

*Selected papers presented at  
the [2nd Network Gender &  
STEM Conference](#), 3–5 July  
2014, in Berlin, Germany*

*In association with*



**NETWORK  
GENDER  
& STEM**  
educational and  
occupational pathways  
and participation

## **The STEM Gender Gap: The Case for Spatial Interventions**

***Lynn S. Liben***

***Department of Psychology, Pennsylvania State University, USA***

### **ABSTRACT**

This paper begins with a brief documentation of the narrowing but still persistent STEM gender gap in the United States. Arguments are then presented for the role of spatial thinking in STEM. Included are brief reviews of early work establishing spatial intelligence as a distinct human capacity and later work aimed at identifying component spatial skills. Despite continuing disagreements about how sub-skills should be categorized and labeled, there is growing consensus that spatial thinking is entailed in STEM disciplines, and that individuals with better spatial skills are more likely to enter, remain, and excel in STEM. Also discussed are data showing male advantages in spatial skills. Combining the role of spatial thinking in STEM with gender differences in spatial skills generates the suggestion that spatial interventions may help reduce the STEM gender gap. Arguments and empirical findings from spatial intervention studies are then reviewed. Extant work permits the conclusion that spatial interventions can advance performance in spatial tasks, but is not yet conclusive about their impact on STEM outcomes in general or on the STEM gender gap in particular. The paper ends by highlighting some remaining questions and recommendations for future research.

### **KEYWORDS**

STEM; gender; spatial thinking; education; spatial curriculum; spatial skills; interventions



## The STEM Gender Gap: The Case for Spatial Interventions

### INTRODUCTION

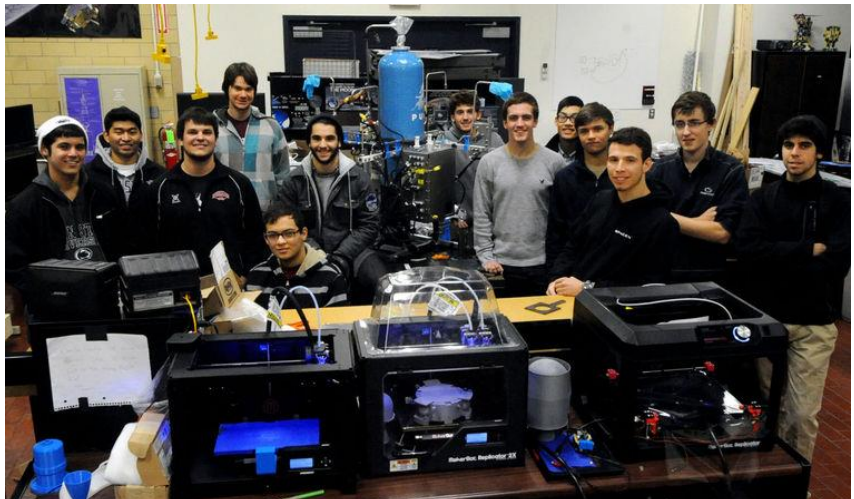
When Sally Ride was among the crew of the US Space Shuttle flight of June 18, 1983, the headlines were no longer focused on only the seeming miracle of a successful launch into space; they were also focused on the seeming miracle of a woman astronaut. The front-page headline of *The New York Times*, for example, read: *Shuttle Rockets to Orbit with 5 Aboard: Physicist is First U.S. Woman in Space*. After a brief description of mission objectives, the story continued:

But what set this flight apart from the 36 other manned [sic] American space missions over the last 22 years was not the cargo but the occupant just behind the two pilots. She was Dr. Sally K. Ride, a 32-year-old physicist who has been in astronaut training since 1978. She is the third woman to fly in space, but the first on an American mission. (Wilford, 1983, p. A1)

By the time the final flight of NASA's 30-year Space Transportation System (STS) program was launched in 2011, the fact that the crew included a woman engineer – Sandra Hall Magnus – was no longer headline news. STS-135 was only one of four shuttle missions Magnus had flown, including two that carried her to and from a five-month stint on the International Space Station.

That historical change, coupled with other gender-related changes over the same years (e.g., between 1983 and 2011 the proportion of women in US medical schools rose from about a third to about half; [Association of American Medical Colleges, 2012]) might suggest that the under-representation of women in science, technology, engineering, and mathematics (STEM) fields is behind us. A recent front-page article in Penn State's student newspaper, however, suggests otherwise. The article describes a campus group's efforts to design a spacecraft to travel to the moon and transmit images back to Earth (Garrity, 2014). The photograph accompanying the article is reproduced in Figure 1. At least as judged from

Figure 1. *Photograph accompanying a news story describing the activities of the Penn State Lunar Lions (Garrity, 2014); photograph by Kevin Kelley. Reproduced with permission of The Daily Collegian.*



appearance, 13 of the 13 pictured club members are male. The group's website likewise implies little involvement of women. For example, among the screen shots that index club videos (Penn State Lunar Lion Team, 2014), only two include a woman (the same in both). One vlog is captioned "Jason works on the layout for the power box, while Kara shows us the mural she painted on the new lab" and the other "Kara takes us through the process of creating the mural in the new lab." It appears that women are under-represented in the efforts to design and launch this moon spacecraft.

Why, in the first quarter of the 21<sup>st</sup> century, do gender differences such as these still exist, and what might be done to reduce them? The latter intervention question provided the theme for the second Gender and STEM Network Conference that led to this special issue of IJGST and thus to the current paper. More specifically, the question guiding the conference was *What can schools, families, and workplaces do?* In an attempt to address this question within a realistic scope for a single paper, I focus here on interventions that target one particular domain – spatial thinking – and draw examples from one geopolitical context – the United States (US).

More specifically, in the section entitled "The STEM Gender Gap," I document by data (rather than anecdote) the existence of a STEM gender gap in the US. In "Why Spatial Thinking?" I explain my rationale for focusing on spatial thinking in the context of STEM. I provide conceptual arguments and illustrative empirical data for the claims that spatial skills are foundational for STEM fields and that – at the group level – girls and women display lower levels of spatial skill than their male peers. Taken together, these observations imply that interventions that facilitate spatial thinking may improve STEM participation and success of students in general and of girls and women in particular. In "Spatial Interventions" I review arguments and empirical data testing the impact of spatial interventions, point to some remaining questions, and briefly describe two ongoing studies to illustrate interventions in school and family contexts. In "Concluding Comments" I offer general suggestions for future work.

## **THE STEM GENDER GAP**

Within the US, it is a relatively simple matter to monitor the existence of a STEM gender gap because the National Science Foundation (NSF) routinely collects and disseminates information about the science and engineering workforce in relation to gender and race/ethnicity. Pie charts reproduced in Figure 2 provide 2010 data. For the population in general, proportions of men and women are roughly equal, but for the science and engineering workforce in particular, men dramatically outnumber women. This contrast is especially telling given that girls in the US outperform boys on many broad academic indicators (e.g., school grades and high-school and college graduation rates, see Leaper, 2015). Historical data show that despite changes over recent decades, the gap persists in many fields. For example, between 1966 and 2006 the proportion of bachelor's degrees awarded to women jumped in biological and agricultural science (25% to 60%); chemistry (19% to 52%); mathematics (33% to 45%); and Earth, atmospheric, and ocean sciences (9% to 41%), but

remained disproportionately low in computer science (15% to 21%); physics (5% to 21%); and engineering (1% to 19%).

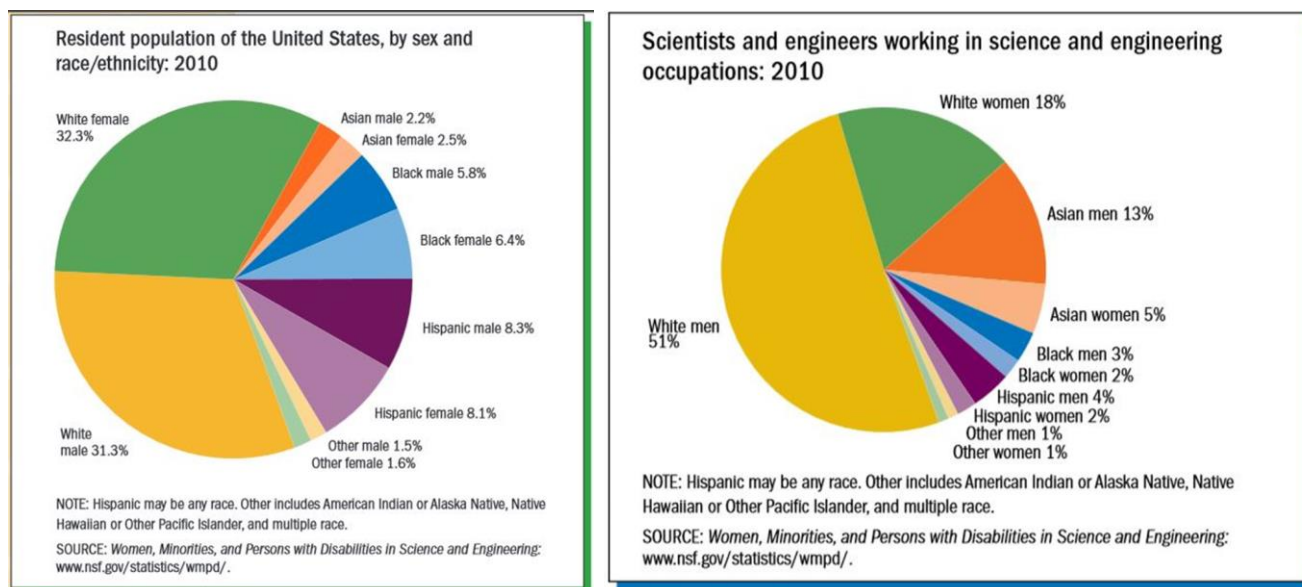


Figure 2. Distributions by gender and race/ethnicity of the general population (left) and of the science and engineering workforce (right) in the US. Reproduced courtesy of the National Science Foundation (2013; p.2 and p. 8, respectively).

Pre-college gaps are also evident. Illustratively, and as reviewed in more detail in another recent paper on the STEM gender gap and interventions (Liben & Coyle, 2014), there remain some striking differences between the proportion of high-school boys versus girls who take STEM Advanced Placement (AP) tests. For example, 2013 AP data (Ericson, 2014) show that although there are some AP tests for which girls' participation is almost as high as boys' (e.g., Calculus AB: 48% vs. 52%; Chemistry: 46% vs. 54%), and one in which girls' participation is greater (Biology: 58% vs. 42%), in others it is dramatically lower (e.g., Physics C: 23% vs. 77%; Computer Science A: 19% vs. 81%). Test scores also show significant gender differences; parallel discrepancies appear in national and international assessments of the general school population as well (Robelen, 2012).

In summary, despite girls' and women's increasing success and participation in STEM-related fields, a gender gap favoring boys and men remains. Societal commitments to improving gender equity and to building a stronger STEM workforce have motivated a range of interventions, including those targeting the domain of spatial thinking. The next section presents the rationale for this domain focus.

## WHY SPATIAL THINKING?

As a first step in explaining the focus on spatial thinking, it is important to define what spatial thinking is, and to document its relevance to gender. The definitional task is, however, challenging (see Liben, 2006), and any single definition cannot capture the full breadth and nuances of the construct. The authors of the National Research Council (NRC) report *Learning to Think Spatially* (NRC, 2006) proposed three components of spatial thinking: knowing about and coordinating information about space, knowing how to create and interpret representations of space, and reasoning about space. Drawing on this NRC report, the Science Education Resource Center of Carleton College (2012) defined spatial thinking as “thinking that finds meaning in the shape, size, orientation, location, direction or trajectory, of objects, processes or phenomena, or the relative positions in space of multiple objects, processes or phenomena” and which “uses the properties of space as a vehicle for structuring problems, for finding answers, and for expressing solutions.”

Much of the case for spatial thinking as an identifiable cognitive skill comes from early psychometric studies of human intelligence conducted in the US during the early 20<sup>th</sup> century. This work was motivated by the perceived need to screen individuals for decisions about who should be allowed to immigrate, join the military, or pursue vocational or college programs. Because some of the people to be screened were not fluent in English and because some jobs required practical skills, diverse tests were developed. Factor analyses of performance across the resulting tests consistently revealed a kind of ability that was identified as spatial intelligence (see Eliot, 1987; Hegarty & Waller, 2005; Lohman, 1994; Thurstone, 1938).

Psychometricians then turned to identifying component spatial skills. One commonly identified sub-skill was *mental rotation* – the ability to imagine how objects look as they are moved through space. Other spatial factors tended to vary across studies and analyses (e.g., see Carroll, 1993). A revealing summary is what Eliot (1987) wrote about this phase of work in his review of alternative models of psychological space: “descriptions of spatial factors were frustrating to other researchers because they appeared at best vague and at worst self-contradictory” (p. 55). In short, there was agreement about the existence of a general spatial ability but relatively little about the number, definition, or labels of component skills.

Perhaps reflecting his pessimistic view of the outcome of these factor-analytic studies, Eliot (1987) judged that the push to identify sub-skills largely ended in the 1960s. His pronouncement appears to have been premature, however, especially from the perspective of a dual interest in spatial development and gender. One of the most useful papers on the structure of spatial skills appeared about two decades later when Linn and Petersen (1985) used meta- and task-analytic approaches to identify component spatial skills and the developmental trajectories of sex differences within each. That there were significant and often considerable sex differences in spatial abilities – favoring men and boys – was becoming more visible at that time. In part these differences were becoming better known because

of a growing focus on the study of gender in general (e.g., Maccoby & Jacklin, 1974) and in part because several studies revealed unexpected and striking sex differences on spatial tasks. Particularly dramatic were findings from research using the Piagetian water-level task (Piaget & Inhelder, 1956), in which respondents are asked to indicate the position that liquid takes in a tipped container. Successfully perceiving and representing the water as horizontal was said to depend on the child having established a Cartesian coordinate axis system, indicative of the more general development of Euclidean concepts. Later studies (e.g., Liben, 1978; Liben & Golbeck, 1980, 1986; Thomas, Jamison & Hummel, 1973) replicated Piaget's observations of age-related improvement in water-level performance, but simultaneously revealed unexpected adult failures and unanticipated (and dramatic) sex differences favoring males.

In their meta-analysis, Linn and Petersen (1985) drew studies from these diverse research traditions and thereby identified three categories of spatial skills. The first, *mental rotation*, is the skill identified in the psychometric literature discussed earlier. Linn and Petersen reported that sex differences in mental rotation were evident by childhood. More recent investigators have shown that these sex differences appear by early childhood (Levine, Huttenlocher, Taylor & Langrock, 1999) and even by early infancy (Quinn & Liben, 2008). The second, *spatial perception*, refers to skill in relating one's bodily position to something in the external space. Performance on spatial perception tasks (such as the water-level task) also showed consistent sex differences beginning in early childhood. The third category, *spatial visualization*, was conceptually the least clear-cut of the skills; tasks falling into this category (e.g., paper-folding tasks) were said to involve multiple-step manipulations of spatial information solved by varied strategies. Probably reflecting the potential for using verbal, analytic strategies in place of (or in addition to) spatial, imagery-dependent ones, sex differences within this category were inconsistent, often entirely absent in individual studies.

Although this three-category system has driven task-selection in some programs of research (e.g., Liben, Susman et al., 2002), its adoption has been far from universal. Terms originally used by Linn and Petersen (1985) to refer to particular categories of spatial skills and particular tasks have since been applied to other skill sets, and entirely new classification systems have been proposed (see Hegarty, 2014; Hegarty & Waller, 2005; Uttal et al., 2013). Simultaneous use of both earlier and new categorization systems and labels in the contemporary literature means that Eliot's (1987) comment about the frustrating nature of the way that spatial factors were conceptualized and labeled during the 1960s remains applicable today. However, despite the continuing disagreements about exactly how to group, distinguish, label, and characterize specific spatial skills or concepts, there is strong agreement that many different kinds of spatial skills and concepts exist, and that, in many of them, girls and women on average perform worse than boys and men. Why should these differences matter for STEM entry and success?

The conceptual connection between spatial thinking and STEM was made compellingly by the Committee on Support for Thinking Spatially. Their report (NRC, 2006) provided extensive examples of the role spatial thinking played in

scientific discovery in fields as diverse as biochemistry, epidemiology, astronomy, geoscience, and geography. Their report also offered illustrations of the relevance of spatial thinking for K-12 curriculum in science and mathematics. Empirical data have also demonstrated links between individuals' performance on spatial tests and their success in scientific and technical domains. One of the most extensive programs of relevant work is research by Benbow, Lubinski, and Humphreys and their colleagues, who have used large, longitudinal data sets to examine the association between scores on spatial tests given during early adolescence, and educational and occupational outcomes observed during adulthood. In one such study, Shea, Lubinski, and Benbow (2001) followed over 500 students who had been identified as intellectually talented via spatial, verbal, and quantitative tests. These investigators found that higher levels of spatial ability at 13 years were associated with greater likelihood of completing STEM-related degrees (both undergraduate and graduate) and of entering STEM occupations during adulthood. Spatial ability accounted for variance in STEM outcomes above what could be accounted for by mathematical ability alone. Wai, Lubinski, and Benbow (2009) examined similar associations in a nationally representative sample of 400,000 students who were given various intelligence tests in grades 9 to 12 (roughly 15 to 18 years) and followed for over a decade. Again, spatial ability predicted later educational and occupational STEM achievement, thus demonstrating that associations between spatial abilities and STEM are evident not only among intellectually talented students, but also within the general student population.

Other investigators have focused on the association between spatial skills and success in individual STEM disciplines such as physics (e.g., Kozhevnikov, Motes & Hegarty, 2007), chemistry (e.g., Wu & Shah, 2004), engineering (e.g., Hsi, Linn & Bell, 1997), and mathematics (Casey, Nuttall & Pezaris, 2001). Some investigators have attempted to correlate spatial skills and STEM mastery at an even finer level of analysis by examining links between specific spatial skills and specific STEM content. Such research can be useful not only for illuminating the spatial processes entailed in particular STEM tasks, but also for suggesting what educational experiences are likely to help students master particular STEM content or practices.

For example, investigators have attempted to link particular spatial skills to understanding diverse geological concepts (e.g., see Alles & Riggs, 2011; Kastens & Ishikawa, 2006; Liben & Titus, 2012; Titus & Horsman, 2009). Illustrative is an investigation of the link between college students' understanding of coordinate axes and their understanding of the geological concepts of strike and dip (Liben, Kastens & Christensen, 2011). Geologists use these concepts as they gather information about rock outcrops (the parts of rock formations that are visible above the surface of the surrounding land) in their quest to draw inferences about hidden geological features. More specifically, geologists record on maps data about each outcrop's orientation relative to north (*strike*) and about the direction and steepness of the outcrop's incline (*dip direction* and *angle*, respectively).

Given that strike is defined by the intersection of the rock surface with a horizontal plane, and that dip angle refers to incline relative to the horizontal surface, one reason these geology concepts might be difficult is that they call upon the student's

understanding of coordinate axes, which, as discussed earlier in the context of the water-level task, is poorly developed in some adults. To test the hypothesized association, Liben et al. (2011) first gave 655 college students a water-level test. Equal numbers of men and women who performed at high, moderate, or low levels were then recruited into the geology study proper. (As expected from findings discussed earlier, there were significant gender differences in water-level performance within the larger group, but – given the selection procedure – not within the sample participating in the geology study itself.)

Geology-study participants (N=125) were escorted individually to a campus field site containing an outcrop, given a short lesson about strike and dip, and asked to record the outcrop's strike and dip lines on a map of the immediate area. Students with low (but not high) water-level scores placed strike lines more or less randomly on the map. Students with high scores were significantly more likely than those with low scores to – correctly – draw strike and dip lines as orthogonal to one another. Findings were thus consistent with the proposal that students would be advantaged in learning the geology lesson on strike and dip if they had horizontality and verticality concepts available.

Taken together, the material covered in this section leads to the overarching conclusions that spatial thinking is relevant to STEM and that girls and women are relatively disadvantaged in spatial thinking. These findings set the stage for developing intervention programs that in some way target spatial skills as a way of improving STEM performance in general, and of reducing the STEM gender gap in particular. Before turning to a discussion of spatial interventions, however, it is important to state explicitly that such interventions cannot be expected to ameliorate the STEM gender gap entirely, or to reduce gender gaps across all STEM disciplines and learning tasks. As recent discussions of the role of spatial education for STEM education have begun to highlight explicitly (see Hegarty, 2014; Liben, 2012), far more research is needed, first, to identify which specific spatial skills support learning of which specific STEM content and practices, and, second, to test whether spatial interventions that enhance one kind of spatial skill also extend benefits to other spatial skills and to learning a diversity of STEM concepts and practices.

## **SPATIAL INTERVENTIONS**

Arguments for the importance of addressing spatial skills in educational contexts are not new. One early proponent was Gardner (1983), who identified spatial intelligence as one of the multiple intelligences traditionally given short shrift in formal schooling. Focusing attention on only verbal and mathematical domains, he argued, undermines many students' opportunities for personal success and deprives society of benefits potentially available from their diverse talents. Similarly, scholars associated with the Study of Mathematically Precocious Youth have argued for the importance of identifying and nurturing spatially talented individuals who have the capacity to develop into adults who make important contributions to STEM (e.g., see Lubinski, 2010).



What is more recent is the argument that spatial education should be offered to all students in support of normative educational, occupational, and life goals. This view is expressed in the following position statement included in the report on *Learning to Think Spatially*: "the committee views spatial thinking as a basic and essential skill that can be learned, that can be taught formally to all students, and that can be supported by appropriately designed tools, technologies, and curricula" (NRC, 2006, p. 231). The committee also observed and warned:

There are neither content standards nor valid and reliable assessments for spatial thinking. Without such standards and assessments, spatial thinking will remain locked in a curious educational twilight zone: extensively relied upon across the K-12 curriculum but not explicitly and systematically instructed in any part of the curriculum. (p. 232)

To address this problem, the committee urged "societal recognition of the importance of spatial thinking and an educational commitment to teaching spatial thinking to all students in all grades" (p. 232). The group recommended identifying which spatial-thinking concepts and tools students need, and then providing educational experiences across grades and disciplines that can promote those concepts and tools.

Consistent with the NRC (2006) call for "societal recognition" of the importance of spatial thinking, there have been increasing efforts to alert teachers, funders, parents, and other members of the public to its importance (e.g., Dewar, 2012a; Liben, 2006; Newcombe, 2006, 2010) and to develop broad guidelines, curricula, and materials to meet those educational goals from preschool through adulthood (e.g., Dewar, 2012b; Janelle, Hegarty & Newcombe, 2014; Sinton, 2011).

There has also been a growing empirical literature that delivers and tests the impact of spatial interventions. Findings from individual studies are mixed. Some concepts and skills appear surprisingly difficult to improve. Again, research using the water-level task is illustrative. Early investigators steeped in Piagetian theory were most concerned with whether training would have a differential impact on children who initially showed no understanding of horizontality whatsoever (i.e., erred on all pre-test items) in comparison to children who initially displayed nascent understanding (i.e., were correct on easy but not difficult pre-test items). Consistent with expectations derived from Piagetian theory, investigators reported that children in the former group generally gained little from training, whereas those from the latter improved significantly. However, as reviewed in detail elsewhere (Liben, 1991a), investigators reporting successful outcomes tended to evaluate training effects by using test items that were identical to training stimuli (i.e., identical bottle shapes and orientations). Improvement in children's scores could thus reflect their memorization of specific angles formed by water lines and bottle sides rather than their general skill in predicting and representing horizontals within a surrounding non-orthogonal frame. Interestingly, research with adults has also demonstrated resistance to instruction: even when told explicitly that "water remains horizontal or level (or straight across) regardless of the position of the

container," many college students continued to make errors (Liben & Golbeck, 1984).

A general and important issue raised by these examples is the importance of selecting appropriate tools to measure intervention effects (see Liben, Kastens & Stevenson, 2002). The narrowest assessment is provided by tasks and items that are identical to those used in training. Of course, demonstrating that one has retained and can reproduce exactly what one has been taught is not unimportant, particularly if the test occurs after a significant delay, but more compelling evidence of impact requires measures using at least moderately different test items (e.g., different container positions) and preferably items tapping the same underlying concept in a different manner (e.g., substituting a cross-bar horizontality test for the water-level test, see Liben, 1991b). However, for the current focus, more critical are measures that tap actual STEM performance. Again, using the horizontality concept to illustrate, one might provide water-level training for some students and test whether their performance on geological strike and dip tasks surpasses that of untrained students. Such an approach would implement the "Real-World Application" assessment approach recommended for curriculum evaluation in general (Liben et al., 2002). Unfortunately, this approach is often overlooked in the spatial training literature. Indeed, as discussed in a section on "missing elements" of their meta-analysis of spatial training studies, Uttal et al. (2013) remarked: "the lack of studies that directly assess the effects of spatial training on performance in a STEM discipline is disappointing" (p. 356).

It is not easy to determine whether a given outcome measure tests "near" versus "far" transfer from the original training task. For example, in their meta-analysis, Uttal et al. (2013) classify as near-transfer tasks those that are drawn from the same spatial-skill category that was taught during the intervention and as far-transfer tasks those drawn from a different spatial-skill category. Importantly, though, the result depends upon what system is used to categorize spatial tasks in the first place. Illustratively, paper-folding and mental rotations tasks fall into the same group ("intrinsic and dynamic") in the system used by Uttal et al. but fall into two different groups ("visualization" vs. "mental rotation") in the system used by Linn and Petersen (1985).

Also varying among past studies is the scope of the spatial skills included in the intervention. For example, programs may target only a single skill with a single type of stimulus (e.g., mental rotation tasks using only block stimuli), a single skill with varied stimuli (e.g., mental rotation with both block and Tetris stimuli), multiple skills (e.g., mental rotation, cross-sections, and perspective-taking), and still others may target not tasks but rather integrated spatial conceptual systems such as the topological, projective, and Euclidean spatial conceptual systems identified by Piaget and Inhelder (1956).

The preceding comments about ways in which individual spatial intervention studies have varied is far from exhaustive, but even this truncated discussion makes it clear that spatial interventions vary simultaneously along many dimensions. Furthermore, despite the large number of individual studies, the extant corpus of

work does not provide all the data one would want because relatively few investigations include assessments of both spatial and STEM outcomes.

The recent meta-analysis by Uttal et al. (2013) referred to earlier has carefully and systematically addressed many of the dimensions along which studies vary, and has led to important conclusions about what this body of work reveals about the effectiveness of spatial training. Three of their conclusions are particularly relevant for the current focus. One is their central conclusion that spatial skills are malleable, summarized in their abstract as follows: "After eliminating outliers, the average effect size (Hedges'  $g$ ) for training relative to control was 0.47 ( $SE = 0.04$ )" (p. 352), a moderate effect size. A second important conclusion highlighted in their abstract is that "[t]raining also transferred to other spatial tasks that were not directly trained" (although again, as discussed earlier, this conclusion is affected by how spatial tasks are categorized, an issue still open to debate). Finally, although not included in the abstract, a third relevant conclusion included in their meta-analysis (replicating one drawn in an earlier meta-analysis by Baenninger & Newcombe, 1989) is that males and females profit about equally from spatial interventions.

Each of these three general conclusions is consistent with the general thesis of this paper that spatial education is a potentially useful route for improving STEM outcomes. Yet these general conclusions also raise important questions. First, with respect to the average effect size for training, will incrementing performance on a spatial task by half a standard deviation be sufficient to lead students to find themselves more strongly attracted to STEM or more successful in mastering it? Second, with respect to the observed transfer effects, do the observed leaps across task categories summarized by the meta-analysis mean that students are able and likely to apply their enhanced spatial skills to the content and practices of STEM? And, third, with respect to gender, is it sufficient to employ spatial interventions if they do not eliminate the gender difference in spatial performance? Do continuing gender differences in spatial skills necessarily imply continuing gender differences in STEM?

I end this section by returning to the question posed for the second Gender and STEM Network Conference – *What can schools, families, and workplaces do?* This question implies that interventions may occur in many contexts; spatial interventions are no exception. To illustrate this point, I briefly describe two of our ongoing projects that are targeted, respectively, for school and home.

The first (Liben, Signorella & Sorby, 2013) is funded by NSF's Gender in Science and Engineering Program, and is co-directed by principal investigators representing the disciplines of developmental psychology (Liben) and engineering education (Sorby). Its goals include providing a sustained spatial curriculum for children in their regular mathematics classes, and testing whether the intervention facilitates not only spatial skills but also students' STEM interests and achievements. An additional goal is to explore the hypothesis that motivational factors identified in expectancy-value theory (Eccles, 2014) may mediate STEM success, perhaps especially so for girls.

We are collecting data from two cohorts of students attending schools serving diverse populations. Some teachers (experimental) are assigned to incorporate a spatial skills curriculum into their seventh-grade math classes. The spatial curriculum (Sorby, 2011) is based on one previously found to increase spatial-test scores and program retention in women engineering students (Sorby & Baartmans, 2000). In our current work, students take a battery of spatial tests before and after the curriculum (waves 1 and 2), and again a year later (wave 3). At those three times they also respond to various surveys that tap job interests, job stereotypes, and their liking, valuing, and self-confidence in various domains and skills (e.g., science, spatial skills, musical skills, language arts). Another year later (wave 4) students respond to surveys inquiring about high-school club and course interests. Finally, we are obtaining multi-year data on math and science achievement including school grades and state achievement-test scores. Although the ongoing nature of data collection precludes presenting findings here, what is already evident is the value of having diverse outcome measures that allow us to examine the program's impact not only on paper-and-pencil spatial assessments, but also on STEM grades and interests.

The second example is aimed at the family context. This project (Borriello & Liben, 2014) begins from the premises that children's interests and foundational skills emerge and expand in early childhood; that play is an important context for developing spatial skills; and that parents are potentially well positioned to facilitate their children's spatial thinking. Past empirical work has already demonstrated correlational associations among parents' behaviors, children's play, and spatial or STEM outcomes (e.g., Pruden, Levine & Huttenlocher, 2011; Szechter & Liben, 2004). The current work was designed to test whether a brief coaching intervention would lead mothers to provide more spatial guidance to their preschool children during dyadic play. Dyads were first given jigsaw puzzles to establish baseline spatial behaviors and language, after which dyads were asked to construct a series of LEGO™-block structures shown in photographs. Mothers randomly assigned to the experimental group were told briefly about spatial thinking and why it might be valuable for their child, and alerted to some ways it could be fostered in play. Control-group mothers were simply asked to play with the block-construction activity as they would at home. Mothers' spatial language has been coded as one index of their spatial guidance. Analyses reveal significantly more spatial language among experimental than control mothers, suggesting that it would be useful to develop and evaluate the effectiveness of sustained home-based intervention programs in future work.

## **CONCLUDING COMMENTS**

By now there is a convincing body of work showing that spatial thinking skills are connected – first – to STEM engagement, persistence, and success and – second – to gender. There is also ample evidence of an often striking, far-reaching, and historically persistent gap between the proportions of girls and women versus boys and men who pursue, succeed, and persist in various STEM fields. These empirical

associations raise the possibility that a potential path toward reducing the STEM gender gap would be intervening to improve girls' and women's spatial skills.

Although I am happy to count myself among those who are enthusiastically designing and implementing spatial education, and although I am optimistic about its ultimate impact, I also believe it important to avoid assuming that spatial education will be a panacea for building STEM participation and success in general, or for reducing the STEM gender gap in particular. To test the effectiveness of these approaches, there must be excellent, well-controlled research studies, not anecdotes and testimonials; impact must be evaluated not only by paper-and-pencil measures of specific spatial skills, but by STEM-outcome measures of STEM participation, mastery, and persistence.

It is also critical to monitor intervention experiences and outcomes with respect to affect as well as cognition. It will do little good to provide education in the spatial domain if the experience is viewed as uninteresting, unpleasant, or unrewarding. Indeed, preliminary analyses of wave 1 and 2 data from one sample tested in the middle-school study described earlier sound a cautionary note (Liben et al., 2013). As one would hope, on almost all spatial-skill measures, we found an interaction between condition (experimental vs. control) and session (wave 1 vs. 2) favoring the experimental group. However, we also found a condition by session interaction on self-reported competence in spatial skills in which only experimental students (especially girls) reported *reduced* confidence in their spatial skills. The effect was small, and perhaps will not be replicated in other samples. However, if the finding holds, it might mean that participating in a spatial curriculum triggers students' realization that a domain they would normally have little reason to contemplate is, in fact, challenging. Such a finding might point to a need to warn students of potential challenges, establish incremental goals, and offer frequent performance feedback (as in athletic-skill training); it might also suggest encouraging students to be meta-cognitive about their own performance, an approach shown to be effective in a variety of educational challenges, including spatial ones (Kastens & Liben, 2007).

Finally, it is important to consider findings from meta-analyses that males and females profit about equally from spatial interventions. Does this mean that those working toward STEM gender equity should offer spatial interventions to only girls and women, or abandon spatial interventions entirely? I would argue that both questions should be answered with a resounding "no". First, what may be most important for attracting and retaining girls and women in STEM fields is achieving some necessary threshold of spatial competence which is independent of gains made by another group. Second, given highly overlapping distributions of males' and females' spatial skills, and assuming that the overarching societal goal is to increase STEM engagement and success among all those who are currently constrained by insufficient spatial skills, targeting spatial interventions by gender is an imprecise and unfair way to assign remedial support (see Bigler, Hayes & Liben, 2014). Furthermore, as discussed in more detail elsewhere (Liben, 2015; Liben & Coyle, 2014), targeting interventions to only one gender runs the risk of

exacerbating gender stereotypes, in turn undermining positive effects that may otherwise accrue from enhancing spatial skills.

In summary, far more work is needed to clarify what intervention qualities are critical; which skills are improved and whether these, in turn, improve STEM outcomes; and what other factors (e.g., motivation, peer contexts) are also critical. It will also be important to confront the value-laden issue of defining the end goal. Actions will differ depending on whether the goal is to achieve numerical parity for males and females in particular, or whether it is to ensure that all individuals – irrespective of gender, ethnicity, nationality, economic status, or any other dimension along which people vary – have access to the full range of opportunities and supports they need to engage and succeed in STEM.

## ENDNOTE

<sup>1</sup>This article is based on a keynote address delivered at the International Gender and STEM Conference held at Technische Universität Berlin, Germany, July 2014. I thank Rebecca Lazarides, Angela Ittel, and Helen M. G. Watt for their invitation to present at the conference and to prepare this paper. Portions of the research described were supported by National Science Foundation grants REC04-11686 and GSE1036715. However, the opinions, findings, conclusions, and recommendations are those of the author and no NSF endorsement should be inferred. Please address correspondence to Lynn S. Liben, Department of Psychology, The Pennsylvania State University, University Park, PA, 16802; [liben@psu.edu](mailto:liben@psu.edu).

## REFERENCES

- Alles, M. & Riggs, E. M. (2011). Developing a process model for visual penetrative ability. In A. D. Feig & A. Stokes (Eds.), *Qualitative Inquiry in Geoscience Education Research*. Geological Society of America, Special paper 474, pp. 63–80.
- Association of American Medical Colleges (2012). U.S. medical school applicants and students 1982–1983 to 2011–2012. Retrieved 17 July 2015 from <https://www.aamc.org/download/153708/data/charts1982to2012.pdf>
- Baenninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles* 20(5–6), 327.
- Bigler, R. S., Hayes, A. R. & Liben, L. S. (2014). Analysis and evaluation of the rationales for single-sex schooling. In L. S. Liben & R. S. Bigler (Eds.), *The role of gender in educational contexts and outcomes* (pp. 225–260). San Diego: Elsevier.
- Borriello, G. A. & Liben, L. S. (2014, May). Mothers' responsiveness to instructions to facilitate preschoolers' spatial thinking during play. Symposium presentation at the 1<sup>st</sup> PA Spatial Cognition Symposium. University Park, PA.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

- Casey, M. B., Nuttall, R. L. & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal of Research in Mathematics Education* 32, 28–57.
- Dewar, G. (2012a). Spatial intelligence in children: Why training matters. Retrieved 17 July 2015 from <http://www.parentingscience.com/spatial-intelligence.html>
- Dewar, G. (2012b). Improving spatial skills in children and teens: Evidence-based activities and tips. Retrieved 17 July 2015 from <http://www.parentingscience.com/spatial-skills.html>
- Eccles, J. S. (2014). Gender and achievement choices. In E. T. Gershoff, R. Mistry & D. Crosby (Eds.) *Societal contexts of child development* (pp. 19–34). New York: Oxford University Press.
- Eliot, J. (1987). *Models of psychological space, psychometric: Developmental, and experimental approaches*. New York: Springer-Verlag.
- Ericson, B. (2014). AP data for the United States. Retrieved 17 July 2015 from <http://home.cc.gatech.edu/ice-gt/321>
- Gardner, H. (1983) *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Garrity, M. (2014, November 19). One small step for Lunar Lions. *The Daily Collegian*, A1.
- Hegarty, M. (2014). Spatial thinking in undergraduate science education. *Spatial Cognition and Computation* 14(2), 142–167.
- Hegarty, M. & Waller, D. (2005). *Individual differences in spatial abilities. The Cambridge handbook of visuospatial thinking*. New York: Cambridge University Press.
- 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(2), 151–158.
- Janelle, D. G., Hegarty, M. & Newcombe, N. S. (2014). Spatial thinking across the college curriculum: A report on a specialist meeting. *Spatial Cognition and Computation* 14(2), 124–141.
- Kastens, K. A. & Ishikawa, T. (2006). Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. In C. A. Manduca & D. W. Mogk (Eds.), *Earth and Mind: How Geologists Think and Learn about the Earth*. Geological Society of America, Special Paper 413, pp. 53–76.
- 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, 45–74.
- Kozhevnikov, M., Motes, M. A. & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science* 31(4), 549–579.

- Leeper, C. (2015). Gender and social-cognitive development. In L. S. Liben & U. Müller (Vol. Eds.), *In Handbook of child development and developmental science*, 7<sup>th</sup> edn (Volume. 2, pp. 806–853). Hoboken, NJ: Wiley.
- Levine, S. C., Huttenlocher, J., Taylor, A. & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology* 35(4), 940–949.
- Liben, L. S. (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill-Palmer Quarterly* 24, 97–110.
- Liben, L. S. (1991a). The Piagetian water-level task: Looking beneath the surface. In R. Vasta (Ed.), *Annals of Child Development* (Vol. 8, pp. 81–143). London: Jessica Kingsley Publishers.
- Liben, L. S. (1991b). Adults' performance on horizontality tasks: Conflicting frames of reference. *Developmental Psychology* 27, 285–294.
- Liben, L. S. (2006). Education for spatial thinking. In K. A. Renninger & I. E. Sigel (Vol. Eds.), *Handbook of child psychology*, 6<sup>th</sup> edn (Vol. 4, pp. 197–247). Hoboken, NJ: Wiley.
- Liben, L. S. (2012). Spatial education across the college curriculum: A psychologist's perspective. *Spatial Thinking Across the Curriculum, Position Papers*, pp. 58–63. Retrieved 17 July 2015 from <http://spatial.ucsb.edu/specialist-meetings/>
- Liben, L. S. (2015). Probability values and human values in evaluating single-sex education. *Sex Roles* 13 doi:http://dx.doi.org/10.1007/s11199-014-0438-9
- Liben, L. S. & Coyle, E. F. (2014). Developmental interventions to address the STEM gender gap: Exploring intended and unintended consequences. In L. S. Liben & R. S. Bigler (Eds.) *The role of gender in educational contexts and outcomes* (pp. 77–116). San Diego: Elsevier.
- Liben, L. S. & Golbeck, S. L. (1980). Sex differences in performance on Piagetian spatial tasks: Differences in competence or performance? *Child Development* 51, 594–597.
- Liben, L. S. & Golbeck, S. L. (1984). Performance on Piagetian horizontality and verticality tasks: Sex-related differences in knowledge of relevant physical phenomena. *Developmental Psychology* 20(4), 595–606.
- Liben, L. S. & Golbeck, S. L. (1986). Adults' demonstration of underlying Euclidean concepts in relation to task context. *Developmental Psychology* 22, 487–490.
- Liben, L. S., Kastens, K. A. & Christensen, A. E. (2011). Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts. *Cognition and Instruction* 29, 45–87.
- Liben, L.S., Kastens, K. A. & Stevenson, L. M. (2002). Real-world knowledge through real-world maps: A developmental guide for navigating the educational terrain. *Developmental Review* 22, 267–322.
- Liben, L. S., Signorella, M. L. & Sorby, S. A. (2013, April). The impact of a spatial skills curriculum in middle school: Cognitive and social-cognitive outcomes in spatial



and STEM domains. Symposium presentation at the biennial meetings of the Society for Research in Child Development, Seattle.

Liben, L. S., Susman, E. J., Finkelstein, J. W., Chinchilli, V. M., Kunselman, S. J., Schwab, J. . . . Kulin, H. E. (2002). The effects of sex steroids on spatial performance: A review and an experimental clinical investigation. *Developmental Psychology* 38, 236–253.

Liben, L. S. & Titus, S. (2012). The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science. In K. A. Kastens & C. A. Manduca (Eds.), *Earth and mind II: A synthesis of research on thinking and learning in the geosciences*. The Geological Society of America, Boulder, CO, Special paper 486, pp. 51–70.

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development* 56, 1479–1498.

Lohman, D. F. (1994). Spatial ability. In R. J. Sternberg (Ed.), *Encyclopedia of intelligence* (Vol. 2, pp. 1000–1007). New York: Macmillan.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences* 49, 344–351.

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

National Research Council (2006). *Learning to Think Spatially*. Washington, D.C.: National Academy Press.

National Science Foundation (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013*. National Center for Science and Engineering Statistics; Directory for Social, Behavioral and Economic Sciences. Washington, D.C.: Retrieved 17 July 2015 from.

<http://www.nsf.gov/statistics/wmpd/2013/downloads.cfm>

Newcombe, N. S. (2006). A plea for spatial literacy. *The Chronicle of Higher Education*. Retrieved 17 July 2015 from

<http://www.iwitts.org/proven-practices/retention-sub-topics/spatial-reasoning/312-a-plea-for-spatial-literacy>

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator, Summer 2010*, 29–43.

Penn State Lunar Lion Team (2014). Penn State Lunar Lions' Team's You Tube videos. Retrieved 27 December, 2014 from

<https://plus.google.com/100178094478432687792/videos>)

Piaget, J. & Inhelder, B. (1956). *The child's conception of space*. New York: Norton.

Pruden, S. M., Levine, S. C. & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science* 14(6), 1417–1430.

Quinn, P. C. & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science* 19(11), 1067–1070.

- Robelen, E. (2012, June 11). Evidence persists of STEM achievement gap for girls. *Education Week*. Retrieved 17 July 2015 from [http://blogs.edweek.org/edweek/curriculum/2012/06/evidence\\_persists\\_of\\_stem\\_a\\_chi.html](http://blogs.edweek.org/edweek/curriculum/2012/06/evidence_persists_of_stem_a_chi.html)
- Science Education Resource Center of Carleton College (2012). Spatial thinking in geosciences. Retrieved 17 July 2015 from [http://serc.carleton.edu/research\\_on\\_learning/synthesis/spatial.html](http://serc.carleton.edu/research_on_learning/synthesis/spatial.html)
- 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, 604–614.
- Sinton, D. S. (2011). A spatial literacy initiative in higher education. *ArcNews* 32(4), 36–37. Retrieved 17 July 2015 from <http://www.esri.com/news/arcnews/winter1011articles/a-spatial-literacy.html>
- Sorby, S. A. (2011). *Developing spatial thinking*. Clifton Park, NY: Cengage Learning.
- 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* 89(3), 301–307.
- Szechter, L. E. & Liben, L. S. (2004). Parental guidance in preschoolers' understanding of spatial-graphic representations. *Child Development* 75(3), 869–885.
- Thomas, H., Jamison, W. & Hummel, D. D. (1973). Observation is insufficient for discovering that the surface of still water is invariably horizontal. *Science* 181, 173–174.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs* 1.
- Titus, S. & Horsman, E. (2009). Characterizing and improving spatial visualization skills. *Journal of Geoscience Education* 57, 242–254.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C. & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin* 139(2), 352–402.
- Wai, J., Lubinski, D. & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101(4), 817.
- Wilford, J. N. (1983, June 19). Shuttle rockets to orbit with 5 aboard: Physicist is first U.S. woman in space. *The New York Times*, A1
- Wu, H.-K. & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education* 88, 465–492.