# The malleability of spatial skills:

## A meta-analysis of training studies

David H. Uttal, Nathaniel G. Meadow, Elizabeth Tipton, Linda L. Hand,
Alison R. Alden, & Christopher Warren
Northwestern University
Nora S. Newcombe
Temple University

(In Press, *Psychological Bulletin*)

This work was supported by the Spatial Intelligence and Learning Center (NSF Grant SBE0541957) and by the Institute for Education Sciences (Department of Education Grant R305H020088). We thank David B. Wilson, Larry Hedges, Loren M. Marulis and Spyros Konstantopoulos for their help in designing and analyzing the meta-analysis. We also thank Greg Ericksson and Kate O'Doherty for assistance in coding, and Kseniya Povod and Kate Bailey for assistance with references and proofreading. Send correspondence to David Uttal (duttal@northwestern.edu), Department of Psychology, Northwestern University, 2029 Sheridan Rd, Evanston, IL 60208-2710 USA.

# Abstract

Having good spatial skills strongly predicts achievement and attainment in STEM fields (e.g., Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). Improving spatial skills is therefore of both theoretical and practical importance. To determine whether and to what extent training and experience can improve these skills, we meta-analyzed 217 research studies investigating the magnitude, moderators, durability and generalizability of training on spatial skills. After eliminating outliers, the average effect size (Hedges' *g*) for training relative to control was 0.47 (*SE* = 0.04). Training effects were stable and were not affected by delays between training and post-testing. Training also transferred to other spatial tasks that were not directly trained. We analyzed the effects of several moderators: Including the presence and type of control groups, sex, age, and type of training. Additionally, we include a theoretically-motivated typology of spatial skills that emphasizes two dimensions: intrinsic versus extrinsic, and static versus dynamic (Newcombe and Shipley, in press). Finally, we consider the potential educational and policy implications of directly training spatial skills. Considered together, the results suggest that spatially enriched education could pay substantial dividends in increasing participation in mathematics, science, and engineering.

*Keywords:* spatial skills, reasoning, training, meta-analysis, sex differences, practice effects, transfer, STEM, malleability, typology.

# The malleability of spatial skills:
# A meta-analysis of training studies

The nature and extent of malleability are central questions in developmental and educational psychology (Bornstein, 1989). To what extent can experience alter people's abilities? Does the effect of experience change over time? Are there critical or sensitive periods for influencing development? What are the origins and determinants of individual variation in response to environmental input? Spirited debate on these matters is long-standing, and still continues. However, there is renewed interest in malleability in behavioral and neuroscientific research on development (e.g., Johnson, Munakata, & Gilmore, 2002; National Research Council (*NRC)*, 2000; Stiles, 2008). Similarly, recent economic, educational, and psychological research has focused on the capacity of educational experiences to maximize human potential, reduce inequality (e.g., Duncan et al., 2007; Heckman & Masterov, 2007), and foster competence in a variety of school subjects, including reading (e.g., Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001), mathematics (e.g., U.S. Department of Education, 2008), and science and engineering (NRC, 2009).

This paper develops this theme further, by focusing on the degree of malleability of a specific class of cognitive abilities: spatial skills. These skills are important for a variety of everyday tasks, including tool use and navigation. They also relate to an important national problem: effective education in the Science, Technology, Engineering, and Mathematics (*STEM*) disciplines. Recent analyses have shown that spatial abilities uniquely predict STEM achievement and attainment. For example, in a long-term longitudinal study, using a nationally representative sample, Wai, Lubinski and Benbow (2009) showed that spatial ability was a significant predictor of achievement in STEM, even after holding constant possible third variables such as mathematics and verbal skills (see also Humphreys, Lubinski, & Yao, 1993; Shea, Lubinski, & Benbow, 2001).

Efforts to improve STEM achievement by improving spatial skills would thus seem logical. However, the success of this strategy is predicated on the assumption that spatial skills are sufficiently malleable to make training effective and economically feasible. Some investigators have argued that training spatial performance leads only to fleeting improvements, limited to cases in which the trained task and outcome measures are very similar (e.g., Eliot, 1987; Eliot & Fralley, 1976; Maccoby & Jacklin, 1974; Sims & Mayer, 2002). In fact, the National Research Council (2006) report, *Learning to Think Spatially*, questioned the generality of training effects and concluded that transfer of spatial improvements to untrained skills has not been convincingly demonstrated. The report called for research aimed at determining how to improve spatial performance in a generalizable way (NRC, 2006).

Prior meta-analyses concerned with spatial ability did not focus on the issue of how, and how much, training influences spatial thinking. Nor did they address the vital issues of durability and transfer of training. For example, Linn and Petersen (1985) conducted a comprehensive meta-analysis of sex differences in spatial skills, but they did not examine the effects of training. Closer to the issues at hand, Baenninger and Newcombe (1989) conducted a meta-analysis aimed at determining whether training spatial skills would reduce or eliminate sex differences in spatial reasoning. However, Baenninger and Newcombe's meta-analysis, which is now quite dated, focused almost exclusively on sex differences. It ignored the fundamental questions of durability and transfer of training, although the need to further explore these issues was highlighted in the Discussion section.

Given the new focus on the importance of spatial skills in STEM learning, the time is ripe for a comprehensive, systematic review of the responsiveness of spatial skills to training and experience. The present meta-analytic review examines the existing literature to determine the size of spatial training effects, as well as whether any such training effects are durable and whether they transfer to new tasks. Durability and transfer of training matter substantially. For spatial training to be educationally relevant, its effects must endure longer than a few days, and must show at least some transfer to non-trained problems and tasks. Thus, examining these issues comprehensively may have a considerable impact on educational policy and the continued development of spatial training interventions. Additionally, it may highlight areas that are as of yet under researched and warrant further study.

Like Baenninger and Newcombe (1989), the current study also examines sex differences in responsiveness to training. Researchers since Maccoby and Jacklin (1974) have identified spatial skills as an area in which males outperform females on many but not all tasks (Voyer, Voyer, & Bryden, 1995). Some researchers (e.g., Fennema & Sherman, 1977; Sherman, 1967) have suggested that females should improve more with training than males because they have been more deprived of spatial experience. However, Baenninger and Newcombe's meta-analysis showed parallel improvement for the two sexes. This conclusion deserves re-evaluation given the many training studies completed since the Baenninger and Newcombe review.

The present study goes beyond the analyses conducted by Baenninger and Newcombe (1989) in evaluating whether those who initially perform poorly on tests of spatial skills can benefit more from training than those who initially perform well. While the idea that this should be the case motivated Baenninger and Newcombe (1989) to examine whether training had differential effects across the sexes, they did not directly examine the impact of initial performance on the size of training effects observed. Notably, there is considerable variation within the sexes in terms of spatial ability (Astur, Ortiz, & Sutherland, 1998; Linn & Petersen, 1985; Maccoby & Jacklin, 1974; Silverman & Eals, 1992; Voyer et al., 1995). Thus, even if spatial training does not lead to greater effects for females as a group (Baenninger & Newcombe, 1989), it might still lead to greater improvements for those individuals who initially perform particularly poorly. In addition, this review examines whether younger children improve more than adolescents and adults, as a sensitive period hypothesis would predict.

## Typology of Spatial Skills

Ideally, a meta-analysis of the responsiveness of spatial skills to training would begin with a precise definition of spatial ability and a clear breakdown of that ability into constituent factors or skills. It would also provide a clear explanation of perceptual and cognitive processes or mechanisms that these different spatial factors demand or tap. The typology would allow for a specification of whether, how, and why the different skills do, or do not, respond to training of various types. Unfortunately, the definition of spatial ability is a matter of contention, and a comprehensive account of the underlying processes is not currently available (Hegary & Waller, 2005).

Prior attempts at defining and classifying spatial skills have mostly followed a psychometric approach. Research in this tradition typically relies on exploratory factor analysis of the relations among items from different tests that researchers believe sample from the domain of spatial abilities (e.g., Carroll, 1993; Eliot, 1987; Lohman, 1988; Thurstone, 1947). However, like most intelligence tests, tests of spatial ability did *not* grow out of a clear theoretical account or even a definition of spatial ability. Thus it is not surprising that the exploratory factor approach has not led to consensus. Instead, it has identified a variety of distinct factors.

Agreement seems to be strongest for the existence of a skill often called *spatial visualization*, which involves the ability to imagine and mentally transform spatial information. Support has been less consistent for other factors, such as *spatial orientation*, which involves the ability to imagine oneself or a configuration from different perspectives (Hegarty & Waller, 2005).

Since a century of research on these topics has not led to a clear consensus regarding the definition and sub-components of spatial ability, a new approach is clearly needed (Hegarty & Waller, 2005; Newcombe & Shipley, in press). Our approach relies on a classification system that grows out of linguistic, cognitive and neuroscientific investigation (Chatterjee, 2008; Palmer, 1978; Talmy, 2000). The system makes use of two fundamental distinctions. The first is between intrinsic versus extrinsic information. Intrinsic information is what one typically thinks about when defining an object. It is the specification of the parts, and the relation between the parts, that defines a particular object (e.g., Biederman, 1987; Hoffman & Singh, 1997; Tversky, 1981). Extrinsic information refers to the relation among objects in a group, relative to each other or to an overall framework. So, for example, the spatial information that allows us to distinguish rakes from hoes from shovels in the garden shed is intrinsic information, while the spatial relations among those tools (e.g., the hoe is *between* the rake and the shovel) are extrinsic, as well as the relations of each object to the wider world (e.g., the rake, hoe, and shovel are all on the north side of the shed, on the side where the brook runs down to the pond). The intrinsic-extrinsic distinction is supported by several lines of research (e.g., Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Huttenlocher & Presson, 1979; Kozhevnikov & Hegarty, 2001; Kozhevnikov, Motes, Rash, & Blajenkova, 2006).

The second distinction is between static and dynamic tasks. So far, our discussion has focused only on fixed, static information. However, objects can also move or be moved. Such movement can change their intrinsic specification, as when they are folded or cut, or rotated in place. In other cases, movement changes an object's position with regard to other objects and overall spatial frameworks. The distinction between static and dynamic skills is supported by a variety of research. For example, Kozhevnikov, Hegarty and Mayer (2002) and Kozhevnikov, Kosslyn and Shepard (2005) found that object visualizers (who excel at intrinsic-static skills in our terminology) are quite distinct from spatial visualizers (who excel at intrinsic-dynamic skills). Artists are very likely to be object visualizers, while scientists are very likely to be spatial visualizers.

Considering the two dimensions together (intrinsic versus extrinsic; dynamic versus static) yields a 2 x 2 classification of spatial skills, as shown in Figure 1. The figure also includes well-known examples of the spatial processes that fall within each of the four cells. For example, the recognition of an object as a rake involves intrinsic, static information. In contrast, the mental rotation of the same object involves intrinsic, dynamic information. Thinking about the relations among locations in the environment, or on a map, involves extrinsic, static information. Thinking about how one's perception of the relations among the object would change as one moves through the same environment involves extrinsic, dynamic relation.

Linn and Petersen's (1985) three categories—*Spatial Perception*, *Mental Rotation*, and *Spatial Visualization*—can be mapped onto the cells in our typology. Table 1 provides a mapping of the relation between our classification of spatial skills and Linn and Petersen's.

Linn and Petersen (1985) described *Spatial Perception* tasks as those that required participants to "determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information" (p. 1482). This category represents tasks that are extrinsic and static in our typology because they require the coding of spatial position in relation to another object, or with respect to gravity. Examples of tests in this category are the Rod and Frame Test and the Water Level Task. Linn and Petersen's *Mental Rotation* tasks involved a dynamic process in which a participant attempts to mentally rotate one stimulus to align it with a comparison stimulus and then make a judgment regarding whether the two stimuli appear the same. This category represents tasks that are intrinsic and dynamic in our typology because they involve the transformation of a single object. Examples of Mental Rotation tests are the Mental Rotations Test (Vandenberg & Kuse, 1978) and the Cards Rotation Test (French, Ekstrom, & Price, 1963).

Linn and Petersen's *Spatial Visualization* tasks, as described by Linn and Petersen (1985), were "those spatial ability tasks that involve complicated, multistep manipulations of spatially presented information…" (p. 1484). This category included Embedded Figures, Hidden Figures, Paper Folding, Paper Form Board, Surface Development, Differential Aptitude Test (spatial relations subtest), Block Design, and Guilford-Zimmerman spatial visualization tests. The large number of tasks in this category reflects its relative lack of specificity. Although all of these tasks require people to think about a single object, and thus are intrinsic in our typology, some tasks, such as the Embedded Figures and Hidden Figures are static in nature, while others, including Paper Folding or Surface Development, require a dynamic mental manipulation of the object. Therefore we feel the 2 x 2 classification provides a more precise description of the spatial skills and their corresponding tests.

**Methodological Considerations**

How individual studies are designed, conducted, and analyzed often turns out to be the key to interpreting the results in a meta-analysis (e.g., Campbell Collaboration, 2001; Lipsey & Wilson, 2001). In this section we describe our approach to dealing with some particularly relevant methodological concerns, including differences in research designs, heterogeneity in effect sizes, and the (potential) analysis of the non-independence and nested structure of some effect sizes. One of the contributions of the present work is the use of a new method for analyzing and understanding the effects of heterogeneity and non-independence.

**Research design and improvement in control groups**. Research design often turns out to be extremely important in understanding variation in effect sizes. A good example in the present work concerns the influences of variation in control groups and control activities on the interpretation of training-related gains. Although control groups do not, by definition, receive explicit training, they often take the same tests of spatial skills as the experimental groups do. For example, researchers might measure a particular spatial skill in both the treatment and control group before, during, and after training. Consequently, the performance of *both groups* could improve due to *retesting effects*—taking a test multiple times in itself leads to improvement, particularly if the multiply-administered tests are identical or similar (Hauschknecht, Halpert, Di Paolo, & Gerrard, 2007). Salthouse and Tucker-Drob (2008) have suggested that retesting effects may be particularly large for measures of spatial skills. Consequently, a design that includes no control group might find a very strong effect of training, but this result would be confounded with retesting effects. Likewise, a seemingly very large training effect could be rendered non-significant if compared to a control group that greatly improved due to retesting effects (Sims & Mayer, 2002). Thus, it is critically important to consider the presence and performance of control groups.

Three different designs have been used in spatial training studies. The first is a simple pre-test, post-test design on a single group, which we label the *within-subjects only* design. The second design involves comparing a training (treatment) group to a control group on a test given after the treatment group receives training. We call this methodology the *between-subjects* design. The final approach is a *mixed* design in which pre- and post-test measures are taken for both the training and control groups and the degree of improvement is determined by the difference between the gains made by each group.

The three research designs differ substantially in terms of their contribution to understanding the possible improvement of control groups. Because the within-subjects design does not include a control group, it confounds training and retesting effects. The between-subjects design does include a control group, but performance is measured only once. Thus only those studies that used the *mixed* design methodology allow us to calculate the effect sizes for the improvement of the treatment and control groups independently, along with the overall effect size of treatment versus control. Fortunately, this design was the most commonly used among the studies in our meta-analysis, accounting for about 60% . We therefore were able to analyze control and treatment groups separately, allowing us to measure the magnitude of improvement as well as investigate possible explanations for this improvement.

**Heterogeneity**. We also considered the important methodological issue of heterogeneity. Classic, fixed-effect meta-analyses assume homogeneity—that all studies estimate the same underlying effect size. However, this assumption is, in practice, rarely met. Because we included a variety of types of training and outcome measures, it is important that we account for heterogeneity in our analyses.

Prior meta-analyses have often handled heterogeneity by parsing the dataset into smaller, more similar groups to increase homogeneity (Hedges & Olkin, 1986). This method is not ideal because to achieve homogeneity, the final groups no longer represent the whole field, and often they are so small that they do not merit a meta-analysis.

We took a different approach that instead accounted for heterogeneity in two ways. First, we used a mixed effects model. In mixed effects models, covariates are used to explain a portion of the variability in effect sizes. We considered a wide variety of covariates, which are addressed in the following sections. Additionally, in mixed models any residual heterogeneity is modeled via random effects, which here account for the variability in true effect sizes. Mixed models are used when there is reason to suspect that variability among effect sizes is not due solely to sampling error (Lipsey & Wilson, 2001).

Second, we used a model that accounted for the nested nature of research studies from the same paper. Effect sizes from the same study or paper are likely to be similar in many ways. For example, they often share similar study protocols, and the participants are often recruited from the same populations, such as an Introductory Psychology Participant Pool at a university. Consequently, effect sizes from the same study or paper can be more similar to one another than effect sizes from different studies. In fact, effect sizes can sometimes be construed as having a nested or hierarchical structure; effect sizes are nested within studies, which are nested within papers, and (perhaps) within authors (Hedges, Tipton, & Johnson, 2010a, 2010b; Lipsey & Wilson, 2001). The nested nature of the effect sizes was important to our meta-analysis because although there are a total of 1,038 effect sizes, these effect sizes are nested within 206 studies.

**Addressing the nested structure of effect sizes.** In the past, analyzing the nested structure of effect sizes has been difficult. Some researchers have ignored the hierarchical nature of the effect sizes and treated them as if they were independent. However, this carries the substantial risk of inflating the significance of statistical tests because it treats each effect size as contributing one unique degree of freedom when in fact the degrees of freedom at different levels of the hierarchy are not unique. Other researchers have averaged or selected at random effect sizes from particular studies, but this approach disregards a great deal of potentially useful information. (See Lipsey and Wilson, 2001, for a discussion of both approaches).

More generally, the problem of nested or multilevel data has been addressed via hierarchical linear modeling (*HLM)* (Raudenbush & Bryk, 2002). Methods for applying HLM theory and estimation techniques to meta-analysis have been developed over the last 25 years (Jackson, Riley, & White, 2011; Kalaian & Raudenbush, 1996; Konstantopoulos, 2010). A shortcoming of these methods, however, is that they can be technically difficult to specify and implement, and can be sensitive to misspecification. This might occur if, for example, a level of nesting had been mistakenly left out, the weights were incorrectly calculated, or if the normality assumptions were violated. A new method for robust estimation was recently introduced by Hedges, Tipton, and Johnson (2010a, 2010b). This method uses an empirical estimate of the sampling variance that is robust to both misspecification of the weights and to distributional assumptions, and is simple to implement, using freely-available, open-source software. Importantly, when the same weights are used, the HLM and robust estimation methods generally give similar estimates of the regression coefficients.

In addition to modeling the hierarchical nature of effect sizes, using a hierarchical meta-regression approach is beneficial because it allows the variation in effect sizes to be divided into two parts: the variation of effect sizes *within* studies and the variation of study-average effect sizes *between* or across studies. The same distinction can be made for the effect of a particular covariate on the effect sizes. The *within*-study effect for a covariate is the pooled within-study correlation between the covariate and the effect sizes. The *between*-study effect is the correlation between the average value of the covariate in a study with the average study effect size. Note that in traditional non-nested meta-analyses, only the *between*-study variation or regression effects are estimable.

Parsing variation into within- and between-study effects is important for two reasons. First, by dividing analyses into these separate parts, we were able to see which protocols (e.g. age, dependent variable) are commonly varied or kept constant within studies. Second, when the values of a covariate vary within a study, the *within* effect estimate can be thought of as the effect of the covariate controlling for other unmeasured study or research-group variables. In many cases this is a better measure of the relationship between the covariate of interest and the effects sizes than using the *between* effect alone.

To illustrate the difference between these two types of effects, imagine two different meta-analyses. In the first, every study has both child and adult respondents. This means that within each study, the outcomes for children and adults can be compared holding constant study or research group variables. This is an example of a *within* meta-regression, which naturally controls for unmeasured covariates within the studies. In the second meta-analysis, none of the studies has both children and adults as respondents. Instead (as is often true in the present meta-analysis), some studies include only children, and others include only adults. The only way that the effect of age can be addressed here is through a comparison across studies, which is a *between* meta-regression model. In such a model, it would be difficult to determine if any effects of age found were a result of actual differences between age groups or of confounds such as systematic differences in the selection criteria or protocols used in studies with children and studies with adults. In the present meta-analysis, we were sometimes able to gain unique insight into sources of variation in effect sizes by considering the contribution of within- and between-study variance.

## Characteristics of the Training Programs

Spatial skills might respond differently to different kinds of training. To investigate this issue, we divided the training program of each study into one of three mutually exclusive categories: (1) those that used video games to administer training, (2) those that used a semester-long or instructional course, and (3) those that trained subjects on spatial tasks through practice, strategic instruction or computerized lessons, often administered in a psychology laboratory. As shown in Table 2, these training categories are similar to Baenninger and Newcombe's (1989) categories. Out of our three categories, course and video game training correspond to what these authors referred to as *indirect* training. We chose to distinguish these two forms of training because of the recent increase in the availability of, and interest in, videogame training of spatial abilities (e.g. Green & Bavelier, 2003). Our third category, spatial task training, involved direct practice or rehearsal (what Baenninger and Newcombe termed *specific* training).

## Missing Elements from this Meta-Analysis

This meta-analysis provides a comprehensive review of work on the malleability of spatial cognition. Nevertheless, it does not address every interesting question related to this topic. Many such questions one might ask are simply so fine-grained that were we to attempt analyses to answer them, the sample sizes of relevant studies would become unacceptably small. For example, it would be nice to know whether men's and women's responsiveness to training differs for each type of skills that we have identified, or how conclusions about age differences in responsiveness to training are affected by study design. However, these kinds of interaction hypotheses could not be evaluated with the present dataset, given the number of effect sizes available. Additionally, the lack of studies that directly assess the effects of spatial training on performance in a STEM discipline is disappointing. To properly measure spatial training's effect on STEM outcomes, we must move away from anecdotal evidence and conduct rigorous experiments testing its effect. Nonetheless, the present study provides important information about whether and how training can affect spatial cognition.

# Method

## Eligibility Criteria

Several criteria were used to determine whether to include a study.

1. The study must have included at least one spatial outcome measure. Examples include, but are not limited to, performance on published psychometric subtests of spatial ability, reaction time on a spatial task (e.g., mental rotation or finding an embedded figure), or measures of environmental learning (e.g., navigating a maze).[1]

2. The study must have used training, education, or another type of intervention that was designed to improve performance on a spatial task.

3. The study must have employed a rigorous, causally-relevant design, defined as meeting at least one of the following design criteria: (a) use of a pre-test, post-test design that assessed performance relative to a baseline measure obtained before the intervention was given; (b) inclusion of a control or comparison group; or (c) a quasi-experiment, such as the comparison of growth in spatial skills among engineering and liberal arts students.

4. The study must have focused on a non-clinical population. For example, we excluded studies that used spatial training to improve spatial skills after brain injury or in Alzheimer's disease. We also excluded studies that focused exclusively on the rehabilitation of high-risk or at-risk populations.

## Literature Search and Retrieval

We began with electronic searches of the PsycInfo, ProQuest, and ERIC databases. We searched for all available records from January 1, 1984 through March 4, 2009 (the day the search was done). We chose this 25-year window for two reasons. First, it was large enough to provide a wide range of studies and to cover the large increase in studies that has occurred recently. Second, the window was small enough to allow us to gather most of the relevant published and unpublished data. The search included foreign-language articles if they included an English abstract.

We used the following search term: (training OR practice OR education OR "experience in" OR "experience with" OR "experience of" OR instruction) AND ("spatial relation" OR "spatial relations" OR "spatial orientation" OR "spatial ability" OR "spatial abilities" OR "spatial task" OR "spatial tasks" OR visuospatial OR geospatial OR "spatial visualization" OR "mental rotation" OR "water-level" OR "embedded figures" OR "horizontality"). After the removal of studies performed on nonhuman subjects, the search yielded 2,545 hits. The process of winnowing these 2,545 papers proceeded in three steps to ensure that each paper met all inclusion criteria (See Figure 2).

Step 1 was designed to eliminate quickly those papers that focused primarily on clinical populations or that did not include a behavioral measure, and involved two raters (postgraduate-level research coordinators and authors of this paper) reading only the titles of the articles. Papers were excluded if the title revealed a focus on atypical human populations, including at-risk or low-achieving populations, or disordered populations (e.g., individuals with Parkinson's, HIV, Alzheimer's, genetic disorder, or mental disorders). Also excluded were studies that did not include a behavioral measure, such as studies that only included physiological or cellular activity. Finally, we excluded papers that did not present original data, such as review papers. We instructed the raters to be inclusive in this first step of the winnowing process. For example, if the title of an article did not include sufficient information to warrant exclusion, raters were instructed to leave it in the sample. In addition, we only eliminated an article at this step if both raters agreed that it should be eliminated. Overall, rater agreement was very good, 82%. Of the 2,545 papers, 649 were excluded at Step 1. In addition, we found that 244 studies were duplicated, so we deleted one of the copies of each. In total, 1,652 studies survived Step 1.

In Step 2, three raters, the same two authors and an incoming graduate student, read the abstracts of all papers that survived Step 1. The goal of Step 2 was to determine whether the papers included training measures and whether they utilized appropriately rigorous (experimental or quasi-experimental) designs. To train the raters, all three first read the same 25% (413) of the abstracts. After discussion, inter-rater agreement was very good (87%), Fleiss' kappa = .78. The remaining 75% of papers (1,239) were then divided into three groups, and each of these abstracts was read by two of the three raters. Inter-rater agreement among the three pairs of raters was high: 84%, 90%, and 88%, and all disagreements were resolved by the third rater. A total of 284 papers survived Step 2.

In Step 3, the remaining papers were read in full. We were unable to obtain seven papers. After reading the papers, we rejected 89, leaving us with 188 papers that met the criteria for inclusion. The level of agreement among raters reading papers in full was good, 87%. The Cohen's kappa was .74, which is typically defined as in the "substantial" to "excellent" range (Capozzoli, McSweeney, & Sinha, 1999; Landis & Koch, 1977). The sample included articles written in several non-English languages, including Chinese, Dutch, French, German, Italian, Japanese, Korean, Romanian, and Spanish. Bilingual individuals who were familiar with psychology translated the articles.

We also acquired relevant papers through directly contacting experts in the field. We contacted 150 authors in the field of spatial learning. We received 48 replies, many with multiple suggestions for papers. Reading through the authors' suggestions led to the discovery of 29 additional papers. Twenty-four of these articles were published in scientific journals or institutional technical reports, and five were unpublished manuscripts or dissertations.

Thus, through both electronic search and communication with researchers, we acquired and reviewed data from a total of 217 papers (188 from electronic searches, and 29 from correspondence).

**Publication bias.** Studies reporting large effects are more likely to be published than those reporting small or null effects (Rosenthal, 1979). We made efforts both to attenuate and to assess the effects of publication bias on our sample and analyses. First, when we wrote to authors and experts, we explicitly asked them to include unpublished work. Second, we searched references lists of our articles for relevant unpublished conference proceedings, and we also looked through the tables of contents of any recent relevant conference proceedings that were accessible online. Third, our search of *ProQuest Dissertations and Theses* yielded many unpublished dissertations, which we included when relevant. If a dissertation was eventually published, we examined both the published article and original dissertation. We augmented the data from the published article if the dissertation provided additional, relevant data. However, we only counted the original dissertation and published article as one (published) study.

We also contacted authors when their papers did not provide sufficient information for calculating effect sizes. For example, we requested separate means for control and treatment groups when only the overall group $F$ or $t$ statistics were reported. Authors responded with usable data in approximately 20% of these cases. We used these data to compute effect sizes separately for males and females and control and treatment groups whenever possible.

## Coding of Study Descriptors

We coded the methods and procedures used in each study, focusing on factors that might shed light on the variability in the effect sizes that we observed. The coding scheme addressed the following characteristics of each study: the publication status, the study design, control group design and characteristics, the type of training administered, the spatial skill trained and tested, characteristics of the sample, and details about the procedure such as the length of delay between the end of training and the post-test. We have provided the full description of the coding procedure in Appendix A. The majority of these characteristics were straightforward to code. Here we discuss in detail two aspects of the coding that are new to the field: the classification of spatial skills based on the 2 x 2 framework and how it relates to the coding of transfer of training.

**The 2 x 2 framework of spatial skills.** We coded each training intervention and outcome measure in terms of both the intrinsic-extrinsic and static-dynamic dimensions. These dimensions are also discussed above in the Introduction; here we focus on the defining characteristics and typical tasks associated with each dimension.

***Intrinsic versus extrinsic.*** Spatial activities that involved defining an object were coded as intrinsic. Identifying the distinguishing characteristics of a single object, for example in the Embedded Figures task, the Paper Folding task, and the Mental Rotations Test, is an intrinsic process because the task requires only contemplation of the object at hand, without consideration of the object's surroundings.

In contrast, spatial activities that required the participant to determine relations among objects in a group, relative to each other or to an overall framework were coded as extrinsic. Classic examples of extrinsic activities are the Water-Level task, and Piaget's Three Mountain Task, as both tasks require the participant to understand how multiple items relate spatially to one another.

***Static versus dynamic.*** Spatial activities in which the main object remains stationary were coded as static. For example, in the Embedded Figures task and the Water-Level task, the object at hand does not change in orientation, location, nor dimension. The main object remains static to the participant throughout the task.

In contrast, spatial activities in which the main object moves, either physically or in the mind of the participant, were coded as dynamic. For example, in the Paper Folding task, the presented object must be contorted and altered to create the three dimensional answer. Similarly, in the Mental Rotations Test and Piaget's Three Mountain Task, the participant must either rotate the object or their own perspective to determine which suggested orientation aligns with the original. These processes require dynamic interaction with the stimulus.

**Transfer.** To analyze transfer of training, we coded both the training task and all outcome measures into a single cell of the 2 x 2 framework (intrinsic and static, or intrinsic and dynamic, etc.).[2] We used the framework to define two levels of transfer. *Within-cell* transfer was coded when the training and outcome measure were (a) not the same, but (b) both in the same cell of the 2 x 2 framework. *Across-cell* transfer was coded when the training and outcome measures were in different cells of the 2 x 2 framework.

## Computing Effect Sizes

The data from each study were entered into the computer program *Comprehensive Meta-Analysis (CMA*; Borenstein, Hedges, Higgins, & Rothstein, 2005). CMA provides a well-organized and efficient format for conducting and analyzing meta-analytic data. (The CMA procedures for converting raw scores into effect sizes can be found in Supplemental Material).

Measures of effect size typically quantify the magnitude of gain associated with a particular treatment relative to the improvement observed in a relevant control group (Morris, 2008). Gains can be conceptualized as an improvement in score. Effect sizes usually are computed from means and standard deviations but they also can be computed from an *F* statistic, *t* statistic or chi-square value as well as from change scores representing the difference in mean performance at two points in time. Thus, in some cases, it was possible to obtain effect sizes without having the actual mean scores associated with a treatment (see Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). All effect sizes were expressed as Hedges' *g,* a slightly more conservative derivative of Cohen's *d* (Cohen, 1992)*;* Hedges' *g* includes a correction for biases due to sample size.

To address the general question of the degree of malleability of spatial skills, we calculated an overall effect size for each study. (The individual effect sizes are reported in Supplemental Material). The definition of the overall effect size depended in part on the design of the study. As discussed above, the majority of studies used a *mixed design,* in which performance was measured both before (pre-test) and after (post-test) training, in both a treatment and a control group. In this case, the overall effect size was the difference between the improvement in the treatment group and the improvement in the control group. Other studies used a *between-only* design, in which treatment and control groups were tested only after training. In this case, the overall effect size represented the difference between the treatment and control groups. Finally, approximately 15% of the studies used a *within-subjects only* design, in which there is no control or comparison group, and performance is assessed before and after training. In this case, the overall effect size was the difference between the post-test and pre-test. We combined the effect sizes from the different designs to generate an overall measure of malleability. However, we also considered the effects of differences in study design and of improvement in control groups in our analysis of moderators.

## Implementing the Hedges et al. Robust Estimation Model

As noted above, we implemented the Hedges et al. (2010a, 2010b) robust variance estimation model to address the nested nature of effects sizes. These analyses were conducted in ***R*** (Hornik, 2011) using the function *robust.hier.se* (http://www.northwestern.edu/ipr/qcenter/RVE-meta-analysis.html) with inverse variance weights, and, when confidence intervals and *p*-values are reported, using a *t*-distribution with *m*-*p* degrees of freedom, where *m* is the number of studies and *p* is the number of predictors in the model.

More formally, the model we used for estimation was

$$T_{ij} = \boldsymbol{X_{ij}\beta} + \theta_i + \eta_{ij} + \boldsymbol{\epsilon_{ij}}$$

where $T_{ij}$ is the estimated effect size from outcome $j$ in study $i$, $\mathbf{X}_{\mathrm{ij}}$ is the design matrix for effect sizes in study j, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, $\theta_i$ is a study level random effect, $\eta_{ij}$ is a within-study random effect, and $\boldsymbol{\epsilon_{ij}}$ is the sampling error. This is a mixed or meta-regression model. It seeks to both explain variation in effect sizes via the covariates in $\mathbf{X}_{\mathrm{ij}}$ and account for unexplained variation via the random effects terms $\theta_i$, $\eta_{ij}$, and $\epsilon_{ij}$. In all of the analyses provided here, we assume that the regression coefficients in $\boldsymbol{\beta}$ are fixed. The covariates in $\mathbf{X}_{\mathrm{ij}}$ include, for instance, an intercept (giving the average effect), dummy variables (when categorical covariates like ‘type of training’ are of interest), and continuous variables.

Using this model, the residual variation of the effect size estimate $T_{ij}$ can be decomposed as

$$V(T_{ij}) = \tau^2 + \omega^2 + v_{ij}$$

where $\tau^2$ is the variance of the between-study residuals $\theta_i$, $\omega^2$ is the variance of the within-study residuals $\eta_{ij}$, and $v_{ij}$ is the known sampling variance of the residuals $\epsilon_{ij}$. This means that there are three sources of variation in the effect size estimates. Although we assume that $v_{ij}$ is known, we estimate both $\tau^2$ and $\omega^2$ using the estimators provided in Hedges et al. (2010b). In all of the results shown here, each effect size was weighted by the inverse of its variance, which gives greater weight to more precise effect size estimates.

Our method controls for heterogeneity without reducing the sample to an inconsequential size. Importantly, this approach also provides a robust standard error for each estimate of interest; the size of the standard error is affected by the number of studies ($m$), the sampling variance within each study ($v_{ij}$), and the degree of heterogeneity ($\tau^2$ and $\omega^2$). This means that when there is a large degree of heterogeneity ($\tau^2$ or $\omega^2$), estimates of the average effect sizes will be more uncertain, and our statistical tests took this uncertainty into account.

Finally, all analyses presented here were estimated using a mixed model approach. In some cases, the design matrix only included a vector of ones; in those cases only the average effect is estimated. In other cases, comparisons between levels of a factor were compared (e.g. post-test delays of one day, less than one week, and less than one month to test durability of training); in those cases the categorical factor with $k$ levels was converted into $k-1$ dummy variables. In a few models we included continuous covariates in the design matrix. In these cases, we centered the within-study values of the covariate around the study-average, enabling the estimation of separate within- and between-study effects. Finally, for each outcome or comparison of interest, following the standard protocol for the robust estimation method used, we present the estimate and p-value. We do not present information on the degree of residual heterogeneity unless it answers a direct question of interest.

# Results

We begin by reporting characteristics of our sample, including the presence of outliers and publication bias. Next, we address the overall question of the degree of malleability of spatial skills, and whether training endures and transfers. We then report analysis of several moderators.

## Characteristics of the Sample of Effect Sizes

**Outliers**. Twelve studies reported very high individual effect sizes, with some as large as 8.33. The most notable commonality among these outliers was that they were conducted in Bahrain, Malaysia, Turkey, China, India and Nigeria; countries that, at the time of analysis, were ranked 39, 66, 79, 92, 134 and 158 respectively on the Human Development Index (*HDI*). The HDI is a composite of standard of living, life expectancy, well-being and education that provides a general indicator of a nation's quality of life and SES (United Nations Development Programme, 2009).[3] For studies with a HDI index over 30, the mean effect size ($g = 1.63$, $SE = 0.44$, $m = 12$, $k = 114$) was more than 3 times the group mean of the remaining sample ($g = 0.47$, $SE = 0.04$, $m = 206$, $k = 1{,}038$), where $m$ represents the number of studies and $k$ represents the total number of effect sizes. Prior research has found that lower SES is associated with larger responses to training or interventions (Ghafoori & Tracz, 2001; Wilson & Lipsey, 2007; Wilson, Lipsey, & Derzon, 2003). The same was true in these data: there was a significant correlation between HDI ranking and effect size, $\rho = .35$, $p < .001$, with the higher rankings (indicating lower standards of living) correlated with larger effect sizes. Because inclusion of these outliers could distort the main analyses, these 12 studies were not considered further.[4]

**Assessing publication bias**. Although we performed a thorough search for unpublished studies, publication bias is always possible in any meta-analysis (Lipsey & Wilson, 1993).

Efforts to obtain unpublished studies typically reduce but do not eliminate the "file drawer" problem (Rosenthal, 1979). We evaluated whether publication bias affected our results in several ways. First, we compared the average effect size of published studies ($g = 0.56$, $SE = 0.05$, $m = 95$, $k = 494$) and unpublished studies ($g = 0.39$, $SE = 0.06$, $m = 111$, $k = 544$) in our sample. The difference was significant at $p < .05$. This result indicates that there is some publication bias in our sample and raises the concern that there could be more unpublished or inaccessible studies that, if included, would render our results negligible (Orwin, 1983) or trivial (Hyde & Linn, 2006). We therefore calculated the fail-safe *N* (Orwin, 1983) to determine how many unpublished studies averaging no effect of training ($g = 0$) would need to exist to lower our mean effect size to trivial levels. Orwin (1983) defined the fail-safe N as follows: $N_{fs} = N_0[(d_0 - d_c)/d_c]$, with $N_{fs}$ as the fail-safe N, $N_0$ as the number of studies, $d_0$ as the overall effect size, and $d_c$ as the set value for a negligible effect size. Adopting Hyde and Linn's (2006) value of 0.10 as a trivial effect size, it would take 762 studies with effect sizes of zero that we overlooked to reduce our results to trivial. Adopting a more conservative definition of a negligible effect size, 0.20, there would still need to be 278 overlooked studies reporting an effect size of zero to reduce our results to negligible levels. Finally, we also created a funnel plot of the results to provide a visual measure of publication bias. Figure 3 shows the funnel plot of each study's mean weighted effect size versus its corresponding standard error. The mostly symmetrical placement of effect sizes in the funnel plot, along with the large fail-safe N calculated above, indicate that although there was some publication bias in our sample, it seems very unlikely that the major results are due largely to publication bias.

**Characteristics of the trimmed sample**. Our final sample consisted of 206 studies with 1,038 effect sizes. The relatively large ratio of effect sizes to studies stems from our goal of analyzing the effects of moderators such as sex and the influence of different measures of spatial skills. Whenever possible we separated the published means by gender, and when different means for different dependent variables were given, we calculated all potential effect sizes for each. Overall, 95 studies (46%) were published in journals, and 111 (54%) were from (unpublished) dissertations, unpublished data, or conference papers. 163 studies (79%) were conducted in the United States. The characteristics of the sample are summarized in Table 3.

## Assessing the Malleability of Spatial Skills

We now turn to the main question of this meta-analysis: How malleable are spatial skills? Excluding outliers, the overall mean weighted effect size relative to available controls was 0.47 (*SE* = 0.04, *m* = 206, *k* = 1,038). This result includes all studies regardless of research design, and suggests that, in general, spatial skills are moderately malleable. Spatial training, on average, improved performance by almost one-half of a standard deviation.

**Assessing and addressing heterogeneity.** It is important to consider not only the average weighted effect size but also the degree of heterogeneity of these effect sizes. By definition, a mixed effects meta-analysis does not assume that each study represents the same underlying effect size, and hence some degree of heterogeneity is expected. But how much was there, and how does this affect our interpretation of the results?

An important contribution of this meta-analysis is the separation of heterogeneity into variability across studies ($\tau^2$) and within studies ($\omega^2$), following the method of Hedges et al. (2010a, 2010b). The between-studies variability, $\tau^2$, was estimated to be 0.185, and the within-studies variability, $\omega^2$, was estimated to be 0.025. These estimates tell us that effect sizes from different studies varied from one another much more than did effect sizes from the same study. It is not surprising that we found greater heterogeneity in effect sizes between studies than in effect sizes that come from the same study, given that studies differ from each other in many ways (e.g., types of training and measures used, variability in how training is administered, participant demographic characteristics, etc.).

As discussed in the Methods, the statistical procedures that we used throughout the manuscript take both sources of heterogeneity into account when estimating the significance of a given effect. In all subsequent analyses, we took both between- and within-study heterogeneity into account when calculating the statistical significance of our findings. Our statistical tests are thus particularly conservative.

**Durability of training**. We have already demonstrated that spatial skills respond to training. It is also very important to consider whether the effects of training are fleeting or enduring. To address this question, we coded the time delay from the end of training until the post-test was administered for each study. Some researchers administered the post-test immediately; some waited a couple of days, some weeks, and a few waited over a month. Comparing post-tests administered immediately after training versus all post-tests that were delayed, collapsing across the delayed post-test time intervals did not show a significant difference, $p > .67$. There were no significant differences between immediate post-test, less than one week delay, and less than one month delay, $p > .19$. Because only four studies involved a delay of more than one month, we did not include this category in our analysis. The similar magnitude of the mean weighted effect sizes produced across the distinct time intervals implies that improvement gained from training can be durable.

**Transferability of training.** The results thus far indicate that training can be effective and that these effects can endure. However, it is also critical to consider whether the effects of training can transfer to novel tasks. If the effects of training are confined to performance on tasks directly involved in the training procedure, it is unlikely that training spatial skills will lead to generalized performance improvements in the STEM disciplines. We approached this issue in two overlapping ways. First, we asked whether there was any evidence of transfer. We separated the studies into those that attempted transfer and those that did not to allow for an overall comparison. For this initial analysis, we considered all studies that reported any information about transfer (i.e., all studies except those coded as “no transfer”). We found an effect size of 0.48, ($SE$ = 0.04, $m$ = 170, $k$ = 764) indicating that training led to improvement of almost one-half a standard deviation on transfer tasks.

Second, we assessed the degree or range of transfer. How much did training in one kind of task transfer to other kinds of tasks? As noted above, we used our 2 x 2 theoretical framework to distinguish within-cell transfer from across-cell transfer, with the latter representing transfer between a training task and a substantially different transfer task. Interestingly, the effect sizes for transfer within cells of the 2 x 2 ($g$ = 0.51, $SE$ = 0.05, $m$ = 94, $k$ = 448), and those for transfer across cells ($g$ = 0.55, $SE$ = 0.10, $m$ = 51, $k$ = 175), both differed significantly from zero, $p < .01$. Thus, for the studies that tested transfer, there was strong evidence of not only within-cell transfer involving similar training and transfer tasks but also of across-cell transfer in which the training and transfer tasks might be expected to tap or require different skills or representations.

## Moderator Analyses

The overall finding of almost one-half of a standard deviation improvement for trained spatial skills raises the question of why there has been such variability in prior findings. Why have some studies failed to find that spatial training works? To investigate this issue, we examined the influences of several moderators that could have produced this variability in the results of studies. Table 4 presents a list of those moderators and the results of the corresponding analyses.

**Study Design.** As previously noted, there were three kinds of study designs: within-subjects only, between-subjects, and mixed. Fifteen percent of studies in our sample used the within-subjects design, 26% used the between-subjects design, and the remaining 59% of the studies used the mixed design. We analyzed differences in overall effect size as a function of design type. In this and all subsequent post-hoc contrasts we set alpha at .01 to reduce the Type I error rate. The difference between design types was significant, $p < .01$. As expected, studies that used a within-subjects only design, which confounds training and retesting effects, reported the highest overall effect size ($g = 0.75$, $SE = 0.08$, $m = 31$, $k = 160$). The within-subjects only mean weighted effect size significantly exceeded those for both the between-subjects ($g = 0.43$, $SE = 0.09$, $m = 55$, $k = 304$, $p < .01$) and the mixed design studies ($g = 0.40$, $SE = 0.05$, $m = 123$, $k = 574$, $p < .01$). The mean weighted effect sizes for the between-subjects and mixed designs did not differ significantly. These results imply that the presence or absence of a control group clearly affects the magnitude of the resulting effect size; and that studies without a control group will tend to report higher effect sizes.

***Control group effects***. Why, and how, do control groups have such a profound effect on the size of the training effect? To investigate these questions, we analyzed control group improvements separately from treatment group improvements. This analysis was only possible for the mixed-design studies, as the within- and between- designs do not include a control group or a measure of control group improvement, respectively. We were unable to separate the treatment and control means for approximately 15% of the mixed design studies because of insufficient information provided in the paper and lack of response from authors to our requests for the missing information. The mean weighted effect size for the control groups ($g = 0.45$, $SE =$

0.04, $m = 106$, $k = 372$) was significantly smaller than that for the treatment groups ($g = 0.62$, $SE = 0.04$, $m = 106$, $k = 365$), $p < .01$.

Two potentially important differences between control groups can be the number of times a participant takes a test, and the number of tests a participant takes. If the retesting effect can appear within a single pre-test and post-test as discussed above, it stands to reason that retesting or multiple distinct tests could generate additional gains. In some studies, control groups were tested only once (e.g., Basham, 2006), whereas in other studies they were tested multiple times (e.g., Heil, Roesler, Link, & Bajric, 1998). To measure the extent of retesting effects on the control group effect sizes, we coded the control group designs into four categories: (a) pre-test and post-test on a single test, (b) pre-test then retest then post-test on a single test (i.e. repeated practice), (c) pre-test and post-test on several different spatial tests (i.e., a battery of spatial ability tests), and (d) pre-test and post-test on a battery of non-spatial tests. As shown in Figure 4, control groups that engaged in repeated practice as their alternate activity produced significantly larger mean weighted effect sizes than those that took a pre-test and post-test only ($p < .01$). These results highlight that a control group can improve substantially without formal training if they receive repeated testing.

*Filler task content.* In addition to retesting effects, control groups can improve through other implicit forms of training. Although by definition control groups do not receive direct training, this does not necessarily mean that the control group did nothing. Many studies included what we will refer to as a *spatial filler* task. These control tasks were designed to determine whether the improvement observed in a treatment group was attributable to a specific kind of training or to simply repeated practice on any form of spatial task. For example, while training was occurring in Feng, Spence, and Pratt’s (2007) study, their control subjects played the

3-D puzzle video game *Ballance*. In contrast, other studies used much less spatially-demanding tasks as fillers, such as playing *Solitaire* (De Lisi & Cammarano, 1996). Control groups that received spatial filler tasks produced a larger mean weighted effect size than control groups that received *non-spatial filler* tasks, with a difference of 0.17. The spatial filler and non-spatial filler control groups did not differ significantly, however, we hypothesized that the large (but non-significant) difference between the two could in fact make a substantial difference on the overall effect size. As mentioned above, a high performing control group can depress the overall effect size reported. Therefore those studies whose control groups received *spatial filler* tasks may report depressed overall effect sizes because the treatment groups are being compared to a highly improving control group. To investigate this possibility, we compared the overall effect sizes for studies in which the control group received a spatial filler task to studies in which the control received a non-spatial filler task. Studies that used a non-spatial filler control group reported significantly higher effect sizes than studies that used a spatial filler control group, $p < .01$. This finding is a good example of the importance of considering control groups in analyzing overall effect sizes: The larger improvement in the spatial filler control groups actually suppressed the *difference* between experimental and control groups, leading to the (false) impression that the training was less effective (Figure 5).

**Type of training.** In addition to control group effects, one would expect that the type of training participants receive could affect the magnitude of improvement from training. To assess the relative effectiveness of different types of training, we divided the training programs used in each study into three mutually exclusive categories: Courses, Videogames, and Spatial Task Training (Table 2). The mean weighted effect sizes for these categories did not differ significantly, $p > .45$. Interestingly, as implied by our mutually exclusive coding for these training programs, no studies implemented a training protocol that included more than one method of training. The fact that these three categories of training did not produce statistically different overall effects largely results from the high degree of heterogeneity for the Course ($\tau^2$ = 0.207) and Videogames ($\tau^2$ = 0.248) training categories. However, overall we can say that each program produced positive improvement in spatial skills as all three of these methods differed significantly from zero at $p < .01$.

**Participant characteristics.** We now turn to moderators involving the characteristics of the participants, including sex, age, and initial level of performance.

***Sex***. Prior work has shown that males consistently score higher than females on many standardized measures of spatial skills (e.g., Erlich, Levine, & Goldin-Meadow, 2006; Geary, Saults, Liu, & Hoard, 2000), with the notable exception of object location memory, in which women sometimes perform better, although the effects for object location memory are extremely variable (Voyer, Postma, Brake, & Imperato-McGinley, 2007). There has been much discussion of the causes of the male advantage, although arguably a more important question is the extent of malleability shown by the two sexes (Newcombe, Mathason, & Terlecki, 2002). Baenninger and Newcombe (1989) found parallel improvement for the training studies in their meta-analysis, so we tested if this equal improvement with training persisted over the last twenty-five years.

We first examined whether there were sex differences in the overall level of performance. Forty-eight studies provided both the mean pre-test and the mean post-test scores for male and female participants separately and thus were included in this analysis. The effect size for this one analysis was *not* a measure of the effect of training but rather of the difference between the level of performance of males and females at pre- and post-test. A positive Hedges' *g* thus represents a male advantage, and a negative Hedges' *g* represents a female advantage. As expected, males on average outperformed females on the pre-test and the post-test in both the control group and the treatment group (See Table 5). All of the reported Hedges' *g* statistics in the table are greater than zero, indicating a male advantage.

Next we examined whether males and females responded differently to training. The mean weighted effect sizes for improvement for males were very similar to that of females, with a difference of only 0.01. Thus males and females improved about the same amount with training. Our findings concur with those of Baenninger and Newcombe (1989) and suggest that while males tend to have an advantage in spatial ability, both genders improve equally well with training.

***Age***. Generally speaking, children's thinking is thought to be more malleable than adults' (e.g., Heckman & Masterov, 2007; Waddington, 1966). Therefore, one might expect that spatial training would be more effective for younger children than for adolescents and adults. Following Linn and Petersen (1985), we divided age into three categories: younger than 13 years (children), 13 to 18 years (adolescents), and older than 18 years (adults). Comparing the mean weighted effect sizes of improvement for each age category showed a difference of 0.17 between children and both adolescents and adults. Nevertheless, the difference between the three categories did not reach statistical significance.

An important question is why this difference was not statistically significant. By accounting for the nested nature of the effect sizes, we were able to isolate two important findings here. First, while the estimated difference between age groups is indeed not negligible, the estimate is highly uncertain; and this uncertainty is largely a result of the heterogeneity in the estimates. For example, within the child group, many of the same-aged participants came from different studies, and the mean effect sizes for these studies differed considerably ($\tau^2 = 0.195$).

This indicates that the average effect for the child group is not as certain as it would have been if the effect sizes were homogenous. This non-significant finding is a good example of the importance of examining heterogeneity and the nested nature of effect sizes.

The high degree of between study variability reflects the nature of most age comparisons in developmental and educational psychology. Individual studies usually do not include participants of widely different ages. In the present meta-analysis, only four studies included both children (younger than age 13) and adolescents (13-18); and no studies compared children to adults or adolescents to adults. Thus, age comparisons can only be made between studies, and it is difficult to tease apart true developmental differences from differences in factors such as study design and outcome measures. Further studies are needed that compare the multiple age groups in the same study.

***Initial level of performance.*** Some prior work suggests that low-performing individuals may show a different trajectory of improvement with training compared to higher performing individuals (Terlecki, Newcombe, & Little, 2008). Thus, we tested whether training studies that incorporated a screening procedure to identify low-scorers yielded higher (or lower) effect sizes compared to those that enrolled all participants, regardless of initial performance level. In all, 19 out of 206 studies used a screening procedure to identify and train low scorers. These 19 studies reported significantly larger effects of training ($g = 0.68$, $SE = 0.09$, $m = 19$, $k = 169$) than the remaining 187 studies ($g = 0.44$, $SE = 0.04$, $m = 187$, $k = 869$), $p = .02$, suggesting that focusing on low-scorers instead of testing a random sample can generate a larger magnitude of improvement.

**Outcome measures.** Our final set of moderators concerned differences in how the effects of training were measured.

***Accuracy versus response time.*** Researchers may use multiple outcome measures to assess spatial skills and responses to training. For example, in Mental Rotation tasks, researchers can measure both accuracy and response time. We investigated whether the use of these different measures influenced the magnitude of the effect sizes. The analysis of accuracy and response time was performed only for studies that used Mental Rotation tasks because only these studies consistently measured and reported both. We used Linn and Petersen's definition of Mental Rotation to isolate the relevant studies. Mental Rotation tests such as the Vandenberg and Kuse's Mental Rotations Test (Alington, Leaf, & Monaghan, 1992), Shepard and Metzler (Ozel, Larue, & Molinaro, 2002), and the Card Rotations Test (Deratzou, 2006) were common throughout the literature.

Both response time ($g = 0.69$, $SE = 0.13$, $m = 15$, $k = 41$), and accuracy ($g = 0.31$, $SE = 0.14$, $m = 92$, $k = 305$) improved in response to training. One sample t-tests indicated that the mean effect size differed significantly from zero ($p < .01$), supporting the malleability of mental rotation tasks established above. Reaction time improved significantly more than accuracy did, $p < .05$.

***2 x 2 spatial skills as outcomes.*** Finally, we examined whether our 2 x 2 typology of spatial skills can shed light on differences in the malleability of spatial skills. Do different kinds of spatial tasks respond differently to training? Table 4 gives the mean weighted effect sizes for each of the 2 x 2 framework's spatial skill cells. The table reveals that each type of spatial skill is malleable; all of the effect sizes differed significantly from zero, $p < .01$. Extrinsic, static measures produced the largest gains. However, the only significant difference between categories at an alpha of .01 was between extrinsic, static measures and intrinsic, static measures. Note that extrinsic, static measures include the Water-level Task or Rod and Frame Task, two tests that ask the participant to apply a learned principle to solve the task. In some cases, teaching participants straightforward rules about the tasks (e.g., draw the line representing the water parallel to the floor) may lead to large improvements, although it is not clear that these improvements always endure—(see Liben, 1977). In contrast, intrinsic, static measures may respond much less to training because the researcher cannot tell the subject what particular form to look for. All that can be communicated is the possibility of finding a form, but it is still up to the subject to determine what shape or form is represented. This more general skill may be harder to teach or train. Overall, despite the variety of spatial skills surveyed here in this meta-analysis, our results strongly suggest that performance on spatial tasks unanimously improved with training and the magnitude of training effects was fairly consistent from task to task.

## Discussion

This is the first comprehensive and detailed meta-analysis of the effects of spatial training. Table 4 provides a summary of the main results. The results indicate that spatial skills are highly malleable, and that training in spatial thinking is effective, durable, and transferable. This conclusion holds true across all of the categories of spatial skill that we examined. In addition, our analyses showed that several moderators, most notably the presence or absence of a control group and what the control group did, help to explain the variability of findings. In addition, our novel meta-analytical statistical methods better control for heterogeneity without sacrificing data. We believe that our findings shed light not only on spatial cognition and its development, but also can help guide policy decisions regarding which spatial training programs can be implemented in economically and educationally feasible ways with particular emphasis on connections to STEM disciplines.

## Effectiveness, Durability, and Transfer of Training

We set several criteria for establishing the effectiveness of training and hence the malleability of spatial cognition. The first was simply that training effects should be reliable and at least moderate in size. We found that trained groups showed an average effect size of 0.62, or well over one-half of a standard deviation in improvement. Even when compared to a control group, the size of this effect was approaching medium, 0.47.

The second criterion was that training should lead to durable effects. Although the majority of studies did not include measures of the durability of training, our results indicate that training can be durable. Indeed, the magnitude of training effects was statistically similar for post-tests given immediately after training and after a delay from the end of training. Of course, it is possible that those studies that included delayed assessment of training were specifically designed to enhance the effects or durability of training. Thus, further research is needed to specify what types of training are most likely to lead to durable effects. In addition, it is important to note that very few studies have examined training of more than a few weeks duration. Although such studies are obviously not easy to conduct, they are essential for answering questions regarding the long-term durability of training effects. The third criterion was that training had to be transferable. This was perhaps the most challenging criterion, as the general view has been that spatial training does not transfer or at best leads to only very limited transfer. However, the systematic summary that we have provided suggests transfer is not only possible, but at least in some cases is not necessarily limited to tasks that very closely resemble the training tasks. For example, Kozhevnikov and Thornton (2006) found that interactive, spatially-rich lecture demonstrations of physics material (e.g., Newton's first two laws) generated improvement on a paper folding post-test. In some cases, the tasks involved different materials, substantial delays, and different conceptual demands, all of which are criteria for meaningful transfers that extend beyond a close coupling between training and assessment (Barnett & Ceci, 2002).

Why did our meta-analysis reveal that transfer is possible when other researchers have argued that transfer is not possible (e.g., National Research Council, 2006; Sims & Mayer, 2002; Schmidt & Bjork, 1992)? One possibility is that studies that planned to test for transfer were designed in such a way to maximize the likelihood of achieving transfer. Demonstrating transfer often requires intensive training (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). For example, many studies that achieved transfer effects administered large numbers of trials during training (e.g., Lohman & Nichols, 1990), trained participants over a long period of time (e.g., Terlecki et al., 2008; Wright et al., 2008), or trained participants to asymptote (Feng et al., 2007). Transfer effects also must be large enough to surpass the test-retest effects observed in control groups (Heil et al., 1998). Thus, although it is true that many studies do not find transfer, our results clearly show that transfer is possible if sufficient training or experience is provided.

## Establishing a Spatial Ability Framework

Research on spatial ability needs a unifying approach, and our work contributes to this goal. We took the theoretically motivated 2 x 2 design outlined in Newcombe and Shipley (in press), deriving from work done by Chatterjee (2008), Palmer (1978), and Talmy (2000), and aligned the pre-existing spatial outcome measure categories from the literature with this framework. Comparing this classification to that used in Linn and Petersen's (1985) meta-analysis, we found that their categories mostly fit into the 2 x 2 design, save one broad category that straddles the static and dynamic cells within the intrinsic dimension. Working from a common theoretical framework will facilitate a more systematic approach to researching the malleability of each category of spatial ability. Our results demonstrate that each category is malleable when compared to zero, although comparisons across categories showed few differences in training effects. We hope that the clear definitions of the spatial dimensions will stimulate further research comparing across categories. Such research would allow for better assessment of whether training one type of spatial task leads to improvements in performance on other types of spatial tasks. Finally, this well-defined framework could be used to investigate which types of spatial training would lead directly to improved performance in STEM-related disciplines.

## Moderator Analyses

Despite the large number of studies that have found positive effects of training on spatial performance, other studies have found minimal or even negative effects of interventions (e.g., Faubion, Cleveland, & Harrel,1942; Gagnon, 1985; Johnson, 1991; Kass, Ahlers & Dugger, 1998; Kirby & Boulter, 1999; Larson, 1996; McGillicuddy-De Lisi, De Lisi & Youniss, 1978; Simmons, 1998; Smith, 1998; Vasta, Knott, & Gaze, 1996). We analyzed the influence of several moderators, and taken together, these analyses shed substantial light on possible causes of variations in the influences of training. Considering the effects of these moderators makes the literature on spatial training substantially clearer and more tractable.

**Study design and control group improvement.** An important finding from this meta-analysis was the large and variable improvement in control groups. In nearly all cases, the size of the training-related improvements depended heavily on whether a control group was used and on the magnitude of gains observed within the control groups. A study that compared a trained group to a control group that improved a great deal (e.g., Sims & Mayer, 2002) may have concluded that training provided no benefit to spatial ability, whereas a study that compared its training to an less active control group (e.g., De Lisi & Wolford, 2002) may have shown beneficial effects of training. Thus, we conclude that the mixed results of past research on training can be attributed, in part, to variations in the types of control groups that were used and to the differences in performance observed among these groups.

What accounts for the magnitude of and variability in the improvement of control groups? Typically, control groups are included to account for improvement that might be expected in the absence of training. Improvement in the control groups, therefore, is seen as a measure of the effects of repeated practice in taking the assessments independent of the training intervention. Such practice effects can result from a variety of factors, including familiarity with the mode of testing (e.g., learning to press the appropriate keys in a reaction time task), improved allocation of cognitive skills such as attention and working memory, or learning of relevant test-taking strategies (e.g., gaining a sense of which kinds of foils are likely to be wrong).

In this case, however, we suggest that the improvement in the control groups may not be attributable solely to improvements that are associated with learning about individual tests. The average level of control group improvement in this meta-analysis, 0.45, was substantially larger than the average test-retest effect in other psychometric measures of 0.29 (Hauschknecht et al., 2007). It is difficult to explain why this should be the case unless control groups were learning more than how to take particular tests. The spatial skills of participants in the control groups may have improved because taking spatial tests, especially multiple spatial tests, can itself be a form of training. For example, the act of taking more than one test could allow items across tests to be compared and, potentially, highlight the similarities and differences in item content (Gentner & Markman, 1994; Gentner & Markman, 1997). This could, in turn, suggest new strategies or approaches for solving subsequent tasks and related spatial problems. Additionally, the spatial content used in some control groups led to greater improvement in those control groups. The finding that the overall mean weighted effect size generated from comparisons to spatial filler control groups was significantly smaller than the overall mean weighted effect size generated from comparisons to non-spatial filler control groups is consistent with the claim that spatial learning occurred in the control groups. In summary, although more work is needed to investigate these claims directly, our results call for a broader conception of what constitutes training. A full characterization of spatial training entails not only examining the content of courses or training regimens but also examining the nature of the practice effect that can result from being enrolled in a training study and being tested multiple times on multiple measures.

***Age.*** We did not find a significant effect of age on level of improvement. This is rather surprising considering the large differential in means when comparing young children to adolescents and adults (a 0.17 difference in both cases). The vast majority of comparisons between ages came from separate studies not necessarily testing exactly the same measures and almost certainly running their participants through different protocols. This large heterogeneity in the developmental literature, represented by the estimate of variance $\tau^2$, generates a large standard error for the individual age groups, especially children. The large standard error in turn reduces the likelihood of finding a significant result when comparing age effects. Thus, our analyses highlight the need for further research involving systematic within-study comparisons of individuals of different ages. While our analyses clearly suggest that spatial skills are malleable across the lifespan, such designs would provide a more rigorous test of whether spatial skills are more malleable during certain periods.

***Type of training.*** We did not find that one type of training was superior to any other. This finding may be analogous to the age effect, in that no studies in this meta-analysis compared distinct methods of training, potentially adding to the heterogeneity of the effects. However, we did find that all of the methods of training studied here improved spatial skills and that all of these effects differed significantly from zero, implying that spatial skills can be improved in a variety of ways. Therefore, although the research to determine which method is best is yet to be done, we can say that there is no wrong way to teach spatial skills.

## Differences in the Response to Training

**Sex.** Both men and women responded substantially to training; however the gender gap in spatial skills did not shrink due to training. Of course, our results do not mean that it is impossible to close the gender gap with additional training. Some studies that have used extensive training have indeed found that the gender gap can be attenuated and perhaps eliminated (e.g., Feng et al., 2007). In addition, many training studies have shown that individual differences in initial level of performance moderate the trajectory of improvements with training (Just & Carpenter, 1985; Terlecki et al., 2008). For example, Terlecki et al. showed that female participants who initially scored poorly improved slowly at first but improved more later in training. In contrast, males and females with initially higher scores improved the most early in training. This study did not include low-scoring males. This difference in learning trajectory is important because it suggests that if training periods are not sufficiently long, female participants will appear to benefit less from training and show smaller training-related gains than male participants will. Additionally, Baenninger and Newcombe (1989) pointed out that improvement among females will likely not close the gender gap until improvement among males has reached asymptote, which is difficult to determine. Therefore whether the gender gap *can* be closed, with appropriate methods of training, still remains very much an open question, but what is clear is that both men and women can improve their spatial skills significantly with training.

More generally, efforts that focus on closing the gender gap of specific spatial skills, such as Mental Rotation, may be misplaced. Differences in performance on isolated spatial skills are of interest for theoretical reasons. However, the recent increases in emphasis on decreasing the gender gap in measures of STEM success (i.e., grades and achievement in STEM disciplines) suggest that training individual spatial skills is desirable only if the training translates into success in STEM. It may be possible that STEM success can be achieved without eliminating the gender gap on basic spatial measures. For example, one possible view is that being able to work in STEM fields is dependent on achieving a threshold level of performance rather than being dependent on achieving absolute parity in performance between males and females. Note that this threshold would be a lower limit, *below* which individuals are not likely to enter a STEM field. Our use of the term threshold contrasts with that of Robertson, Smeets, Lubinski, and Benbow (2010), who have argued that there is no *upper* threshold for the relation between various cognitive abilities and STEM achievement and attainment at the highest levels of eminence. The goal of future research perhaps should not be to focus on remediation in order to close the gender gap in basic spatial skills but rather to close the gap in STEM interest and entry into STEM-based occupations.

**Initial level of performance.** Finally, we found that initial level of spatial skills affected the degree of malleability. Participants who started at lower levels of performance improved more in response to training than those who started at higher levels. In part, this effect could stem from a ceiling that limits the improvement of participants who begin at high levels. However, in some studies (e.g., Terlecki et al., 2008), scores were not depressed by ceiling effects, so it is possible that we are seeing the beginnings of asymptotic performance. Nevertheless, it is important to note that improvement was not limited only to those who began at particularly low levels.

### Contributions of the novel meta-analytic approach.

The approach developed by Hedges et al. (2010a, 2010b) helps to control for the fact that most studies in this meta-analysis report results from multiple experiments. Importantly, this method does not require any effect sizes to be disregarded, while correctly taking into account the levels of nesting. The estimation method provided by this approach is robust in many important ways; for example, unlike most estimation routines for hierarchical meta-analyses, it is robust to any misspecification of the weights and does not require the effect sizes to be normally distributed.

By taking nesting into account, the calculations appreciate that there are multiple types of variance across the literature. In addition to taking into account sampling variability, Hedges et al. (2010a, 2010b) estimates the variance between effect sizes from experiments within a single study, $\omega^2$, and the variance between average effect sizes in different studies, $\tau^2$. By using all three factors to calculate the standard error for a mean weighted effect size, this methodology reflects the heterogeneity in the literature. That is, the larger the heterogeneity, the larger the standard error produced, and the less likely comparison groups will be found to be statistically significantly different. It is the combination of this weighting and the robust estimation routine that allows us to be very confident in the significant differences found within our dataset. Two examples from our analyses illustrate well the importance of taking these parameters into account; we found no significant effect of age and no significant differences between the types of training, but the lack of differences may stem in part from the fact that studies tend to include only one (or occasionally two) age groups and to include only one type of training.

## Mechanisms of Learning and Improvement

The evidence suggests that a wide range of training interventions improve spatial skills. The findings of the present analysis suggest that comparing and attempting to optimize different methods of training may serve as an important focus for future research. This process of optimizing training should be informed by our empirical and theoretical knowledge of the mechanisms through which training leads to improvements. Considering the basic cognitive processes, such as attention and memory, required to perform spatial tasks, may inform our efforts to understand how individuals improve on these processes and facilitate relevant training.

Mental rotation is one example of a domain in which the mechanisms of improvement are reasonably well understood. Part of the mechanism is simply that participants become faster at rotating the objects in their minds (Kail & Park, 1992). This source of improvement is reflected in the slope of the line that relates response time to the angular disparity between the target and test figures, but other aspects of performance improve as well. The y-intercept of the line that relates response time to angular disparity also decreases (Terlecki et al., 2008; Heil et al., 1998) and may change more consistently then the slope of this line (Wright et al., 2008). Researchers initially assumed that changes in the y-intercept reflected basic changes in reaction time (e.g., shortened motor response to press the computer key) as opposed to substantive learning. However, recent work suggests that these changes actually may be meaningful and important. For example, Wright et al. (2008) argued that intercept changes following training may reflect improved encoding of the stimuli. They suggest that training interventions need not focus exclusively on training the mental transformation process, which targets the slope, but should also focus on facilitating initial encoding since this should also improve mental rotation performance (Amorim, Isableu, & Jarraya, 2006). Individual differences also moderate the impact of training on mental rotation (e.g., Terlecki et al., 2008). For example, Just and Carpenter (1985) found that as training progresses, individuals who were high in spatial ability were more adept and flexible at performing rotations of items about non-principal axes, suggesting that they were able to adapt to the variety of coordinate systems represented by the test items.

Training-related improvements in mental rotation and other spatial tasks also likely occur through some basic cognitive pathways, such as improved attention and memory. Spatial skills are obviously affected by the amount of information that can be held simultaneously in memory. Many spatial tasks require holding in working memory the locations of different objects, landmarks, etc. Research indicates that individual differences in (spatial) working memory capacity are responsible for some of the observed differences in performance on spatial tasks. As Hegarty and Waller (2005) suggest, individuals who cannot hold information in working memory may “lose” the information that they are trying to transform. Several lines of research indicate that spatial attentional capacity improves with relevant training (e.g., Castel, Pratt, & Drummond, 2005; Feng et al., 2007; Green & Bavelier, 2007). Instructions or training that improves working memory and attentional capacities are therefore likely to enhance the amount of information that participants can think about and act on.

A good example of interventions that improve working memory capacity comes from research on the effect of video game playing on performance on a host of spatial attention tasks. Video-game players performed substantially better in several tasks that tap spatial working memory, such as a subitization/enumeration task that requires participants to estimate the number of dots shown on a screen in a brief presentation. Most people can recognize (subitize) five or fewer dots without counting; after this point, performance begins to decline in typical non-video game players and counting is required. However, video-game players can subitize a larger number of dots—approximately seven or eight (e.g, Green & Bevalier, 2003, 2007). Thus, the additional subitization capacity suggests that videogame players can hold a greater number of elements in working memory and act upon them. Likewise, video-game players appear to have a smaller attentional blink, allowing them to take in and use more information across the range of spatial attention. In addition, many spatial tasks and transformations can be accomplished by the application of rules and strategies. For example, Just and Carpenter (1985) found that many participants completed mental rotations not by rotating the entire stimulus mentally but by comparing specific aspects of the figure and checking to determine whether the corresponding elements would match after rotation. Likewise, advancement in the well-known spatial game Tetris is often accomplished by the acquisition of specific rules and strategies that help the participant learn when and how to fit new pieces into existing configurations. Furthermore, spatial transformations in chemistry (Stieff, 2007) are often accomplished by learning and applying specific rules that specify the spatial properties of molecules. In summary, part of learning and development (and hence, one of the effects of training) may be the acquisition and appropriate application of strategies or rules.

## Educational and Policy Implications

The present meta-analysis has important implications for policy decisions. It suggests that spatial skills are moderately malleable and that a wide variety of training procedures can lead to meaningful and durable improvements in spatial ability. Although our analysis appears to indicate that certain recreational activities, such as video games, are comparable to formal courses in the extent to which they can improve spatial skills, we cannot assume that all students will engage in this type of spatial skills training in their spare time. Our analysis of the impact of initial performance on the effect of training suggests that those students with initially poor spatial skills are most likely to benefit from spatial training. In summary, our results argue for the implementation of formal programs targeting spatial skills.

Prior research gives us a way to estimate the consequences of administering spatial training on a large scale in terms of producing STEM outcomes. Wai, Lubinski, Benbow and Steiger (2010) have established that STEM professionals often have superior spatial skills, even after holding constant correlated abilities such as math and verbal skills. Using a nationally representative sample, Wai et al. (2009) found that the spatial skills of individuals who obtained at least a bachelor's degree in engineering were 1.58 SD greater than the general population (Lubinski, personal communication, August 14, 2011; Wai, personal communication, August 17, 2011). The very high level of spatial skills that seems to be required for success in engineering (and other STEM fields) is one important factor that limits the number of Americans who are able to become engineers (Wai et al., 2009; Wai et al., 2010) and thus contributes to the severe shortage of STEM workers in the United States.

In this paper, we have demonstrated that spatial skills can be improved. To put this finding in context, we asked how much difference this improvement would make in the number of students whose spatial skills meet or exceed the average level of engineers' spatial skills. We calculated the expected percentage of individuals who would have a Z-score of + 1.58 before and after training. To provide the most conservative estimate, we used the effect of spatial training that was derived from the most rigorous studies, the mixed design, which included control groups and measured spatial skills both before and after training. The mean effect size for these studies was 0.40. As shown in Figure 6, increasing the population level of spatial skills by 0.40

standard deviations would approximately double the number of people who would have levels of spatial abilities equal to or greater than that of current engineers.

We recognize that our estimate entails many assumptions. Perhaps most importantly, our estimate of the impact of increased spatial training implies a causal relationship between spatial training and improvement in STEM learning or attainment. Unfortunately, this assumption has rarely been tested, and has, to our knowledge, never been tested using a rigorous experimental design, such as randomized control trials. Thus, the time is ripe to conduct full, prospective, and randomized tests of whether and how spatial training can enhance STEM learning.

**An example of when and how spatial training might benefit STEM learning.** There are many ways in which spatial training may facilitate STEM attainment, achievement, and learning. Comprehensive discussions on these STEM topics have been offered elsewhere (e.g., Newcombe, 2010; Sorby & Baartmans, 1996). Here we concentrate on one example that we believe highlights particularly well the potential of spatial training to improve STEM achievement and attainment.

Although it certainly may be useful to ground early STEM learning in spatially-rich approaches, our results suggest that it may be possible to help students even after they have finished high school. Specifically, we suggest that spatial training might increase the retention of students who have expressed interest in and perhaps already begun to study STEM topics in college. One of the most frustrating challenges of STEM education is the high dropout rate among STEM majors. For example, in Ohio public universities, more than 40% of the students who declare a STEM major leave the STEM field before graduation (Price, 2010). The relatively high attrition rates among self-identified STEM students are particularly disappointing because these students are the “low hanging fruit” in terms of increasing the number of STEM workers in the U.S. They have already attained the necessary prerequisites to pursue a STEM field at the college level, yet even many of these highly qualified students do not complete a STEM degree. Thus an intervention that could help prevent early dropout among STEM majors might prove to be particularly helpful.

We suggest that spatial training might help lower the dropout rate among STEM majors. The basis for this claim comes from analyses (e.g., Hambrick & Meinz, 2011; Hambrick et al., in press; Uttal & Cohen, in press) of the trajectory of importance for spatial skills in STEM learning. Psychometrically-assessed spatial skills strongly predict performance early in STEM learning. However, psychometrically-assessed spatial skills actually become *less* important as students progress through their STEM coursework and move toward specialization. For example, Hambrick et al. (in press) showed that psychometric tests of spatial skills predicted performance among novice geologists but did *not* predict performance in an expert-level geology mapping task among experts. Likewise, psychometrically-assessed spatial skills predicted initial performance in physics coursework but became less important after learning was complete (Khozhenikov, Motes, & Hegarty, 2007). Experts can rely on a great deal of semantic knowledge of the relevant spatial structures and thus can make judgments without having to perform classic mental spatial tasks such as rotation or two- to three-dimensional visualization. For example, experts know about many geological sites and might know the underlying structure simply from learning about it in class or via direct experience. At a more abstract level, geology experts might be able to solve spatial problems by analogy, thinking about how the structure of other well-known outcrops might be similar or different from the one they are currently analyzing. Similarly, expert chemists often do not need to rely on mental rotation to reason about the spatial properties of two molecules because they may know, semantically, that the target and stimulus are chiral (i.e. mirror images) and hence can respond immediately without having to rotate the stimulus mentally to match the target. This decision could be made quickly, regardless of the degree of angular disparity, because the chemist knows the answer as a semantic fact and hence mental rotation is not required.

Such findings and theoretical analyses led Uttal and Cohen (in press) to propose what that they termed the Catch-22 of spatial skills in early STEM learning. Students who are interested in STEM but have relatively low levels of spatial skills may face a frustrating challenge: They may have difficulty performing the mental operations that are needed to understand chemical molecules, geological structures, engineering designs, etc. Moreover, they may also have difficulty understanding and using the many spatially-rich paper or computer-based representations that are used to communicate this information (See Cohen & Hegarty, 2007; Stieff, 2007; Uttal & Cohen, in press). If these students could just get through the early phases of learning that appear to be particularly dependent on decontextualized spatial skills, then their lack of spatial skills might become less important as semantic knowledge increased. Unfortunately their lack of spatial skills keeps them from getting through the early classes, and many drop out. Thus spatial skills may act as a gatekeeper for students interested in STEM; those with low spatial skills may have particular problems getting through the very challenging introductory-level classes.

Spatial training of the form reviewed in this paper could be particularly helpful for STEM students with low spatial skills. Even a modest increase in the ability to rotate figures, for example, could help some students solve more organic chemistry problems and thus be less likely to drop out. In fact, some research (e.g, Sorby & Baartmans, 2000; Sorby, 2009) does suggest that spatial training focusing on engineering students who self-identify as having problems with spatial tasks can be particularly helpful, resulting both in very large gains in psychometrically-assessed spatial skills and lower drop-out rates in early engineering classes that appear to depend heavily on spatial abilities.

Of course, spatial training at earlier ages might be even more helpful. For example, Cheng and Mix (2011; see also Mix and Cheng, in press) recently demonstrated that practicing spatial skills improved math performance among first- and second-graders. Results indicate that spatial training is effective at a variety of skill levels and ages; further research is needed to determine how effective this training will be in improving STEM learning.

**Selecting an intervention.** There is not a single answer to the question of "what works best?" or "what should we do to improve spatial skills?" Perhaps the most important finding from this meta-analysis is that several different forms of training can be highly successful. Decisions about what types of training to use depend on one's goals as well as the amount of time and other resources that can be devoted to training. Here we give two examples of training that has worked well and that may not require substantial resources.

One example of a highly effective but easy to administer form of training comes from the work of McAuliffe, Pibrum, Reynolds, and colleagues. They have demonstrated that adding spatially-challenging activities to standard courses (e.g., high school physics) can further improve spatial skills. For example, in one study (McAuliffe, 2003), two days of training students in a physics class to use two- and three-dimensional representations consistently led to improvement and transfer to a spatially-demanding task, reading a topographical map. Improvement was compared to students performing normal course work. This treatment did not require extensive intervention or the use of expensive materials, and it was incorporated into standard classes. It was administered in two consecutive class periods on different days, and the post-test was a visualizing topography test administered the day following the completion of the training. McAuliffe found an overall effect size of 0.64. Therefore, with relatively simple interventions, implemented in a traditional high school course, participants improve on spatially challenging post-tests.

The positive returns gained from classroom instruction should not limit the teaching tools available for spatial ability. For example, there is great excitement about the possibility of using videogames in both formal and informal education (e.g., Federation of American Scientists, 2006; Foreman et al., 2004; Gee, 2003). Our results highlight the relevance of videogames for improving spatial skills. For example, Feng et al. (2007) investigated the effects of videogame playing on spatial skills, including transfer to mental rotation tasks. They focused on action (*Medal of Honor*; single user role playing) versus non-action (*Ballance*; 3-D puzzle solving) videogames. A total of 10 hours of training was administered over four weeks. Participants who played the action game performed much better than those who played the control game. The average effect size was 1.19. This result indicates that playing active games has the potential to enhance spatial thinking substantially, even when compared to a strong control group. These activities can be done outside of school and hence do not need to displace in-school activities. The policy question here is how to encourage this kind of game playing.

The above examples address the training needs of adolescents and adults. While elementary school children also play videogames, there are likely different ways to enhance the spatial ability of very young children: Including block play, puzzle play, the use of spatial language by parents and teachers, and the use of gesture. For an overview of spatial interventions for younger children, see Newcombe (2010).

## Conclusion

Most efforts aimed at educational reform focus on reading or science. This focus is appropriate because achievement in these areas is readily measured and of great interest to educators and policy makers. However one potentially negative consequence of this focus is that it misses the opportunity to train basic skills, such as spatial thinking, that can underlie performance in multiple domains. Recent research is beginning to remedy this deficit, with an increase in work examining the link between spatial thinking and performance in STEM disciplines such as biology (Lennon, 2000), chemistry (Coleman & Gotch, 1998; Wu & Shah, 2004) and physics (Kozhevnikov, et al., 2007); as well as the relation of spatial thinking to skills relevant to STEM performance in general (Gilbert, 2005), such as reasoning about scientific diagrams (Stieff, 2007). Our hope is that our findings on how to train spatial skills will inform future work of this type and, ultimately, lead to highly effective ways to improve STEM performance.

For many years, much of the focus of research on spatial cognition and its development has been on the biological underpinnings of these skills (e.g., Eals & Silverman, 1994; Kimura, 1992; Kimura, 1996; McGillicuddy-De Lisi & De Lisi, 2001; Silverman & Eals, 1992). Perhaps as a result, relatively little research has focused on the environmental factors that influence spatial thinking and its improvement. Our results clearly indicate that spatial skills are malleable. Even a small amount of training can improve spatial reasoning in both males and females, and children and adults. Spatial training programs therefore may play a particularly important role in the education and enhancement of spatial skills and mathematics and science more generally.

## References

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

* Alderton, D. L. (1989, March). *The fleeting nature of sex differences in spatial ability.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. Retrieved from ERIC database. (ED307277)

* Alington, D. E., Leaf, R. C., & Monaghan, J. R. (1992). Effects of stimulus color, pattern, and practice on sex differences in mental rotations task performance. *Journal of Psychology*, *126*(5), 539-553.

Amorim, M.-A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: “Body analogy” for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327-347. doi:10.1037/0096-3445.135.3.327

* Asoodeh, M. M. (1993). *Static visuals vs. computer animation used in the development of spatial visualization* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9410704)

Astur, R. S., Ortiz, M. L., & Sutherland, R. J. (1998). A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference. *Behavioural Brain Research, 93*(1-2), 185-190. doi:10.1016/S0166-4328(98)00019-9

* Azzaro, B. Z. H. (1987). *The effects of recreation activities and social visitation on the cognitive functioning of elderly females* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8810733)

Baenninger, M., & Newcombe, N. S. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*(5-6), 327-344. doi:10.1007/BF00287729

* Baldwin, S. L. (1984). *Instruction in spatial skills and its effect on math achievement in the*

*intermediate grades.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8501952)

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637. doi:10.1037//0033-2909.128.4.612

* Barksy, R. D., & Lachman, M. E. (1986). Understanding of horizontality in college women: Effects of two training procedures. *International Journal of Behavioral Development*, *9*(1), 31–43. doi:10.1177/016502548600900103

* Bartenstein, S. (1985). *The effects of drawing instruction on spatial abilities and academic achievement of year I medical students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 0558939)

* Basak, C., Boot, W., Voss, M., & Kramer, A. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, *23*(4), 765-777. doi:10.1037/a0013494

* Basham, J. (2006). *Technology integration complexity model: A systems approach to technology acceptance and use in education* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3202062)

* Batey, A. H. (1986). *The effects of training specificity on sex differences in spatial ability*. (Unpublished doctoral dissertation). College Park, MD: University of Maryland.

* Beikmohamadi, M. (2006). *Investigating visuospatial and chemistry skills using physical and computer models.* (Doctoral dissertation). Available from ProQuest Dissertations & Theses database. (UMI No. 3234856)

* Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial

visualization skills of middle school boys and girls. *American Educational Research Journal*, *25*, 51-71. doi:10.3102/00028312025001051

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*(2), 115-147. doi:10.1037/0033-295X.94.2.115

* Blatnick, R. A. (1986). *The effect of three-dimensional models on ninth-grade chemistry scores.* Available from ProQuest Dissertations and Theses database. (UMI No. 8626739)

* Boakes, N. J. (2006). *The effects of origami lessons on students' spatial visualization skills and achievement levels in a seventh-grade mathematics classroom.* Available from ProQuest Dissertations and Theses database. (UMI No. 3233416)

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis.* Englewood, NJ: Biostat, Inc.

Bornstein, M. H. (1989). Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin, 105*(2), 179-197. doi:10.1037/0033-2909.105.2.179

* Boulter, D. R. (1992). *The effects of instruction on spatial ability and geometry performance.* Available from ProQuest Dissertations and Theses database. (UMI No. MM76356)

* Braukmann, J. (1991). *A comparison of two methods of teaching visualization skills to college students*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 9135946)

* Brooks, L. A. (1992). *The effects of training upon the increase of transactive behavior*. Available from ProQuest Dissertations and Theses database. (UMI No. 9304639)

* Calero, M., & García, T. (1995). Entrenamiento de la competencia espacial en ancianos [Training of spatial competence in the elderly]. *Anuario de Psicología*, *64*(1), 67-81.

Retrieved from http://dialnet.unirioja.es/servlet/articulo?codigo=2947065

The Campbell Collaboration. (2001). *Guidelines for Preparation of Review Protocols.* Retrieved from http://www.campbellcollaboration.org/artman2/uploads/1/C2_Protocols_guidelines_v1.pdf

Capozzoli, M. V., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, *27*(1), 3-23. doi:10.2307/3315487

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Castel, A. D., Pratt, J., Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica 119,* 217-230. doi:10.1016/j.actpsy.2005.02.004

* Cathcart, W. G. (1990). Effects of Logo instruction on cognitive style. *Journal of Educational Computing Research, 6*(2)*,* 231-242. doi:10.1016/0747-5632(94)00034-F

* Center, L. E. (2004). *Issues of gender and peer collaboration in a problem-solving computer activity.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3135894)

Chatterjee A. (2008). The neural organization of spatial thought and language. *Seminars in Speech and Language*, 29, 226-238. doi:10.1055/s-0028-1082886

* Chatters, L. B. (1984*). An assessment of the effects of video game practice on the visual motor perceptual skills of sixth grade children*. (Unpublished doctoral dissertation). University of Toledo, Toledo, OH.

Cheng, Y., & Mix, K. (2011). Does spatial training improve children's mathematics ability.

Presented at the bi-annual meetings of the Society for Research in Educational Effectiveness, Washington.

* Cherney, I. (2008). Mom, let me play more computer games: They improve my mental rotation skills. *Sex Roles, 59*(11-12), 776-786. doi:10.1007/s11199-008-9498-z

* Chevrette, P. A. (1987). *A comparison of college students' performances with alternate simulation formats under cooperative or individualistic class structures.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8808743)

* Chien, S.-C. (1986). *The effectiveness of animated and interactive microcomputer graphics on children's development of spatial visualization ability/mental rotation skills (educational instruction, technology).* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8618765)

* Clark, S. K. (1996). *Effect of spatial perception level on achievement using two and three dimensional graphics.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9702272)

* Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal*, *98*(2), 171-186.

* Cockburn, K. S. (1995). *Effects of specific toy playing experiences on the spatial visualization skills of girls ages 4 and 6.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9632241)

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Cohen, C. A., & Hegarty, M. (2007). Individual differences in use of external visualizations to perform an internal visualization task. *Applied Cognitive Psychology, 21,* pages701-711. doi:10.1002/acp.1344

Coleman, S. L., & Gotch, A. J. (1998). Spatial perception skills of chemistry students. *Journal of Chemical Education*, *75*(2), 206-209. doi:10.1021/ed075p206

* Comet, J. J. K. (1986). *The effect of a perceptual drawing course on other perceptual tasks*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8609012)

* Connolly, P. E. (2007). *A comparison of two forms of spatial ability development treatment.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3291192)

* Crews, J. W. (2008). *Impacts of a teacher geospatial technologies professional development project on student spatial literacy skills and interests in science and technology in grade 5--12 classrooms across Montana.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3307333)

* Curtis, M. A. (1992). *Spatial visualization and leadership in teaching multiview orthographic projection: An alternative to the glass box.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9222441)

* D'Amico, A. (2006). Potenziare la memoria di lavoro per prevenire l'insuccesso in matematica. [Training of working memory for preventing mathematical difficulties]. *Età Evolutiva*, *83, 90*-99.

* Dahl, R. D. (1984). *Interaction of field dependence independence with computer-assisted instruction structure in an orthographic projection lesson.* (Doctoral dissertation).

Available from ProQuest Dissertations & Theses database. (UMI No. 8423698)

* Day, J. D., Engelhardt, J. L., Maxwell, S. E., & Bolig, E. E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. *Journal of Educational Psychology*, *89*(2), 358-368. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022066302003446

* Del Grande, J. J. (1986). *Can grade two children's spatial perception be improved by inserting a transformation geometry component into their mathematics program?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. NL31402)

* De Lisi, R., & Cammarano, D. M. (1996). Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behavior*, *12*(3), 351-361. doi:10.1016/0747-5632(96)00013-1

* De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *Journal of Genetic Psychology*, *163*(3), 272-282. doi:10.1080/00221320209598683

* Deratzou, S. (2006). *A qualitative inquiry into the effects of visualization on high school chemistry students' learning process of molecular structure*. (Unpublished doctoral dissertation), Drexel University, Philadelphia, PA.

* Dixon, J. K. (1995). Limited English proficiency and spatial visualization in middle school students' construction of the concepts of reflection and rotation. *Bilingual Research Journal*, *19*(2), 221-247. Retrieved from http://www.ncela.gwu.edu

* Dorval, M. & Pepin, M. (1986). Effect of playing a video game on a measure of spatial visualization. *Perceptual and Motor Skills*, *62*(1), 159-162. Retrieved from

http://archive.ammonsscientific.com/

* Duesbury, R. T. (1992). *The effect of practice in a three dimensional CAD environment on a learner's ability to visualize objects from orthographic projections.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 0572601)

* Duesbury, R. T., & O'Neil, H. F. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*(3), 249-260. doi:10.1037/0021-9010.81.3.249

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. doi:10.1037/0012-1649.43.6.1428

* Dziak, N. (1985). *Programming computer graphics and the development of concepts in geometry.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8510566)

Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, *15*(2), 95-105. doi:10.1007/s10508-006-9168-6

* Edd, L. A. (2001). *Mental rotations, gender, and experience.* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1407726)

* Ehrlich, S., Levine, S., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*(6), 1259-1268. doi:10.1037/0012-1649.42.6.1259

* Eikenberry, K. (1988). *The effects of learning to program in the computer language Logo on the visual skills and cognitive abilities of third grade*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8827343)

Eliot, J. (1987). *Models of psychological space: Psychometric, developmental and experimental approaches.* New York: Springer-Verlag.

Eliot, J., & Fralley, J. S. (1976). Sex differences in spatial abilities. *Young Children*, *31*(6), 487-498.

* Embretson, S. E. (1987). Improving the measurement of spatial aptitude by dynamic testing. *Intelligence*, *11*(4), 333-358. doi:10.1016/0160-2896(87)90016-X

* Englehardt, J. L. (1987). *A Comparison of Static and Dynamic Assessment Procedures and their relation to independent performance in preschool children.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8721823)

* Eraso, M. (2007). *Connecting visual and analytic reasoning to improve students' spatial visualization abilities: A constructivist approach.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3298580)

* Fan, C-F. D. (1998). *The effects of dialogue between a teacher and students on the representation of partial occlusion in drawings by Chinese children.*(Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9837575)

Faubion, R. W., Cleveland, E. A., & Harrell, T. W. (1942). The influence of training on mechanical aptitude test scores. *Educational and Psychological Measurement*, *2*, 91-94. doi:10.1177/001316444200200109

Federation of American Scientists. (2006). Harnessing the power of video games for learning.

*Summit on Educational Games.* Retrieved from http://www.fas.org/gamesummit/Resources/Summit%20on%20Educational%20Games.pdf

* Feng, J. (2006). *Cognitive training using action video games: A new approach to close the gender gap*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. MR21389)

* Feng, J., Spence, I., & Pratt, J. (2007). Playing and action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850-855. doi:10.1111/j.1467-9280.2007.01990.x

* Ferguson, C. (2008). *A comparison of instructional methods for improving the spatial-visualization ability of freshman technology seminar students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3306534)

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal*, *14*(1), 51-71. doi:10.3102/00028312014001051

* Ferrara, M. M. (1992). *The effects of imagery instruction on ninth graders' performance of a visual synthesis task.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9232498)

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

* Fitzsimmons, R. W. (1995). *The relationship between cooperative student pairs' van Hiele levels and success in solving geometric calculus problems following graphing calculator-*

*assisted spatial training.* (Doctoral dissertation). Available ProQuest Dissertations and Theses database. (UMI No. 9533552)

Foreman, J., Gee, J. P., Herz, J. C., Hinrichs, R., Prensky, M., & Sawyer, B. (2004). Game-based learning: How to delight and instruct in the 21st century. *EDUCAUSE Review, 39*(5), 50-66. Retrieved from http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume39/GameBasedLearningHowtoDelighta/157927

* Frank, R. E. (1986). *The emergence of route map reading skills in young children*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8709064)

French, J., Ekstrom, R. and Price, L. (1963) *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

* Funkhouser, C. P. (1990). *The effects of problem-solving software on the problem-solving ability of secondary school students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9026186)

* Gagnon, D. (1985). Video games and spatial skills: An exploratory study. *Educational Communication and Technology*, *33*(4), 263-275. doi:10.1007/BF02769363

* Gagnon, D. M. (1986). *Interactive versus observational media: The influence of user control and cognitive styles on spatial learning.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 8616765)

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*(4), 337-353. doi:10.1006/jecp.2000.2594

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave/Macmillan.

* Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, *36*(6), 556-563. doi:10.1016/j.intell.2007.12.003

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5*(3), 152-158. doi:10.1111/j.1467-9280.1994.tb00652.x

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45-56. doi:10.1111/j.1467-9280.1994.tb00652.x

* Gerson, H. B. P., Sorby, S. A., Wysocki, A., & Baartmans, B. J. (2001). The development and assessment of multimedia software for improving 3-D spatial visualization skills. *Computer Applications in Engineering Education*, 9(2), 105-113. doi:10.1002/cae.1012

* Geva, E., & Cohen, R. (1987). *Transfer of spatial concepts from Logo to map-reading* (Report No. 143). Toronto: Ontario Department of Education.

Ghafoori, B., & Tracz, S. M. (2001). Effectiveness of cognitive behavioral therapy in reducing classroom disruptive behaviors: A meta-analysis. *ERIC Document Reproduction Service* (ED457182). Retrieved from http://www.eric.ed.gov/PDFS/ED457182.pdf

* Gibbon, L. W. (2007). *Effects of LEGO Mindstorms on convergent and divergent problem-solving and spatial abilities in fifth and sixth grade students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3268731)

Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization Science Education* (pp. 9-27). Dordrecht, the Netherlands: Springer.

* Gillespie, W. H. (1995). *Using solid modeling tutorials to enhance visualization skills.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9606359)

* Gitimu, P. N., Workman, J. E., Anderson, M. A. (2005). Influences of training and strategical information processing style on spatial performance in apparel design. *Career and Technical Education Research, 30*(3), 147-168. doi:10.5328/CTER30.3.147

* Gittler, G., & Gluck, J. (1998). Differential transfer of learning: Effects of instruction in descriptive geometry on spatial test performance. *Journal for Geometry and Graphics*, *2*(1), 71-84. Retrieved online from http://www.heldermann.de/JGG/jgg02.htm

* Godfrey, G. S. (1999). *Three-dimensional visualization using solid-model methods: A comparative study of engineering and technology students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9955758)

* Golbeck, S. L. (1998). Peer collaboration and children's representation of the horizontal surface of liquid. *Journal of Applied Developmental Psychology*, *19*(4), 571-592. doi:10.1016/S0193-3973(99)80056-2

* Goodrich, G. A. (1992). *Knowing which way is up: Sex differences in understanding horizontality and verticality.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9301646)

* Goulet, C., Talbot, S., Drouin, D., & Trudel, P. (1988). Effect of structured ice hockey training on scores on field-dependence/independence. *Perceptual and Motor Skills*, *66*(1), 175-181. Retrieved from http://archive.ammonsscientific.com/

* Guillot, A. & Collet, C. (2004). Field dependence-independence in complex motor skills. *Perceptual and Motor Skills 98*(2), 575-583. Retrieved from

http://archive.ammonsscientific.com/

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423(29)*, 534-537. Retrieved from http://www.nature.com/nature/journal/v423/n6939/full/nature01647.html

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94. doi:10.1111/j.1467-9280.2007.01853.x

* Gyanani, T., & Pahuja, P. (1995). Effects of peer tutoring on abilities and achievement. *Contemporary Educational Psychology 20(4), 469-475.* doi:10.1006/ceps.1995.1034

Hambrick, D.Z., Libarkin, J.C., Petcovic, H.L., Baker, K.M., Elkins, J., Callahan, C., … Ladue, N. D. (in press). A test of the circumvention-of-limits hypothesis in geological bedrock mapping. *Journal of Experimental Psychology: General.*

Hambrick, D. Z., Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science, 20*(5) 275-279. doi: 10.1177/0963721411422061

Hauschknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Gerrard, M. O. M. (2007). Retesting in selection: A meta-analysis of practice effects for tests of cognitive ability. *Journal of Applied Psychology*, *92*(2), 373-385. doi:10.1037/0021-9010.92.2.373

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, *29*(3), 446-493.

Hedges, L. V. & Olkin, I. (1986). Meta-analysis: A review and a new view. *Educational Researcher, 15*(8), 14-21.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010a). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

doi:10.1002/jrsm.5

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010b). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(2), 164-165. doi:10.1002/jrsm.17

* Hedley, M. (2008). *The use of geospatial technologies to increase students' spatial abilities and knowledge of certain atmospheric science content.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3313371)

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah, A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121-169). New York, NY: Cambridge University Press.

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. doi:10.1016/j.intell.2005.09.005

* Heil, M., Rossler, F., Link, M., & Bajric, J. (1998). What is improved if a mental rotation task is repeated--the efficiency of memory access, or the speed of a transformation routine? *Psychological Research*, *61*(2), 99-106. doi:10.1007/s004260050016

* Higginbotham, C. A. (1993). *Evaluation of spatial visualization pedagogy using concrete materials vs. computer-aided instruction.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1355079)

Hoffman, D., & Singh, M. (1997). Salience of visual parts. Cognition, 63, 29–78. doi:10.1016/S0010-0277(96)00791-3

Hornik, K. (2011). The R FAQ. Retrieved from http://CRAN.R-project.org/doc/FAQ/R-FAQ.html ISBN 3-900051-08-9.

* Hozaki, N. (1987). *The effects of field dependence/independence and visualized instruction in a lesson of origami, paper folding, upon performance and comprehension.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8804052)

* Hsi, S., Linn, M. C., & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education, 86,* 151-158. Retrieved from www.jee.org/1997/april/505.pdf

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*(2), 250-261. doi:10.1037/0021-9010.78.2.250

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications, Inc.

Huttenlocher, J. & Presson, C. C. (1979). The coding and transformation of spatial information. *Cognitive Psychology*, *11*(3), 375-394. doi:10.1016/0010-0285(79)90017-3

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, *314*, 599-600. doi:10.1126/science.1132154

* Idris, N. (1998). *Spatial visualization, field dependence/independence, van Hiele level, and achievement in geometry: The influence of selected activities for middle school students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9900847)

Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine, 30*. doi:10.1002/sim.4172

* Janov, D. R. (1986). *The effects of structured criticism upon the perceptual differentiation and*

*studio compositional skills displayed by college students in an elementary art education course.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8626321)

* Johnson, J. E. (1991). *Can spatial visualization skills be improved through training that utilizes computer-generated visual aids?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9134491)

* Johnson-Gentile, K., Clements, D., & Battista, M. (1994). Effects of computer and noncomputer environments on students' conceptualizations of geometric motions. *Journal of Educational Computing Research*, *11*(2), 121-140. doi:10.2190/49EE-8PXL-YY8C-A923

Johnson, M. H., Munakata, Y., & Gilmore, R. O. (Eds.). (2002). *Brain development and cognition: A reader* (2nd ed.). Oxford: Blackwell Publishers.

* July, R. A., (2001). *Thinking in three dimensions: Exploring students' geometric thinking and spatial ability with Geometer's sketchpad.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3018479).

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137-172. doi:10.1037/0033-295X.92.2.137

Kail, R., & Park, Y.-S. (1992). Global developmental change in processing time. *Merrill-Palmer Quarterly, 38*(4), 525-541.

Kalaian, H.A., & Raudenbush, S.W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*(3), 227-235.

* Kaplan, B. J., & Weisberg, F. B. (1987). Sex differences and practice effects on two visual-

spatial tasks. *Perceptual and Motor Skills, 64*(1), 139-142. Retrieved from http://archive.ammonsscientific.com/

* Kass, S. J., Ahlers, R. H., & Dugger, M. (1998). Eliminating gender differences through practice in an applied visual spatial task. *Human Performance, 11*(4), 337-349. doi:10.1207/s15327043hup1104_3

* Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction, 25*(1), 45–74. doi:10.1080/07370000709336702

* Kastens, K. A., Kaplan, D., & Christie-Blick, K. (2001). Development and evaluation of "Where are We?" map skills software and curriculum. *Journal of Geoscience Education*, *49*(3), 249-266.

* Keehner, M., Lippa, Y., Motello, D. R., Tendick, F., & Hegarty, M. (2006). Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology, 20*(4), 487–503. doi:10.1002/acp.1198

Kimura, D. (1992). Sex differences in the brain. *Scientific American*, *267*(3), 118-125.

Kimura, D. (1996). Sex, sexual orientation and sex hormones influence on human cognitive function. *Current Opinion in Neurobiology*, *6*(2), 259-263. doi:10.1016/S0959-4388(96)80081-X

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

* Kirchner, T., Forns, M., Amador, J. A. (1989). Effect of adaptation to the task on solving the Group Embedded Figures Test. *Perceptual and Motor Skills, 68*(3), 1303-1306.

Konstantopoulos, S. (2010) Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*, 61-76.

* Kovac, R. J. (1985). *The effects of microcomputer assisted instruction, intelligence level, and gender on the spatial ability skills of junior high school industrial arts students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8608967)

Kozhevnikov, M. & Hegarty, M. (2001). A dissociation between object-manipulation and perspective-taking spatial abilities. *Memory & Cognition, 29,* 745-756. doi:10.3758/BF03200477

Kozhevnikov, M., Hegarty, M. & Mayer, R. E. (2002). Revising the visualizer/verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, *20*, 47-77. doi:10.1207/S1532690XCI2001_3

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, *33*, 710-726. doi:10.3758/BF03195337

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*(4), 549-579. doi:10.1080/15326900701399897

Kozhevnikov, M., Motes, M. Rasch, B. & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*, 397–417. doi:10.1002/acp.1192

* Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education and Technology*, *15*(1), 111-132. doi:10.1007/s10956-006-0361-0

* Krekling, S. and Nordvik, H. (1992). Observational training improves adult women's

performance on Piaget's water-level task. *Scandinavian Journal of Psychology, 33*(2), 117-124. doi:10.1111/j.1467-9450.1992.tb00891.x

* Kwon, O.-N., Kim, S.-H., & Kim, Y. (2002). Enhancing spatial visualization through Virtual Reality (VR) on the web: Software design and impact analysis. *Journal of Computers in Mathematics and Science Teaching*, *21*(1), 17-31.

* Kwon, O.-N. (2003). Fostering spatial visualization ability through web-based virtual-reality program and paper-based program. In C.-W. Chung, C.-K. Kim, W. Kim, T.-W. Ling, & K.-H. Song (Eds.), *Web and Communication Technologies and Internet-related Social Issues* (pp. 710-706). Berlin: Springer.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. doi:10.2307/2529310

* Larson, K. L. (1996). *Persistence of training effects on the coordination of perspective tasks*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9700682)

* Larson, P., Rizzo, A. A., Buckwalter, J. G., Van Rooyen, A., Kratz, K., Neumann, U., . . . Van Der Zaag, C. (1999). Gender issues in the use of virtual environments. *CyberPsychology & Behavior*, *2*(2), 113-123. doi:10.1089/cpb.1999.2.113

* Lee, K.-O. (1995). The effects of computer programming training on the cognitive development of 7-8 year-old children. *Korean Journal of Child Studies*, *16*(1), 79-88.

* Lennon, P. (1996). *Interactions and enhancement of spatial visualization, spatial orientation, flexibility of closure, and achievement in undergraduate microbiology*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9544365)

Lennon, P. A. (2000). Improving students' flexibility of closure while presenting biology content. *The American Biology Teacher*, *62*(3), 177-180. doi:10.1662/0002-7685(2000)062[0177:ISFOCW]2.0.CO;2

Liben, L. S. (1977). The facilitation of long-term memory improvement and operative development. *Developmental Psychology, 13*(5), 501-508. doi:10.1037/0012-1649.13.5.501

* Li, C. (2000). Instruction effect and developmental levels: A study on water-level task with Chinese children ages 9-17. *Contemporary Educational Psychology, 25*(4), 488-498. doi:10.1006/ceps.1999.1029

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479-1498. doi:10.1111/1467-8624.ep7252392

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist*, *48*(12), 1181-1201. doi:10.1037/0003-066X.48.12.1181

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

* Lizarraga, M. L. S., & Garcia Ganuza, J. M. (2003). Improvement of mental rotation in girls and boys. *Sex Roles*, *49*(5-6), 277-286. doi:10.1023/A:1024656408294

* Lohman, D. F. (1988). Spatial abilities as traits, processes and knowledge. In R. J. Stermberg (Ed.), *Advances in the Psychology of Human Intelligence*, (Vol. 4, pp. 181-248). Hillsdale, NJ: Erlbaum.

* Lohman, D. F., & Nichols, P. D. (1990). Training spatial abilities: Effects of practice on

rotation and synthesis tasks. *Learning and Individual Differences*, *2*(1), 67-93. doi:10.1016/1041-6080(90)90017-B

* Longstreth, L. E., & Alcorn, M. B. (1990). Susceptibility of Wechsler spatial ability to experience with related games. *Educational and Psychological Measurement*, *50*(1), 1-6. doi:10.1177/0013164490501001

* Lord, T. R. (1985). Enhancing the visuo-spatial aptitude of students. *Journal of Research in Science Teaching*, *22*(5), 395-405. doi:10.1002/tea.3660220503

* Luckow, J. J. (1984). *The effects of studying logo turtle graphics on spatial ability.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8504287)

* Luursema, J., Verwey, W. B., Kommers, P. A. M., Geelkerken, R. H., & Vos, H. J. (2006). Optimized conditions for computer-assisted anatomical learning. *Interacting with Computers*, *18*(5), 1123-1138. doi:10.1016/j.intcom.2006.01.005

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

* Martin, D. J. (1991). *The effect of concept mapping on biology achievement of field-dependent students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9201468)

* McAuliffe, C. (2003). *Visualizing topography: Effects of presentation strategy, gender, and spatial ability.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3109579)

* McClurg, P. A., & Chaille, C. (1987). Computer games: Environments for developing spatial cognition? *Journal of Educational Computing Research*, *3*(1), 95-111.

* McCollam, K. M. (1997). *The modifiability of age differences in spatial visualization.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9827490)

* McCuiston, P. J. (1991). Static vs. dynamic visuals in computer-assisted instruction. *Engineering Design Graphics Journal*, *55*(2), 25-33.

McGillicuddy-De Lisi, A. V., De Lisi, R., & Youniss, J. (1978). Representation of the horizontal coordinate with and without liquid. *Merrill-Palmer Quarterly, 24*(3), 199-208.

McGillicuddy-De Lisi, A. V., & De Lisi, R. (Eds.). (2001). *Biology, society, and behavior: The development of sex differences in cognition*. Westport, CT: Ablex Publishing.

* McKeel, L. P. (1993). *The effects of drawing simple isometric projections of self-constructed models on the visuo-spatial abilities of undergraduate females in science education.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9334783)

* Merickel, M. (1992, October). *A study of the relationship between virtual reality (perceived realism) and the ability of children to create, manipulate, and utilize mental images for spatially related problem solving.* Paper presented at the Annual Convention of the National School Boards Association, Orlando, FL. Retrieved from ERIC database. (ED 352942).

* Miller, E. A. (1985). *The use of logo-turtle graphics in a training program to enhance spatial visualization.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. MK68106)

* Miller, G. G., & Kapel, D. E. (1985). Can non-verbal, puzzle type microcomputer software affect spatial discrimination and sequential thinking skills of 7th and 8th graders?

*Education*, *106*(2), 160-167.

* Miller, J. (1995). *Spatial ability and virtual reality: Training for interface efficiency.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1377036)

* Miller, R. B., Kelly, G. N., Kelly, J. T. (1988). Effects of Logo computer programming experience on problem solving and spatial relations ability. *Contemporary Educational Psychology*, *13*(4), 348-357. doi:10.1016/0361-476X(88)90034-3

Mix, K. S. & Cheng, Y. L. (in press). The relation between space and math: Developmental and educational implications. In J. Benson (Ed.),z *Advances in Child Development and Behavior* (vol. 42). Elsevier.

* Mohamed, M. A. (1985). *The effects of learning LOGO computer language upon the higher cognitive processes and the analytical/global cognitive styles of elementary school students*. (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

* Moody, M. S. (1998). *Problem-solving strategies used on the Mental Rotations Test: Their relationship to test instructions, scores, handedness, and college major*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9835349)

* Morgan, V. R. L. F. (1986). *A comparison of an instructional strategy oriented toward mathematical computer simulations to the traditional teacher-directed instruction on measurement estimation*. (Unpublished doctoral dissertation). Boston University, Boston, MA.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386. doi:10.1177/1094428106291059

* Mowrer-Popiel, E. (1991). *Adolescent and adult comprehension of horizontality as a function of cognitive style, training, and task difficulty*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9033610)

* Moyer, T. O. (2004). *An investigation of The Geometer's Sketchpad and van Hiele levels*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3112299)

* Mshelia, A.Y. (1985). *The influence of cognitive style and training on depth picture perception among non-western children* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8523208)

* Mullin, L. N. (2006). *Interactive navigational learning in a virtual environment: Cognitive, physical, attentional, and visual components*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3214688)

National Research Council. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

National Research Council. (2006). *Learning to think spatially*. Washington D. C.: The National Academies Press.

National Research Council. (2009). *Engineering in K-12 education: Understanding the status and improving the Prospects*. Washington, D.C.: The National Academies Press.

* Newman, J. L. (1990). *The mediating effects of advance organizers for solving spatial relations problems during the menstrual cycle in college women.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1341060)

Newcombe, N. S., Mathason, L., & Terlecki, M. (2002). Maximization of spatial competence: More important than finding the cause of sex differences. In A. McGillicuddy-De Lisi &

R. De Lisi (Eds.), *Biology, society and behavior: The development of sex differences in cognition* (183-206). Westport, CT:Ablex.

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, 29-35, 43. doi:10.1037/A0016127

Newcombe, N. S. & Shipley, T. F. (in press). Thinking about spatial thinking: New typology, new assessments. In J. S. Gero (Ed.), *Studying visual and spatial reasoning for design creativity*. New York, NY: Springer.

* Noyes, D. L. K. (1997). *The effect of a short-term intervention program on the development of spatial ability in middle school.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9809155)

* Odell, M. R. L. (1993). *Relationships among three-dimensional laboratory models, spatial visualization ability, gender and earth science achievement.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9404381)

* Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*(1), 33-58. doi:10.1016/0193-3973(94)90005-1

* Olson, J. (1986). *Using Logo to supplement the teaching of geometric concepts in the elementary school classroom (mathematics, locus-of-control, spatial visualization, computer equity, sex differences)*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8611553)

Orwin, R. G. (1983). A fail safe N for effect size in meta-analysis. *Journal for Educational Statistics*, *8*(2), 157-159. doi:10.3102/10769986008002157

* Ozel, S., Larue, J., & Molinaro, C. (2002). Relation between sport activity and mental rotation:

Comparison of three groups of subjects. *Perceptual and Motor Skills, 95*(3), 1141-1154. Retrieved from http://archive.ammonsscientific.com/

* Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching, 21*(5), 507-516. doi:10.1002/tea.3660210508

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259-303). Hillsdale, NJ: Lawrence Erlbaum.

* Parameswaran, G. (1993). *Dynamic assessment of sex differences in spatial ability.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9328925)

* Parameswaran, G. (2003). Age, gender and training in children's performance of Piaget's horizontality task. *Educational Studies, 29*(2-3), 307-319. doi:10.1080/030556903200009207

* Parameswaran, G., & De Lisi, R. (1996). Improvements in horizontality performance as a function of type of training. *Perceptual and Motor Skills, 82*(2), 595-603. Retrieved from http://archive.ammonsscientific.com/

* Pazzaglia, F., & De Beni, R. (2006). Are people with high and low mental rotation abilities differently susceptible to the alignment effect? *Perception, 35*(3), 369-383.

* Pearson, R. C. (1991). *The effect of an introductory filmmaking course on the development of spatial visualization and abstract reasoning in university students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9204546)

* Pennings, A. H. (1991). Altering the strategies in embedded-figure and water-level tasks via

instruction: A neo-Piagetian learning study. *Perceptual and Motor Skills*, *72*(2), 639-660. doi:10.2466/PMS.72.2.639-660

* Perez-Fabello, M., & Campos, A. (2007). Influence of training in artistic skills on mental imaging capacity. *Creativity Research Journal*, *19*(2-3), 227-232. doi:10.1080/10400410701397495

* Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, *28*(1), 39-58.

* Philleo, T. J. (1997). *The effect of three-dimensional computer microworlds on the spatial ability of preadolescents.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9801756)

* Piburn, M. D., Reynolds, S. J., McAuliffe, C., Leedy, D. E., Birk, J. P., & Johnson, J. K. (2005). The role of visualization in learning from computer-based images. *International Journal of Science Education*, *27*(5), 513-520. doi:10.1080/09500690412331314478

* Pleet, L. J. (1991). *The effects of computer graphics and Mira on acquisition of transformation geometry concepts and development of mental rotation skills in grade eight.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9130918)

* Pontrelli, M. J. (1990). *A study of the relationship between practice in the use of a radar simulation game and ability to negotiate spatial orientation problems.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9119893)

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields.

*Economics of Education Review. 29*(6). 901-910. doi:10.1016/j.econedurev.2010.07.00

* Pulos, S. (1997). Explicit knowledge of gravity and the water-level task. *Learning and Individual Differences, 9*(3), 233-247. doi:10.1016/S1041-6080(97)90008-X

* Qiu, X. (2006). *Geographic information technologies: An influence on the spatial ability of university students?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3221520)

* Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences*, *40*(3), 609-619. doi:10.1016/j.paid.2005.07.015

* Rafi, A., Samsudin, K. A., & Said, C. S. (2008). Training in spatial visualization: The effects of training method and gender. *Educational Technology & Society, 11*(3), 127-140. Retrieved from www.ifets.info/journals/11_3/10.pdf

* Ralls, R. (1998). *Age and computer training performance: A test of training enhancement through cognitive practice*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9808657)

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31-74. doi:10.1111/1529-1006.00004

Raudenbush, S.W., & Bryk, A.S. (2002) *Hierarchical linear models: Applications and data analysis methods.* Sage Publications, Thousand Oaks, CA.

* Robert, M., & Chaperon, H. (1989). Cognitive and exemplary modelling of horizontality representation on the Piagetian water-level task. *International Journal of Behavioral Development, 12*(4), 453-472. doi:10.1177/016502548901200403

* Rosenfield, D. W. (1985). *Spatial relations ability skill training: The effect of a classroom exercise upon the skill level of spatial visualization of selected vocational education high school students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8518151)

Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, *19*(6), 346-351. doi:10.1177/0963721410391442

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3) 638-461. doi:10.1037/0033-2909.86.3.638

* Rush, G., & Moore, D. M. (1991). Effects of restructuring training and cognitive style. *Educational Psychology, 11*(3-4), 309-321. doi:10.1080/0144341910110307

* Russell, T. L. (1989). *Sex differences in mental rotation ability and the effects of two training methods.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8915037)

* Saccuzzo, D. P., Craig, A. S., Johnson, N. E., Larson, G. E. (1996). Gender differences in dynamic spatial abilities. *Personality and Individual Differences, 21*(4), 599-607. doi:10.1016/0191-8869(96)00090-6

Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, *22*(6), 800–811. doi:10.1037/a0013091

* Savage, R. (2006). *Training wayfinding: Natural movement in mixed reality.* (Doctoral

dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3233674)

* Schaeffer, P. D., & Thomas, J. (1998). Difficulty of a spatial task and sex difference in gains from practice. *Perceptual and Motor Skills*, *87*(1), 56-58. Retrieved from http://archive.ammonsscientific.com/

* Schaie, K., & Willis, S. L. (1986). Can decline in adult intellectual functioning be reversed? *Developmental Psychology*, *22*(2), 223-232. doi:10.1037/0012-1649.22.2.223

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*(4), 207-217. doi:10.1111/j.1467-9280.1992.tb00029.x

* Schmitzer-Torbert, N. (2007). Place and response learning in human virtual navigation: Behavioral measures and gender differences. *Behavioral Neuroscience*, *121*(2), 277-290. doi:10.1037/0735-7044.121.2.277

* Schofield, N. J., & Kirby, J. R. (1994). Position location on topographical maps: Effects of task factors, training and strategies. *Cognition and Instruction*, *12*(1), 35-60. doi:10.1080/09593330.2010.543930

* Scribner, S. A. (2004). *Novice drafters' spatial visualization development: Influence of instructional methods and individual learning styles.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3147123)

* Scully, D. F. (1988). *The use of a measure of three dimensional ability as a tool to evaluate three dimensional computer graphic learning ability.* Available from ProQuest Dissertations and Theses database. (UMI No. 0564844)

* Seddon, G. M., & Shubber, K. E. (1984). The effects of presentation mode and colour in

teaching the visualization of rotation in diagrams of molecular structures. *Research in Science in Technological Education*, *2*(2), 167-176. doi:10.1080/0263514840020208

* Seddon, G. M., & Shubber, K. E. (1985a). The effects of colour in teaching the visualization of rotation in diagrams of three-dimensional structures. *British Educational Research Journal*, *11*(3), 227-239. doi:10.1111/j.2044-8279.1986.tb02655.x

* Seddon, G. M., & Shubber, K. E. (1985b). Learning the visualization of three-dimensional spatial relationships in diagrams at different ages in Bahrain. *Research in Science and Technological Education*, *3*(2), 97-108. doi:10.1080/0263514850030104

* Seddon, G. M., Eniaiyeju, P. A., & Jusoh, I. (1984). The visualization of rotation in diagrams of three-dimensional structures. *American Educational Research Journal*, *21*(1), 25-38. doi:10.3102/00028312021001025

* Sevy, B. A. (1984). *Sex-related differences in spatial ability: The effects of practice*. (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

* Shavalier, M. (2004). The effects of CAD-like software on the spatial ability of middle school students. *Journal of Educational Computing Research*, *31*, 37-49. doi:10.2190/X8BU-WJGY-DVRU-WE2L

Shea, D. L., Lubinski, D., Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604-614. doi:10.1037/0022-0663.93.3.604

Sherman, J. A. (1967). Problem of sex differences in space perception and aspects of intellectual functioning. *Psychological Review*, *74*(4), 290-299. doi:10.1037/h0024723

* Shubbar, K. E. (1990). Learning the visualisation of rotations in diagrams of three dimensional

structures. *Research in Science and Technological Education*, *8*(2), 145–154. doi:10.1080/0263514900080206

* Shyu, H-Y. (1992). *Effects of learner control and learner characteristics on learning a procedural task.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9310304)

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and Generation of Culture* (pp. 533-549). New York, NY: Oxford Universities Press.

* Simmons, N. A. (1998). *The effect of orthographic projection instruction on the cognitive style of field dependence-independence in human resource development graduate students*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9833489)

* Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, *16*(1), 97-115. doi:10.1002/acp.759

* Smith, G. G. (1999). *Computers, computer games, active control and spatial visualization strategy.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9837698)

* Smith, G. G., Gerretson, H., Olkun, S., Yuan, Y., Dogbey, J., Erdem, A. (2009). Stills, not full motion, for interactive spatial training: American, Turkish and Taiwanese female pre-service teachers learn spatial visualization. *Computers & Education*, *52*(1), 201-209. doi:10.1016/j.compedu.2008.07.011

* Smith, J. P. (1998). *A quantitative analysis of the effects of chess instruction on the*

*mathematics achievement of southern, rural, black secondary students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9829096)

* Smith, J. and M. Sullivan (1997, November). *The Effects of Chess Instruction on Students' Level of Field Dependence/Independence.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Memphis, TN. Retrieved from ERIC database. (ED 415257).

* Smith, R. W. (1996). *The effects of animated practice on mental rotation tests.* Available from ProQuest Dissertations and Theses database. (UMI No. 1380535)

* Smyser, E. M. (1994). *The effects of "The Geometric Supposers:" Spatial ability, van Hiele levels, and achievement.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9427802)

* Snyder, S. J. (1988). *Comparing training effects of field independence-dependence to adaptive flexibility on nurse problem-solving.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8818546)

* Sokol, M. D. (1986). *The effect of electromyographic biofeedback assisted relaxation on Witkin embedded figures test scores.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8615337)

* Sorby, S. A. (2007). *Applied educational research in developing 3-D spatial skills for engineering students.* Unpublished manuscript. Michigan Tech University: Houghton, MI.

Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education, 31*(3), 459-480.

doi:10.1080/09500690802595839

* Sorby, S. A., & Baartmans, B. J. (1996). A course for the development of 3-D spatial visualization skills. *Engineering Design Graphics Journal*, *60*(1), 13-20.

Sorby, S. A., & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first-year engineering students. *Journal of Engineering Education,* 301-307.

* Spangler, R. D. (1994). *Effects of computer-based animated and static graphics on learning to visualize three-dimensional objects.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9423654)

* Spencer, K. (2008). *Preservice elementary teachers' two-dimensional visualization and attitude toward geometry: Influences of manipulative format.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3322956)

* Sridevi, K., Sitamma, M., & Krishna-Rao, P. V. (1995). Perceptual organisation and yoga training. *Journal of Indian Psychology, 13*(2), 21-27.

* Stewart, D. R. (1989). *The relationship of mental rotation to the effectiveness of three-dimensional computer models and planar projections for visualizing topography.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8921422)

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219-234. doi:10.1016/j.learninstruc.2007.01.012

Stiles, J. (2008). *The fundamentals of brain development: Integrating nature and nurture.* Cambridge, MA: Harvard University Press.

* Storey Vasu, E., & Tyler, D. (1997). A comparison of the critical thinking skills and spatial

ability of fifth grade children using simulation software or logo. *Journal of Computing in Childhood Education, 8*(4), 345-363. Retrieved from http://vnweb.hwwilsonweb.com/hww/Journals/getIssues.jhtml?sid=HWW:EDI&id=-2479

* Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology, 15*(1), 13-32. doi:10.1016/0193-3973(94)90004-3

* Sundberg, S. E. (1994). *Effect of spatial training on spatial ability and mathematical achievement as compared to traditional geometry instruction.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9519018)

*Talbot, K., & Haude, R. (1993). The relationship between sign language skill and spatial visualization ability: Mental rotation of three-dimensional objects. *Perceptual and Motor Skills*, *77*(3), 1387-1391. Retrieved from http://archive.ammonsscientific.com/

Talmy, L. (2000). *Toward a cognitive semantics. Volume 1: Concept structuring systems.* Cambridge, Massachusetts: MIT Press.

* Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*(7), 996-1013. doi:10.1002/acp.1559

* Thomas, D. A. (1996). *Enhancing spatial three-dimensional visualization and rotational ability with three-dimensional computer graphics*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9702192)

* Thompson, K., & Sergejew, A. A. (1998). Two-week test-retest of the WAIS-R and WMS-R subtests, the Purdue pegboard, trailmaking, and the mental rotation test in women.

*Australian Psychologist*, *33*(3), 228-230. doi:10.1080/00050069808257410

* Thomson, V. L. R. (1989). *Spatial ability and computers elementary children's use of spatially defined microcomputer software*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. NL51052)

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago, Illinois: University of Chicago Press.

* Tillotson, M. L. (1984). *The effect of instruction in spatial visualization on spatial abilities and mathematical problem solving*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8429285)

* Tkacz, S. (1998). Learning map interpretation: Skill acquisition and underlying abilities. *Journal of Environmental Psychology*, *18*(3), 237-249. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0272494498900949

* Trethewey, S. O. (1990). *Effects of computer-based two and three dimensional visualization training for male and female engineering graphics students with low, middle, and high levels of visualization skill as measured by mental rotation and hidden figures tasks*. (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

* Turner, G. F. W. (1997). *The effects of stimulus complexity, training, and gender on mental rotation performance: A model-based approach*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9817592)

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, *13*, 407- 433. doi:10.1016/0010-0285(81)90016-5

United Nations Development Programme. (2009). *Human development report*. New York, NY: Palgrave Macmillan.

* Ursyn-Czarnecka, A. Z. (1994). *Computer art graphics integration of art and science.*

(Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9430792)

Uttal, D. H., & Cohen, C. A. (In press). Spatial thinking and STEM education: When, why and how? In B. Ross (Ed.), *Psychology of Learning and Motivation* (vol. 57). New York, NY: Academic Press.

U.S. Department of Education (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Retrieved online from http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-604. Retrieved from http://archive.ammonsscientific.com/

* Vasta, R., Knott, J. A., & Gaze, C. E. (1996). Can spatial training erase the gender differences on the water-level task? *Psychology of Women Quarterly*, *20*(4), 549-567. doi:10.1111/j.1471-6402.1996.tb00321.x

* Vazquez, J. L. (1990). *The effect of the calculator on student achievement in graphing linear functions.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9106488)

* Verner, I. M. (2004). Robot manipulations: A synergy of visualization, computation and action for spatial instruction. *International Journal of Computers for Mathematical Learning, 9*(2), 213-234. doi:10.1023/B:IJCO.0000040892.46198.aa

Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic Bulletin and Review*, *14*(1), 23-38. doi:10.3758/BF03194024

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. doi:10.1037/0033-2909.117.2.250

Waddington, C. H. (1966). *Principles of development and differentiation*. New York, NY: Macmillan Company.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101(4)*, 817-835. doi:10.1037/a0016127

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in Science, Technology, Engineering, and Mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102*(4), 860-871. doi:10.1037/a0019454

* Wallace, B., & Hofelich, B. G. (1992). Process generalization and the prediction of performance on mental imagery tasks. *Memory & Cognition, 20*(6), 695-704. doi:10.3758/BF03202719

* Wang, C.-H., Chang, C.-Y., & Li, T.-Y. (2006). The comparative efficacy of 2D-versus 3D-based media design for influencing spatial visual skills. *Computers in Human Behavior*, *23*(4), 1943-1957. doi:10.1016/j.chb.2006.02.004

* Werthessen, H. W. (1999). *Instruction in spatial skills and its effect on self-efficacy and achievement in mental rotation and spatial visualization.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9939566)

* Wideman, H. H., & Owston, R. D. (1993). Knowledge base construction as a pedagogical activity. *Journal of Educational Computing Research, 9*(2), 165-196. doi:10.2190/LUHJ-

5MRH-WT8T-1J6N

* Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction, 18*(1), 30-41. doi:10.1016/j.learninstruc.2006.09.009

* Weidenbauer, G., Schmid, J., & Jansen-Osmann, P. (2007). Manual training of mental rotation. *European Journal of Cognitive Psychology*, *19*(1), 17-36. doi:10.1080/09541440600709906

Wilson, S. J. & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventative Medicine*, *33*(2), S130-S143. doi:10.1016/j.amepre.2007.04.011

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive and disruptive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*(1), 136-149. doi:10.1037/0022-006X.71.1.136

* Workman, J. E., Caldwell, L. F., & Kallal, M. J. (1999). Development of a test to measure spatial abilities associated with apparel design and product development. *Clothing and Textiles Research Journal*, *17*(3), 128-133. doi:10.1177/0887302X9901700303

* Workman, J. E. & Lee, S. H. (2004). A cross-cultural comparison of the apparel spatial visualization test and paper folding test. *Clothing and Textiles Research Journal*, *22*(1-2), 22-30. doi:10.1177/0887302X0402200104

* Workman, J. E. & Zhang, L. (1999). Relationship of general and apparel spatial visualization ability. *Clothing and Textiles Research Journal*, *17*(4), 169-175. doi:10.1177/0887302X9901700401

* Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 14*(4), 763-771.

doi:10.3758/PBR.

Wu, H.-K. & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465-492. doi:10.1002/sce.10126

* Xuqun, Y., & Zhiliang, Y. (2002). The properties of cognitive process in visuospatial relations encoding. *Acta Psychologica Sinica*, *34*(4), 344-350.

* Yates, L. G. (1986). Effect of visualization training on spatial ability test scores. *Journal of Mental Imagery, 10*(1), 81-91.

* Yates, B. C. (1988). *The computer as an instructional aid and problem-solving tool: An experimental analysis of two instructional methods for teaching spatial skills to junior high school students.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8903857)

* Yeazel, L. F. (1988). *Changes in spatial problem-solving strategies during work with a two-dimensional rotation task*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8826084)

* Zaiyouna, R. (1995). *Sex differences in spatial ability: The effect of training in mental rotation.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. MM04165)

* Zavotka, S. L. (1987). Three dimensional computer graphics animation: A tool for spatial skill instruction. *Educational Communication and Technology Journal*, *35*(3), 133-144. doi:10.1007/BF02793841

# Footnotes

[1] Cases in which the training outcome was a single summary score from an entire psychometric test (e.g., WPPSI-R or the Kit of Factor Referenced Tests) and provided no breakdown of the subtests were excluded. We were concerned that the high internal consistency of standardized test batteries would inflate improvement, overstating the malleability of spatial skills. Therefore this exclusion is a conservative approach to analyzing the malleability of spatial skills, and ensures that any effects found are not due to this confound.

[2] In some cases training could not be classified into a 2 x 2 cell, for example studies that used experience in athletics as training (Guillot & Collet, 2004; Ozel, Larue, & Molinaro, 2002). Experiments such as these were not included in the analyses of transfer within and across cells of the 2 x 2 framework.

[3] For the analyses involving the Human Development Index, the rankings for each country were taken from the Human Development Report 2009 (United Nations Development Programme, 2009). The 2009 HDI ranking goes from 1 (best) to 182 (worst) and is created by combining indicators of life expectancy, educational attainment and income. Norway had an HDI ranking of 1, Niger had an HDI ranking of 182, and the U.S. had an HDI of 13. HDI rankings were first published in 1990, and therefore it was not possible to get the HDI at the time of publication for each paper. Therefore to be consistent, we used the 2009 (year the analyses were performed) HDI rankings to correlate with the magnitude of the effect sizes.

[4] The studies that we excluded were Gyanani & Pahuja (1995, India), Li (2000, China), Mshelia (1985, Nigeria), Rafi, Samsudin, & Said (2008, Malaysia), Seddon, Eniaiyeju, & Jusoh (1984, Nigeria), Seddon & Shubber (1984, Bahrain), Seddon & Shubber (1985a, 1985b, Bahrain), Shubbar (1990, Bahrain), Smith et al. (2009, Turkey), Sridevi, Sitamma, & Krishna-

Rao (1995, India), and Xuqun & Zhiliang (2002, China).

Table 1

*Defining characteristics of the outcome measure categories and their correspondence to categories used in prior research.*

| Spatial Skills described by the 2 x 2 classification | Description | Examples of measures | Linn and Petersen (1985) | Carroll (1993) |
|---|---|---|---|---|
| **Intrinsic and Static** | Perceiving objects, paths, or spatial configurations amidst distracting background information. | Embedded Figures Tasks, Flexibility of Closure, Mazes | Spatial Visualization | Visuospatial Perceptual Speed |
| **Intrinsic and Dynamic** | Piecing together objects into more complex configurations, visualizing and mentally transforming objects, often from 2D to 3D or vice versa. Rotating 2D or 3D objects | Form Board, Block Design, Paper Folding, Mental Cutting, Mental Rotations Test, Cube Comparison, Purdue Spatial Visualization test, Card Rotation Test | Spatial Visualization, Mental Rotation | Spatial Visualization, Spatial Relations/ Speeded Rotation |
| **Extrinsic and Static** | Understanding abstract spatial principles, such as horizontal invariance or verticality | Water-level, Water-clock, Plumb-line, Cross-bar, Rod and Frame test | Spatial perception | Not included |

| | | | | |
|---|---|---|---|---|
| **Extrinsic and Dynamic** | Visualizing an environment in its entirety from a different position | Piaget's Three Mountains Task, Guilford-Zimmerman-spatial orientation | Not included | Not included |

Table 2

*Defining characteristics of training categories and their correspondence to the training categories used by Baenninger and Newcombe (1989).*

| **Type of training** | **Description** | **Baenninger and Newcombe (1989)** |
|---|---|---|
| **Video game training** | Video game used during treatment to improve spatial reasoning | Indirect training |
| **Course training** | Semester long spatially relevant course used to improve spatial reasoning | Indirect training |
| **Spatial task training** | Training utilizes spatial task to improve spatial reasoning | Specific training |

Table 3

*Characteristics of the 206 studies included in the meta-analysis after the exclusion of outliers.*

| **Coded variables** | **n (studies)** | **% of studies** |
|---|---|---|
| Participant characteristics | | |
| Gender composition | | |
| All males | 10 | 5 |
| All females | 18 | 9 |
| Both males and females | 48 | 24 |
| Not specified* | 130 | 62 |
| Age of participants in years^ | | |
| Younger than 13 | 53 | 26 |
| 13 – 18 (inclusive) | 39 | 19 |
| Older than 18 | 118 | 57 |
| Study methods and procedures | | |
| Study design^ | | |
| Mixed design | 123 | 59 |
| Between-subjects | 55 | 27 |

| | | |
|---|---|---|
| Within-subjects only | 31 | 15 |
| Days from end of training to post-test^ | | |
| None (immediate post-test) | 137 | 67 |
| 1-6 | 17 | 8 |
| 7-31 | 37 | 18 |
| More than 31 | 4 | 1 |
| Transfer^ | | |
| No Transfer | 45 | 22 |
| Transfer within 2 x 2 cell | 94 | 46 |
| Transfer across 2 x 2 cells | 51 | 25 |
| 2 x 2 spatial skill cells as outcome measures^ | | |
| Intrinsic, Static | 52 | 25 |
| Intrinsic, Dynamic | 189 | 92 |
| Extrinsic, Static | 14 | 6 |
| Extrinsic, Dynamic | 15 | 7 |
| Training categories | | |
| Video Games | 24 | 12 |

| | | |
|---|---|---|
| Courses | 42 | 21 |
| Spatial Task training | 138 | 67 |
| Prescreened to include only low-scorers | 19 | 9 |
| Study characteristics | | |
| Published | 95 | 46 |
| Publication year (for all articles) | | |
| 1980s | 55 | 27 |
| 1990s | 93 | 45 |
| 2000s | 58 | 28 |
| Location of study^ | | |
| Australia | 2 | 1 |
| Austria | 1 | 1 |
| Canada | 14 | 7 |
| France | 2 | 1 |
| Germany | 5 | 2 |
| Greece | 1 | 1 |
| Israel | 1 | 1 |

| | | |
|---|---|---|
| Italy | 3 | 1 |
| Korea | 6 | 3 |
| Norway | 1 | 1 |
| Spain | 4 | 2 |
| Taiwan, Republic of China | 2 | 1 |
| The Netherlands | 2 | 1 |
| United Kingdom | 1 | 1 |
| United States | 163 | 79 |

*Note*. [*]Data were not reported in a way that separate effect sizes could be obtained for each sex. [^]Percentages do not sum to 100% because of studies that tested multiple age groups, used multiple study designs, used life experience as the intervention, included outcome measures from multiple cells of the 2 x 2, or tested participants from more than one country.

Table 4

*Summary of the moderators considered and corresponding results. Mean weighted effect size g (SE), m, k.*

| | | | | |
|---|---|---|---|---|
| Malleability of Spatial Skills | | | | |
| • Malleable | Overall | Treatment Only | Control Only | |
| | .47 (.04), 206, 1038 | .62 (.04), 106, 365$^{a}$ | .45 (.04), 106, 372$^{b}$ | |
| • Durable | Immediate Post-Test | Delayed Post-Test | | |
| | .48 (.05), 137, 611 | .44 (.08), 65, 384 | | |
| • Transferable | No Transfer | Within 2 x 2 Cell | Across 2 x 2 Cells | |
| | .45 (.09), 45, 272 | .51 (.05), 94, 448 | .55 (.10), 51, 175 | |
| Study Design | Within-Subject | Between-Subject | Mixed Design | |
| | .75 (.08), 31, 160* | .43 (.09), 55, 304 | .40 (.05), 123, 574 | |
| Control Group Activity | | | | |
| • Retesting Effect | Pre-test / Post-test on a Single Test | Repeated Practice | Pre-test / Post-test on Spatial Battery | Pre-test / Post- test Non-spatial Battery |
| | .33 (.04), 34, 111$^{a}$ | .75 (.17), 7, 27$^{b}$ | .46 (.07), 34, 109 | .40 (.11), 9, 36 |
| • Spatial Filler | Spatial Filler (Control Group) | Non-Spatial Filler (Control Group) | Overall for Spatial Filler Controls | Overall for Non-spatial Filler Controls |
| | .51 (.06), 49, 160 | .37 (.05), 46, 159 | .33 (.05), 70, 315$^{a}$ | .56 (.06), 69, 309$^{b}$ |
| Type of Training | Course | Videogames | Spatial Task | |
| | .41 (.11), 42, 154 | .54 (.12), 24, 89 | .48 (.05), 138, 786 | |
| Sex | Male Improvement | Female Improvement | | |
| | .54 (.08), 63, 236 | .53 (.06), 69, 250 | | |

| | | | |
|---|---|---|---|
| Age | Children | Adolescents | Adults |
| | .61 (.09), 53, 226 | .44 (.06), 39, 158 | .44 (.05), 118, 654 |
| Initial Level of Performance | Studies that used only low scoring subjects | Studies that did not separate subjects | |
| | .68 (.09), 19, 169[a] | .44 (.04), 187, 869[b] | |
| Accuracy vs RT | Accuracy | Response Time | |
| | .31 (.14), 99, 347[c] | .69 (.14), 15, 41[d] | |
| 2 x 2 Spatial Skills Outcomes | All categories differed from zero, *p* < .01 | | |
| • Intrinsic, Static | .32 (.10), 52, 166 | | |
| • Intrinsic, Dynamic | .44 (.05), 189, 637 | | |
| • Extrinsic, Static | .69 (.10), 14, 148 | | |
| • Extrinsic, Dynamic | .49 (.13), 15, 45 | | |

*Note.* Key to significance markings:

[ab] indicates the two groups differ at *p* < .01;

[cd] indicates the two groups differ at *p* < .05;

* indicates that group differs at *p* < .01 from all other groups;

Table 5

*Mean weighted effect sizes favoring males, g (SE), m, k, for the sex separated comparisons.*

| | **Pre-test** | **Post-test** |
|---|---|---|
| **Control Group** | .29 (.07), 29, 79* | .24 (.06), 29, 79* |
| **Treatment Group** | .37 (.08), 29, 79* | .26 (.05), 29, 79* |

*Note*. * $p < .01$ when compared to zero, indicating a male advantage. $g$ = effect size indicating male advantage; $m$ = the number of effect sizes, and $k$ = number of effect sizes.

# Appendix A

Coding scheme used to classify studies included in the meta-analysis.

## Publication Status

Articles from peer-reviewed journals were considered to be published, as were research articles in book chapters. Papers presented at conferences, agency and technical reports, and dissertations were all considered to be unpublished work. If we found both the dissertation and the published version of a paper, we counted the study only once as a published article. If any portion of a dissertation appeared in press, the work was considered published.

## Study Design

The experimental design was coded into one of three mutually exclusive categories. *Within-subject,* defined as a pre-test/post-test for only one subject group. *Between-subject,* defined as a post-test only for a control and treatment group; and *mixed design,* defined as a pre-test/post-test for both control and treatment groups.

## Control Group Design and Details

For each control group, we noted whether a single test or a battery of tests was administered. We also noted the frequency with which participants received the tests (i.e. repeated practice or pre-test/post-test only). Finally, if the control group was administered a filler task, we determined if it was spatial in nature.

## Type of Training

We separated the studies into three training categories:

**Video game training.** In these studies, the training involved playing a video or computer game (e.g. Zaxxon, in which players navigate a fighter jet through a fortress while shooting at enemy planes). Because many types of spatial training use computerized tasks that share some similarities with video games, we defined a video game as one designed primarily with an entertainment goal in mind rather than one designed specifically for educational purposes. For example, we did not include interventions that involved learning the programming language Logo because they typically are not presented as games.

**Course training.** These studies tested the effect of being enrolled in a course that was presumed to have an impact on spatial skills. Inclusion in the *Course* category indicated that either the enrollment in a semester long course was the training manipulation (e.g. Engineering graphics course) or the participant took part in a course that required substantial spatial thinking (e.g. chess lessons, geology).

**Spatial task training**. Spatial training was defined as studies that used practice, strategic instruction, or computerized lessons. Spatial training often was administered in a psychology laboratory.

## Typology: Spatial Skill Trained and Tested (See Methods and Table 1)

## Sex

Whenever possible, effect sizes were calculated separately for males and females, but many authors did not report differences broken down by sex. In cases where separate means were not provided for each sex, we contacted authors for the missing information and received responses in eight cases. If we did not receive a reply from the authors, or the author reported the information was not available, we coded sex as not specified.

## Age

The age of the participants used was identified and categorized as either children (under 13 years old), adolescent (13 to 18 inclusive), or adult (over 18 years old).

## Screening of Participants

Because some intervention studies focus on the remediation of individuals who score relatively poorly on pre-tests, we used separate codes to distinguish studies in which low-scoring individuals were trained exclusively and those in which individuals received training regardless of pre-test performance.

## Durability

We noted how much time elapsed from the end of training to the administration of the post-test. We incorporated data from any follow-ups that were conducted with participants to assess the retention and durability of training effects.

## Supplemental Materials

Method for the calculation of effect sizes (Hedges' *g*) by CMA.

To calculate Hedges' *g* when provided with the Means and SD for the pre-test and post-test of the Treatment and Control group, the Standardized Mean Difference (d) is multiplied by the correction factor (J).

Example raw data:

| Treatment (T) | | | | | Control (C) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pre-test | SD | Post-test | SD | Sample Size (N) | Pre-test | SD | Post-test | SD | Sample Size (N) | Pre-Post Corr. |
| 7.87 | 4.19 | 16.0 | 4.07 | 8 | 5.22 | 3.96 | 8.0 | 5.45 | 9 | 0.7 |

Calculation of the Standardized Mean Difference (d)

$$d = \frac{\text{Raw Difference between the Means}}{\text{SD Change Pooled}}$$

| | |
|---|---|
| Raw Difference between the Means | = Mean Change T – Mean Change C |
| • Mean Change T | = Mean Post T – Mean Pre T<br>= 16.0 – 7.87<br>= 8.13 |
| • Mean Change C | = Mean Post C – Mean Pre C<br>= 8.0 – 5.22<br>= 2.78 |
| Raw Difference between the Means | = 8.13 – 2.78<br>= **5.35** |
| SD Change Pooled | $= \sqrt{\frac{(N_T - 1)(\text{SD Change T})^2 + (N_C - 1)(\text{SD Change C})^2}{(N_T + N_C - 2)}}$ |
| • SD Change T | $= \sqrt{[(\text{SD Pre T})^2 + (\text{SD Post T})^2 - 2(\text{Pre Post Corr})(\text{SD Pre T})(\text{SD Post T})]}$ |

$$= \sqrt{[(4.19)^2 + (4.07)^2 - 2(0.7)(4.19)(4.07)]}$$

$$= 3.20$$

- SD Change C

$$= \sqrt{[(\text{SD Pre C})^2 + (\text{SD Post C})^2 - 2(\text{Pre Post Corr})(\text{SD Pre C})(\text{SD Post C})]}$$

$$= \sqrt{[(3.96)^2 + (5.45)^2 - 2(0.7)(3.96)(5.45)]}$$

$$= 3.89$$

SD Change Pooled

$$= \sqrt{\frac{(8-1)(3.20)^2 + (9-1)(3.89)^2}{(8+9-2)}}$$

$$= \mathbf{3.58}$$

$$d = \frac{5.35}{3.58} = \mathbf{1.49}$$

Standard Error d, SE (d)

$$= \sqrt{\frac{1}{N_T} + \frac{1}{N_C} + \frac{d^2}{2(N_T + N_C)}}$$

$$= \sqrt{\frac{1}{8} + \frac{1}{9} + \frac{1.49^2}{2(8+9)}}$$

$$= 0.55$$

Calculation of the correction factor (J)

$$J = 1 - \frac{3}{4 * df - 1}$$

$$df = (N_{total} - 2) = (8 + 9 - 2) = 15$$

$$= 1 - \frac{3}{4 * 15 - 1}$$

$$= 0.95$$

Calculation of Hedges' *g*

$$g = d * J$$

$$= 1.49 * 0.95$$

$$= \mathbf{1.42}$$

Standard Error *g*, SE (*g*)

$$= SE(d) * J$$

$$= 0.55 * 0.95$$

$$= 0.52$$

Variance of *g*

$$= SE(g)^2$$

$$= 0.52^2$$

$$= 0.27$$

Hedges' *g* was also calculated if the data were reported as an *F* value for the difference in Change between Treatment and Control groups. The equations are provided here:

$$\text{Standard Change Difference} = \sqrt{\frac{N_T + N_C}{N_T * N_C}}$$

$$\text{Standard Change Difference SE} = \sqrt{\frac{1}{N_T} + \frac{1}{N_C} + \frac{\text{Standard Change Difference}^2}{2 * N_T + N_C}}$$

The calculations for the correction factor (J) and Hedges' *g* are the same as above.

Hedges' *g* was also calculated if the data were reported as a t-value for the difference between the Treatment and Control groups. The equations are provided here:

$$\text{Standard Paired Difference} = \frac{t}{\sqrt{N}}$$

$$\text{Standard Paired Difference SE} = \frac{t}{\sqrt{N}} * \sqrt{\frac{1 + \text{Standard Paired Difference}^2}{2}}$$

The calculations for the correction factor (J) and Hedges' *g* are the same as above. Standard Paired Difference replaces d and Standard Paired Difference SE replaces SE (d).

Mean-weighted effect sizes and key characteristics of studies included in the meta-analysis.

| Authors | Training description | Training category [a] | Outcome measure | Outcome category [b] | Sex [c] | Age [d] | g | k |
|---|---|---|---|---|---|---|---|---|
| Alderton (1989) | Repeated practice on the test used for pre- and post-test | 3 | Integrating Details task, Mental Rotations tests, Intercept tasks | 1,2 | 1,2 | 2 | .383 | 16 |
| Alington, Leaf & Monaghan (1992) | Repeated practice on Mental Rotation tasks | 3 | V-K Mental Rotation Test (*V-K MRT)* | 2 | 1,2 | 3 | .517 | 6 |
| Asoodeh (1993) | Animated graphics used to present orthographic projection treatment module | 3 | Orthographic projection quiz, V-K MRT | 2 | 3 | 3 | .765 | 5 |
| Azzaro (1987) - Overall | Recreation activities with emphasis on spatial orientation | 3 | STAMAT - Object Rotation | 2 | 1 | 4 | .297 | 6 |
| Azzaro (1987) - Control |  |  |  |  |  |  | .251 | 4 |
| Azzaro (1987) - Treatment |  |  |  |  |  |  | .412 | 2 |
| Baldwin (1984) - Overall | 10 lessons in spatial orientation and spatial visualization tasks | 3 | Group Embedded Figures Test (*GEFT*) and Differential Aptitude test (*DAT*) combined score | 1,2 | 1,2 | 1 | .704 | 2 |
| Baldwin (1984) - Control |  |  |  |  |  |  | .393 | 2 |
| Baldwin (1984) - Treatment |  |  |  |  |  |  | .832 | 2 |
| Barsky & Lachman (1986) - Overall | Physical knowledge vs. Reference system vs. control (observe and think only) | 3 | Rod-and-Frame task (*RFT)*, Water-Level task (*WLT)*, Plumb-Line task, Embedded Figures Test (*EFT*), Primary Mental Abilities-Space Relations (*PMA-SR*) | 1,2,3 | 1 | 3 | .288 | 10 |
| Barsky & Lachman (1986) - Control |  |  |  |  |  |  | -.018 | 4 |
| Barsky & Lachman (1986) - Treatment |  |  |  |  |  |  | .372 | 8 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Bartenstein (1985) | Visual skills training program -- 8 hours of drawing activities | 3 | Monash Spatial test, Space Thinking (Flags), Career Ability Placement Survey Spatial Relations subtest (*CAPS-SR*), Closure Flexibility | 1,2 | 3 | 3 | .185 | 4 |
| Basak, Boot, Voss & Kramer (2008) - Overall | Quick Battle Solo Mission in RON -- Video Game | 1 | Battery of Mental Rotation Cognitive Assessment from RON | 2 | 3 | 3 | .566 | 2 |
| Basak, Boot, Voss & Kramer (2008) - Control | | | | | | | .256 | 2 |
| Basak, Boot, Voss & Kramer (2008) - Treatment | | | | | | | .564 | 1 |
| Basham (2006) | Pro/Desktop CADD solid modeling software | 3 | Purdue Spatial Visualization Test (*PSVT*) | 2 | 3 | 2 | .308 | 6 |
| Batey (1986) | Highly-specific training vs. non-specific training (instruction in orthographic projection) vs. control (no training) | 3 | Spatial Relations-DAT (*SR-DAT)*, Horizontality test, V-K MRT, GEFT | 1,2,3 | 1,2 | 2 | .502 | 16 |
| Beikmohamadi (2006) | Web-based tutorial on Valence Shell Electron Repulsion Theory and molecular visualization skills | 3 | PSVT, Shape Identification test | 1,2 | 3 | 3 | .181 | 12 |
| Ben-Chaim, Lappan & Houang (1988) | 5th, 6th, 7th, and 8th graders from inner city, rural, and suburban schools trained in spatial visualization (concrete activities, building, drawing solids) | 3 | Middle Grades Mathematics Project (Spatial Visualizations test) | 2 | 1,2 | 1 | 1.08 | 14 |
| Blatnick (1986) - Overall | Verbal instruction, demo, and assembly of molecular molecules | 3 | General Aptitude Test Battery - Spatial | 2 | 1,2 | 2 | -.080 | 2 |
| Blatnick (1986) - Control | | | | | | | .500 | 2 |
| Blatnick (1986) - Treatment | | | | | | | .333 | 2 |
| Boakes (2006)  Overall | Origami Lessons | 3 | Paper Folding task, Surface Development test, Card Rotation | 2 | 1,2 | 1 | .277 | 6 |
| Boakes (2006) - Control | | | | | | | .446 | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Boakes (2006) - Treatment | | | test | | | | .378 | 6 |
| Boulter (1992) | Transformational Geometry with Object Manipulation and Imagery | 2 | Surface Development test, Card Rotations test, Hidden Patterns test | 1,2 | 3 | 2 | .384 | 3 |
| Braukmann (1991) - Overall | Along with lecture on orthographic projection, 3D Computer-Aided Design vs. control (traditional 2D manual drafting) training | 3 | Test of Orthographic Projection Skills, Shepard and Metzler cube test | 2 | 3 | 3 | .301 | 3 |
| Braukmann (1991) - Control | | | | | | | .664 | 2 |
| Braukmann (1991) - Treatment | | | | | | | .430 | 3 |
| Brooks (1992) - Overall | Interaction strategy of explaining disagreement | 3 | Structural Index Score - Placing a house in spatial relations scene | 4 | 1,2,3 | 1 | .540 | 6 |
| Brooks (1992) - Control | | | | | | | .182 | 6 |
| Brooks (1992) - Treatment | | | | | | | .585 | 6 |
| Calero & Garcia (1995) | Instrumental enrichment program to improve subject's orientation of own body | 3 | Practical Spatial Orientation, Thurstone's spatial ability test from PMA | 2,4 | 3 | 3 | .600 | 4 |
| Cathcart (1990) - Overall | Course training in Logo | 2 | GEFT | 1 | 3 | 1 | .428 | 1 |
| Cathcart (1990) - Control | | | | | | | .527 | 1 |
| Cathcart (1990) - Treatment | | | | | | | .941 | 1 |
| Center (2004) - Overall | Building Perspective Deluxe software | 3 | V-K MRT | 2 | 3 | 1 | .296 | 1 |
| Center (2004) - Control | | | | | | | .455 | 1 |
| Center (2004) - Treatment | | | | | | | .312 | 1 |
| Chatters (1984) | Groups comparable in visual-motor perceptual skill given video game training (Space Invaders) vs. control (no training) | 1 | Block Design, Mazes (both WISC subtests) | 1,2 | 3 | 1 | .559 | 2 |
| Cherney (2008) - Overall | *Antz Extreme Racing* in | 1 | V-K MRT | 2 | 1,2 | 3 | .352 | 12 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Cherney (2008) - Control | 3D space with joystick | | | | | | .471 | 8 |
| Cherney (2008) - Treatment | | | | | | | .645 | 4 |
| Chevrette (1987) | Computer simulation game asking subjects to locate urban land uses in 3 cities | 1 | GEFT | 1 | 3 | 3 | -.078 | 2 |
| Chien (1986) - Overall | Computer graphics spatial training | 3 | Author created Mental Rotation test | 2 | 1,2 | 1 | .281 | 6 |
| Chien (1986) - Control | | | | | | | .320 | 6 |
| Chien (1986) - Treatment | | | | | | | .479 | 6 |
| Clark (1996) | Computer graphics designed to aide spatial perception | 3 | Restaurant Spatial Comparison test | 1 | 1,2 | 3 | -.646 | 2 |
| Clements, Battista, Sarama & Swaminathan (1997) | Geometry training in slides, flips, turns etc. using video game Tumbling Tetronimoes | 1 | Wheatley Spatial test (MRT) | 2 | 1,2 | 1 | 1.226 | 2 |
| Cockburn (1995) - Overall | Play with LEGO Duplo blocks and build objects | 3 | Kinesthetic Spatial Concrete Building test, Kinesthetic Spatial Concrete Matching test, Motor Free Visual Perception test | 1,2 | 1 | 1 | .649 | 6 |
| Cockburn (1995) - Control | | | | | | | .456 | 6 |
| Cockburn (1995) - Treatment | | | | | | | .895 | 6 |
| Comet (1986) - Overall | Art to improve realistic drawing skills | 3 | EFT | 1 | 3 | 3 | .541 | 1 |
| Comet (1986) - Control | | | | | | | .269 | 1 |
| Comet (1986) - Treatment | | | | | | | .104 | 1 |
| Connolly (2007) - Overall | Practice converting 2D to 3D, 3D to 2D and Boolean operations to combine objects spatially | 3 | Paper Folding task, Spatial Orientation Cognitive Factor test | 2 | 3 | 3 | .085 | 4 |
| Connolly (2007) - Control | | | | | | | .550 | 4 |
| Connolly (2007) - Treatment | | | | | | | .546 | 4 |
| Crews (2008) | Teacher participation in Geospatial Technologies professional development | 2 | Spatial Literacy Skills | 2 | 1,2 | 2 | -.103 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Curtis (1992) | Orthographic principles with glass box and bowl/hemisphere imagery | 3 | Multi-view orthographic projection | 2 | 3 | 3 | .191 | 3 |
| D'Amico (2006) - Overall | Verbal and visuo-spatial working memory | 3 | Visuo-spatial Working Memory test | 2 | 3 | 1 | 2.118 | 1 |
| D'Amico (2006) - Control | | | | | | | -.773 | 1 |
| D'Amico (2006) - Treatment | | | | | | | .820 | 1 |
| Dahl (1984) | Computer aided orthographic training – projection problems | 2 | Multiple Aptitudes Test 8/9 – Choose the correct piece to finish the figure, GEFT | 1,2 | 3 | 3 | -.157 | 3 |
| Day, Engelhardt, Maxwell & Bolig (1997) | Block design training | 3 | Block Design (WPPSI subtest) | 2 | 3 | 1 | 1.420 | 2 |
| DelGrande (1986) | Geometry intervention | 3 | Author designed tests that span across outcome categories | 5 | 3 | 1 | .984 | 1 |
| De Lisi & Cammarano (1996) - Overall | Video game training with Blockout vs. control (Solitaire) | 1 | V-K MRT | 2 | 1,2 | 3 | .689 | 2 |
| De Lisi & Cammarano (1996) - Control | | | | | | | .227 | 2 |
| De Lisi & Cammarano (1996) - Treatment | | | | | | | .573 | 2 |
| De Lisi & Wolford (2002) - Overall | Video game training with Tetris vs. control (Carmen Sandiago) | 1 | French Kit Card Rotation test | 2 | 1,2 | 1 | 1.341 | 2 |
| De Lisi & Wolford (2002) - Control | | | | | | | -.058 | 2 |
| De Lisi & Wolford (2002) - Treatment | | | | | | | .591 | 2 |
| Deratzou (2006) | Visualization training with problems sets, journals, videos, lab experiments, computers | 2 | Card Rotation test, Cube Comparison test, Form Board test, Paper folding task, Surface Development test | 2 | 1,2 | 2 | .583 | 10 |
| Dixon (1995) | Geometer Sketchpad spatial skills training | 3 | Paper folding task, Card Rotation test, Computer and Paper-Pencil Rotation/Reflection instrument | 2 | 3 | 2 | .543 | 4 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dorval & Pepin (1986) - Overall | Zaxxon video game playing vs. control (no game play) | 1 | EFT | 1 | 1,2 | 3 | .354 | 2 |
| Dorval & Pepin (1986) - Control | | | | | | | .260 | 2 |
| Dorval & Pepin(1986) - Treatment | | | | | | | .549 | 2 |
| Duesbury (1992) | Orthographic techniques, Line and feature matching, Instruction and practice visualizing | 3 | Surface Development test, Flanagan Industrial test, Paper Folding task, Test of 3D shape visualization | 2 | 2 | 3 | 1.520 | 12 |
| Duesbury & O'Neil (1996) | Wireframe Computer-Aided Design training on orthographic projection, line-feature matching, 2 and 3D visualization vs. control (traditional blueprint reading course) | 3 | Flanagan Industrial Tests Assembly, SR-DAT, Surface Development test, Paper Folding task | 2 | 2 | 3 | .648 | 4 |
| Dziak (1985) | Instruction in BASIC graphics | 3 | Card Rotations test | 2 | 3 | 2 | .115 | 1 |
| Edd (2001) | Practice with rotating/handling MRT models | 3 | Shepard Metzler MRT | 2 | 1 | 3 | .383 | 1 |
| Ehrlich et al. (2006) - Overall | Imagine and actually move pieces with instruction | 3 | Mental Rotations test | 2 | 1,2 | 1 | .620 | 6 |
| Ehrlich et al. (2006) - Control | | | | | | | 1.091 | 4 |
| Ehrlich et al. (2006) - Treatment | | | | | | | .873 | 2 |
| Eikenberry (1988) - Overall | Learning to program in LOGO | 2 | Space Thinking Flags test - MR, Children's GEFT | 1,2 | 1,2 | 1 | .311 | 4 |
| Eikenberry (1988) - Control | | | | | | | .518 | 4 |
| Eikenberry (1988) - Treatment | | | | | | | .514 | 4 |
| Embretson (1987) | Paper folding training vs. control (clerical training) | 3 | SR-DAT | 2 | 3 | 3 | .686 | 3 |
| Engelhardt (1987) | Guided instruction to construct block designs from a model | 3 | Block Design (WPPSI), DAT Spatial Folding task, SR-DAT | 2 | 3 | 1 | 1.190 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Eraso (2007) - Overall | Geometer's Sketchpad interactive computer program | 2 | PSVT | 2 | 1,2 | 2 | .282 | 6 |
| Eraso (2007) - Control | | | | | | | .486 | 4 |
| Eraso (2007) - Treatment | | | | | | | .375 | 2 |
| Fan (1998) | Drawing with instructional verbal cues and visual props | 3 | Correct responses to selection task, Representation of size relationship, hidden outlines, and occlusion in drawing | 2 | 3 | 1 | .505 | 12 |
| Feng (2006) - Overall | Training using action vs. control (nonaction video game) | 1 | V-K MRT | 2 | 1,2 | 3 | 1.137 | 2 |
| Feng (2006) - Control | | | | | | | .186 | 2 |
| Feng (2006) - Treatment | | | | | | | 1.136 | 2 |
| Feng, Spence & Pratt (2007) - Overall | Training using action vs. control (nonaction video game) | 1 | V-K MRT | 2 | 1,2 | 3 | 1.194 | 4 |
| Feng, Spence & Pratt (2007) - Control | | | | | | | .400 | 4 |
| Feng, Spence & Pratt (2007) - Treatment | | | | | | | 1.347 | 4 |
| Ferguson (2008) - Overall | Engineering drawing with dissection of hand held and computer-generated manipulatives | 3 | PSVT - Rotations | 2 | 3 | 3 | .257 | 3 |
| Ferguson (2008) - Control | | | | | | | .197 | 2 |
| Ferguson (2008) - Treatment | | | | | | | .107 | 1 |
| Ferrara (1992) | Imagery instruction with visual synthesis task | 3 | Draw shapes from training by hand and on computer | 2 | 3 | 2 | -.424 | 1 |
| Ferrini-Mundy (1987) - Overall | Audiovisual spatial visualization training, with or without tactual practice, vs. Control 1 (post-test only group) | 2 | SR-DAT | 2 | 3 | 3 | .344 | 4 |
| Ferrini-Mundy (1987) - Control | Audiovisual spatial visualization training, with or without tactual practice vs. Control 2 (calculus course as usual) | | | | | | .655 | 1 |
| Ferrini-Mundy (1987) - Treatment | | | | | | | .620 | 2 |
| Fitzsimmons (1995) - Overall | Solving 3D geometric problems in calculus | 2 | PSVT - Rotations and visualization of views | 2 | 3 | 3 | .232 | 8 |
| Fitzsimmons (1995) - Control | | | | | | | .116 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Fitzsimmons (1995) - Treatment | | | | | | | .150 | 8 |
| Frank (1986) - Overall | Map reading, memetic or itinerary | 3 | Post-test score in locating animal using a map, Symbol recognition, Representational correspondence - route items and landmark items | 2 | 1,2,3 | 1 | .490 | 10 |
| Frank (1986) - Control | | | | | 3 | | .790 | 4 |
| Frank (1986) - Treatment | | | | | | | .687 | 4 |
| Funkhouser (1990) | Geometry course or second year algebra plus computer problem solving activities | 2 | Problem solving test - spatial subtest | 1 | 1,2 | 1 | .420 | 2 |
| Gagnon (1985) - Overall | Playing 2D Targ and 3D Battlezone video games vs. control (no video game playing) | 1 | Guilford-Zimmermann Spatial Orientation and Visualization; Employee Aptitude Survey: Visual Pursuit test | 1,2,4 | 3 | 3 | .310 | 3 |
| Gagnon (1985) - Control | | | | | | | .346 | 3 |
| Gagnon (1985) - Treatment | | | | | | | .507 | 3 |
| Gagnon (1986) | Video game training - Interactive versus observational | 1 | Guilford-Zimmerman MRT | 2 | 1 | 3 | .156 | 3 |
| Geiser, Lehman, & Eid (2008) | Administered MRT test twice as practice | 3 | MRT described in Peters et. al (1995) | 2 | 1,2 | 1 | .674 | 2 |
| Gerson, Sorby, Wysocki & Baartmans (2001) - Overall | Engineering course with lecture and spatial modules vs. control (course and modules without lecture) | 2 | SR-DAT, Mental Cutting test, MRT, 3DC (Cube test), PSVT-R | 2 | 3 | 3 | .087 | 5 |
| Gerson, Sorby, Wysocki & Baartmans (2001) - Control | | | | | | | .692 | 5 |
| Gerson, Sorby, Wysocki & Baartmans (2001) - Treatment | | | | | | | .984 | 5 |
| Geva & Cohen (1987) - Overall | 2nd vs. 4th graders - 7 months of LOGO instruction vs. control (regular computer use in school) | 2 | Map reading task - Rotate, Start and Turn | 2 | 3 | 1 | 1.094 | 6 |
| Geva & Cohen (1987) - Control | | | | | | | .249 | 6 |
| Geva & Cohen (1987) - Treatment | | | | | | | .368 | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gibbon (2007) | LEGO Mindstorms Robotics Invention System | 3 | Raven's Progressive Matrices | 1 | 3 | 1 | .246 | 3 |
| Gillespie (1995) - Overall | Engineering graphics course with solid modeling tutorials | 2 | Paper Folding task, V-K MRT, Rotated Blocks | 2 | 3 | 3 | .417 | 6 |
| Gillespie (1995) - Control | | | | | | | .326 | 2 |
| Gillespie (1995) - Treatment | | | | | | | .881 | 3 |
| Gitimu, Workman & Anderson (2005) | Pre-existing fashion design experience: level of experience based on credit hours | -- | Apparel Spatial Visualization test | 2 | 3 | 3 | .698 | 6 |
| Gittler & Gluck (1998) - Overall | Training in Descriptive Geometry vs. control (no course) | 2 | 3D cube test | 2 | 1,2 | 2 | .377 | 2 |
| Gittler & Gluck (1998) - Control | | | | | | | .296 | 2 |
| Gittler & Gluck (1998) - Treatment | | | | | | | .622 | 2 |
| Godfrey (1999) - Overall | 3D Computer-aided modeling, Draw in 2D and build 3D models | 2 | PSVT | 2 | 3 | 3 | .198 | 3 |
| Godfrey (1999) - Control | | | | | | | .619 | 1 |
| Godfrey (1999) - Treatment | | | | | | | .409 | 1 |
| Golbeck (1998) - Overall | 4th vs. 6th grade, matched ability vs. unmatched-high vs. umatched-low vs. control (worked alone) | 3 | WLT | 2 | 3 | 1 | .176 | 6 |
| Golbeck (1998) - Control | | | | | | | .304 | 2 |
| Golbeck (1998) - Treatment | | | | | | | .528 | 6 |
| Goodrich (1992) - Overall | Watched training video versus watched placebo video of a cartoon | 3 | Verticality-Horizontality test | 3 | 3 | 3 | .591 | 6 |
| Goodrich (1992) - Control | | | | | | | .734 | 4 |
| Goodrich (1992) - Treatment | | | | | | | .917 | 2 |
| Goulet & Talbot (1988) - Overall | Those with hockey training versus those without training | 3 | GEFT | 1 | 2 | 1 | .173 | 1 |
| Goulet & Talbot (1988) - Control | | | | | | | .689 | 1 |
| Goulet & Talbot (1988) - Treatment | | | | | | | .625 | 1 |
| Guillot & Collet (2004) | Acrobatic sport training | 3 | GEFT | 1 | 3 | 3 | 1.582 | 1 |
| Gyanani & Pahuja (1995) | Course lectures plus peer tutoring | 2 | Spatial geography ability | 2 | 3 | 1 | .317 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Hedley (2008) | Course training using geospatial technologies | 2 | Spatial Abilities test - Map skills | 2 | 3 | 2 | .691 | 4 |
| Heil, Rossler, Link & Bajric (1998) | Practice group with additional, specific practice vs. control (3 sessions of MR practice without additional specific practice) | 3 | RT mental rotations-familiar objects in familiar orientations | 2 | 3 | 3 | .562 | 2 |
| Higginbotham (1993) - Overall | Computer based versus concrete visualization instruction | 3 | MGM-PSVT (Spatial Visualization), Non-standardized SVT | 2 | 3 | 3 | .770 | 2 |
| Higginbotham (1993) - Control | | | | | | | .623 | 2 |
| Higginbotham (1993) - Treatment | | | | | | | .597 | 2 |
| Hozaki (1987) | Instruction and practice visualizing 2D and 3D objects with CAD software | 3 | Paper Folding task | 2 | 1,2 | 3 | 1.233 | 6 |
| Hsi, Linn & Bell, (1997) | Pretest vs. posttest after strategy instruction using Block Stacking and Display Object software modules with isometric vs. orthographic items | 3 | Paper Folding task, cube counting, matching rotated objects, spatial battery of orthographic, isometric views | 2,5 | 1, 2 | 3 | .517 | 4 |
| Idris (1998) - Overall | Instructional activities to visualize geometric constructions, relate properties and disembed simple geometric figures | 3 | Spatial Visualization test, GEFT | 1,2 | 3 | 2 | .916 | 6 |
| Idris (1998) - Control | | | | | | | .340 | 6 |
| Idris (1998) - Treatment | | | | | | | 1.229 | 6 |
| Janov (1986) - Overall | Instruction in drawing and painting accompanied by artistic criticism | 3 | GEFT | 1 | 3 | 3 | .260 | 6 |
| Janov (1986) - Control | | | | | | | .959 | 1 |
| Janov (1986) - Treatment | | | | | | | .470 | 3 |
| Johnson (1991) | Isometric drawing aid vs. 3D rendered model vs. animated wireframe vs. control (no aid, practice with drawings only) | 3 | SR-DAT | 2 | 3 | 3 | .016 | 4 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Johnson-Gentile, Clements & Battista (1994) - Overall | LOGO geometry motions unit | 3 | Geometry motions post-test | 2 | 3 | 1 | .544 | 3 |
| Johnson-Gentile, Clements & Battista (1994) - Control | | | | | | | .239 | 2 |
| Johnson-Gentile, Clements & Battista (1994) - Treatment | | | | | | | .321 | 1 |
| July (2001) | Course to teach 3D spatial ability using Geometer's Sketchpad | 2 | Surface Development test, MRT | 2 | 3 | 2 | .606 | 2 |
| Kaplan & Weisberg (1987) | Pretest vs. posttest for 3rd vs. 5th graders vs. control (no feedback) | 3 | Purdue Perceptual Screening test (embedded and successive figures) | 1 | 3 | 1 | .430 | 2 |
| Kass, Ahlers & Dugger (1998) | Practice Angle on the Bow task with feedback and read instruction manual vs. control (read manual only) | 3 | Angle on the Bow measure | 1 | 1,2 | 3 | .501 | 8 |
| Kastens & Liben (2007) | Explaining condition vs. control (did not explain sticker placement) | 3 | Sticker Map task (representational correspondence, errors, offset) | 2 | 3 | 2 | .791 | 3 |
| Kastens, Kaplan & Christie-Blick (2001) | Training using Where are We? video game vs. control (completed task without assistance) | 3 | Reality-to-Map (Flag-Sticker) test | 2 | 1,2 | 1 | .320 | 2 |
| Keehner & Lippa (2006) - Overall | Learned to use an angled laparoscope (tool used by surgeons) | 3 | Laproscopic simulation task | 1 | 3 | 3 | 1.748 | 1 |
| Keehner & Lippa (2006) - Control | | | | | | | 1.248 | 1 |
| Keehner & Lippa (2006) - Treatment | | | | | | | .547 | 1 |
| Kirby & Boulter (1999) - Overall | Training in object manipulation and visual imagery vs. paper-pencil instruction vs. control (test-only) | 3 | Factor referenced tests (Hidden Patterns, Card Rotations, Surface Development test) | 5 | 3 | 1 | .125 | 4 |
| Kirby & Boulter (1999) - Control | | | | | | | .055 | 1 |
| Kirby & Boulter (1999) - Treatment | | | | | | | .123 | 2 |
| Kirchner, Forns & Amador (1989) - Overall | Repeated practice of the GEFT | 3 | GEFT | 1 | 1 | 3 | .976 | 1 |
| Kirchner, Forns & Amador (1989) - Control | | | | | | | .521 | 1 |
| Kirchner, Forns & Amador (1989) - Treatment | | | | | | | .242 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Kovac (1985) | Microcomputer assisted instruction with mechanical drawing | 2 | SR-DAT | 2 | 3 | 2 | .166 | 1 |
| Kozhevnikov & Thornton (2006) - Overall | Added Interactive Lecture Demonstrations (ILDs) to physics instruction for Dickinson vs. Tufts science and non-science majors and middle-school and high school science teachers | 2,3 | Paper Folding task, MRT | 2 | 3 | 3 | .474 | 16 |
| Kozhevnikov & Thornton (2006) - Control | | | | | | | .399 | 9 |
| Kozhevnikov & Thornton (2006) - Treatment | | | | | | | .424 | 9 |
| Krekling & Nordvik (1992) | Observation training to perform WLT | 3 | Adjustment error in WLT | 3 | 1 | 3 | 1.008 | 2 |
| Kwon, Kim & Kim (2002) - Overall | Visualization software using Virtual Reality vs. control (standard 2D text and software) | 1,3 | Middle Grades Mathematics Project Spatial Visualization test | 5 | 1 | 2 | .387 | 1 |
| Kwon, Kim & Kim (2002) - Control | | | | | | | .521 | 1 |
| Kwon, Kim & Kim (2002) - Treatment | | | | | | | .703 | 1 |
| Kwon (2003) - Overall | Spatial visualization instructional program using Virtual Reality vs. Paper-based instruction vs. control (no training) | 3 | Middle Grades Mathematics Project Spatial Visualization test | 5 | 3 | 2 | 1.088 | 2 |
| Kwon (2003) - Control | | | | | | | .150 | 1 |
| Kwon (2003) - Treatment | | | | | | | .915 | 2 |
| Larson (1996) | Commentary and movement vs. control (no commentary and no movement) | 3 | View point task (based on 3 Mountains) | 4 | 3 | 1 | 1.423 | 1 |
| Larson et al. (1999) | Repeated Virtual Reality Spatial Rotation training vs. control (filler task) | 3 | V-K MRT | 2 | 1,2 | 3 | .303 | 2 |
| Lee (1995) - Overall | LOGO training vs. control (no LOGO training) for 2nd graders | 3 | WLT | 3 | 3 | 1 | .921 | 1 |
| Lee (1995) - Control | | | | | | | .210 | 1 |
| Lee (1995) - Treatment | | | | | | | .280 | 1 |
| Lennon (1996) - Overall | Spatial enhancement course - spatial visualization and orientation tasks | 2 | Surface Development test, Paper Folding task, Cube Comparison test, Card Rotation test, | 2 | 3 | 3 | .280 | 4 |
| Lennon (1996) - Control | | | | | | | .845 | 1 |
| Lennon (1996) - Treatment | | | | | | | .924 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Li (2000) - Overall | Explanatory statement of physical properties of water versus no explanatory statement | 3 | Proportion correct on WLT | 3 | 3 | 1 | .314 | 1 |
| Li (2000) - Control | | | | | | | .205 | 1 |
| Li (2000) - Treatment | | | | | | | .435 | 1 |
| Lizarraga & Garcia Ganuza (2003) - Overall | Mental rotation training worksheet vs. control (regular math course) | 3 | SR-DAT – visualization and mental rotation | 2 | 3 | 2 | 1.244 | 2 |
| Lizarraga & Garcia Ganuza (2003) - Control | | | | | | | .256 | 2 |
| Lizarraga & Garcia Ganuza (2003) - Treatment | | | | | | | .791 | 2 |
| Lohman (1988) | Repeated practice mental rotation problems like Shepard-Metzler | 3 | Paper Folding test, Form Board test, Figure Rotation task, Card Rotation task | 2 | 3 | 3 | .255 | 12 |
| Lohman & Nichols (1990) - Overall | Train with repeated practice on 3D MRT - test on speeded rotation task vs. control (test-retest, without the repeated practice) | 3 | Form Board test, Paper Folding task, Card Rotations, Thurstone's Figures, MRT | 2 | 3 | 3 | .291 | 4 |
| Lohman & Nichols (1990) - Control | | | | | | | 1.039 | 5 |
| Lohman & Nichols (1990) - Treatment | | | | | | | .894 | 4 |
| Longstreth & Alcorn (1990) - Overall | Play with blocks different vs. same in color as those used in WPSSI vs. control (play with non-block toys) | 3 | Block Design, Mazes (both WPPSI subtests) | 1,2 | 3 | 1 | .677 | 4 |
| Longstreth & Alcorn (1990) - Control | | | | | | | .328 | 2 |
| Longstreth & Alcorn (1990) - Treatment | | | | | | | .618 | 4 |
| Lord (1985) - Overall | Imagining planes cutting through solid training vs. control (regular biology class with lecture, seminar and lab) | 3 | Planes of Reference, Factor Referenced Tests- Spatial Orientation and Visualization, Flexibility of Closure | 1,2,5 | 3 | 3 | .920 | 4 |
| Lord (1985) - Control | | | | | | | .057 | 4 |
| Lord (1985) - Treatment | | | | | | | .795 | 4 |
| Luckow (1984) - Overall | Course in LOGO Turtle Graphics | 2 | Paper Form Board test | 2 | 3 | 3 | .564 | 3 |
| Luckow (1984) - Control | | | | | | | .480 | 2 |
| Luckow (1984) - Treatment | | | | | | | .785 | 2 |
| Luursema et al. (2006) | Study 3D stereoptic and 2D anatomy stills vs. control (study only typical 2D biocular stills) | 3 | Identification of anatomical structures and localization of cross-sections in | 1,2 | 3 | 3 | .465 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | frontal view | | | | | |
| Martin (1991) | Learning concept mapping skills for biology | 3 | GEFT | 1 | 3 | 2 | .347 | 3 |
| McAuliffe (2003) - Overall | 2D static visuals, 3D animated visuals, and 3D interactive animated visuals to display a topographic map | 3 | Visualizing Topography test | 4 | 1,2 | 2 | .642 | 12 |
| McAuliffe (2003) - Control | | | | | | | .476 | 6 |
| McAuliffe (2003) - Treatment | | | | | | | .479 | 3 |
| McClurg & Chaille (1987) - Overall | 5th vs. 7th vs. 9th grade: Factory themed vs. Stellar 7 mission video games vs. control (no video game play) | 1 | Mental Rotations test | 2 | 3 | 1,2 | 1.157 | 6 |
| McClurg & Chaille (1987) - Control | | | | | | | .339 | 3 |
| McClurg & Chaille (1987) - Treatment | | | | | | | .796 | 6 |
| McCollam (1997) | Paper folding manipulatives | 3 | Spatial Learning Ability test | 2 | 3 | 3 | .371 | 4 |
| McCuiston (1991) | Computer assisted descriptive geometry lesson with animation and 3D views vs. control (static lessons with text, no animation) | 3 | V-K MRT | 2 | 3 | 3 | .503 | 1 |
| McKeel (1993) - Overall | Construct machines and make sketches using LEGO Dacta Technic | 2 | Paper Folding test | 2 | 1 | 3 | .047 | 1 |
| McKeel (1993) - Control | | | | | | | .464 | 1 |
| McKeel (1993) - Treatment | | | | | | | .386 | 1 |
| Merickel (1992) - Overall | Autocade vs. Cyberspace spatial skills training | 3 | SR-DAT, Paper Form Board test, Displacement and Transformation | 2 | 3 | 1 | .717 | 5 |
| Merickel (1992) - Control | | | | | | | .612 | 3 |
| Merickel (1992) - Treatment | | | | | | | .582 | 3 |
| Miller E (1985) - Overall | Training in LOGO Turtle graphics | 3 | Eliot Price Spatial test | 4 | 1,2 | 3 | .076 | 2 |
| Miller E (1985) - Control | | | | | | | .219 | 2 |
| Miller E (1985) - Treatment | | | | | | | .262 | 2 |
| Miller & Kapel (1985) - Overall | 7th vs. 8th grade gifted vs. control (average ability) students trained with problem solving video | 1 | Wheatly Spatial test (MRT) | 2 | 3 | 1 | .468 | 4 |
| Miller & Kapel (1985) - Control | | | | | | | .750 | 4 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Miller & Kapel (1985) - Treatment | game | | | | | | .939 | 4 |
| Miller J (1995) - Overall | Virtual Reality spatial orientation tests | 1 | Virtual reality navigation, Spatial Menagerie | 4,5 | 3 | 1 | 1.026 | 2 |
| Miller J (1995) - Control | | | | | | | .967 | 1 |
| Miller J (1995) - Treatment | | | | | | | 1.911 | 1 |
| Miller, Kelly et al (1988) - Overall | One academic year of LOGO programming | 2 | Primary Mental Abilities | 2 | 3 | 1 | .738 | 1 |
| Miller, Kelly et al (1988) - Control | | | | | | | .136 | 1 |
| Miller, Kelly et al (1988) - Treatment | | | | | | | .650 | 1 |
| Mohamed (1985) - Overall | LOGO programming course | 2 | Developing Cognitive Abilities Test-Spatial, Children's EFT | 1,2 | 3 | 1 | .743 | 2 |
| Mohamed (1985) - Control | | | | | | | .579 | 2 |
| Mohamed (1985) - Treatment | | | | | | | 1.138 | 2 |
| Moody (1998) | Strategy instructions for solving Mental Rotations test | 3 | V-KMRT | 2 | 3 | 3 | .113 | 1 |
| Morgan (1986) - Overall | Computer estimation instructional strategy for mathematical simulations | 3 | Area estimation, Length estimation with and without scale | 1 | 1,2 | 1 | .408 | 6 |
| Morgan (1986) - Control | | | | | | | .349 | 6 |
| Morgan (1986) - Treatment | | | | | | | .318 | 6 |
| Mowrer-Popiel (1991) | Explicit explanation of horizontality principle vs. demo of the principle | 3 | Crossbar and Tilted Crossbar WLT, Spherical and Square water bottle task | 3 | 1,2 | 3 | .519 | 8 |
| Moyer (2004) - Overall | Geometry course with Geometer's Sketch Pad | 2 | PSVT | 2 | 3 | 2 | .030 | 1 |
| Moyer (2004) - Control | | | | | | | .333 | 1 |
| Moyer (2004) - Treatment | | | | | | | .444 | 1 |
| Mshelia (1985) | Depth perception task and field-independence training | 3 | Mshelia's Picture Depth Perception task, GEFT | 1,2 | 3 | 1 | 1.165 | 4 |
| Mullin (2006) | Physical vs. cognitive vs. no physical control over navigation, with attention vs. distracted during wayfinding | 3 | Wayfinding to target, pointing to target, recalling object locations | 1 | 3 | 3 | .392 | 32 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Newman (1990) - Overall | Educational intervention spread across 2 consecutive menstrual cycles | 3 | Primary Mental Abilities-Space Relations (*PMA-SR*) | 2 | 1 | 3 | .079 | 2 |
| Newman (1990) - Control | | | | | | | .251 | 2 |
| Newman (1990) - Treatment | | | | | | | .207 | 2 |
| Noyes (1997) - Overall | Lessons to develop ability to perceive, manipulate, and record spatial information | 3 | Developing Cognitive Abilities Test-Spatial | 2 | 3 | 1 | 1.132 | 1 |
| Noyes (1997) - Control | | | | | | | .080 | 1 |
| Noyes (1997) - Treatment | | | | | | | .880 | 1 |
| Odell (1993) - Overall | Earth Science course with or without 3D laboratory models | 2 | Surface Development test, Paper Folding task | 2 | 3 | 2 | -.392 | 2 |
| Odell (1993) - Control | | | | | | | .175 | 2 |
| Odell (1993) - Treatment | | | | | | | .042 | 2 |
| Okagaki & Frensch (1994) - Overall | Tetris video game training vs. control (no video game) | 1 | Form Board test, Card Rotation test, Cube Comparison test (from French kit) | 2 | 1,2 | 3 | .643 | 6 |
| Okagaki & Frensch (1994) - Control | | | | | | | .239 | 6 |
| Okagaki & Frensch (1994) - Treatment | | | | | | | .420 | 6 |
| Olson (1986) - Overall | Geometry course supplemented with CAI in geometry or LOGO training | 2 | Monash Spatial Visualization test | 2 | 1,2 | 1 | .523 | 6 |
| Olson (1986) - Control | | | | | | | .819 | 4 |
| Olson (1986) - Treatment | | | | | | | .622 | 2 |
| Ozel, Larue & Molinaro (2002) | Gymnastics training | 3 | Shepard-Metzler MRT (RT and Rotation speed) | 2 | 2 | 3 | .330 | 6 |
| Pallrand & Seeber (1984) - Overall | Draw scenes outside, locate objects relative to fictitious observer, reorientation exercises, geometry lessons | 3 | Paper Folding test, Surface Development test, Card Rotation test, Cube Comparison test, Hidden Figures test | 1,2 | 3 | 3 | .679 | 15 |
| Pallrand & Seeber (1984) - Control | | | | | | | .330 | 15 |
| Pallrand & Seeber (1984) - Treatment | | | | | | | .831 | 5 |
| Parameswaran (1993) - Overall | Interactive and rule training on horizontality task | 3 | Water Clock verticality and horizontality score, Crossbar test, Verticality test, Water Bottle test | 3 | 1,2 | 3 | .633 | 30 |
| Parameswaran (1993) - Control | | | | | | | .354 | 4 |
| Parameswaran (1993) - Treatment | | | | | | | .790 | 2 |
| Parameswaran (2003) | Ages 5, 6, 7, 8, 9: Graduated training vs. demonstration training vs. | 3 | WLT, Verticality task | 3 | 1,2 | 1 | .870 | 40 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | control (completed task with no feedback) | | | | | | | |
| Parameswaren & De Lisi (1996) - Overall | Tutor guided direct instruction in principle vs. learner guided self-discovery vs. control (no feedback) | 3 | Van verticality test, Water-clock and cross-bar tests of horizontality,<br>WLT | 3 | 1,2 | 3 | .703 | 16 |
| Parameswaren & De Lisi (1996) - Control | | | | | | | .103 | 2 |
| Parameswaren & De Lisi (1996) - Treatment | | | | | | | .525 | 4 |
| Pazzaglia (2006) | Repeated practice through four learning phases | 3 | Map reading an pointing task | 4 | 3 | 3 | .386 | 4 |
| Pearson (1991) - Overall | Intensive introductory film production course | 2 | SR-DAT | 2 | 3 | 3 | .301 | 3 |
| Pearson (1991) - Control | | | | | | | .748 | 3 |
| Pearson (1991) - Treatment | | | | | | | .482 | 3 |
| Pennings (1991) - Overall | Conservation of horizontality training and restructuring in perception training | 3 | WLT, Diagnostic EFT | 1,3 | 1,2 | 1 | .532 | 12 |
| Pennings (1991) - Control | | | | | | | .336 | 8 |
| Pennings (1991) - Treatment | | | | | | | .465 | 4 |
| Perez-Fabello & Campos (2007) | Years of training in artistic skills (Academic year used to approximate) | 3 | Visual Congruence - SR, Spatial Representation test | 4 | 3 | 3 | 1.397 | 4 |
| Peters, Laeng et al. (1995) - Overall | Repeated practice once a week for four weeks | 3 | V-K MRT | 2 | 1 | 3 | .118 | 1 |
| Peters, Laeng et al. (1995) - Control | | | | | | | 3.683 | 1 |
| Peters, Laeng et al. (1995) - Treatment | | | | | | | 3.558 | 1 |
| Philleo (1997) | Produced 2D diagram from 3D view with Microworlds virtual reality or paper and pencil | 3 | Author created "Where am I standing?" task | 4 | 3 | 1 | .154 | 1 |
| Piburn et al. (2005) | Computer enhanced geology module vs. control (regular geology course with standard written manuals) | 3 | Surface Development test- Visualization and Orientation (Cube Rotation test ) | 2 | 3 | 3 | .539 | 3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pleet (1991) - Overall | Transformational geometry training with Motions computer program or hands-on manipulatives | 3 | Card Rotations test | 2 | 1,2 | 2 | .086 | 6 |
| Pleet (1991) - Control | | | | | | | .568 | 4 |
| Pleet (1991) - Treatment | | | | | | | .501 | 2 |
| Pontrelli (1990) - Overall | TRACON -- Terminal Radar Approach Control computer simulation | 3 | Author created spatial perception test based on exam for ATCs | 1 | 3 | 3 | .783 | 1 |
| Pontrelli (1990) - Control | | | | | | | .116 | 1 |
| Pontrelli (1990) - Treatment | | | | | | | .887 | 1 |
| Pulos (1997) | Demonstration of WLT without description of the phenomenon | 3 | WLT | 3 | 3 | 3 | .662 | 3 |
| Qiu (2006) - Overall | College course in different Geographic Information Technologies | 2 | Author created tests: Spatial Visualization, Spatial Orientation, Spatial Relations | 2,4 | 3 | 3 | .351 | 9 |
| Qiu (2006) - Control | | | | | | | .181 | 3 |
| Qiu (2006) - Treatment | | | | | | | .143 | 9 |
| Quaiser-Pohl & Geiser (2006) | Preference for video games: Action and simulation vs. logic and skill vs. non-players | 1 | V–K MRT | 2 | 1,2 | 2 | .084 | 6 |
| Rafi, Samsudin, et al. (2008) | Virtual environment training | 3 | Author created test of assembly and transformation | 2 | 1,2 | 2 | .737 | 6 |
| Ralls (1998) | Computer based instruction in logical and spatial ability tasks | 3 | Paper Folding task | 2 | 3 | 3 | .577 | 1 |
| Robert & Chaperon (1989) - Overall | Watched video tape demonstration of correct responses to WLT, with and without discussion | 3 | WLT -- Acquisition,<br>WLT -- Proximal Generalization | 3 | 3 | 3 | .656 | 12 |
| Robert & Chaperon (1989) - Control | | | | | | | .428 | 4 |
| Robert & Chaperon (1989) - Treatment | | | | | | | .997 | 8 |
| Rosenfield (1985) - Overall | Exercise in spatial visualization and rotation | 3 | CAPS-SR | 2 | 2 | 2 | .304 | 1 |
| Rosenfield (1985) - Control | | | | | | | .759 | 1 |
| Rosenfield (1985) - Treatment | | | | | | | .454 | 1 |
| Rush & Moore (1991) | Restructuring strategies, finding hidden patters, visualizing paper folding, finding path through a maze | 3 | Paper Folding task, GEFT | 1,2 | 3 | 3 | .366 | 6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Russell (1989) - Overall | Mental rotation practice with feedback | 3 | V-K MRT, Depth Plane Object Rotation test, Rotating 3D Objects test | 2 | 1,2 | 3 | .171 | 18 |
| Russell (1989) - Control | | | | | | | .491 | 12 |
| Russell (1989) - Treatment | | | | | | | .463 | 6 |
| Saccuzzo, Craig, et al (1996) | Repeated practice on computerized and paper and pencil tests | 3 | Surface Development test, Computerized Cube test, PMA Space Relations test, Computerized MRT | 2 | 3 | 3 | .609 | 4 |
| Savage (2006) | Repeated practice using virtual reality to traverse a maze | 3 | Time required to traverse each tile of maze | 1 | 3 | 3 | .723 | 6 |
| Schaeffer & Thomas (1998) | Repeated practice on rotated EFT | 3 | Gottschadlt Hidden Figures, EFT | 1 | 1,2 | 3 | .828 | 2 |
| Schaie & Willis (1986) | Spatial training vs. control (inductive reasoning training) | 3 | Alphanumeric rotation, Object rotation, PMA-Spatial Orientation | 2 | 3 | 3 | .417 | 3 |
| Schmitzer-Torbert (2007) - Overall | Place vs. response learning of transfer target vs. control (training target) | 3 | Percent correct and Route stability on maze learning for first vs. last trial | 1 | 1,2 | 3 | .699 | 8 |
| Schmitzer-Torbert (2007) - Control | | | | | | | .882 | 8 |
| Schmitzer-Torbert (2007) - Treatment | | | | | | | 1.645 | 8 |
| Schofield & Kirby (1994) | Area (restricted search space) vs. area and orientation provided, vs. spatial training (identify features and visualize contour map) vs. verbal training (verbalize features) vs. control (no instructions) | 3 | Location time to mark placement on map, Surface Development test, Card Rotations (S-1 Ekstrom kit of factor referenced cognitive tests) | 2 | 2 | 3 | .491 | 12 |
| Scribner (2004) | Drafting instruction tailored to students | 2 | PSVT | 2 | 3 | 3 | .126 | 4 |
| Scully (1988) - Overall | CADAM 3D computer graphics design | 3 | Guilford-Zimmerman 3D Visualization | 2 | 3 | 3 | .058 | 1 |
| Scully (1988) - Control | | | | | | | .469 | 1 |
| Scully (1988) - Treatment | | | | | | | .571 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Seddon & Shubber (1984) | All vs. half colored slides vs. monochrome slides shown simultaneously vs. cumulatively vs. individually vs. control | 3 | Rotations test (author created) | 2 | 2 | 2 | .758 | 9 |
| Seddon & Shubber (1985)a | 13-14, 15-16, vs. 17-18 year-olds, with 0, 6, 9,15 or 18 colored structures, with and without 3, 6, or 9 diagrams | 3 | Mental rotations test (author created) | 2 | 2 | 2 | 1.886 | 36 |
| Seddon & Shubber (1985)b | 13-14, 15-16 vs. 17-18 year-olds | 3 | Framework test, Cues test—Overlap, Angles, Relative size, Foreshortening; Mental rotation (author created) | 2 | 2 | 2 | .995 | 18 |
| Seddon, Eniaiyeju & Jusoh (1984) | D vs. SMD vs. MD training for those failing 1, 2, 3, or 4 cue tests. Compared 10° vs. 60°, abrupt vs. dissolving, diagram change for children remediated in Stage 1 vs. control (no remediation) | 3 | Mental rotations test (author created) | 2 | 2 | 3 | 1.742 | 24 |
| Sevy (1984) | Practice with 3D tasks on Geometer's Sketch Pad | 3 | Paper Folding task, Paper Form Board test, V-K MRT, Card Rotations test, Cub Comparison test, Hidden Patterns test, CAB -- flexibility of closure | 1,2 | 3 | 3 | 1.008 | 7 |
| Shavalier (2004) - Overall | Trained with Virtus WalkThrough Pro software vs. Control group (no treatment) | 3 | Paper Folding test, Eliot Price test (adaptation of 3 mountains), V-K MRT | 2,4 | 3 | 1 | .211 | 3 |
| Shavalier (2004) - Control | | | | | | | .435 | 3 |
| Shavalier (2004) - Treatment | | | | | | | .491 | 3 |
| Shubbar (1990) | 3 vs. 6 vs. 30 second rotation speed, with or | 3 | Mental rotations test (author created) | 2 | 2 | 2 | 2.260 | 6 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | without shadow | | | | | | | | |
| Shyu (1992) | Origami instruction and prior knowledge | 3 | Building an Origami Crane | 2 | 3 | 3 | -.686 | 1 |
| Simmons (1998) - Overall | Took pre-test and post-test on both Visualization and GEFT vs. only Visualization vs. no Visualization (GEFT posttest only) | 3 | Visualization test, GEFT | 1,2 | 3 | 3 | .359 | 3 |
| Simmons (1998) - Control | Self-paced instruction booklet in orthographic projection vs. control (professor-led discussion of professional issues) | 3 | Visualization test, GEFT | 1,2 | 3 | 3 | .565 | 1 |
| Simmons (1998) - Treatment | | | | | | | .646 | 1 |
| Sims & Mayer (2002) - Overall | Tetris players vs. non-Tetris players vs. control (no video game play) | 1 | Paper Folding test, Form Board and MRT (with tetris vs. nontetris shapes or letters), Card Rotations test | 2 | 1 | 3 | .316 | 9 |
| Sims & Mayer (2002) - Control | | | | | | | 1.111 | 9 |
| Sims & Mayer (2002) - Treatment | | | | | | | 1.193 | 9 |
| Smith, G.G. (1998) - Overall | Active (used computer) vs. passive participants (watched actives use the computer) | 3 | Visualization puzzles, Polynomial assembly | 2 | 2 | 1 | -.630 | 1 |
| Smith, G.G. (1998) - Control | | | | | | | .220 | 1 |
| Smith, G.G. (1998) - Treatment | | | | | | | -.254 | 1 |
| Smith, Gerretson, et al. (2009) - Overall | Solving interactive Tetronimo problems | 2,3 | Accuracy on visualization puzzles | 2 | 1 | 3 | .330 | 8 |
| Smith, Gerretson, et al. (2009) - Control | | | | | | | .360 | 8 |
| Smith, Gerretson, et al. (2009) - Treatment | | | | | | | .483 | 8 |
| Smith, J.P. (1998) | Instruction in chess | 2 | Guilford-Zimmerman test of Spatial Orientation, G-Z test of Spatial Visualization, GEFT | 1,2,4 | 3 | 2 | 1.037 | 3 |
| Smith & Sullivan (1997) | Instruction in chess | 3 | GEFT | 1 | 3 | 2 | .380 | 1 |
| Smith, R.W. (1996) | Animated versus non-animated feedback | 3 | Author created MRT -- Accuracy and RT | 2 | 3 | 3 | .506 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Smyser (1994) - Overall | Computer program for spatial practice | 2 | Card Rotations test | 2 | 3 | 2 | .166 | 3 |
| Smyser (1994) - Control | | | | | | | 1.502 | 3 |
| Smyser (1994) - Treatment | | | | | | | 1.688 | 3 |
| Snyder (1988) - Overall | Training to increase field Independence | 3 | GEFT | 1 | 3 | 3 | .274 | 2 |
| Snyder (1988) - Control | | | | | | | 1.207 | 2 |
| Snyder (1988) - Treatment | | | | | | | .851 | 2 |
| Sokol (1986) - Overall | Biofeedback-assisted relaxation | 3 | EFT | 1 | 3 | 3 | .319 | 1 |
| Sokol (1986)- Control | | | | | | | .135 | 1 |
| Sokol (1986) - Treatment | | | | | | | .431 | 1 |
| Sorby (2007) | Pre-test vs. post-test scores for those in initial spatial skills course (1 quarter) or those in multimedia software course (1 semester) | 2 | Mental Cutting test (*MCT*), SR-DAT, PSVT-Rotation | 2 | 3 | 3 | 1.718 | 9 |
| Sorby & Baartmans (1996) | Freshman engineering students (male and female) | 2 | PSVT-Rotation - Score - Identify 3D irregular solid in a different orientation | 2 | 3 | 3 | .926 | 1 |
| Spangler (1994) | Computer lesson converting 3D objects to 2D and 2D objects to 3D, Mental Rotation of 3D objects | 3 | 2D Sketching, 3D Sketching, MRT | 2 | 3 | 3 | .573 | 30 |
| Spencer (2008) - Overall | Practice with physical, digital or choice of physical or digital geometric manipulatives | 3 | Test of Spatial Visualization in 2D Geometry, Wheatley Spatial Ability test | 2 | 3 | 3 | -.005 | 10 |
| Spencer (2008) - Control | | | | | | | .463 | 2 |
| Spencer (2008) - Treatment | | | | | | | .191 | 6 |
| Sridevi, K., M. Sitamma, et al. (1995) - Overall | Yoga Practice | 3 | Perceptual Acuity test, GEFT | 1 | 3 | 3 | .967 | 2 |
| Sridevi, K., M. Sitamma, et al. (1995) - Control | | | | | | | -.110 | 2 |
| Sridevi, K., M. Sitamma, et al. (1995) - Treatment | | | | | | | .626 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Stewart (1989) | Lecture on map interpretation/terrain analysis | 3 | Map Relief Assessment exam | 5 | 3 | 3 | .381 | 3 |
| Storey Vasu, E. & Kennedy Tyler, D. (1997) - Overall | Course training using LOGO | 2 | Developing Cognitive Abilities test -- Spatial | 2 | 3 | 1 | .216 | 3 |
| Storey Vasu, E. & Kennedy Tyler, D. (1997) - Control | | | | | | | .577 | 2 |
| Storey Vasu, E. & Kennedy Tyler, D. (1997) - Treatment | | | | | | | .488 | 1 |
| Subrahmanyam & Greenfield (1994) | Playing spatial video game Marble Madness vs. control (quiz show game Conjecture) | 1 | Computer based test of dynamic spatial skills | 1 | 1,2 | 1 | 2.176 | 2 |
| Sundberg (1994) | Spatial training with physical materials and geometry instruction | 3 | Middle Grades Mathematics Project | 5 | 3 | 2 | 1.143 | 4 |
| Talbot & Haude (1993) | Experience with Sign Language | -- | MRT | 2 | 1 | 3 | .416 | 3 |
| Terlecki, Newcombe & Little (2008) - Overall | Playing tetris along with repeated practice vs. control (repeated practice only) | 1,3 | Paper Folding task, Surface Development test, Guilford-Zimmerman Clock task, MRT | 2 | 1,2,3 | 3 | .305 | 6 |
| Terlecki, Newcombe & Little (2008) - Control | | | | | | | .629 | 6 |
| Terlecki, Newcombe & Little (2008) - Treatment | | | | | | | .852 | 6 |
| Thomas, D. (1996) - Overall | 3D CADD instruction vs. control (2D CADD instruction) | 2 | Cube rotation (author created) | 2 | 2 | 3 | .299 | 2 |
| Thomas, D. (1996) - Control | | | | | | | .745 | 2 |
| Thomas, D. (1996) - Treatment | | | | | | | 1.074 | 2 |
| Thompson & Sergejew (1998) | Practice of the Wechsler Adult Intelligence Scale-Rotations (*WAIS-R*) Block Design test, MRT | 3 | WAIS-R Block Design test, MRT | 2 | 3 | 3 | .380 | 2 |
| Thomson (1989) | Transformational geometry and mapping computer programs | 3 | Map Relief Assessment test | 4 | 3 | 1 | .442 | 3 |
| Tillotson (1984) - Overall | Classroom training in spatial visualization skills | 2 | Punched Holes test, Card Rotation test, Cube Comparison | 2 | 3 | 1 | .762 | 3 |
| Tillotson (1984) - Control | | | | | | | .127 | 3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tillotson (1984) - Treatment | | | test | | | | .667 | 3 |
| Tkacz (1998) | Map Interpretation and Terrain Association Course | 2 | Perspective Orientation, Shepard-Metzler MRT, 2D Rotation, GEFT | 1,2,4 | 3 | 3 | -.063 | 12 |
| Trethewey (1990) - Overall | Paired with partner vs. control (worked alone): High vs. mid vs. low scorers on PFT placement test | 3 | MRT, Flexibility of Closure | 1,2 | 3 | 3 | .481 | 4 |
| Trethewey (1990) - Control | | | | | | | .604 | 4 |
| Trethewey (1990) - Treatment | | | | | | | .591 | 3 |
| Turner (1997) - Overall | Computer-Aided Design mental rotation training using same or different, old or new item types, for Cooper Union vs. Penn State engineering students vs. control (standard wireframe CAD) | 3 | V-K MRT | 2 | 1,2 | 3 | .141 | 8 |
| Turner (1997) - Control | | | | | | | .117 | 8 |
| Turner (1997) - Treatment | | | | | | | .219 | 8 |
| Ursyn-Czarnecka (1994) - Overall | Geology Course with computer art graphics training | 2 | V-K MRT | 2 | 3 | 3 | .037 | 1 |
| Ursyn-Czarnecka (1994) - Control | | | | | | | .154 | 1 |
| Ursyn-Czarnecka (1994) - Treatment | | | | | | | .169 | 1 |
| Vasta, Knott & Gaze (1996) | Self-discovery (problems ranked in difficulty and competing cues) vs. control (equal practice with non-ranked problem set) | 3 | WLT, Plumb-Line task | 3 | 1,2 | 3 | .214 | 4 |
| Vazquez (1990) - Overall | Training on Spatial Visualization with the aid of a graphing calculator | 3 | Card Rotations test | 2 | 3 | 2 | .459 | 1 |
| Vazquez (1990) - Control | | | | | | | .325 | 1 |
| Vazquez (1990) - Treatment | | | | | | | .852 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Verner (2004) | Practice with RoboCell computer learning environment | 3 | Spatial Visualization test -- Author designed and Eliot & Smith, MRT -- Author designed and Eliot & Smith, Spatial Perception test -- Author designed and Eliot & Smith | 1,2 | 3 | 2 | .532 | 6 |
| Wallace & Hofelich (1992) | Training in Geometric analogies | 3 | MRT -- Accuracy and RT | 2 | 3 | 3 | 1.073 | 3 |
| Wang, Chang & Li (2006) - Overall | 3D media presenting interactive visualization exercises showing different perspectives, manipulation and animation of objects vs. control (2D media) | 3 | Purdue Visualization of Rotation test | 2 | 3 | 3 | .389 | 1 |
| Wang, Chang & Li (2006) - Control | | | | | | | -.211 | 1 |
| Wang, Chang & Li (2006) - Treatment | | | | | | | .080 | 1 |
| Werthessen (1999) - Overall | Hands on construction of 3D figures | 2 | SR-DAT, V-K MRT | 2 | 1,2 | 1 | 1.282 | 4 |
| Werthessen (1999) - Control | | | | | | | .32 | 4 |
| Werthessen (1999) - Treatment | | | | | | | 1.345 | 4 |
| Wideman & Owston (1993) | Weather system prediction task | 2 | SR-DAT | 2 | 3 | 2 | .229 | 3 |
| Wiedenbauer & Jansen-Osmann (2008) | Manually operated rotation of digital images | 3 | Author created computerized MRT -- Accuracy and RT | 2 | 3 | 1 | .506 | 2 |
| Wiedenbauer, Schmid & Jansen-Osmann (2007) - Overall | Virtual Manual MRT training with joystick vs. control (play computer quiz show game). Compared rotations of $22.5^{o}$, $67.5^{o}$, $112.5^{o}$, $157.5^{o}$ | 3 | V-K MRT – RT, Errors | 2 | 1,2 | 3 | .462 | 16 |
| Wiedenbauer, Schmid & Jansen-Osmann (2007) - Control | | | | | | | .245 | 16 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Wiedenbauer, Schmid & Jansen-Osmann (2007) - Treatment | | | | | | | .373 | 16 |
| Workman, Caldwell & Kallal (1999) | Training in clothing construction and pattern making vs. control (no training) | 2 | Apparel Spatial Visualization test (author created), SR-DAT | 2 | 1 | 3 | 1.015 | 2 |
| Workman & Lee (2004) | Flat pattern apparel training | 2 | Apparel Spatial Visualization test (author created), Paper Folding task | 2 | 3 | 3 | .373 | 2 |
| Workman & Zhang (1999) | Computer-Aided Design vs. Manual Pattern making vs. control (course in CAD instead of Pattern making) | 2 | Apparel Spatial Visualization test (author created). Surface Development test | 2 | 3 | 3 | 1.937 | 4 |
| Wright, Thompson, Ganis, Newcombe & Kosslyn (2008) - Overall | Practice vs. transfer on MRT or Paper Folding task | 3 | Mental Paper Folding task and MRT - RT, slope, intercept , errors | 2 | 3 | 3 | .373 | 12 |
| Wright, Thompson, Ganis, Newcombe & Kosslyn (2008) - Control | | | | | | | 1.201 | 12 |
| Wright, Thompson, Ganis, Newcombe & Kosslyn (2008) - Treatment | | | | | | | .581 | 12 |
| Xuqun & Zhiliang (2002) | Cognitive processing of image rotation tasks | 3 | MRT - Accuracy and RT, Assembly and Transformation task - - Accuracy and RT | 2 | 2 | 3 | .619 | 4 |
| Yates (1986) | Spatial visualization training vs. control (no training) | 3 | Paper Folding task, Cube Comparison test | 2 | 2 | 3 | .619 | 2 |
| Yates, B (1988) | Computer based teaching of spatial skills | 3 | DAT | 2 | 3 | 2 | .490 | 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Yeazel (1988) | Air Traffic Control task, repeated practice | 3 | Angular Error on the Air Traffic Control task | 1 | 3 | 2 | .869 | 8 |
| Zaiyouna (1995) | Computer training on MRT | 3 | V-K MRT, Accuracy and Speed | 2 | 1,2,3 | 3 | 1.754 | 3 |
| Zavotka (1987) - Overall | Animated films of rotating objects changing from 3D to 2D | 3 | Orthographic drawing task, V-K MRT | 2 | 3 | 3 | .597 | 9 |
| Zavotka (1987) - Control | | | | | | | -.048 | 1 |
| Zavotka (1987) - Treatment | | | | | | | .523 | 3 |

[a] 1 = Video game
2 = Course
3 = Spatial Task training

[b] 1 = Intrinsic, Static
2 = Intrinsic, Dynamic
3 = Extrinsic, Static
4 = Extrinsic, Dynamic
5 = Measure that spans cells

[c] 1 = Females
2 = Males
3 = Not specified

[d] 1 = Younger than 13
2 = 13 - 18 years
3 = Adults (over 18)