in situ. Mathematics itself may offer fertile grounds in which to develop and practice a variety of spatial skills. It seems advisable that interdisciplinary teams of experts, including educators, cognitive scientists, and mathematicians, work together to better understand how and when spatial and mathematical thinking interact and potentially codevelop in practice.

Our findings also suggest the use of 'hands-on' concrete materials are more effective than spatial instruction that does not use concrete materials. However, the practical implications that follow from this are not as straightforward as they might appear. For example, there were several studies in which computerized training did result in gains in mathematics performance (Bower, 2021; Gilligan et al., 2019; Hung et al., 2012). As noted above, more research is needed to understand the reasons why, and under what conditions, concrete materials facilitate or fail to facilitate spatial and mathematics learning compared to approaches which do not use concrete materials. Moreover, with the advent of new computerized technologies, including innovative touch screen technologies, there are reasons to be cautious in adhering to the conclusion that concrete materials are associated with better learning outcomes than nonconcrete approaches.

Next Steps

Our ability to make practice and policy recommendations relies on the extent to which we understand when, why, and how spatial training transfer to mathematics. To quote Lewin (1951), "There is nothing more practical than a good theory" (p. 169). Indeed, our review of the literature suggests few efforts to test specific theories as to why spatial training is expected to transfer to math. Instead, the majority of studies (including our own), appeared to rationalize their training studies on the basis of previously reported correlations between spatial and mathematical thinking. We found no studies designed with the explicit intent to reveal how training influences the strategies and specific processes used by individuals to approach and solve mathematics problems. Moving forward, we highlight the need to be more specific in why transfer is expected. For example, if shared processing is theorized to underlie transfer, what might these shared processes be? Are they believed to be domain-general or domain-specific? On what mathematics measures would these shared processes have the largest impact and why? Does the degree of shared processing change over time and with experience with the math task in question? How does training influence the strategies, and in turn, shared processes, one uses to approach and solve the mathematics tasks in question? A better understanding of these underlying mechanisms could have important implications for instruction. For example, for some individuals, spatial instruction may be best directed at highlighting the role and potential use of spatial strategies during mathematical problem solving (i.e., "spatializing" mathematics). This may offer an additional way of seeing, conceptualizing, and reasoning through a problem. For other individuals, however, low spatial skill levels may be standing in the way of successful problem solving in the first place. In this case, it may make more sense to intervene at the level of spatial processing and target the development of spatial skills, prior to highlighting their use in practice. In short, more nuanced and theoretically guided spatial interventions are recommended moving forward.

An improved understanding of the mechanisms that underlie transfer may prove useful in explaining two of our more puzzling findings. That is, the absence of dose-response effects and failure to obtain evidence that spatial gains were associated with transfer effects. Priming and threshold effects offer possible reasons for observing these findings. As discussed above, one explanation for the observed transfer effects is that the training may have encouraged (primed) individuals to adapt or adopt new mathematical problem-solving strategies. For example, spatial training that involves practicing spatial transformation skills, such as transforming one shape into another shape, might encourage individuals to use a similar strategy when solving linear and area measurement problems (e.g., see Hawes et al., 2017). In this case, even a very short bout of spatial training may be enough to encourage the use of more effective mathematical problem-solving strategies. This is one reason we may not have observed a linear dose-response relationship.

Moreover, this account also suggests that gains in spatial thinking may not always be necessary to observe changes in mathematics performance. This may help explain why overall gains in spatial training may not have been linked to changes in mathematics performance. This finding may also be explained in part due to threshold effects. It is possible that spatial training improves mathematics up to a point, after which further training has less impact (Freer, 2017). However, achieving such a threshold will vary immensely from individual to individual. While a short amount of training and/or small gains are needed for some individuals to reach a certain threshold, more extensive training duration and gains in spatial thinking may be necessary for others to reach this same threshold. Future research efforts are needed to more formally test these conjectures. Doing so will require more concerted efforts to understand how spatial training is linked to changes in strategy recruitment and effective mathematics problem-solving. It will also require a better understanding of whether and how individual differences in one's baseline spatial skills (and strategy use) relate to the amount of transfer observed.

The majority of studies included in this meta-analysis had an explicit focus on training spatial visualization skills. While focusing on a single type of spatial skill makes the evidence more interpretable, it also leaves us with many unanswered questions. Are there other spatial skills that might also support mathematics performance? We believe the answer is yes. From the studies included in this meta-analysis there is some limited evidence showing that targeting spatial scaling, spatial perspective taking, and form perception skills may also be linked to gains in mathematics performance (Gilligan et al., 2019; Lowrie, Logan, et al., 2019; Mix et al., 2020). From a shared-processing perspective, there are reasons to theorize strong links between these spatial skills and specific types of mathematical reasoning. For example, spatial scaling may relate to how one thinks about multiplication through area models, understanding how to place and reason about numbers on a number line (e.g., zooming in and out of scale), and conceptualizing geometric invariance. These sorts of natural links between specific types of spatial and mathematical reasoning have scarcely been explored, and yet, potentially offer the largest opportunities for transfer. Given the importance of training-outcome alignment, there is reason to believe that these sorts of theoretically guided training approaches may yield even larger effects than those reported here.

The present study revealed several methodological concerns, suggesting important next steps moving forward. First, many of the studies used small sample sizes, suggesting the need for future training studies to include a priori power analyses and sample size justification. Second, not one study included a delayed follow-up posttest at least one month following training. Therefore, we know very little about the long-term effects of spatial training on mathematics. On the one hand, the effects of spatial training may be short-lived and "fade-out," as is typical of many cognitive and educational interventions (see Bailey et al., 2020). On the other hand, the effects may build over time as children become more strategic and successful at recruiting spatial processes when confronted with new and unfamiliar mathematics content. According to Bailey et al. (2017; 2020), interventions that target "trifecta" skills—those that are malleable, fundamental, and unlikely to develop in the absence of the intervention—hold the most promise for achieving longer lasting and impactful effects. Arguably, spatial skills meet these first two criteria, but whether the third criterion is met likely depends on the specific type of spatial intervention employed (e.g., domain-general spatial training vs. domain-specific spatial training). Moving forward, the emerging literature on fade-out effects may provide a useful framework for predicting whether and to what extent the effects of spatial training on mathematics are durable.

Third, only one study measured participants' motivation and training-related expectations. These and other training-related factors, including participants' awareness of the intervention's intent, are critical to control, as they may influence the ways in which members of the intervention and control groups respond to the experimental demands, including the amount of effort exerted during posttests (Green et al., 2019). Together, these issues represent major gaps in the literature and until systematically addressed, will continue to limit the reliability, reproducibility, and ultimately, inferences that can be made about the efficacy of spatial training on mathematics.

Conclusion

There are reasons to be both optimistic and skeptical about the future of space-to-math training studies. Perhaps the spatial training literature is destined for the same fate as the cognitive training literature, generally, and the working memory literature, specifically (e.g., see Gobet & Sala, 2020; Sala & Gobet, 2019). As researchers address the limitations above and more carefully controlled studies accumulate, the initial effects observed here may in fact represent an overestimation of the true effects. Alternatively, there may be something unique about the relation between spatial reasoning and mathematics that holds genuine promise (e.g., see Hawes, Moss, et al., 2019). The findings that space-math relations may be linked through a combination of domain-general, domain-specific, and strategy-level variables provide reasons to suspect that this may be the case (Mix, 2019). The same cannot so easily be said of other cognitive skills (e.g., spatial approaches to mathematical problem-solving are routine practice; working memory approaches are not). Indeed, the multiple ways in which space and mathematics are related may make their relations particularly difficult to study, while at the same time, offering myriad possibilities for intervention. We have arrived at a critical juncture in the effort to improve mathematics through spatial training. While the present study provides evidence that transfer is possible, continued efforts are now needed to determine how to make these effects stronger and more consistent. The time is ripe to capitalize on space-math relations, conducting studies that address the limitations above, while also systematically testing and revealing the mechanisms that underly these relations.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Akayuure, P., Asiedu-Addo, S. K., & Alebna, V. (2016). Investigating the effect of origami instruction on preservice teachers' spatial ability and geometric knowledge for teaching. *International Journal of Education in Mathematics. Science and Technology*, *4*(3), 198–209. https://doi.org/10.18404/ijemst.78424

Al-Khateeb, M. A., & Atoum, A. Y. (2008). The effect of cognitive style and training on social learning and visual spatial representation strategies on the development of mathematical and social problems skills. *Journal of Educational & Psychological Sciences*, *9*(4), 1–28.

*Arici, S., & Aslan-Tutak, F. (2015). The effect of origami-based instruction on spatial visualization, geometry achievement, and geometric reasoning. *International Journal of Science and Mathematics Education*, *13*(1), 179–200. https://doi.org/10.1007/s10763-013-9487-8

Arjmandnia, A. A., Sharifi, A., & Rostami, R. (2014). The effectiveness of computerized cognitive training on the performance of visual-spatial working memory of students with mathematical problems. *Journal of Learning Disabilities*, *3*(4), 6–24.

Asli, A. M., & Yarmohamadeian, A. (2012). The effect of metacognition training and perception of spatial relationship on improvement of mathematical operation in children with mathematics learning disabilities. *Journal of Clinical Psychology*, *4*(2), 61–70.

Azad, M. A., Abedi, A., & Yarmohammadian, A. (2013). The effectiveness of Visual-Spatial Skills Training on the Mathematical Performance of Students with Math-learning Disabilities. In 6th International Congress on Child and Adolescent Psychiatry, Tabtiz University of Medical Sciences.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

*Baldwin, S. L. (1984). *Instruction in spatial skills and its effect on math achievement in the intermediate grades* (Doctoral dissertation, University of Northern Colorado, Department of Curriculum and Instruction).

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328. https://doi.org/10.1080/19345740802400072

*Boakes, N. J. (2009). Origami instruction in the middle school mathematics classroom: Its impact on spatial visualization and geometry knowledge of students. *RMLE Online: Research in Middle Level Education*, *32*(7), 1–12. https://doi.org/10.1080/19404476.2009.11462060

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Borenstein, M., Rothstein, D., & Cohen, J. (2005). Comprehensive meta-analysis: A computer program for research synthesis (Version 3) [Computer software]. Biostat.

Boulter, D. R. (1992). *The effects of instruction on spatial ability and geometry performance* (Unpublished master's thesis). Queen's University.

Bower, C. A., Zimmermann, L., Verdine, B. N., Pritulsky, C., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). *Enhancing spatial skills of* *preschoolers from under-resourced backgrounds* (Unpublished manuscript). Temple University.

*Bower, C., Zimmermann, L., Verdine, B., Toub, T. S., Islam, S., Foster, L., Evans, N., Odean, R., Cibischino, A., Pritulsky, C., Hirsh-Pasek, K., & Golinkoff, R. M. (2020). Piecing together the role of a spatial assembly intervention in preschoolers' spatial and mathematics learning: Influences of gesture, spatial language, and socioeconomic status. *Developmental Psychology*, *56*(4), 686–698. https://doi.org/10.1037/dev0000899

Bruce, C., Sinclair, N., Moss, J. Z., Caswell, B., & The Spatial Reasoning Study Group. (2015). Spatializing the mathematics curricula. In B. Davis (Ed.), *Spatial reasoning in the early years: Principles, assertions and speculations* (pp. 85–106). Routledge.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence*, *3*(4), 345–354. https://doi.org/10.1016/0160-2896(79)90003-5

Burte, H., Gardony, A. L., Hutton, A., & Taylor, H. A. (2017). Think3d!: Improving mathematics learning through embodied spatial training. *Cognitive Research: Principles and Implications*, *2*(1), Article 13. https://doi.org/10.1186/s41235-017-0052-9

Campos, T. (2018). *Using visual thinking strategies to improve mathematics instruction* (Unpublished doctoral dissertation). University of Missouri.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380–400. https://doi.org/10.1037/a0031084

Casey, B. M., & Fell, H. (2018). Spatial reasoning: A critical problem-solving tool in children's mathematics strategy tool-kit. In K. Mix & M. Battista (Eds.), *Visualizing mathematics: The role of spatial reasoning in mathematical thought* (pp. 47–75). Springer Nature Publishing. https://doi.org/10.1007/978-3-319-98767-5_3

Casey, B. M., Lombardi, C. M., Pollock, A., Fineman, B., & Pezaris, E. (2017). Girls' spatial skills and arithmetic strategies in first grade as predictors of fifth-grade analytical math reasoning. *Journal of Cognition and Development*, *18*(5), 530–555. https://doi.org/10.1080/15248372.2017.1363044

*Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

*Cheung, C. N., Sung, J. Y., & Lourenco, S. F. (2020). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, *84*(7), 2000–2017. https://doi.org/10.1007/s00426-019-01202-5

*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), 1–21. https://doi.org/10.1080/10888691.2016.1276835

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25–32. https://doi.org/10.1016/S0160-2896(03)00061-8

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fias, W., & Fischer, M. H. (2005). Spatial representation of numbers. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 43–54). Psychology Press.

Fisher, Z., & Tipton, E. (2015). *Robumeta: An R-package for robust variance estimation in meta-analysis*. http://arxiv.org/abs/1503.02220

Freer, D. (2016). *Effects of spatial training on mathematics* (unpublished) Michigan State University.

Freer, D. (2017). *The threshold hypothesis applied to spatial skill and mathematics* (Unpublished doctoral dissertation). Michigan State University.

*Freina, L., Bottino, R., Ferlino, L., & Tavella, M. (2017). December Training of spatial abilities with digital games: Impact on mathematics performance of primary school students. In *International Conference on Games and Learning Alliance* (pp. 25–40). Springer.

Frick, A. (2019). Spatial transformation abilities and their relation to later mathematics performance. *Psychological Research*, *83*(7), 1465–1484. https://doi.org/10.1007/s00426-018-1008-5

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Galton, F. (1880). Visualised numerals. *Nature*, *21*(533), 252–256. https://doi.org/10.1038/021252a0

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2019). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science*, *22*(4), Article e12786

*Gilligan, K. A., Thomas, M. S., & Farran, E. K. (2019). First demonstration of effective spatial training for near transfer to spatial performance and far transfer to a range of mathematics skills at 8 years. *Developmental Science*, *23*(4), Article e12909. https://doi.org/10.1111/desc.12909

Glenberg, A. M. (2015). Few believe the world is flat: How embodiment is changing the scientific understanding of cognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *69*(2), 165–171. https://doi.org/10.1037/cep0000056

Gobet, F., & Sala, G. (2020). *Cognitive training: Methodological and substantive issues*. PsyArxiv. https://doi.org/10.31234/osf.io/vxzq

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., . . . Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement: Towards the Integration of Theory and Practice*, *3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, *146*(12), 1059–1083. https://doi.org/10.1037/bul0000301

Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Routledge.

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review*, *27*(3), 465–482. https://doi.org/10.3758/s13423-019-01694-7

*Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction*, *35*(3), 236–264. https://doi.org/10.1080/07370008.2017.1323902

*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*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001

Hawes, Z., Moss, J., Caswell, B., Seo, J., & Ansari, D. (2019). Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach. *Cognitive Psychology*, *109*, 68–90. https://doi.org/10.1016/j.cogpsych.2018.12.002

Hawes, Z., Sokolowski, H. M., Ononye, C. B., & Ansari, D. (2019). Neural underpinnings of numerical and spatial cognition: An fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neuroscience and Biobehavioral Reviews*, *103*, 316–336. https://doi.org/10.1016/j.neubiorev.2019.05.007

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences, USA, USA of the United States of America*, *104*(33), 13250–13255. https://doi.org/10.1073/pnas.0701362104

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176. https://doi.org/10.1016/j.intell.2005.09.005

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

*Honour, L. A. (2020). *Children's mental representation of number, their number line estimations and maths achievement: exploring the role of 3D mental rotation skills* [Doctoral dissertation, University of Southampton].

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448. https://doi.org/10.1038/nrn1684

*Hung, P. H., Hwang, G. J., Lee, Y. H., & Su, I. H. (2012). A cognitive component analysis approach for developing game-based spatial learning tools. *Computers & Education*, *59*(2), 762–773. https://doi.org/10.1016/j.compedu.2012.03.018

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, *123*(3), 284–296. https://doi.org/10.1037/0096-3445.123.3.284

Johnson, M. A. (1983). *Spatial visualization ability: Effects of long term practice and relationship to mathematical ability* [Unpublished doctoral dissertation]. The University of Arizona.

Kirby, J. R., & Boulter, D. R. (1999). Spatial ability and transformational geometry. *European Journal of Psychology of Education*, *14*(2), 283–294. https://doi.org/10.1007/BF03172970

Kohen, Z., Amram, M., Dagan, M., & Miranda, T. (2019). Self-efficacy and problem-solving skills in mathematics: The effect of instruction-based dynamic versus static visualization. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2019.1683588

*Krisztián, Á., Bernáth, L., Gombos, H., & Vereczkei, L. (2015). Developing numerical ability in children with mathematical difficulties using origami. *Perceptual and Motor Skills*, *121*(1), 233–243. https://doi.org/10.2466/24.10.PMS.121c16x1

Kyttälä, M., & Kanerva, K. (2013). Training mathematical skills in preschool children: Specific training of phonological, and visuospatial working memory. *NMI-bulletin*, *23*, 12–22.

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, *23*(1), 77–94. https://doi.org/10.1007/BF03173141

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *63*(10), 1953–1968. https://doi.org/10.1080/17470211003787619

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences*, *23*, 123–130. https://doi.org/10.1016/j.lindif.2012.08.001

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.).

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist*, *48*(12), 1181–1209. https://doi.org/10.1037/0003-066X.48.12.1181

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

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). Elsevier.

Lowrie, T., & Kay, R. (2001). Relationship between visual and nonvisual solution methods and difficulty in elementary mathematics. *The Journal of Educational Research*, *94*(4), 248–255. https://doi.org/10.1080/00220670109598758

*Lowrie, T., Logan, T., & Hegarty, M. (2019). The influence of spatial visualization training on students' spatial reasoning and mathematics performance. *Journal of Cognition and Development*, *20*(5), 729–751. https://doi.org/10.1080/15248372.2019.1653298

*Lowrie, T., Harris, D., Logan, T., & Hegarty, M. (2019). The impact of a spatial intervention program on students' spatial reasoning and mathematics performance. *Journal of Experimental Education*, 1–19. https://doi.org/10.1080/00220973.2019.1684869

*Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *The British Journal of Educational Psychology*, *87*(2), 170–186. https://doi.org/10.1111/bjep.12142

Lowrie, T., Logan, T., Harris, D., & Hegarty, M. (2018). The impact of an intervention program on students' spatial reasoning: Student engagement through mathematics-enhanced learning activities. *Cognitive Research: Principles and Implications*, *3*(1), Article 50. https://doi.org/10.1186/s41235-018-0147-y

Meir, B. H. (1988). *The differential effects of spatial visualisation training on male and female adolescents with learning and behaviour problems: its relationship to mathematics achievement* [Unpublished doctoral dissertation]. Columbia University Teachers College.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and*

Individual Differences, *26*, 141–152. https://doi.org/10.1016/j.lindif.2012.03.012

Mix, K. (2019). Why are spatial skill and mathematics related? *Child Development Perspectives*, *13*(2), 121–126. https://doi.org/10.1111/cdep.12323

Mix, K. S. (2010). Spatial tools for mathematical thought. In K. S. Mix, L. B. Smith, & M. Gasser (Eds.), *Space and language* (pp. 41–66). Oxford University Press.

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197–243. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y. L., Young, C. J., Hambrick, D. Z., & Konstantopoulos, S. (2017). The latent structure of spatial skills and mathematics: A replication of the two-factor model. *Journal of Cognition and Development*, *18*(4), 465–492. https://doi.org/10.1080/15248372.2017.1346658

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*(9), 1206–1227. https://doi.org/10.1037/xge0000182

*Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2020). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000494

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Möhring, W., Newcombe, N. S., Levine, S. C., & Frick, A. (2016). Spatial proportional reasoning is associated with formal knowledge about fractions. *Journal of Cognition and Development*, *17*(1), 67–84. https://doi.org/10.1080/15248372.2014.996289

Moss, J., Bruce, C. D., Caswell, B., & Flynn, T. Z. (2016). *Taking shape: Activities to develop geometric and spatial thinking*. Pearson Canada.

Mundy, J. F. (1980). *Spatial ability, mathematics achievement, and spatial training in male and female calculus students* (Unpublished doctoral dissertation). University of New Hampshire.

Mundy, J. F. (1987). Spatial training for calculus students: Sex differences in achievement and in visualization ability. *Journal for Research in Mathematics Education*, *18*(2), 126–140. https://doi.org/10.2307/749247

National Research Council. (2006). *Learning to think spatially: GIS as a support system in the K-12 curriculum*. National Academic Press.

Nelwan, M., & Kroesbergen, E. H. (2016). Limited near and far transfer effects of jungle memory working memory training on learning mathematics in children with attentional and mathematical difficulties. *Frontiers in Psychology*, *7*, Article e1384. https://doi.org/10.3389/fpsyg.2016.01384

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, *34*(2), 29–35.

Newcombe, N. S. (2013). Seeing relationships: Using spatial thinking to teach science, mathematics, and social studies. *American Educator*, *37*(1), 26

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* National Research and Development Centre for Adult Literacy and Numeracy, Institute of Education.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499968

Perry, P. C. (2013). *Influences on Visual Spatial Rotation: Science, Technology, Engineering, and Mathematics (STEM) Experiences, Age, and Gender* (unpublished doctoral dissertation). Notre Dame of Maryland University.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Pustejovsky, J. (2017). clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections. R package version 0.2.3. https://cran.r-project.org/web/packages/clubSandwich/index.html

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

*Rodán, A., Gimeno, P., Elosúa, M. R., Montoro, P. R., & Contreras, M. J. (2019). Boys and girls gain in spatial, but not in mathematical ability after mental rotation training in primary education. *Learning and Individual Differences*, *70*, 1–11. https://doi.org/10.1016/j.lindif.2019.01.001

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160.

Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences*, *23*(1), 9–20. https://doi.org/10.1016/j.tics.2018.10.004

Sala, G., & Gobet, F. (2020). Working memory training in typically developing children: A multilevel meta-analysis. *Psychonomic Bulletin & Review*, *27*(3), 423–434. https://doi.org/10.3758/s13423-019-01681-y

*Sala, G., Bolognese, M., & Gobet, F. (2017). Spatial training and mathematics: The moderating effect of handedness. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society, 3030–3045.*

*Schmitt, S. A., Korucu, I., Napoli, A. R., Bryant, L. M., & Purpura, D. J. (2018). Using block play to enhance preschool children's mathematics and executive functioning: A randomized controlled trial. *Early Childhood Research Quarterly*, *44*, 181–191. https://doi.org/10.1016/j.ecresq.2018.04.006

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274

Seron, X., Pesenti, M., Noël, M. P., Deloche, G., & Cornet, J. A. (1992). Images of numbers, or "When 98 is upper left and 6 sky blue". *Cognition*, *44*(1–2), 159–196. https://doi.org/10.1016/0010-0277(92)90053-K

Smith, I. M. (1964). *Spatial ability: Its educational and social significance*. Knapp.

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

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, *20*(5), 498–505. https://doi.org/10.2307/749423

*Sundberg, S. E. (1994). *Effect of spatial training on spatial ability and mathematical achievement as compared to traditional geometry instruction* (Unpublished doctoral dissertation). University of Missouri.

Tam, Y. P., Wong, T. T. Y., & Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: The mediating role of mental number line representation. *Contemporary Educational Psychology*, *56*, 14–24. https://doi.org/10.1016/j.cedpsych.2018.10.007

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and spss. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT Press.

*Thompson, A. (2012). *The effect of enhanced visualization instruction on first grade students' scores on the North Carolina standard course assessment* (Unpublished doctoral dissertation). Liberty University.

*Tillotson, M. L. (1984). *The effect of instruction in spatial visualization on spatial abilities an mathematical problem solving* (Unpublished doctoral dissertation). The University of Florida.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2009). A structural model of algebra achievement: Computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology*, *29*(2), 239–266. https://doi.org/10.1080/01443410802708903

Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of spatial-numerical associations: Evolutionary and cultural factors co-construct the mental number line. *Neuroscience and Biobehavioral Reviews*, *90*, 184–199. https://doi.org/10.1016/j.neubiorev.2018.04.010

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation*, *57*, 147–181. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

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., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, *22*(5), 367–373. https://doi.org/10.1177/0963721413484756

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*(1), 7–13. https://doi.org/10.1016/j.tine.2014.02.005

Vernon, P. E. (1950). *The structure of human abilities*. Methuen.

Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: An individual differences approach. *PLoS ONE*, *9*(4), e95756. https://doi.org/10.1371/journal.pone.0095756

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *The British Journal of Educational Psychology*, *82*(Pt. 1), 157–181. https://doi.org/10.1111/j.2044-8279.2011.02049.x

Xie, F., Zhang, L., Chen, X., & Xin, Z. (2020). Is spatial ability related to mathematical ability: A meta-analysis. *Educational Psychology Review*, *32*(1), 113–155. https://doi.org/10.1007/s10648-019-09496-y

*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*(6), 854–866. https://doi.org/10.1037/dev0000124

Zelinsky, N. A. M., & Shadish, W. (2018). A demonstration of how to do a meta-analysis that combines single-case designs with between-groups experiments: The effects of choice making on challenging behaviors performed by people with disabilities. *Developmental Neurorehabilitation*, *21*(4), 266–278. https://doi.org/10.3109/17518423.2015.1100690

Appendix A

Call for Papers

Request for Papers: The effects of spatial training on mathematics outcomes: Meta-Analysis

Spatial skills have been linked to mathematics achievement in many studies. However, fewer studies have explored the transfer of spatial training effects to mathematics domains. To further understand the relations between these domains, we are conducting a meta-analysis examining the effect of spatial interventions/training on mathematics outcomes.

We have identified papers through PUBMED, PsycINFO, ERIC and ProQuest, but to ensure that the findings are as accurate as possible, we are in search of any unpublished data, including results presented at conferences.

Appendix B

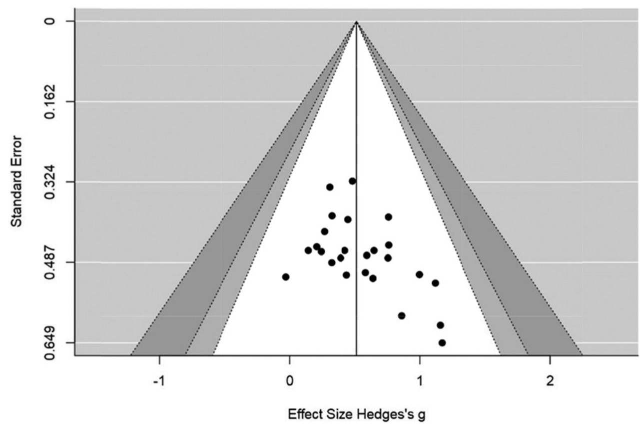
Publication Bias for the Effect of Spatial Training on Spatial Outcomes

A moderator analysis found similar effect sizes for published (mean effect (g) = .512) compared to unpublished studies (mean effect (g) = .567). A Wald test confirmed no significant effect of publication type as a moderator of the effects of spatial training on spatial outcomes, F (5.73) = .161, p = .703, I^2 = 66.61, T^2 = .08. Second, we generated a contour-enhanced funnel plot (see Figure B1). Visual inspection of the plot revealed clustering on one side of the vertical

(*Appendices continue*)

Figure B1

Contour-Enhanced Funnel Plot for Studies Investigating the Effects of Spatial Training on Spatial Outcomes



Note. Key for color codes: White indicates $.01 < p < .05$; light gray indicates $.05 < p < .01$; Dark gray indicates $p < .01$.

line. that is, bottom right quadrant of the plot. We tested asymmetry statistically and found that while the Rank Correlation test was significant ($r_T = .31$, $p = .031$), Egger's (1997) Regression test was not ($z = 1.40$, $p = .162$). It is noteworthy that compared to the Rank Correlation Test, Egger's Regression test is deemed a more accurate measure of asymmetry for plots with fewer than 25 studies (which is the case here) (Begg & Mazumdar, 1994). Taken together, the findings suggest that the funnel plot is slightly asymmetrical, that is, some publication bias may be evident. For this reason, we next applied the trim and fill method and found that approximately 2 studies were missing given the asymmetry in the plot. The results indicated that adding these two additional studies would reduce the effect size from $g = .521$ ($SE = .06$), to $g = .487$ ($SE = .09$). However, the overall effect of spatial training on spatial outcomes would remain significant ($p < .001$).

Received May 11, 2021
Revision received August 6, 2021
Accepted August 20, 2021 ■