Early Puzzle Play:
A Predictor of Preschoolers' Spatial Transformation Skill

Susan C. Levine, Kristin R. Ratliff, Janellen Huttenlocher, and Joanna Cannon
University of Chicago

Individual differences in spatial skill emerge prior to kindergarten entry. However, little is known about the early experiences that may contribute to these differences. The current study examined the relation between children's early puzzle play and their spatial skill. Children and parents ($n = 53$) were observed at home for 90 min every 4 months (6 times) between 2 and 4 years of age (26 to 46 months). When children were 4 years 6 months old, they completed a spatial task involving mental transformations of 2-dimensional shapes. Children who were observed playing with puzzles performed better on this task than those who did not, controlling for parent education, income, and overall parent word types. Moreover, among those children who played with puzzles, frequency of puzzle play predicted performance on the spatial transformation task. Although the frequency of puzzle play did not differ for boys and girls, the quality of puzzle play (a composite of puzzle difficulty, parent engagement, and parent spatial language) was higher for boys than for girls. In addition, variation in puzzle play quality predicted performance on the spatial transformation task for girls but not for boys. Implications of these findings as well as future directions for research on the role of puzzle play in the development of spatial skill are discussed.

Keywords: spatial skill, mental rotation, puzzle play, gender differences, preschool children

As early as the preschool years and persisting into adulthood, there are individual and gender differences on certain spatial tasks, notably those involving mental rotation (e.g., Halpern et al., 2007; Hyde, 1981; Levine, Huttenlocher, Taylor, & Langrock, 1999; Linn & Petersen, 1985). These variations are of considerable interest because of their reported relation to mathematics achievement (e.g., Casey, Nutall, & Pezaris, 1997) and to entry into science, technology, engineering, and mathematics disciplines (e.g., Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Hedges & Chung, 2011; Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). Although spatial skills have been shown to be sensitive to input variations (Baenninger & Newcombe, 1989, 1995; Casey & Braebeck, 1990; Huttenlocher, Levine, & Vevea, 1998; Terlecki, Newcombe, & Little, 2008), little is known about the kinds of naturally occurring early experiences that are related

to the development of these skills. This paucity of information stands in marked contrast to the detailed studies of naturalistic parent–child interactions that have illuminated the understanding of the kinds of early experiences that are related to the development of language and literacy skills. For example, many studies have shown that amount of parent speech is related to young children's vocabulary growth (e.g., Hart & Risley, 1995; Hoff, 2003; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991), that complexity of caregiver syntax is related to children's syntactic growth (e.g., Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002), and that book reading and exposure to print predicts later language and literacy skills (e.g., DeLoache & DeMendoza, 1987; Pan, Rowe, Singer, & Snow, 2005; Wells, 1987; Whitehurst et al., 1994).

In the current study, we begin an examination of individual variations in puzzle play, a kind of early experience that may be related to the development of spatial skills. We do this by adopting a research approach that forms the backbone of research on language development, observing naturalistic parent–child interactions. In particular, we focus on naturally occurring puzzle play, which many children engage in from an early age, and ask whether individual variation in amount and quality of puzzle play predicts individual variation in spatial transformation skill at the start of kindergarten. We focus on puzzle play as a potentially important early experience related to individual variations in spatial skill for several reasons. First, this kind of play provides a potentially rich context for developing mental rotation skill (e.g., Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005; Williams, 2004). That is, puzzle play typically involves both mentally and physically transforming pieces to fit into particular locations and provides immediate feedback about whether a piece fits. This

feedback allows children to see whether the outcomes of their mental and physical transformations are accurate. Second, puzzle play, unlike activities such as block play, is a spatial experience that is not gender stereotyped as traditionally masculine or traditionally feminine (Serbin & Connor, 1979). Thus, both girls and boys have the potential to gain spatial skill from this activity during play.

In addition to engaging children in a spatial activity that involves physical and mental transformations, puzzle play may increase children's exposure to spatial language (e.g., *edge, flat, straight, corner, curve, side, top, bottom, long, short, inside, outside, between, upside down, flip*) as parents frequently use such terms to guide children's efforts during puzzle play. Thus, we also ask whether the amount of spatial language children hear during puzzle play is related to their later skill on a task involving mental rotation and mental translation. As pointed out by Gentner (2003), exposure to spatial language may be particularly useful in "the learning and retention [of spatial concepts by] . . . inviting children to store the information and its label" (pp. 207–208). Supporting the role of spatial language in the acquisition of spatial concepts, Gentner found that children who heard specific spatial labels during a laboratory experiment that involved hiding objects (e.g., "I'm putting this *on/in/under* the box") were better able to find the objects in a mapping task than children who heard a general reference to location ("I'm putting this *here*"). Moreover, this advantage held 2 days later, without further exposure to the spatial language provided (Loewenstein & Gentner, 2005). Similarly, Szechter and Liben (2004) observed parents and children in the laboratory as they read a children's book with spatial–graphic content. These researchers found a relation between the frequency with which parents drew children's attention to the spatial–graphic content in the book (e.g., "the rooster is really *tiny* now") and children's performance on spatial–graphic comprehension tasks. Further, parent spatial language input has been shown to be related to children's performance on spatial tasks, with this relation mediated by the child's own use of spatial language (Pruden, Levine, & Huttenlocher, in press).

Several studies have examined the relation of young children's play activities to their spatial skill levels. Many of these studies rely on questionnaire responses about children's activities by parents, teachers or the children themselves (see Baenninger & Newcombe, 1989, 1995; Serbin, Zelkowitz, Doyle, Gold, & Wheaton, 1990). Some of these studies have reported a relation between frequency of participation in a variety of spatial activities that is related to mental rotation skill (Signorella, Jamison, & Krupa, 1989). Several studies have reported that block play, as measured by a combination of teacher/parent reports, is related to spatial visualization skills, such as the ability to reproduce abstract patterns, recognize geometric figures embedded within more complex pictures, and reproduce block constructions (Caldera et al., 1999; Serbin et al., 1990). However, reports of spatial activity participation may be influenced by respondents' perceptions of children's spatial skill; that is, respondents might overreport spatial activity participation for children they perceive as high spatial and underreport activity for those they perceive as low spatial, leading to a spurious correlation. Moreover, parents', teachers', and even children's reports of play activities may be influenced by social norms for the kinds of play that boys and girls engage in (Huston, 1983; Silvern, 1978).

Other studies have directly observed children's play. Connor and Serbin (1977), for example, observed children during their free-play time during preschool and characterized their activities as masculine or feminine sex typed. In addition, they assessed children's performance on the Preschool Embedded Figures Test (Witkin, Oltman, Raskin, & Karp, 1971), a task involving spatial skills, and on the Block Design subtest of the Wechsler Preschool and Primary Intelligence Scale (WPPSI; Wechsler, 1967). Results showed that boys' but not girls' performance on the Preschool Embedded Figures Test was related to time spent in male sex-typed activities, which included spatial activities (e.g., blocks, Lincoln Logs) as well as a variety of other play activities (e.g., vehicles, balls, magnetic letters and numbers, a magnifying glass, and an abacus). Play activities were not related to performance on the Block Design task for either boys or girls. In another study, Serbin and Connor (1979) reported that boys who scored above the median of their sex on masculine toy play and below the median on feminine toy play scored higher on the WPPSI Block Design subtest than on the Vocabulary subtest, whereas girls who scored above the median of their sex on feminine toy play and below the median on masculine toy play showed the opposite pattern of scores. Neither of these studies provides information about which of the many activities that composed the masculine–feminine toy play categories specifically relates to performance on the cognitive tasks. Another study by Caldera et al. (1999) showed that the complexity of preschoolers' free play with blocks relates to their performance on the WPPSI Block Design subtest (Wechsler, 1967) and that their structured block play, which involved measuring their accuracy in copying a complex three-dimensional (3-D) block structure, relates to their performance on the WPPSI Block Design subtest and their performance on the block portion of the Copying Blocks subscale of the Stanford–Binet Intelligence Scale (4th ed., Thorndike, Hagen, & Sattler, 1986). Other observational studies of children's play at home and during free play at school have revealed that boys are more likely to engage in block play and to play with other toys that involve two-dimensional (2-D) and 3-D spatial transformations (e.g., Legos; Farrell, 1957; Saracho, 1994, 1995; Tracy, 1987). Although it seems likely that this kind of play could relate specifically to the development of spatial skills, such as mental rotation, these studies did not assess children's spatial skill.

Training studies that involve engaging children in particular spatial activities provide a powerful way to test whether particular kinds of spatial activities lead to improvements in children's spatial skill. One study (Casey, Erkut, Ceder, & Young, 2008) found that kindergarteners whose training involved teacher-guided block play showed improved performance on the Block Design subtest of the Wechsler Intelligence Scale for Children, Fourth Edition, but not on a 3-D mental rotation test, possibly because of the difficulty of 3-D mental rotation for young children (Voyer, Voyer, & Bryden, 1995). In another study, Casey, Andrews, et al. (2008) found that presenting children with part–whole geometric puzzles in a story context led to more improvement in the part–whole spatial thinking of kindergarten children than presenting these puzzles alone. Further, across conditions, girls improved more than boys. Results of other studies have indicated that experience playing action video games improves children's mental rotation skills (e.g., McClurg & Chaille, 1987; Terlecki & New-

combe, 2005), particularly for those starting with lower spatial skill levels (Subrahmanyam & Greenfield, 1994).

The current study takes a somewhat different approach to examining the relation of spatial activities and spatial skill, an approach more akin to that taken in the language development literature. In particular, we directly observed children's naturally occurring engagement in puzzle play in the context of a longitudinal study during six home visits that occurred when children were between 26 and 46 months of age. We then related individual variation in the frequency and quality of children's play to their spatial skill at 54 months of age, as assessed by a mental transformation task (Levine et al., 1999). Because we videotaped the parent–child interactions, we were able to code and examine various aspects of puzzle play including time devoted to puzzle play, frequency of puzzle play, difficulty of puzzles, level of engagement of child and parent during puzzle play, and use of spatial language by child and parent during puzzle play.

These data allowed us to address a number of questions. First, we examined whether puzzle play frequency or quality varied with the child's age, sex, or socioeconomic status (SES) background across the 26- to 46-month age range that we examined. Second, we examined whether children who engaged in puzzle play had higher scores on a spatial transformation task at 54 months, the youngest age at which significant individual and sex differences have been found on this task (Levine et al., 1999), than did children who did not engage in puzzle play. Further, among those children who engaged in puzzle play, we examined whether the frequency of their puzzle play and/or the quality of their puzzle play predicted their spatial transformation skill at 54 months of age. Moreover, because we had various other kinds of information about the families participating in our study, including demographic information and measures of overall parent language input to children, we were able to ask whether puzzle play frequency and/or quality were related to performance on the spatial transformation task controlling for these variables.

Given that we were examining spontaneous, naturalistic puzzle play, all of our findings are correlational. Thus, the finding of a relation between puzzle play and spatial skill may indicate that puzzle play positively impacts spatial skills, that children who have higher spatial skills are more likely to engage in spatial activities, such as puzzle play, or that both spatial skill and engagement in spatial activities could be related to some other factor or factors (e.g., parental encouragement). Although our naturalistic study is agnostic with respect to causality, it can provide information that helps narrow the hypothesis space for the design of experiments aimed at testing whether particular kinds of puzzle play experiences impact the development of children's spatial thinking.

Method

Participants

The 53 parent–child dyads that contributed to this study (27 boys, 26 girls) were drawn from a larger sample of 64 families in a longitudinal study of language development. Children were videotaped at home with their primary caregiver, who for 47 of the 53 children was the mother (for five children in our sample, the mother and father reported that they were dual primary caregivers,

and for one child the father was the primary caregiver). Occasionally, both the mother and father were present during a visit. Additionally, there were sometimes siblings or other adults present. However, during all but one instance of puzzle play that we observed, it was the mother who engaged in the puzzle play with the child.

Recruitment for the longitudinal study was conducted through an advertisement in a free parent magazine and a mailing to approximately 5,000 families living in the Chicago metropolitan area and surrounding suburbs. Parents who responded completed a screening interview over the phone in which they were asked about their income, education, occupation, race, ethnicity, language or languages spoken in the home, and their child's gender. Sixty-four English-speaking families were selected to participate in the longitudinal study so that the sample represented the diversity of the greater Chicago population as measured through the 2000 census data on family income and ethnicity. As is typical for children in this age range, the majority of primary caregivers (36 of 53, 68%) worked full- or part-time when their child was 26 to 46 months of age. Two of those who worked part-time were also students, and one other who did not work was a full-time student. The remaining primary caregivers (16/53, 30%) were full-time homemakers. Observation sessions were scheduled at the convenience of the families.

Children were included in the present study if they had data for each of six visits between 26 and 46 months and if they completed the spatial transformation task at 54 months (see task description later). Of the 64 families included in the larger sample, five families were excluded because they did not complete all six observation sessions. Six additional families were excluded because they did not complete the mental transformation task either because they refused ($n = 3$) or because they were not available to do the 54-month visit ($n = 3$). The 53 remaining families represented six income levels (less than $15,000; $15,000–$34,999; $35,000–$49,999; $50,000–$74,999; $75,000–$99,999; $100,000 or more) and four educational levels (high school; some college; bachelor's degree; graduate degree) as self-reported in a demographic questionnaire. The excluded families did not distort the sample as they were evenly distributed across income and education groups. Given the diversity of our sample, we expected that the findings would generalize to broad segments of the U.S. population.

Materials

Children completed a shortened, 10-item version of a spatial transformation task developed by Levine et al. (1999) because of time constraints during the visits. On this task, children are first shown an array of four shapes and two target pieces (Figure 1). On each item, children received the following instruction: "Look at the pieces. Look at the shapes. If you put the pieces together they can make one of the shapes. Point to the shape the pieces make." Half of the target shapes (five items) were unilaterally symmetrical around the vertical axis and the other half were bilaterally symmetrical around the horizontal and vertical axes. The test items comprised four kinds of transformations, all in the 2-D plane of the page: direct translation in which the pieces are horizontally displaced, diagonal translation in which the pieces are diagonally displaced, direct rotation in which the pieces are horizontally
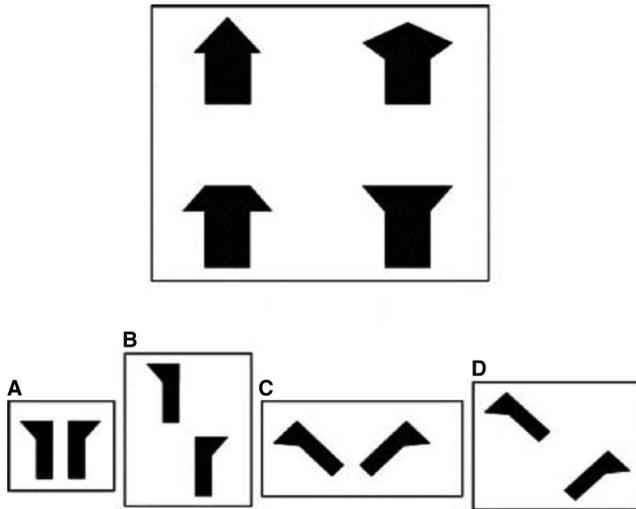
F1

Figure 1. Example item from the spatial transformation task (Levine et al., 1999). Shown is the choice array (top) and four possible configurations of the target pieces: (a) direct translation, (b) diagonal translation, (c) direct rotation, and (d) diagonal translation. Note that only one configuration is displayed for each item.

displaced and rotated by 45 degrees, and diagonal rotation in which the pieces are diagonally displaced and rotated by 45 degrees. Four of the items involved rotational transformations, and the other six involved translational transformations. Prior studies have shown this task to be a reliable measure of spatial transformation skill and have found that boys outperform girls on both rotation and translation items, with no significant interaction of gender and problem type (Ehrlich, Levine, & Goldin-Meadow, 2006; Levine et al., 1999). A split-half reliability test of odd and even items on this abbreviated 10-item task showed that it maintains reliability, $r = .55$ adjusted using the Spearman–Brown formula.

Procedure

Parent–child dyads were visited in the home every 4 months beginning at child age 14 months. To examine the relation of early puzzle play to children's later spatial transformation skill, this study focused on six of the home visits: when the children were 26, 30, 34, 38, 42, and 46 months old (ages 2 years 2 months through 3 years 10 months). At each visit, dyads were videotaped for 90 min while engaging in their ordinary activities. Toy play, book reading, and meal or snack time were common activities during visits. If parents and children were engaging in separate activities (e.g., the child was playing with toys while the parents washed dishes), the camera focused on the child. The families were not given any materials to play with or direction about engaging in particular activities—rather, parents were asked to spend their time as they ordinarily would. When children were 54 months old (4 [1/2] years), they were administered the spatial transformation task.

Transcription of the observational data and basic linguistic analyses followed the procedures outlined by Huttenlocher, Vasilyeva, Waterfall, Vevea, and Hedges (2007). In short, the speech of both parents and children was transcribed by trained research assistants and parsed into utterances (a sequence of words delimited by pauses, a change in conversational turn, or a shift in intonation pattern). A second research assistant independently transcribed 20% of the videotapes. The reliability criterion was 95%. Our procedures for defining and coding spatial language are explained in more detail later.

Puzzle Play Coding System

Puzzle play was coded when the child played with puzzles that involved individual pieces that interlock to form an interconnected whole or played with puzzles that involved placing pieces attached to pegs in corresponding holes on a board, typically played with at younger ages. We included only puzzles that had one, and only one, correct location for a given puzzle piece (e.g., unique interconnecting pieces of a jigsaw puzzle or a corresponding hole on a peg-piece puzzle board, such as the cat in the cat hole, the dog in the dog hole, etc.). In addition, to be coded as a puzzle play episode, children had to actually attempt to put the puzzle together and not just play with the pieces in other ways.

Our transcripts included not only parent and child speech but also detailed notes. We therefore predicted that, if children and parents played with a puzzle, there would be a high likelihood that the word *puzzle* would be mentioned in the transcript notes or the transcript itself. Thus, we searched all transcripts in each of the six sessions for any mention of the word *puzzle* (either in parent speech, child speech, or transcript notes). To ensure that this was a valid system for identifying potential puzzle play, a random sample of 60 videotapes (10 from each of the six visits) that did not contain the word *puzzle* were viewed in their entirety. None of these videotapes contained any puzzle play, indicating that our system for identifying children's engagement in puzzle play was valid.

Measures of puzzle play frequency. For each child, we calculated three measures of the frequency of puzzle play: the total number of sessions during which the child played with at least one puzzle, the total number of puzzle play episodes over all sessions, and the total length of time (rounded to the closest half minute) for all puzzle play episodes. Each puzzle play episode was defined as the child playing with one or more puzzles for at least 30 s (episodes under this time limit consisted of the child deciding not to play with the puzzle), with all the play time counted until there was a 30-s break in the play.

Measures of puzzle play quality. We calculated three measures of the quality of puzzle play: puzzle difficulty, level of engagement during puzzle play, and spatial language during puzzle play. Mean puzzle difficulty for each child was calculated on the basis of the proportion of puzzles played with that fell into the following four categories: (a) peg-board puzzles, (b) jigsaw puzzles with nine or fewer pieces, (c) jigsaw puzzles with 10 to 24 pieces, (d) jigsaw puzzles with more than 24 pieces. For example, if a child played with four puzzles, two at the first level of difficulty, one at the second level of difficulty, and one at the fourth level of difficulty, his or her average difficulty scores would be

$$\text{difficulty score} = [2(1) + 1(2) + 1(4)]/4 = 2.00.$$

The possible difficulty scores ranged from 1 (all puzzle play scored as lowest difficulty) to 4 (all puzzle play scored as highest difficulty).

We also coded parent and child engagement levels during each puzzle play episode as follows: low engagement (scored as 0) was coded if a parent was not present during puzzle play or if the child or parent was present but not engaged in completing the puzzle and the other member of the dyad was working on the puzzle alone; medium engagement (scored as 1) was coded if the participant showed some interest in the puzzle but was easily distracted (e.g., child requires several prompts from parent to stay on task; child requests help from the parent to stay on task); high engagement (scored as 2) was coded if the participant (parent/child) was verbally and/or physically participating in trying to solve the puzzle. Mean engagement scores were calculated for the child and for the parent on the basis of the proportion of puzzles on which they showed high engagement, medium engagement, or low engagement. For example, if a child played with four puzzles across all the sessions and was engaged at a low level on one puzzle, at a medium level on one puzzle, and at a high level on two puzzles, he or she would receive an engagement score of 1.25, calculated as follows:

$$\text{engagement score} = [1(0) + 1(1) + 2(2)]/4 = 1.25.$$

The range of possible engagement scores was from 0 (all puzzle play scored as low engagement) to 2 (all puzzle play scored as high engagement).

Spatial Language Coding System

Our first step in coding parent and child spatial language from the transcripts was to create a comprehensive list of spatial terms parents would likely use when talking to their young children about spatial concepts. We next reviewed a subset of puzzle play episodes ($n = 18$, with three for each session) to refine our list. Only three words occurred in these data that were not in the original list (*hole, bump, border*); we added these words to the master list.

In addition, we imposed three criteria for the contextual and semantic use of the listed spatial terms during puzzle play. First, we only included usages of spatial terms that concerned the construction of puzzles but did not include any events that occurred during play that were unrelated to the puzzle (e.g., we included "this piece goes *under* the bird" but did not include "the bird flew *under* the chair"). Second, we did not include homonyms of the listed terms (e.g., we included "put this piece to the *left* of the cat" but did not include "the cat *left* his hat in the car"). Third, we only included terms that referred to specific spatial labels during puzzle play (e.g., "This piece goes at the *top/bottom/side*") rather than deictic or vague references to location or space (e.g., "This piece goes somewhere over *here/there*" or "Put it *in/on* the puzzle"). As previously reviewed, research suggests that using more specific spatial labels impacts children's understanding of spatial concepts more so than using general spatial terms (Loewenstein & Gentner, 2005). In other words, although there are many aspects of speech that could be considered spatial, some of this language is not particularly germane to the task of identifying and labeling specific spatial aspects of the puzzle array.

Thus, three categories of spatial language were coded:

1. *Dimensions, features,* and *shapes of objects* were defined as terms that describe the size, geometric features, and shape names of two- and three-dimensional objects. Examples of such words are *long, short, corner, straight, square*, and *triangle*.

2. *Orientation* and *transformations* were defined as words that describe the relative orientation or act of moving objects and people in space. Examples of such words are *upside-down, sideways, turn,* and *flip*.

3. *Location* and *direction* were defined as words that describe the spatial locations of puzzle pieces. Examples of such words are *top, under, between, right,* and *left*.

We transcribed and coded all parent and child utterances during each puzzle episode. We then calculated for both the parent and the child the average number of spatial words (*spatial tokens*) per puzzle episode to control for variation in frequency of puzzle play. To control for the effects of language input in general compared with any unique impact of spatial language in our analyses, we also coded the total number of nonspatial words (*cumulative word tokens*) and the total number of different nonspatial words (*cumulative word types*) produced by parents and children over all six observational sessions.

Coding Reliability

Reliability was conducted on approximately 25% of the data marked as potentially containing puzzle play (23 out of 89 transcripts). Intercoder agreement on the three measures of puzzle play frequency was as follows: number of sessions with puzzle play, 100%; number of total puzzles played with, 100%; length of time for each puzzle episode, 82.61% (Cohen's $\kappa = .65$, 95% CI [.34, .96]). The lower reliability for coding time for each puzzle episode compared with number of puzzle sessions and total number of puzzles played with reflects the greater difficulty of deciding the time when each puzzle episode starts and stops as well as timing of interruptions that occur during puzzle episodes. Intercoder agreement on puzzle difficulty was 100%, whereas judgments of level of engagement were more subjective and had lower levels of reliability: child engagement, 84.03% (Cohen's $\kappa = .70$, 95% CI [.43, .96]); parent engagement, 89.28% (Cohen's $\kappa = .79$, 95% CI [.58, .99]). Intercoder agreement on the spatial language coding was 96.4% for parent language (Cohen's $\kappa = .88$, 95% CI [.83, .92]) and 98.1% for child language (Cohen's $\kappa = .90$, 95% CI [.84, .96]). All disagreements were resolved through discussion.

Results

First, we examine children's performance on the spatial transformation task in relation to the child's gender, parent demographic characteristics (parent speech, self-reported education, and income), and the type of problems on the test (translation, rotation) for all 53 children in the study. Next, we compare performance on the spatial transformation task for children who did and did not engage in puzzle play. Finally, we focus only on those children who played with puzzles to examine the relation between various quantitative and qualitative aspects of puzzle play and performance on the spatial transformation task. Our quantitative mea-

sures include the total number of observation sessions during which the child played with at least one puzzle, the total number of puzzles played with across sessions, and the total time spent playing with puzzles across sessions. Our qualitative measures include puzzle difficulty, level of parent engagement during puzzle play, and parent spatial language during puzzle play. We analyzed data using three ordinary least squares moderated regressions and, in the case of analyses involving two dichotomous variables, an analysis of covariance.

Child Gender and Parent Demographics in Relation to Spatial Transformation Performance

Consistent with prior research, child gender was significantly related to spatial transformation skill such that boys outperformed girls, $\beta = .34, p < .05, R^2 = .12, f^2 = .13$ ($M_{boys} = 0.58, SD_{boys} = 0.18; M_{girls} = 0.45, SD_{girls} = .19$). We also examined the relation of children's spatial transformation task performance to the socioeconomic variables of parent demographics (income and education) and the home language environment. Parent education and income were positively but not significantly related to spatial transformation skill. Additionally, the number of total words (cumulative word tokens) and the number of unique words (cumulative word types) spoken by parents across the six observation sessions were related to children's performance on the spatial transformation task (see Table 1). To examine whether the effects of child sex and parent demographics differed for different kinds of spatial transformation problems (translation vs. rotation items; see Figure 1), we conducted an analysis of variance with problem type (as a within-subject variable) by gender and parent education, while controlling for overall word types. There were no main effects of problem type and no interactions of problem type with child gender or parent education, all $ps > .72$. Thus, for all further analyses, we used the children's overall spatial transformation score as a measure of general spatial transformation (including rotation) skill.

Puzzle Play in Relation to Spatial Transformation Performance

Approximately half of the children in our sample (27 out of 53; 50.9%) played with a puzzle at least once during the six observation sessions that occurred between 26 and 46 months of age. There was no significant difference between the number of boys and girls who played with puzzles (14 boys, 13 girls) and those that never played with a puzzle (13 boys, 13 girls), $p = 1.0$, two-tailed

Fisher's exact test. However, the parents of children who played with puzzles had significantly higher incomes, cumulative word types, and cumulative word tokens compared with the parents of children who did not play with puzzles (see Tables 2 and 3). Given these differences, we controlled for these parent characteristics in our analysis of the relation between children's spatial transformation performance and puzzle play.

We conducted a two-way analysis of covariance to examine differences in spatial transformation performance, with child gender and puzzle play (children who did vs. did not play with puzzles) as predictor variables. Parent cumulative word types, income, and education were entered as covariates. Given that parent cumulative word types and tokens were highly correlated ($r = .59, p < .001$), we excluded cumulative tokens to reduce multicollinearity. We found significant main effects of puzzle play and child gender but no interaction effect. Thus, overall, boys performed better than girls on the spatial transformation task, $F(1, 46) = 5.70, p = .02, d = .70$, and children who played with puzzles had higher spatial scores than children who did not play with puzzles, $F(1, 46) = 6.24, p = .02, d = .89$ (see Figure 2).

Analyses of Puzzle Playing Children Only

For the remaining analyses, only the 27 children (51% of the sample) who played with puzzles (14 boys and 13 girls) were included to examine whether puzzle play frequency and puzzle play quality predicted children's spatial transformation skill above and beyond just playing with a puzzle. Parent cumulative word types and education were entered as covariates. We chose these covariates because they were the only parent characteristics significantly associated with the dependent measure in the subset of children who played with puzzles and that did not have any interactive effects with gender or any other independent variable in the current set of analyses.

Puzzle play frequency and relation to spatial transformation performance. We coded three measures of puzzle play frequency: (a) total number of sessions where puzzle play occurred, $M = 3.81, SD = 2.94$; (b) total number of puzzle play episodes over all sessions, $M = 4.22, SD = 3.75$; and (c) total amount of time spent playing with puzzles over all sessions (in minutes), $M = 20.64, SD = 18.69$. To provide an understanding of the time children spent playing with puzzles (among those who engaged in this activity), we used the observed time children spent playing with puzzles to estimate the number of minutes that children would spend playing with puzzles per week. On the basis of the average of 20.64 min of puzzle play over the total of 540 min of observation across the six observation sessions, the mean number of minutes of puzzle play per week would be 128.42 min or slightly over 2 hr per week, assuming 8 waking hours per day and 56 waking hr (3,360 waking min) per week.

To contextualize the amount of time spent on puzzle play, we also did some comparisons to another parent–child activity widely considered to have effects on later language and literacy outcomes—book reading. We have book-reading data for this same group of 53 children at the 30-month session, which was one of our time points. We compared the number of minutes spent on puzzle play at 30 months to the number of minutes spent on book reading. The average number of minutes spent on book reading was only 1 min 33 s among parent–child dyads who read books, whereas the

Table 1

Regression Models Predicting Mental Transformation Score Based on Parent Characteristics

Predictor	Model 1	Model 2	Model 3	Model 4
Parent cumulative word tokens	.41**	.25**	.23*	.23*
Parent cumulative word types		.39**	.25*	.27*
Parent education			.26†	.18
Parent income				.12
Cumulative R^2 statistic	.17	.32	.39	.40

† $p < .10$. * $p < .05$. ** $p < .01$.

Table 2
Distribution of Parent Income and Education for Children Who Did and Did Not Play With Puzzles (N = 53)

Family income	Primary caregiver's education				
	High school or GED	Some college or trade school	Bachelor's degree	Advanced degree	Total
	Puzzle play				
Less than $15,000	1	1	0	0	2
$15,000–$34,999	0	0	1	0	1
$35,000–$49,999	0	0	0	1	1
$50,000–$74,999	0	1	3	3	7
$75,000–$99,999	1	0	5	2	8
$100,000 or more	0	1	2	5	8
Total	2	3	11	11	27
	No puzzle play				
Less than $15,000	1	0	0	1	2
$15,000–$34,999	3	4	0	0	7
$35,000–$49,999	1	2	0	1	4
$50,000–$74,999	0	0	4	0	4
$75,000–$99,999	1	0	3	2	6
$100,000 or more	0	0	0	3	3
Total	6	6	7	7	26

Note. GED = general equivalency diploma.

average spent on puzzle play was 8 min 33 s among those who played with puzzles. However, 23 parent–child dyads engaged in book reading at 30 months, whereas only 10 parent–child dyads engaged in puzzle play at 30 months. Thus, fewer parent–child dyads engage in puzzle playing than book reading at 30 months of age, but dyads that engage in these activities spend more time on puzzle play than on book reading at this age. On the basis of this comparison, it seems feasible that the amount of puzzle play children engage in might be sufficient to have an impact on the development of spatial skill just as book reading is thought to have an impact on later language and literacy skills (e.g., Bus, van IJzendoorn, & Pellegrini, 1995; Demir, Applebaum, Levine, Petty, & Goldin-Meadow, 2011; Sénéchal & Lefevre, 2001). However, an experimental study is required to determine whether this amount of puzzle play would lead to improvements in children's spatial thinking.

A principal components analysis revealed that all three of our measures of puzzle play frequency were significantly correlated and loaded onto one factor measuring the same latent variable (Eigenvalue 2.50; correlations: number of sessions and number of puzzle play episodes: $r = .78$, $p < .001$; number of sessions and

total time playing puzzles: $r = .74$, $p < .001$; number of puzzle play episodes and time playing puzzles: $r = .74$, $p < .001$). In all subsequent analysis, we used the composite score computed from the principal components analysis as a measure of puzzle play frequency for each child.

Puzzle play frequency did not significantly differ for the earlier (26, 30, and 34 months) versus the later observation sessions (38, 42, and 46 months; $M_{early} = 0.22$, $SD = 1.19$; $M_{late} = 0.54$, $SD = 0.95$). In addition, puzzle play frequency did not differ for boys versus girls, $t(25) = 1.85$, $p = .41$ ($M_{boys} = 0.22$, $SD_{boys} = 0.30$; $M_{girls} = 0.23$, $SD_{girls} = 0.22$, respectively). A regression analysis indicated that mental transformation skill was predicted by the frequency of puzzle play, $\beta = .41$, $p < .05$, $R^2 = .43$, $\Delta R^2 = .33$, $f^2 = .76$, and there was no interaction of puzzle play frequency and child gender. Thus, among the children who played with puzzles, higher mental transformation scores were associated with more frequent puzzle play.

Puzzle play quality and relation to spatial transformation performance. We next examined whether the quality of puzzle play related to performance on the spatial transformation task and whether quality of play varied between boys and girls. Table 4 **T4**

Table 3
Parent Characteristics of Children Who Did and Did Not Play With Puzzles

Characteristic	Puzzle play (N = 27)		No puzzle play (N = 26)		Cohen's d
	M	SE	M	SE	
Parent word tokens**	37,452.11	2,279.89	24,731.00	2,781.61	1.00
Parent word types*	1,559.00	55.86	1,270.96	86.43	0.82
Parent education†	3.15	0.17	2.58	0.21	0.54
Parent income**	4.56	0.59	3.54	0.24	0.66

† $p < .10$. * $p < .05$. ** $p < .01$.

displays the qualitative measures of puzzle play we coded from the observed sessions: puzzle difficulty, level of child and parent engagement during play, and child and parent spatial language during puzzle play. First we present a description of these qualitative measures, and then we present analyses examining the relation of these measures to children's performance on the spatial transformation task.

AQ: 1   ***Quality of puzzle play.***   Puzzle difficulty was lower at the earlier three observation sessions (26, 30 and 34 months) than the later three sessions 38, 42, and 46 months) but did not significantly differ (M_{early} = 1.90, SD = 1.12; M_{late} = 2.73, SD = 1.00; $t(10)$ = 2.075, p = .065. Puzzle difficulty did not significantly differ for girls versus boys across the four categories of difficulty from the simplest peg-board puzzles to the most complex jigsaw puzzles, $t(25)$ = 1.59, p = .13.

The spatial language spoken by the parent and child during puzzle play was analyzed as previously described using spatial tokens per puzzle episode. Descriptive statistics for parent and child spatial language during puzzle play and overall word types used during the six sessions are provided in Table 5. Parent

T5   and child spatial tokens during puzzle play were correlated (r = .785, p < .001), with the most frequent spatial category being location and direction terms, followed closely by dimensional adjectives, spatial features, and shape terms, and then by orientation and transformation terms. Some representative examples of parent utterances that occurred in the context of puzzle play are provided in the Appendix.

We next examined whether the amount of spatial language that occurred during puzzle play differed for boys versus girls or for the parents of boys versus the parents of girls. Because of the low frequency with which spatial language is used and the small sample size of children who played with puzzles, we used a nonparametric test to examine whether there were gender-associated differences in parents' use of spatial language or in children's use of spatial language during puzzle play. We used a median split of spatial tokens spoken per puzzle play episode to divide children categorically into high or low spatial language input groups on the basis of parent spatial talk (Mdn = 6 words;

T6,AQ2   see Table 6). The distribution of boys and girls who heard higher amounts of parent spatial language during puzzle play was not significantly different than those who heard lower amounts of parent spatial talk, p = .057, two-tailed Fisher's exact test. We then did a parallel analysis, dividing children categorically into
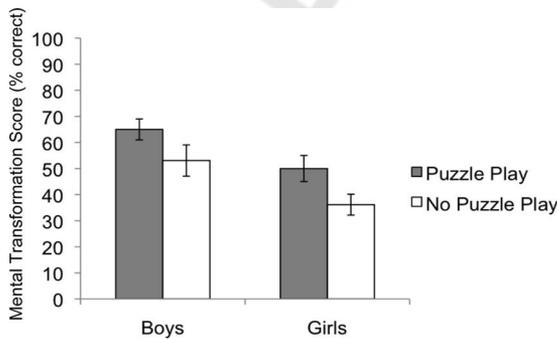
Table 4

Mean Proportion of Total Puzzles Played With (Standard Error) by Puzzle Difficulty, Levels of Child and Parent Engagement, and Spatial Language Spoken by the Parent and Child (Mean Spatial Tokens Per Puzzle) for Boys and Girls

Puzzle play quality	Boys (n = 14)		Girls (n = 13)	
	M	SE	M	SE
Puzzle difficulty				
Peg board	0.28	0.09	0.56	0.13
<9-piece jigsaw	0.23	0.09	0.12	0.07
10- to 24-piece jigsaw	0.28	0.10	0.22	0.11
24-piece jigsaw	0.21	0.08	0.10	0.08
Child engagement				
High	0.52	0.10	0.53	0.11
Medium	0.19	0.08	0.36	0.11
Low	0.29	0.10	0.11	0.05
Weighted average	1.23	0.18	1.41	0.14
Parent engagement				
High	0.40	0.10	0.19	0.08
Medium	0.40	0.12	0.50	0.11
Low	0.16	0.08	0.16	0.06
Not present	0.04	0.02	0.15	0.08
Weighted average	1.21	0.14	0.88	0.13
Child spatial language	4.38	1.45	1.88	0.52
Parent spatial language	22.73	6.65	10.35	3.53

high and low spatial language production groups on the basis of child spatial talk (Mdn = 2 words). The distributions of boys and girls who themselves said a higher versus a lower amount of spatial language did not significantly differ, p = .70, two-tailed Fisher's exact test (see Table 6).

Finally, the level of engagement during puzzle play did not significantly differ for girls and boys, $t(25)$ = 0.81, p = .42, or for the parents of boys compared with the parents of girls, $t(25)$ = 1.69, p = .11. Overall engagement was not significantly correlated between parent and child, r = .21, p = .30.

Quality of puzzle play.   A principal components analysis revealed that our three measures of quality of puzzle play input (puzzle difficulty; parent engagement in puzzle play; number of parent spatial tokens per puzzle episode) were significantly correlated and loaded onto one factor measuring the same latent variable (Eigenvalue 2.02; correlation of puzzle difficulty and parent engagement, r = .44, p < .05; puzzle difficulty and parent spatial language, r = .68, p < .001; parent engagement and parent spatial language, r = .38, p < .05). In all further analysis, we used the composite score computed from the principal components analysis as a measure of puzzle play quality for each child.[1]   **Fn1**

Puzzle play quality was significantly higher for boys than girls, $t(25)$ = 2.08, p = .05, d = 1.05 (M_{boys} = 0.36, SD_{boys} = 0.91; M_{girls} = −0.39, SD_{girls} = 0.97), even though none of the individ-

Figure 2. Mean spatial transformation task performance for girls and
AQ: 5   boys who did or did not engage in puzzle play. Error bars represent ???..

---

[1] We calculated a composite score of puzzle play quality separately for parent and child. However, puzzle difficulty was a constant variable included in both composite scores, and spatial language during puzzle play was significantly correlated between parent and child. Thus, we used only the parent composite score to examine the relation between quality of puzzle play and spatial transformation performance, because the child quality score would have been redundant.

Table 5
Descriptive Statistics for Parent and Child Language (Mean Tokens Per Puzzle Episode)

Feature of speech	Boys (n = 14)					Girls (n = 13)				
	M	SE	Mdn	Min	Max	M	SE	Mdn	Min	Max
Parent										
Overall parent speech (cumulative word types)	1,549.00	66.57	1,530.00	1,241.00	2,145.00	1,568.00	94.10	1,497.00	993.00	2,246.00
Mean spatial tokens per puzzle	22.73	6.65	18.96	1.71	95.00	10.35	3.53	3.50	0.00	34.50
Location and direction (e.g., "We want to put it next to there, right?")	10.56	3.25	6.63	0.00	43.00	5.46	1.82	2.75	0.00	19.50
Dimensions, features, and shape (e.g., "Here's all my straight pieces.")	10.24	3.59	5.50	0.00	50.00	4.10	1.68	0.67	0.00	17.33
Orientation and transformation (e.g., "You just have to turn it a little.")	1.93	0.49	1.33	0.00	5.25	0.789	0.28	0.67	0.00	3.50
Child										
Overall child speech (cumulative word types)	818.19	37.68	783.00	456.00	1,229.00	790.48	36.97	770.00	456.00	1,117
Mean spatial tokens per puzzle	4.38	1.45	2.75	0.00	18.00	1.62	0.52	1.00	0.00	5.75
Location and direction (e.g., "We want to put it next to there, right?")	2.45	1.03	1.00	0.00	13.00	0.96	0.31	0.75	0.00	3.50
Dimensions, features, and shape (e.g., "Here's all my straight pieces.")	1.57	0.44	1.29	0.00	5.00	0.50	0.24	0.00	0.00	2.89
Orientation and transformation (e.g., "You just have to turn it a little.")	0.36	1.10	0.20	0.00	1.08	0.15	0.10	0.00	0.00	1.25

Note. Min = minimum; Max = maximum.

ual measures that compose this score were individually significantly different for boys versus girls. In the Appendix, we provide four examples of parent–child dyads playing with puzzles. Two dyads interacted in the context of relatively difficult jigsaw puzzles (Examples 1 and 2), and two interacted in the context of easier puzzles, first a pegboard puzzle (Example 3) and then a jigsaw puzzle with only six pieces (Example 4). These examples give a sense of how parent spatial language varied as a function of puzzle difficulty. For harder jigsaw puzzles, parents tended to provide a lot of spatial language. It appears that the need to construct a frame is associated with greater use of spatial language and that harder jigsaw puzzles may lead to greater amounts of scaffolding through the use of spatial language. For pegboard puzzles (Example 3), parents tended to use object labels, and for easy jigsaw puzzles (Example 4), the language parents provided tended to include a mixture of object labels and spatial terms. These examples show that when parents and children played with simpler puzzles, the interactions tended to include less parent spatial language, a find-

ing that is reflected by the results of our principle components analysis on aspects of puzzle play quality, described earlier.

Relation of puzzle play quality to child performance on the spatial transformation task. We next examined the relation of puzzle play quality (composite score from factor analysis) to children's performance on the spatial transformation task by carrying out a regression analysis with education and cumulative parent word types across all observation sessions entered as covariates. Additionally, because of the main effect of puzzle frequency, we included puzzle play frequency as a covariate. For puzzle playing children, there was a significant interaction effect (but no main effects) of child gender and puzzle play quality on spatial transformation skill, $\beta = 1.55$, $p < .01$, $R^2 = .62$, $\Delta R^2 = .18$, $f^2 = 1.60$. However, there was no interaction between puzzle play quality and puzzle play frequency, $\beta = 0.22$, $p = .45$, $r = .15$. Thus, higher scores on the spatial transformation task were associated with more frequent puzzle play for both girls and boys but with higher puzzle play quality only for girls (Figure 3). **F3**

Table 6
Distribution of Boys and Girls Into High and Low Spatial Language Groups Based on a Median Split of Spatial Tokens Per Puzzle Episode Spoken by Parents (Mdn = 6 Words) and Children (Mdn = 2 Words)

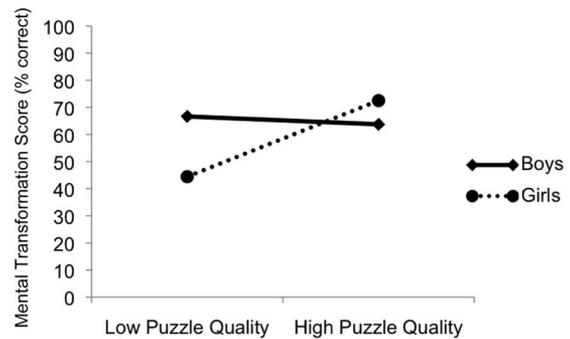|	Child spatial speech		Parent spatial speech	
	Low	High	Low	High
No. of girls	7	6	9	4
No. of boys	6	8	4	10

Figure 3. Mean spatial transformation task performance for girls and boys with high- or low-quality puzzle play.

Discussion

Our study examined children's puzzle play over six observation sessions that took place when the children were between 26 and 46 months of age. We also examined children's performance on a spatial transformation task at 54 months of age and the relation of puzzle play frequency and quality to performance on the spatial transformation task. Consistent with previous findings, boys performed significantly better than girls on the spatial transformation task (Levine et al., 1999). Additionally, children from higher SES groups performed better on the spatial transformation task than children from lower SES groups (as indexed by parent education; Levine et al., 2005).

Our findings regarding children's puzzle play are more novel. Our coding of naturalistic parent–child interactions showed that children varied in whether they engaged in puzzle play, with about half of our sample (27/53 children) playing with a puzzle during at least one observation session. The parents of children who played with puzzles, on average, had more education and income than the parents of children who did not play with puzzles. Moreover, children who played with puzzles also received more overall language input from their parents than the children who did not play with puzzles. Of particular interest, we found that children who played with puzzles during our observation sessions performed better on the spatial transformation task than those who did not, and this was true for both boys and girls. In addition, puzzle play predicted performance on the spatial transformation task even when we controlled for parents' SES and overall amount of parent speech to children. Additionally, among those children who played with puzzles, frequency of puzzle play was significantly related to their spatial transformation scores.

Consistent with prior findings that puzzle play is not sex stereotyped (Baenninger & Newcombe, 1989, 1995), we found that the frequency of puzzle play did not differ for boys and girls. The absence of a gender difference in frequency of puzzle play contrasts with other spatially relevant activities, such as block play and video game play, which are engaged in more by boys (Connor & Serbin, 1977; De Lisi & Wolford, 2002; Terlecki & Newcombe, 2005). However, we did find that the quality of puzzle play, as assessed by our composite measure, was higher for boys than girls between 26 and 46 months of age. There are, of course, many possible explanations for this difference, ranging from those that rest on the child's own interest and/or ability to complete puzzles to differences in parents' beliefs about the interests and abilities of boys and girls to complete puzzles, perhaps because of stereotypes about sex differences in spatial skill.

Although the present findings cannot explain the reason for the gender-related difference in the quality of puzzle play, the existence of this difference raises the possibility that girls' spatial thinking might be improved through play with more challenging puzzles. An experimental study that systematically varies the difficulty of puzzles presented to girls and boys could test the hypothesis that spatial skills improve more when children play with more difficult puzzles. Such a study also could provide important information about the optimal level of puzzle difficulty for children of particular ages and skill levels.

Our examination of naturally occurring parent–child puzzle interactions revealed that parents provide their children with more spatial language and are more engaged in puzzle play when their children try to put together more difficult puzzles. By experimentally manipulating the challenge children face in completing puzzles, we can assess whether it is the variation in puzzle difficulty that leads to these differences in parent input and to differences in the child's opportunity for spatial learning.

Unexpectedly, our analyses also showed that the quality of puzzle play was positively related to girls' but not boys' performance on the spatial transformation task. It is possible that this interaction is spurious and will not generalize to larger samples. It is also possible that puzzle play quality relates to both boys' and girls' performance on spatial tasks, such as our spatial transformation task, but that the boys in our sample received such high-quality puzzle play that we were not able to observe a relation between quality of play and their spatial transformation scores. Alternatively, it is possible that particular aspects of high-quality puzzle interactions are more helpful to the spatial thinking of girls than boys. Girls have been reported to depend more on verbal strategies than boys when performing spatial tasks, such as mental rotation (e.g., Ehrlich et al., 2006; Folk & Luce, 1987; Heil & Jansen-Osmann, 2008; Kail, Carter, & Pellegrino, 1979; Pezaris & Casey, 1991; Ratliff, Levine, & Saunders, 2011), raising the possibility that the parent spatial language input that occurs during high-quality puzzle interactions may be particularly helpful in supporting their spatial thinking. Given the unexpected nature of this interaction, it warrants further study.

In summary, our study shows that the frequency and quality of early puzzle play varies across children and that engagement in puzzle play is associated with demographic variables as well as high levels of parent language input. It also shows that puzzle play predicts children's later performance on a spatial transformation task, even controlling for demographic factors and parent language input. However, it remains an open question whether puzzle play promotes children's spatial transformation skill. Experimental studies that manipulate the quantity and quality of children's puzzle play experiences, including the spatial language children hear during puzzle play, are needed to examine whether puzzle play is causally related to the development of young children's spatial thinking. If this turns out to be the case, then engaging children in puzzle play would be a relatively easy and inexpensive way to support the development of an aspect of cognition that has been implicated in success in the science, technology, engineering, and mathematics.

References

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

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical skill. *Learning and Individual Differences, 7,* 363–379. doi:10.1016/1041-6080(95)90007-1

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning skill at age 13: Their status 20 years later. *Psychological Science, 11,* 474–480. doi:10.1111/1467-9280.00291

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65,* 1–21.

Caldera, Y. M., Culp, A. M., O'Brien, M., Truglio, R. T., Alvarez, M., & Huston, A. C. (1999). Children's play preferences, construction play with blocks, and visual–spatial skills: Are they related? *International Journal of Behavioral Development, 23,* 855–872. doi:10.1080/016502599383577

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

Casey, M. B., & Brabeck, M. M. (1990). Women who excel on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition, 12,* 73–84. doi:10.1016/0278-2626(90)90005-9

Casey, M. B., Erkut, S., Ceder, I., & Young, J. M. (2008). Use of storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29,* 29–48. doi:10.1016/j.appdev.2007.10.005

Casey, M. B., Nuttall, R., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33,* 669–680. doi:10.1037/0012-1649.33.4.669

Connor, J. M., & Serbin, L. A. (1977). Behaviorally based masculine- and feminine-activity-preference scales for preschoolers: Correlates with other classroom behaviors and cognitive tests. *Child Development, 48,* 1411–1416. doi:10.2307/1128500

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

DeLoache, J. S., & DeMendoza, O. A. P. (1987). Joint picturebook reading of mothers and one-year-old children. *British Journal of Developmental Psychology, 5,* 111–123. doi:10.1111/j.2044-835X.1987.tb01047.x

Demir, Ö. E., Applebaum, L., Levine, S. C., Petty, K., & Goldin-Meadow, S. (2011). The story behind parent–child book-reading interactions: Relations to later language and reading outcomes. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Boston University Conference on Language Development* (Vol. 1, pp. 157–169). Somerville, MA: Cascadilla Press.

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

Farrell, M. (1957). Gender differences in block play in early childhood education. *Journal of Educational Research, 51,* 279–284.

Folk, M. D., & Luce, R. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 395–404. doi:10.1037/0096-1523.13.3.395

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8,* 1–51.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children.* Baltimore, MD: Brooks.

Hedges, L. V., & Chung, V. (2011). *Does spatial ability predict STEM college major and employment: An examination of two longitudinal studies.* Manuscript in preparation.

Heil, M., & Jansen-Osmann, P. (2008). Sex difference in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology, 61,* 683–689. doi:10.1080/17470210701822967

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1368–1378. doi:10.1111/1467-8624.00612

Huston, A. (1983). Sex typing. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, 3rd ed., pp. 387–467). New York, NY: Wiley.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248. doi:10.1037/0012-1649.27.2.236

Huttenlocher, J., Levine, S. C., & Vevea, J. (1998). Environmental effects on cognitive growth. *Child Development, 69,* 1012–1029. doi:10.2307/1132360

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. C. (2002). Language input at home and at school: Relation to child syntax. *Cognitive Psychology, 45,* 337–374. doi:10.1016/S0010-0285(02)00500-5

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology, 43,* 1062–1083. doi:10.1037/0012-1649.43.5.1062

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and d. *American Psychologist, 36,* 892–901. doi:10.1037/0003-066X.36.8.892

Kail, R., Carter, P., & Pellegrino, J. (1979). The locus of sex differences in spatial skill. *Perception & Psychophysics, 26,* 182–186. doi:10.3758/BF03199867

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

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16,* 841–845. doi:10.1111/j.1467-9280.2005.01623.x

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498. doi:10.2307/1130467

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

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

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76,* 763–782.

Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental effects. *Brain and Cognition, 17,* 1–22. doi:10.1016/0278-2626(91)90062-D

Pruden, S., Levine, S. C., & Huttenlocher, J. (in press). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science.*

Ratliff, K. R., Levine, S. C., & Saunders, J. (2011). *A cognitive basis of individual differences in children's spatial problem-solving strategies.* Manuscript submitted for publication.

Saracho, O. N. (1994). The relationship of preschool children's cognitive style to their play preferences. *Early Child Development and Care, 97,* 21–33. doi:10.1080/0300443940970103

Saracho, O. N. (1995). Preschooler's cognitive style and their selection of academic areas in their play. *Early Child Development and Care, 112,* 27–42. doi:10.1080/0300443951120103

Sénéchal, M., & Lefevre, J.-A. (2001). Storybook reading and parent teaching: Links to language and literacy development. In J. Brooks-Gunn & P. Rebello (Eds.), *Sourcebook on emergent literacy* (pp. 39–52). San Francisco, CA: Jossey-Bass.

Serbin, L. A., & Connor, J. M. (1979). Sex-typing of children's play preferences and patterns of cognitive performance. *Journal of Genetic Psychology, 134,* 315–316.

Serbin, L. A., Zelkowitz, P., Doyle, A., Gold, D., & Wheaton, B. (1990). The socialization of sex-differentiated skills and academic performance:

A meditational model. *Sex Roles, 23,* 613–628. doi:10.1007/BF00289251

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

Signorella, M., Jamison, W., & Krupa, M. (1989). Predicting spatial performance from gender stereotyping in activity preferences and in self-report. *Experimental Psychology, 25,* 89–95.

Silvern, L. E. (1978). Masculinity–femininity in children's self concepts: The relationship to teachers' judgments, social adjustment and academic ability, classroom behavior and popularity. *Sex Roles, 4,* 929–949. doi:10.1007/BF00287712

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

Szechter, L. E., & Liben, L. S. (2004). Parental guidance in preschoolers' understanding of spatial–graphic representations. *Child Development, 75,* 869–885. doi:10.1111/j.1467-8624.2004.00711.x

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53,* 433–441. doi:10.1007/s11199-005-6765-0

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,* 996–1013. doi:10.1002/acp.1420

Thorndike, R. L. Hagen, E. P., & Sattler, M. (1986). *Stanford–Binet Intelligence Scale Fourth Edition: Technical manual.* Chicago, IL: Riverside.

Tracy, D. M. (1987). Toys, spatial ability, and science and mathematics achievement: Are they related? *Sex Roles, 17,* 115–138. doi:10.1007/BF00287620

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,* 250–270. doi:10.1037/0033-2909.117.2.250

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101,* 817–835. doi:10.1037/a0016127

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI).* New York, NY: Psychological Corporation/Harcourt, Brace, Jovanovich.

Wells, G. (1987). Apprenticeship in literacy. *Interchange, 18,* 109–123. doi:10.1007/BF01807064

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68,* 77–94. doi:10.1016/S0010-0277(98)00032-8

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30,* 679–689. doi:10.1037/0012-1649.30.5.679

Williams, A. D. (2004). *The jigsaw puzzle: Piece together a history.* New York, NY: Penguin Group.

Witkin, H. S., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the Embedded Figures Tests.* Palo Alto, CA: Consulting Psychologists Press.

AQ: 3

(*Appendix follows*)

Appendix

Examples of Parent–Child Puzzle Play Interactions

Example 1: Mother and 2 [1/2]-Year-Old Son Interacting Over a 30-Piece Jigsaw Puzzle

This mother provided rich spatial language to her child while he worked on a puzzle from our most difficult puzzle category. Here we focus on the language the mother used to scaffold the process of building the frame of the puzzle, which was entirely put together by the little boy. At the beginning of the puzzle play session, the mother mainly focused on encouraging the child to start with the edges and corners of the puzzle (e.g., she says, "You have to find the edge pieces first"). At one point, when the little boy was trying to put an edge piece in the middle of the puzzle, the mother says, "The straight part goes at the bottom. See how these are all straight? The straight part goes at the bottom." Later she reiterates, "The straight part goes with the other straight parts. . . . Does that have a straight side on it?" and even outright tells the child, "You need one that has a straight edge; that's not it. Put that one down and look for one with a straight edge." Even cursory review of the videotape shows that without the mother's involvement, this boy would not have been able to put this puzzle together. With his mother's help, however, he was able to succeed and had a positive experience. Reflecting the positive nature of the experience, when he finished the frame of the puzzle, his mother said, "You did the whole outside of the puzzle, now we need to do the inside." At that point, he replied with great pride, "I did it!"

Example 2: Mother and 46-Month-Old Son Interacting Over a 10- to 24-Piece Puzzle

This dyad provides a similar example of rich parent spatial language in the context of a somewhat less difficult 20-piece jigsaw puzzle. Similar to the mother in the first example, this mother provided rich spatial language to her 46-month-old son to scaffold his efforts. For example, she said, "I think it might go over here though. Let's leave it here. It's a corner. See how it has two straight ends? It's a corner." Later on she said, "This is straight. Where do you think the straight piece goes?" Toward the end of the puzzle interaction she said, "Straight lines on the outside," clearly trying to teach her son about how to find the pieces he needed to construct the frame of a puzzle.

Example 3: Mother and 30-Month-Old Daughter Interacting Over a Pegboard Puzzle

Most of the verbal input the mother provided to her daughter during the completion of a 20-piece pegboard puzzle consisted of labeling objects rather than labeling spatial relations, shapes, or features. For example, the mother asked the child, "Where are we going to put the bird?" Then, after the child put the bird in its place, the mother asked, "Where are we going to put the kittycat?" Later, the mother asked, "Is someone going to ride the horse? You going to put him on the board?" Of note, the only pegboard puzzle interaction with a lot of spatial language involved a 30-month-old girl and her mother working on a puzzle that involved putting geometric shapes into the board. In this case, the mother repeatedly asked for the names of shapes and then typically provided the answer by stating, "It's a square," "Rectangle," or "Triangle," when the child was unable to answer.

Example 4: Mother and 34-Month-Old Son Interacting Over a Jigsaw Puzzle With Fewer Than Nine Pieces

This mother provided her son with a mixture of object labels and spatial language to guide his construction of a simple six-piece jigsaw puzzle. She started out by stating, "Now should we find the seahorse?" Later she said, "I think we have to find the other piece with the letters on it." Still later she said, "But let's find the yellow piece with the letters." Toward the end of the interaction, the mother said, "You're close. I think it goes right—maybe in this corner. Let's turn it around. So it fits the shape like that." There was a great deal of talk about the necessity of turning pieces around (e.g., "You got to turn it around"; "Turn it that way", etc.). Still later, the mother said, "Put the flat side—See the flat side? Straight side. Put that on this side. If you want turn it around."

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

AQ1: Author: The heading "Quality of puzzle play" appears twice in a row. Should one of these headings be reworded or should one be deleted?

AQ2: Author: Please provide a heading for the left column of Table 6.

AQ3: Author: Please indicate where Wexler et al. (1998) should be cited in the text or delete it from the reference list.

AQ4: Author: Is Department of Psychology correct for all authors? If not, please provide correct departmental affiliations.

AQ5: Author: Please complete the note for Figure 2 defining the error bars in the figure.