Archives of Sexual Behavior
https://doi.org/10.1007/s10508-019-1415-8

ORIGINAL PAPER

Early Gender Differences in Spatial and Social Skills and Their Relations to Play and Parental Socialization in Children from Hong Kong

Wang Ivy Wong[1,2] · **Sui Ping Yeung**[1]

Received: 23 June 2017 / Revised: 31 January 2019 / Accepted: 2 February 2019

Abstract
Children's play preferences are highly gender-typed. At the same time, much research revolves around spatial and social skills that sometimes show male and female advantages, respectively. There is evidence that play with masculine toys is associated with better spatial skills and emerging evidence suggests that play with feminine toys is associated with better social skills. However, several research gaps limit current knowledge on these aspects of gender development. First, the study of childhood gender development has been largely Eurocentric; second, the link between gender-typed play and social skills development is not well supported. We tested 644 5-year-old Hong Kong Chinese children on five gender-typed skills, play preferences, and parental gender socialization. The pattern of gender differences was remarkably similar to those in the West. Boys preferred masculine toys more than girls and were better at mental transformation and targeting accuracy, while girls preferred feminine (and neutral) toys more than boys and were better at empathy and were less aggressive, although there was no significant gender difference in comforting skill. There was little evidence that these gender differences varied with socioeconomic status (parental income and education). Play correlated with some outcomes in expected ways. This is in contrast to parents' gender socialization, which showed some expected differences by child gender but minimally correlated with children's skills. These findings shed light on the generalizability of current knowledge on early gender differences and may facilitate gender developmental research outside the West. Although the study did not test the direction of effects, they substantiate the growing discourse on gender-typed play as an important learning mechanism.

Keywords Gender differences · Gender-typed play · Social and spatial abilities · Socialization · Socioeconomic status

Introduction

Gender-Typed Play

Gender differences can vary across context, age, and method. Although some researchers suggest that the majority of gender differences are small (Hyde, 2005), some differences are nevertheless more consistent and larger than others. Much effort has been spent to identify these differences, especially those that emerge early.

The differences in boys' and girls' play preferences have been demonstrated in a large number of studies using different measures (see review by Zosuls & Ruble, 2018). Such preferences show one of the largest behavioral gender differences (with Cohen's d sometimes > 2.00; Hines, 2010). Recent research has found gender differences in the interest for specific toys in infants approximately 5–6 months (Alexander, Wilcox, & Woods, 2009). Gender differences in toy preferences are apparent across a broader range of toys in preschool and across childhood (Etaugh & Liss, 1992; Todd et al., 2018). By age 2 years, boys and girls have distinctive gender-typed play preferences (Zosuls & Ruble, 2018).

Gender Differences in Spatial and Social Skills

Although perhaps less consistent and strong than the gender differences in play, the tendency for males to excel on some spatial skills and for females to excel on some social skills is also the basis for much research and debate. With regard to

✉ Wang Ivy Wong
iwwong@cuhk.edu.hk

[1] Department of Psychology, University of Hong Kong, Pokfulam, Hong Kong

[2] Present Address: Gender Studies Program, The Chinese University of Hong Kong, Sino Building, Shatin, N.T., Hong Kong

Published online: 23 May 2019

spatial skills, a meta-analysis found that mental rotation shows a male advantage ($d = .56$) that is larger than those in tasks traditionally categorized under spatial visualization (e.g., Block Design, i.e., a subtest of the Wechsler Intelligence Scale, and the Embedded Figures Test [EFT]) and spatial perception (e.g., the Water Level Task [WLT]) (Voyer, Voyer, & Bryden, 1995). While the male advantages in all three categories of spatial skills were larger with increasing age (Voyer et al., 1995), the male advantage in mental rotation or mental transformation appears to emerge very early, at least in some studies using age-appropriate measures (see null findings in Miller & Halpern, 2014). For example, some habituation studies found that 3- to 5-month-old males already outperform female infants in recognizing rotated images (Moore & Johnson, 2008; Quinn & Liben, 2008). Many studies found a moderate male advantage around age 4.5 years to 5.5 years using 2D measures (Ehrlich, Levine, & Goldin-Meadow, 2006; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Pruden, Levine, & Huttenlocher, 2011). Some found a smaller male advantage in preschoolers (Levine, Huttenlocher, Taylor, & Langrock, 1999; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). Gender differences in other spatial measures such as Block Design are either less frequently studied or do not show as consistent a gender difference in children (e.g., Casey et al., 2008; Pruden et al., 2011). Certain spatial skills that do not readily fit into the traditional typology of spatial skills may be avenues in which gender differences are also apparent. For example, targeting accuracy shows a male advantage ($d > 1$) that appears to be larger than that in mental rotations in adulthood (Hines et al., 2003). A meta-analysis of targeting accuracy in children also reported a large d of .96 that was invariant across ages 6–11 years (Thomas & French, 1985).

In contrast, some social skills tend to show female advantages. A meta-analysis found a small female advantage in overall prosocial behaviors such as helping, sharing, and comforting in young children and adolescents ($d = .20$), which tended to be larger in older samples (Fabes & Eisenberg, 1998). Empathy shows large gender differences, at least when measured by self-report scales, in both children and adults (Eisenberg & Lennon, 1983), perhaps partly because females are more willing to report being prosocial than males. A popular measure of children's empathy in earlier studies, the picture/story technique, showed only a small female advantage, but this measure has been criticized on methodological grounds (Eisenberg & Lennon, 1983). Recent studies employing parent reports of children found a moderate female advantage (Auyeung et al., 2009, $d = .56$ in 4–11-year-olds; Chapman et al., 2006, $d = .76$ in 6–8-year-olds; Escovar, Rosenberg-Lee, Uddin, & Menon, 2016, $d \sim .65$ in 7–12-year-olds; Owen-Anderson, Jenkins, Bradley, & Zucker, 2008, $d \sim .77$ in 4–8-year-olds). Girls also seem to be more capable of inhibiting physical aggression. A meta-analysis found that across cultures and at all ages sampled, males were more physically aggressive than females (Archer, 2004).

These gender differences are consequential. For example, since spatial skills are fundamental for achievement in STEM fields, gender differences in spatial skills may lead to gender disparity in STEM careers (Liben, Schroeder, Borriello, & Weisgram, 2018). On the other hand, social skills such as responsiveness to social cues, which tends to be higher in girls, have direct effects on academic achievement (Serbin, Zelkowitz, Doyle, Gold, & Wheaton, 1990). Moreover, in today's service-oriented labor market, social skills are increasingly vital for career success (Deming, 2015).

Socialization of Gender Differences in Spatial and Social Skills

Gender differences may be due partly to differential socialization. In the context of the socialization of skills, parents may interact with sons and daughters in different ways, potentially encouraging boys to develop better spatial skills and girls better social skills (for a review, see Ruble, Martin, & Berenbaum, 2006). However, the majority of studies relating parental socialization with child outcomes focused on parental attitudes. A meta-analysis found a small positive association between parents' gender attitudes and children's gender schemas (Tenenbaum & Leaper, 2002). In another study, parents' gender-typed attributes and attitudes correlated with children's stereotype knowledge and attitude, but minimally with children's actual behavior or interests (Turner & Gervai, 1995).

Compared to parental socialization, a socialization mechanism that perhaps more directly relates to gender differences in skills is gender-typed play. Toys favored by boys, such as blocks and Lego, often involve construction (Blakemore & Centers, 2005). These toys require children to understand spatial relations and rotate objects to different orientations (Casey et al., 2008), thus offering children spatial training.

Correlational research using parent- or self-reports found that exposure to masculine toys involving construction or spatial manipulation such as blocks and Lego associates positively with spatial skills assessed by Picture Rotation Task, WLT, EFT, or Block Design in preschool and elementary school children (Oostermeijer, Boonen & Jolles, 2014; Robert & Heroux, 2004; Serbin et al., 1990), sometimes regardless of gender and socioeconomic status (Jirout & Newcombe, 2015). Male and female college students' retrospectively reported childhood play with masculine toys also correlated with their current mental rotation performance (Doyle, Voyer, & Cherney, 2012). Observational studies in preschoolers found similar results. Observed frequency of masculine play positively correlated with performance on EFT or Block Design (Connor & Serbin, 1977; Serbin & Connor, 1979) and observed involvement in

and complexity of block play correlated with EFT and Block Design scores, respectively (Caldera et al., 1999).

We are aware of only one experimental study on the consequences of traditional childhood gender-typed toys which involved a control group and typically developing samples (Sprafkin, Serbin, Denier, & Connor, 1983; see Ben-Chaim, Lappan, & Houang, 1988 for a training study without a control group, Coxon, 2012 for a study on gifted children, and Casey et al., 2008 which focused on the instructional style of block training). In Sprafkin et al. (1983), the training group was taught and encouraged by teachers to play with masculine, construction toys over a six-week period, whereas the control group was not exposed to these toys. Boys and girls in the training group showed greater improvement in scores than the control group on geometric design and the Punched Holes test (a spatial visualization test).

The link between feminine toys and social skills is less clear. Research on play and social outcomes mostly focused on the characteristics of the play, sometimes of single toys that do not adequately reflect masculine or feminine toys in general (see review by Murnen, 2018). In studies of large arrays of toys, feminine toys are rated as more nurturing and domestically oriented and less aggressive (Blakemore & Centers, 2005). Observational studies suggest that feminine toys such as dolls elicit more nurturance and masculine toys such as toy weapons elicit more aggression (Caldera & Sciaraffa, 1998; Hellendoorn & Harinck, 1997; Watson & Peng, 1992). For example, in Caldera and Sciaraffa, children engaged in more caretaking behavior (e.g., feeding, combing hair, and dressing) and parents more caretaking and nurturing behaviors (e.g., hugging, kissing, and cuddling) when playing with a doll (feminine) than with a clown (less gender specific). It is therefore expected that play with feminine toys promotes social skill development.

Although the social impacts of gender-typed play have received attention (Murnen, 2018), we are aware of only one study that directly examined the correlation between gender-typed play and social skills. Li and Wong (2016) observed first graders' structured play with feminine, masculine, and gender-neutral toys. Comforting skill was measured by the number of strategies proposed to comfort a crying infant ostensibly described as the researcher's baby and shown on a baby monitor and empathy by the child version of the Empathy Quotient (EQ-C) (Auyeung et al., 2009). Girls' though not boys' play with feminine toys positively correlated with comforting skill, which showed an expected female advantage, but not empathy, which did not differ between boys and girls.

Research Gaps

Limitations of Studies on Gender-Typed Play

Although masculine and feminine play has each been hypothesized to provide different benefits, the majority of studies have focused on the spatial benefits of masculine toys and neglected the proposed benefits of feminine toys (e.g., nurturance and empathy). Another limitation is that many studies on the correlates or outcomes of play, especially those in children, did not employ outcome measures that were sensitive to gender differences. Some did not focus on gender and did not find gender differences in the spatial measures (e.g., Oostermeijer et al., 2014). Some focused on gender but found no gender differences in their tasks (e.g., Caldera et al., 1999; Robert & Heroux, 2004; Serbin & Connor, 1979; Sprafkin et al., 1983) or did not report whether there was any(e.g., Connor & Serbin, 1977). Using measures insensitive to gender differences may undermine the basis of hypotheses in studies that set out to test how gender differences in one domain relate to gender differences in another domain. It may also limit the power to find hypothesized relations between different domains of gender-typing (Wong & Hines, 2016).

Eurocentric Tradition of Studies on Gender Development

A more general limitation is that the knowledge on gender development has largely been derived from Western samples. For example, in a recent meta-analysis of gender-typed play, 13 of 16 studies focused on Western Caucasian samples (Todd et al., 2018). In the behavioral sciences, researchers have cautioned that current psychological knowledge is based largely on Western, especially American, samples and yet conclusions are stated as if they are readily generalizable to all humans (Heinrich, Heine, & Norenzayan, 2010). To the extent that gender development differs across cultures (e.g., Gibbons, 2000; Turner, Gervai, & Hinde, 1993), research on early gender-typing conducted outside the West would be valuable for theory testing and for facilitating research in these other regions. For example, gender development researchers often need to consider the existence, magnitude, and developmental pattern of gender differences in their sample. In studies of the correlates of sexual differentiation, researchers sometimes establish expected gender differences in the variables as a prerequisite for interpretation (e.g., Hines et al., 2003) and the lack of which may make it harder to detect predicted relations (Wong & Hines, 2016).

So, can we make the same predictions about gender differences and their correlates for Western and non-Western samples such as Chinese? Several large-scale studies of adolescents and adults showed that the direction of gender differences in some academic and spatial abilities, at least when significant, is remarkably consistent across cultures (e.g., Guiso, Monte, Sapienza, & Zingales, 2008; Lippa, Collaer, & Peters, 2010; Stoet & Geary, 2013). There are few systematic cross-cultural comparisons of gender development in childhood, but the available evidence suggests that variations mainly concern the magnitude and timing of gender differences or the specific content of stereotypes, while the broad pattern and direction of gender differences and gender roles are similar (for a review on gender

identity and gender stereotypes, see Gibbons, 2000). A meta-analysis on play found large gender differences in 16 studies regardless of geographical location and methodology, providing support for widespread gender differences in play (Todd et al., 2018). However, this conclusion is limited as most of the studies included relied on Western samples.

Nevertheless, Chinese children do show large gender differences in preferences for toys that are stereotypically masculine or feminine in the West (Li & Wong, 2016; Yu, Winter, & Xie, 2010). Also, gender differences in empathy (Geng, Xia, & Qin, 2012), comforting (Li & Wong, 2016), and mental transformation (Lippa et al., 2010) have been reported in Chinese samples. Like Western parents, Hong Kong (HK) Chinese parents wish their sons and daughters to play gender-stereotyped roles; for example, more parents wished their sons than daughters to attain tertiary education and sons were more often anticipated to be doctors while daughters were anticipated to be nurses (Lam, 1982); media in HK portrays gender in ways highly similar to those in the West, with women being depicted as stay-at-home housewife while men as capable leaders in the company (The Women's Foundation, 2015). Finally, in a HK sample, 6-year-old girls' preference for feminine toys correlated with their social skills (Li & Wong, 2016), suggesting that the correlates of gender-typing in Chinese are consistent with those hypothesized in the West.

Variability of Gender Differences Across Socioeconomic Status

Recent research suggested that gender differences may vary even within the same culture, with socioeconomic status (SES) being a possible moderator (Miller & Halpern, 2014). For example, in the U.S., the male advantage in two spatial skills (Levine et al., 2005) and in math (Reardon, Fahle, Kalogrides, Podolsky, & Zárate, 2018) was larger in school-aged children from wealthier than poorer districts, possibly due to more abundant gender-differentiated experiences in wealthy families or local norms. However, some of these wealth differences may be confounded by differences in ethnicity and level of urbanization across districts. Many studies found no significant or consistent relation between SES and gender differences, including some studies on math, reading, and science literacy (van Langen, Bosker, & Dekkers, 2006), math and spatial assembly (Verdine et al., 2014), spatial abilities (Wai, Lubinski, & Benbow, 2009), leisure activity behavior (Leversen, Torsheim, & Samdal, 2012), and toy collections (Nelson, 2005). Even parents' gender socialization toward their own children often did not interact with SES (Block, 1983; Endendijk, Groeneveld, Bakermans-Kranenburg, & Mesman, 2016), although higher SES parents reported more egalitarian attitude on scales not about their own offspring (Marks, Lam, & McHale, 2009).

Aims and Hypotheses

Given the importance of early gender differences and the limitations of drawing conclusions based on limited cultural settings (Gibbons, 2000; Heinrich et al., 2010), our first aim was to test the gender differences in play, spatial and social skills, and parental socialization in Chinese children. HK is a blend of both Western culture (due to British colonial history and industrialization) and Eastern Confucian culture (due to the continuing practice of traditional Chinese customs). However, patterns of gender differences and gender socialization are similar between Chinese and Western samples (e.g., Geng et al., 2012; Lam, 1982; Li & Wong, 2016; Lippa et al., 2010; The Women's Foundation, 2015; Yu et al., 2010). Thus, we hypothesized that the gender differences would manifest in similar ways across the West and HK.

We further explored how generalizable the gender differences were across SES. We challenged the hypothesis that SES moderates gender differences because while certain gender differences have been found to be larger at higher SES levels (Levine et al., 2005; Reardon et al., 2018), other research provided no support for this pattern for a variety of outcomes (e.g., Block, 1983; Endendijk et al., 2016; Leversen et al., 2012; Nelson, 2005; van Langen et al., 2006; Wai et al., 2009). Note that finding no support for SES × Gender interactions does not contradict a main effect of SES, such that higher SES may correlate with better skills (Conger, Conger, & Martin, 2010), and higher SES parents can afford more toys of any type.

Finally, given that prior studies on the correlates of play have focused on spatial skills, we tested how play correlated with both spatial and social skills. We hypothesized that spatial skills would relate positively to masculine play, empathy and comforting would relate positively to feminine play, and aggression would relate positively to masculine play. Although not the focus of the study, gender-neutral toys have been rated as most educational and conducive of a more diverse range of skills than gender-typed toys (Blakemore & Centers, 2005), so we hypothesized that it would correlate with more desirable outcomes in both spatial and social domains. In addition, to assess the relative importance of play, we compared the correlations between the skills and play to the correlations between the skills and parental gender socialization. Prior research suggested that parental gender socialization is correlated minimally with children's actual behavior or interests (Tenenbaum & Leaper, 2002; Turner & Gervai, 1995). Thus, we hypothesized that there would be more correlations of the skills to gender-typed play than to parental gender socialization.

In order to address these hypotheses with appropriate measures, we chose measures that assess domains that have been found to be gender-differentiated. Whenever possible, we chose measures that have shown gender differences in similar-aged children.

Method

Participants and Procedure

Participants were 644 K3 (i.e., final year of kindergarten) children from demographically diverse kindergartens in HK. This sample size is larger than that in many studies on early gender differences and is sufficient to detect small to moderate group differences with > 80% power. There were 349 boys and 295 girls (both $M_{age} = 67$ months, SD 5.37 and 5.80, respectively). Primary caretakers (509 mothers, 135 fathers; $M_{age} = 37$ years) also took part. All but 5 children and 6 parents were Chinese.

We chose family income and parental education as the SES indicators because they provide reasonably good coverage of SES and relate to many cognitive and socioemotional outcomes (Conger et al., 2010). Although annual income may better reflect the stability of participants' financial situation, reporting monthly income is the official governmental practice in HK, thus participants reported monthly instead of annual family income (in Hong Kong dollars, HKD) (USD1/HKD7.78). The median income (USD3856) was close to the city's (USD3702 for economically active households) (Census and Statistics Department, 2015). Income was categorized as low (< USD3021, $n = 248$), middle (USD3021-7712, $n = 276$), or high (≥ USD7713, $n = 120$). The cutoff points reflect the city's median household income, perception of middle class, and the two highest-earning percentiles (Census and Statistics Department, 2011, 2015). Parents' educational attainment was categorized as low (lower secondary school or below, $n = 168$), middle (senior secondary school to A-level, $n = 282$), or high (post-secondary or higher, $n = 194$). We used univariate measures because each SES indicator has unique stability and relations with outcomes, and composite measures have drawbacks (Oakes & Andrade, 2017). There were no significant differences in child age by gender, income groups, or parental education (all t tests $ps > .05$).

Children completed measures of mental transformation, comforting, targeting accuracy, and an observed play paradigm, in this order, individually at school. Parents completed a questionnaire assessing demographic information, children's empathy, aggression, play experience, and their child gender socialization.

Measures

Mental Transformation

The Mental Transformation Task (Levine et al., 1999) is an untimed spatial task designed to suit preschoolers and is sensitive to the male advantage in children as young as 4.5-year-olds. It is a popular measure for studying gender differences in mental transformation in young children. At least several studies using this measure or measures with similar task requirements found a male advantage in young children. In Levine et al. (1999), the effect size was largest at 5 to 5.5 years ($d \sim .64$). Others have found a d of 0.59 to 0.7 from 4.5 years to 5.5 years (Ehrlich et al., 2006; Levine et al., 2012; Pruden et al., 2011).

There were 4 practice items and 32 test items ($\alpha = .79$), each requiring the child to choose one of four whole shapes (i.e., one target shape and three foils in a 2 × 2 choice array) that could be formed by two given halves divided along the vertical axis of the target shape. The two halves were separated apart either horizontally or diagonally. In half of the trials, they were also rotated 60° clockwise and counterclockwise, respectively. All shapes were two-dimensional. Correct answers were summed (0–32). One-sample t tests showed that both boys, $t(348) = 39.53$, and girls, $t(294) = 32.67$, $ps < .001$, scored significantly above 8, the score that corresponds with answering correctly at chance level.

Targeting Accuracy

Targeting accuracy shows a large male advantage ($d = .96$) that was age-invariant in children aged 6–11 years (Thomas & French, 1985). Adapting from typical measures (e.g., Hines et al., 2003), the task required participants to throw Velcro balls at the center of a 1 m × 1 m Velcro board mounted vertically 1.2 m away. Children used their dominant hand but could throw overhand or underhand as they preferred. There were 4 practice trials and 10 test trials. The distance (in mm) from the center was averaged. Shorter distance indicates higher accuracy.

Comforting

Comforting shows a small gender difference ($d = .17$) across a variety of measures in children and adolescents combined (Fabes & Eisenberg, 1998). Some studies employing an infant cry paradigm found a moderate female advantage in terms of how many comforting strategies were offered (e.g., Catherine & Schonert-Reichl, 2011, $d \sim 0.41$ for caregiving strategies in 5–13-year-olds; Li & Wong, 2016, $d = 0.52$ at $M_{age} = 5.83$ years). In this study, the researcher showed a photograph of a crying infant with gender-neutral appearance and asked, "The baby is crying. What can be done to make the baby more comfortable or happier, so it will stop crying?" If the child stopped, the researcher gave standard prompts (e.g., "If the baby keeps crying, what would you do?"). The task ended when the child indicated no more ideas twice consecutively. Responses were recorded verbatim. The total number of strategies as coded by two raters blind to participants' gender correlated highly, $r = .99$, $p < .001$ (for details of the coding, see Li & Wong, 2016). Total comforting strategies were log-transformed to correct for non-normality.

Empathy

Recent studies employing parent reports of empathy reported a consistent moderate female advantage (Chapman et al., 2006; Escovar et al., 2016; Owen-Anderson et al., 2008); few were specific to preschoolers, but one study using a large sample size ($N = 1256$) including this age group found a moderate effect size ($d = .56$; Auyeung et al., 2009). The EQ correlates with performance measures of empathy and is unrelated to social desirability (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). The Chinese version of the Empathy Quotient-Child (EQ-C; Auyeung et al., 2009) was used in this study. Parents indicated how much they agreed (1–4) with 27 statements (e.g., "My child gets upset at seeing others crying or in pain") ($\alpha = .81$). Responses were scored 0, 1 or 2 depending on the level of empathy (Auyeung et al., 2009). Higher scores (0–54) indicate more empathy.

Aggression

Peer reports seem to show the largest gender difference in physical aggression ($d \sim 0.80$ across all ages), and self-reports are rarely used for preschool children (Archer, 2004). Reports from adults such as teachers and parents have sometimes been used and show a gender difference in young children (e.g., Pasterski et al., 2007; Sanson, Prior, Smart, & Oberklaid, 1993). Not many provided an effect size but Pasterski et al. (2007) reported a d of 1.34 using a parent report. Thus, following Pasterski et al. (2007), parents completed a 17-item temperament questionnaire (Zucker & Bradley, 1995) and a principal components analysis (PCA) was performed. This yielded a two-factor (aggression and activity level) solution explaining 42% of the variance. Items were assigned to the factor for which the loading was higher. A 5-item factor with an eigenvalue > 1 and explaining 14% of the variance was used to measure aggression ($\alpha = .85$). The PCA results resembled Pasterski et al. (2007) except that the item "My child likes rough-and-tumble play" was assigned to the aggression factor. Two research assistants translated and back-translated the scale into Chinese. Parents indicated on a 5-point scale how much their child resembled each statement.

Play Experience

Chinese children showed gender differences in play in observation at 5.83 years (Li & Wong, 2016, $d \sim 1.5$) and in a parental questionnaire that included items on games and toys in 6–12-year-olds (Yu et al., 2010). Unlike most studies, we assessed play using both standardized observation and parent report to provide both objective behavioral data and data on daily behavior. In the observation, 5 masculine toys (dart gun involving targeting, blocks, toy vehicles, toy tools, transformer requiring configuration), 5 feminine toys (baby doll, dress-up costume set, cooking toys, Barbie doll, makeup kit), and 5 gender-neutral toys (xylophone, book, drawing board, puzzle, toy animal) were selected based on adult perceptions and actual child preferences (Blakemore & Centers, 2005; Pasterski et al., 2005; Zosuls et al., 2009). Thirty-three adults also made gender assignments for each toy. Their perception confirmed the gender-typing of all toys except blocks: Masculine toys were assigned "for boys" 67–91% of the time; feminine toys were assigned "for girls" 70–100% of the time; gender-neutral toys were assigned "gender-neutral" 80–97% of the time (ps of all z tests < .05). Blocks were perceived as more "gender-neutral" (64%) than "for boys" (36%) ($z = 3.35$, $p < .001$), but was included as a masculine toy because it is typically boy-preferred (e.g., Pasterski et al., 2005). The toys were arranged in a semi-circle in random order. The first 6 scorable minutes of 8 min of recorded free play were coded for the number of seconds the child played with each toy (purposeful interaction with or without physical contact). Total play time with each toy category (masculine, feminine, neutral) was calculated. Reliability between two raters blind to the hypotheses was high for masculine ($r = 1.00$), feminine ($r = .94$), and neutral toys ($r = 1.00$). The total number of seconds for play with each category was divided by 360 to derive the proportion of time playing with each category to adjust for individual differences in total play time (Pasterski et al., 2005; Zosuls et al., 2009).

Adapting from Serbin et al. (1990), parents reported the frequency (1 = almost never, 5 = almost all the time) their child played with each of the 15 types of toys used in the observation task and the absolute number of each type of toys at home. Because the reported absolute numbers of toys had extreme values, the absolute number of each toy was rank-transformed in SPSS by ranking the reported number against that of other participants. Since there were 644 children in our study, a child with an absolute number for a given toy that was ranked as 644 had more of that toy relative to children with a lower rank (e.g., 643). Reported frequency of play and rank-transformed number of toys were then averaged within each toy category, respectively. The types of masculine, feminine, and neutral toys reported to be available were also calculated (e.g., if all 5 types of masculine toys were available, then the score for "types of masculine toys at home" was 5).

Finally, PCAs combined observed play and parent-reported play to derive composite scores of play. Observed masculine play, reported frequency of masculine play, reported number of masculine toys at home (rank-transformed), and reported types of masculine toys at home were entered into a PCA using direct oblimin rotation and requesting 1 factor. The same procedure was repeated for feminine and neutral play. The observed and reported play variables together explained 69%, 74%, and 48% of the variances of the composite masculine, feminine, and gender-neutral play scores, respectively. All factors had eigenvalues > 1.

Parental Gender Socialization

Few well-structured measures exist to measure parents' attitudes about gender roles in their own children instead of general gender-role attitudes. We adapted an exception, the Child Gender Socialization scale (Blakemore & Hill, 2008). Parents rated how negative or positive (1–7) they would feel if their child engaged in certain activities, and how strongly they disagreed or agreed (1–7) that they would discourage cross-gender characteristics in their child. The scale was translated and back-translated into Chinese by two research assistants. The original scale consists of 28 items, constituting 5 subscales and 1 filler scale. Only "Toys and Activities Stereotyped for Girls" (CGS-G, 8 items, $\alpha = .89$), "Toys and Activities Stereotyped for Boys" (CGS-B, 7 items, $\alpha = .88$), and "Disapproval of Other-Gender Characteristics" (CGS-D, 2 items, $\alpha = .61$) were analyzed because they have a clearer theoretical basis. These three scales showed gender differences of ds ~ − 1.42, 0.49, and 0.64, respectively, in parents of children 2–8 years old in Blakemore and Hill (2008).

Results

To test for gender differences in play preferences, spatial and social skills, and parental socialization, and to explore their (in)consistency across SES, we conducted a series of Gender × Income × Education ANOVAs. The focus was on the main effect of gender, and whether it interacted with the two indicators of SES. To test how well play correlated with children's skills, and to compare the role of play to that of parental gender socialization, we conducted correlational analyses within each gender (to avoid its confounding effect) between play and the skills, and between parental gender socialization and the skills. To reduce the number of analyses, we focused on the predicted correlations, that is, the correlations of spatial skills to masculine play and parental socialization of boy-typical activities, the correlations of comforting and empathy to feminine play and parental socialization for girl-typical activities, the correlations of aggression to masculine play and parental socialization for boy-typical activities, and the correlations of both spatial and social skills to gender-neutral play. To reduce reliance on normality, 1000 bootstrap samples (Efron & Tibshirani, 1993) were requested in SPSS. Cohen's ds are based on means (M) and standard deviations (SD).

Gender Differences and Their (In)Consistency Across SES (Fig. 1)

Spatial Skills

For mental transformation, there was a main effect of gender, $F(1, 627) = 4.06$, $p = .044$, $\eta_p^2 = .006$, such that boys ($M = 19.20$, SD 5.29) scored higher than girls ($M = 18.29$, SD 5.41), $d = .17$. A main effect of income was also found, $F(2, 627) = 4.50$, $p = .011$, $\eta_p^2 = .014$. Bonferroni-corrected post hoc tests showed that the high-income group ($M = 20.43$, SD 5.61) outperformed the middle-income group ($M = 18.60$, SD 5.27), $p = .005$, $d = .34$, bootstrapped $CI_{.95}$ of M_{diff} (.46, 3.02), and the low-income group ($M = 18.20$, SD 5.21), $p < .001$, $d = .41$, bootstrapped $CI_{.95}$ of M_{diff} (1.01, 3.43).

For targeting accuracy, there was a main effect of gender, $F(1, 627) = 6.61$, $p = .010$, $\eta_p^2 = .010$. Boys' targeting average distance from the center ($M = 168.82$, SD 54.44) was shorter

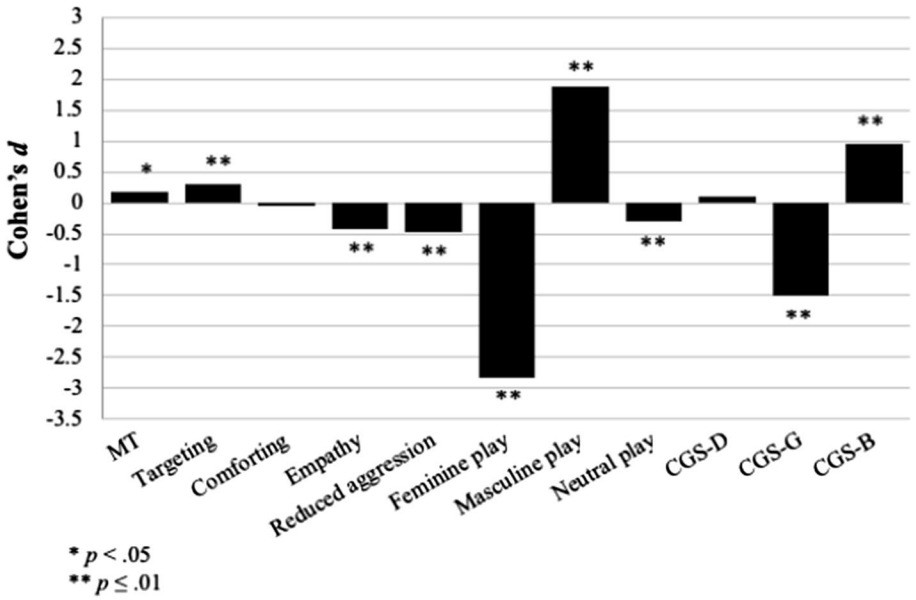

Fig. 1 Magnitude of gender differences. Positive effect sizes indicate male advantage; negative effect sizes indicate female advantage. *MT* mental transformation, *CGS-D* parental disapproval of other-gender characteristic, *CGS-G* parental socialization of toys and activities stereotyped for girls, *CGS-B* parental socialization of toys and activities stereotyped for boys. All except comforting and CGS-D showed hypothesized gender differences. Except for the gender differences in masculine play and CGS-B, no other gender differences interacted with SES (family income or parental education)

than girls' ($M = 186.52$, SD 58.40), $d = .31$, indicating higher targeting accuracy of boys. There was no significant three-way or two-way interaction between gender, income, and education for either spatial skill.

Social Skills

There was no significant main effect or interaction between gender, income, and education for the number of comforting strategies children suggested. However, main effects of gender, $F(1, 627) = 22.45$, $p < .001$, $\eta_p^2 = .035$, and income, $F(2, 627) = 3.79$, $p = .023$, $\eta_p^2 = .012$, were found for empathy. Girls ($M = 31.86$, SD 7.35) scored higher on the EQ-C than boys ($M = 28.66$, SD 7.43), $d = -.43$. Bonferroni-corrected post hoc tests showed that the high-income group ($M = 31.89$, SD 8.20) scored higher than the middle-income group ($M = 29.62$, SD 7.54), $p = .014$, $d = .29$, bootstrapped $CI_{.95}$ of M_{diff} (.57, 3.89), and the low-income group ($M = 29.83$, SD 7.14), $p = .036$, $d = .27$, bootstrapped $CI_{.95}$ of M_{diff} (.36, 3.68).

There was also a main effect of gender for aggression, $F(1, 627) = 30.17$, $p < .001$, $\eta_p^2 = .046$, with girls ($M = 1.44$, SD .67) being less aggressive than boys ($M = 1.81$, SD .87), $d = -.47$. No significant three-way or two-way interaction between gender, income, and education was found for any indicators of social skills.

Play Experience

For feminine play, there was a main effect of gender, $F(1, 627) = 862.53$, $p < .001$, $\eta_p^2 = .579$. Girls ($M = .89$, SD .59) played with feminine toys more than boys ($M = -.75$, SD .57), $d = -2.83$.

For masculine play, there were main effects of gender, $F(1, 627) = 409.28$, $p < .001$, $\eta_p^2 = .395$, and income, $F(2, 627) = 7.97$, $p < .001$, $\eta_p^2 = .025$. Boys ($M = .63$, SD .68) played with masculine toys more than girls ($M = -.74$, SD .78), $d = 1.88$. Bonferroni-corrected post hoc tests showed that, compared to the low-income group ($M = -.18$, SD 1.00), the high-income group ($M = .15$, SD .94), $p < .001$, $d = .34$, bootstrapped $CI_{.95}$ of M_{diff} ($-.55$, $-.11$), and the middle-income group ($M = .09$, SD 1.01), $p < .001$, $d = .27$, bootstrapped $CI_{.95}$ of M_{diff} ($-.43$, $-.09$), engaged in more masculine play. There was an interaction between gender and education, $F(2, 627) = 4.63$, $p = .010$, $\eta_p^2 = .015$. Bonferroni-adjusted pairwise comparisons showed that, for girls, the high-education group ($M = -.47$, SD .76) played with masculine toys more than the middle-education group ($M = -.82$, SD .73), $p = .013$, $d = .47$, and the low-education group ($M = -.90$, SD .84), $p = .017$, $d = .54$. For boys, the amount of masculine play did not differ across the low-, middle-, and high-education groups. The gender difference was largest in the middle-education group ($d = 2.10$), followed by the low-education group ($d = 1.78$), and was smallest in the high-education group ($d = 1.64$).

For neutral play, a main effect of gender was found, $F(1, 627) = 8.61$, $p = .003$, $\eta_p^2 = .014$, with girls ($M = .16$, SD .95) playing with neutral toys more than boys ($M = -.14$, SD 1.03), $d = -.30$. There were also main effects of income, $F(2, 627) = 4.69$, $p = .010$, $\eta_p^2 = .015$, and education, $F(2, 627) = 6.29$, $p = .002$, $\eta_p^2 = .020$. Bonferroni-corrected post hoc tests showed that the high-income group ($M = .49$, SD .77) played with neutral toys more than the middle-income group ($M = .06$, SD 1.00), $p < .001$, $d = .48$, bootstrapped $CI_{.95}$ of M_{diff} (.24, .61), and the low-income group ($M = -.30$, SD 1.00), $p < .001$, $d = .89$, bootstrapped $CI_{.95}$ of M_{diff} (.60, .97). The middle-income group in turn engaged in more neutral play than the low-income group, $p < .001$, $d = .36$, bootstrapped $CI_{.95}$ of M_{diff} (.19, .52). Similarly, the high-education group ($M = .43$, SD .80) engaged in more neutral play than the middle-education group ($M = -.04$, SD .97), $p < .001$, $d = .53$, bootstrapped $CI_{.95}$ of M_{diff} (.31, .61), and low-education group ($M = -.43$, SD 1.07), $p < .001$, $d = .91$, bootstrapped $CI_{.95}$ of M_{diff} (.65, 1.04), and the middle-education group engaged in more neutral play than the low-education group, $p < .001$, $d = .38$, bootstrapped $CI_{.95}$ of M_{diff} (.17, .59). No significant three-way or two-way interaction between gender, income, and education was found.

Parental Gender Socialization

There was no significant main effect or interaction between gender, income, and education for CGS-D. A main effect of gender was found for both CGS-G, $F(1, 627) = 207.66$, $p < .001$, $\eta_p^2 = .249$, and CGS-B, $F(1, 627) = 114.72$, $p < .001$, $\eta_p^2 = .155$. Parents of girls ($M = 5.87$, SD .73) were more positive about their child's engagement in girl-typical activities than were parents of boys ($M = 4.55$, SD 1.02), $d = -1.49$, whereas parents of boys ($M = 5.63$, SD .79) were more positive than were parents of girls ($M = 4.77$, SD 1.03) when their child engaged in boy-typical activities, $d = .94$.

A gender × education interaction was found for CGS-B, $F(2, 627) = 4.26$, $p = .015$, $\eta_p^2 = .013$. Bonferroni-adjusted pairwise comparisons showed that parents of boys in the high-education group ($M = 5.84$, SD .71) were more positive than those in the low-education group ($M = 5.35$, SD .86), $p = .015$, $d = .62$. As for parents of girls, the attitudes did not differ with education. The gender difference was largest in the high-education group ($d = 1.25$), followed by the middle-education group ($d = .85$), and was smallest in the low-education group ($d = .70$).

Correlates of the Skills

Correlations Between Play and Skills

We tested whether spatial skills and masculine play were correlated. Mental transformation, but not targeting accuracy, correlated positively with masculine play in boys ($r[347] = .11$, $p = .036$, bootstrapped $CI_{.95}$ [.00, .23]), but not in girls.

We next tested whether masculine play was correlated with aggression while feminine play was correlated with comforting and empathy. Aggression correlated positively with masculine play in boys ($r[347] = .14$, $p = .009$, bootstrapped $CI_{.95}$ [.02, .24]). Unexpectedly, feminine play correlated negatively with empathy ($r[347] = -.19$, $p < .001$, bootstrapped $CI_{.95}$ [−.29, −.08]) in boys. Comforting did not correlate with any type of play, and girls' social skills did not correlate with play, except for a marginally significant correlation between empathy and feminine play ($r[293] = .11$, $p = .064$).

Finally, we tested whether neutral play correlated with both spatial and social skills. Neutral play correlated positively with empathy in both boys ($r[347] = .12$, $p = .024$, bootstrapped $CI_{.95}$ [.02, .22]) and girls ($r[293] = .21$, $p < .001$, bootstrapped $CI_{.95}$ [.09, .32]). It also correlated negatively with aggression ($r[293] = -.12$, $p = .041$, bootstrapped $CI_{.95}$ [−.26, .03]) in girls. It did not correlate with spatial skills.

Correlations Between Parental Gender Socialization and Skills

We tested whether CGS-B correlated with spatial skills and whether CGS-G correlated with social skills. CGS-G correlated positively with empathy in girls ($r[293] = .15$, $p = .009$, bootstrapped $CI_{.95}$ [.03, .27]). However, CGS-B ($r[293] = .15$, $p = .013$, bootstrapped $CI_{.95}$ [.03, .27]) correlated unexpectedly with worse targeting in girls. No other significant correlations were found.

Discussion

To understand the generalizability of early gender differences across domains and cultures and to extend the study of the developmental correlates of gender-typed play, we examined a range of skills and experiences in a large sample of Chinese children aged 5 years. Included were some measures of spatial and social skills, play experience, and parental socialization for which gender differences have frequently been reported in the literature on Western samples. We also examined how gender-typed play and parental gender socialization related to the skills.

Our first aim was to test whether the hypothesized early gender differences can be found in Chinese children. As hypothesized, boys were better than girls in mental transformation and targeting accuracy, while girls were better in empathy and lower in physical aggression. These differences were consistent with prior studies of Western children on mental transformation or rotation (Casey et al., 2008; Ehrlich et al., 2006; Levine et al., 1999, 2005, 2012; Pruden et al., 2011), targeting accuracy (Thomas & French, 1985), empathy (Auyeung et al., 2009; Chapman et al., 2006; Escovar et al., 2016; Owen-Anderson et al., 2008), and physical aggression (Archer, 2004; Pasterski et al., 2007; Sanson et al., 1993). Comforting was the only skill that did not show the hypothesized gender difference, which may be surprising since other studies (e.g., Catherine & Schonert-Reichl, 2011), including one on HK Chinese children (Li & Wong, 2016), found a female advantage. The discrepancy is possibly due to the children being older (Catherine & Schonert-Reichl, 2011) or the paradigm more realistic (Li & Wong, 2016) in prior studies.

On a composite of observed and parent-reported play, boys and girls also differed on their preferences for masculine and feminine toys, consistent with many studies (Etaugh & Liss, 1992; Pasterski et al., 2005; Serbin et al., 1990; Zosuls et al., 2009), although caution should be taken on the generalizability of such gender differences to all gender-typed toys. Our girls also showed a preference for stereotypically neutral toys. The finding that girls prefer neutral toys is not unprecedented and may reflect girls' more flexible preferences (Wood, Desmarais, & Gugula, 2002). Parents' reported socialization toward their child's behavior was also similar to that in prior studies. Like U.S. parents (Blakemore & Hill, 2008), HK parents were more positive with their girls engaging in girl-typical activities and with their boys engaging in boy-typical activities, though they were not more disapproving of their boys' than their girls' cross-gender behavior.

It has been hypothesized that certain gender differences are moderated by SES (Levine et al., 2005; Miller & Halpern, 2014; Reardon et al., 2018). With a sample rather uniform in ethnicity and urbanization, we found little support for this notion. There were main effects of SES on the skills, with higher income associated with better mental transformation and higher empathy, consistent with the generally positive effect of high SES (Conger et al., 2010). There were also main effects of SES on play, with higher income associated with more masculine play, and both higher income and higher parental education associated with more gender-neutral play. These main effects may reflect the preference of higher SES parents for their children, regardless of gender, to play with certain types of toys that differ in nature; for example, neutral toys are more educational and masculine toys are more risky (Blakemore & Centers, 2005). The more important finding is that only two gender differences—those in masculine play and parental encouragement of boy-typical activities—were moderated by SES, parental education specifically.

These gender × SES interactions seem only partially consistent with the hypothesis that higher SES associates with larger gender differences (Levine et al., 2005; Miller & Halpern, 2014); while the child gender difference in parental encouragement of boy-typical activities increased with parental education, the gender difference in children's actual masculine play was smallest in the high-education group. It should also be noted that the hypothesis was often made in relation to skills, but then no variables of skills and no other variables of play and parental socialization showed gender differences moderated by SES in this study. In fact, other studies also often found little to no association between SES and gender differences (van Langen et al., 2006; Verdine et al., 2014; Wai et al., 2009) or gender-typing (Block, 1983; Endendijk et al., 2016; Leversen et al., 2012; Nelson, 2005) across a variety of outcomes. Prior research has pointed out that the prediction of gender × SES interactions was based on a few small studies (Wai et al., 2009). Our findings further suggest that gender differences in child behavior, skills, and parental socialization may, in fact, be fairly generalizable across the social strata. Whether this consistency speaks to strong biological roots, uniform socialization and learning experiences across SES, or both, requires further investigation.

While some psychological phenomena may be universal, this is not always the case (Gibbons, 2000; Heinrich et al., 2010). Thus, the consistency in the gender differences across a range of variables between the current sample and typical, Western samples is noteworthy given how little is known about gender development outside the West. Of course, the current study does not show that any gender differences are universal, nor does it exclude the possibility that the gender differences may be larger or smaller if they were measured differently. Indeed, to aid our purpose of examining gender-differentiated development, we chose variables and measures on the basis that they showed more consistent and larger gender differences than some other measures in prior studies. Nevertheless, the high level of resemblance in gender-typing between our study and other studies on similar-aged children using similar measures adds credibility to the theory that there are gender differences at an early age in outcomes such as mental transformation, targeting accuracy, empathy, aggression, play, and parental socialization, at least in an industrialized society that resembles the West in many ways.

Although the findings speak to a high resemblance in gender differences, at least in their direction, with those in Western samples, they also indicate potential cultural differences. Specifically, while the gender differences in play and parental socialization were comparable in magnitude to those in prior studies (e.g., Blakemore & Hill, 2008; Hines, 2010), those in the spatial and social outcomes, even when significant, appeared to be systematically smaller (*d*s: .17–.47) than those in some prior studies using similar measures (for example, cf. Levine et al., 2012 for mental transformation; cf. Thomas & French, 1985 for targeting accuracy; cf. Auyeung et al., 2009 for empathy; cf. Pasterski et al., 2007 for aggression). High consistency in direction but inconsistency in magnitude has also been found in large cross-cultural studies of gender differences (Guiso et al., 2008; Lippa et al., 2010; Stoet & Geary, 2013). Future studies may include a concurrent comparison group from the West to examine whether the smaller gender differences in childhood spatial and social skills in the Chinese sample may be the result of cultural-specific practices. For example, Chinese character writing may reduce the gender gap in spatial skills (Sakamoto & Spiers, 2013).

Our second aim was to expand the understanding of the relations between gender-typed play and spatial and social skills, a topic that has been extensively discussed but on which empirical research is limited. The first question was whether spatial skills would correlate positively with masculine play (Liben et al., 2018). Play with masculine toys is suggested to offer children spatial training and promote spatial skills because masculine toys often involve construction (Blakemore & Centers, 2005) and require children to identify spatial relations and manipulate objects to different orientations (Casey et al., 2008). Consistent with the hypothesis, masculine play associated with better mental transformation skill in boys. However, the same association was not found in girls. While some studies have found an association between masculine or spatial play and spatial skills in both males and females (e.g., Jirout & Newcombe, 2015; Serbin et al., 1990), some have found the association only in boys (e.g., Connor & Serbin, 1977). Even when they play with the same toys, the quality of play (Levine et al., 2012) and their playmates (Ruble et al., 2006) may differ for boys and girls, so one possible reason why masculine play did not correlate significantly with spatial skills as strongly in girls may be that when they play with masculine toys, they play in a less spatial way. Although targeting accuracy is a gender-differentiated and highly spatial skill, it did not correlate significantly with masculine play. This is the first study to test targeting accuracy as a correlate of play and prior studies that included multiple spatial tasks (such as mental rotation, WLT, EFT, and Block Design) have also found the correlations to be non-generalizable (e.g., Connor & Serbin, 1977; Robert & Heroux, 2004; Sprafkin et al., 1983), so this null finding is not surprising and warrants further study. However, it is also possible that the role of play behavior will not appear until later in development.

The second question was whether empathy and comforting would correlate positively with feminine play and aggression would correlate positively with masculine play (Murnen, 2018). Masculine toys, such as action figures and weapons, model and foster the emulation of aggressive acts. Along this line, physical aggression in boys, such as getting into fights and rough-and-tumble play, correlated positively with masculine play, validating adults' perceptions (Blakemore & Centers, 2005) and observations (Hellendoorn & Harinck, 1997; Watson & Peng, 1992) indicating the aggressive nature of masculine toys.

Feminine toys seem to be more nurturing and domestically oriented (Blakemore & Centers, 2005), encouraging caretaking and nurturing behavior while children play (Caldera & Sciaraffa, 1998). However, the hypothesis concerning empathy and comforting was not supported perhaps except for a marginally significant positive correlation between empathy and feminine play in girls. These findings contradict the positive correlation between girls' feminine play and comforting skill (Li & Wong, 2016), which may again be related to the lack of a gender difference of the current comforting task. These findings fail to support the hypothesis and perception that feminine toys promote nurturance and domestic skills (e.g., Blakemore & Centers, 2005; Caldera & Sciaraffa, 1998). However, it should be noted that only this study and one prior study (Li & Wong, 2016) have directly tested empathy and comforting as correlates of feminine toys. Given the young age of the child participants, it is possible that feminine toys would promote social skills only when children are old enough to focus on the nurturing and perspective-taking aspects rather than some other aspects (e.g., appearance) of these toys.

Lastly, we asked whether gender-neutral play would correlate positively with both spatial and social outcomes. Little attention has been paid to the role of gender-neutral toys, but adults perceive neutral toys, more than strongly gender-typed toys, as educational and likely to foster a wide range of skills (Blakemore & Centers, 2005). Results suggested that gender-neutral play may be more beneficial, at least for social development, than either masculine or feminine play. Empathy in boys and girls, and reduced aggression in girls, all correlated with more neutral play. Similarly, in Li and Wong (2016), boys' comforting skill correlated with gender-neutral play but not feminine play. These findings coincided with the perception that gender-neutral toys support optimal development, although the perception that they support learning of diverse types of skills (Blakemore & Centers, 2005) awaits further examination. Nevertheless, the finding that play with gender-neutral toys was associated with better social outcomes and no undesirable outcomes (such as higher aggression) suggests that neutral toys may be considered as an avenue for social skills intervention, especially because boys' avoidance of cross-gender activities (Doering, Zucker, Bradley, & MacIntyre, 1989; Ruble et al., 2006) may make it difficult to encourage them to play with feminine toys.

The importance of play stood out particularly as parental gender socialization showed very few significant correlations with child outcomes. Girls whose parents were more positive about girl-typical activities had better empathy, but girls whose parents were more positive about boy-typical activities had unexpectedly worse targeting accuracy. There were no other significant correlations, which is consistent with prior studies. For example, parental gender schemas correlated weakly with children's gender cognitions and minimally with children's gender-related interests (Tenenbaum & Leaper, 2002). Similarly, parents' gender-typed attributes and attitudes correlated with children's stereotype knowledge and attitude, but minimally with children's actual behavior or interests (Turner & Gervai, 1995). Parental socialization is multi-dimensional (Turner & Gervai, 1995), so these findings did not rule out relations with other forms of parental socialization, but they suggested that, for spatial and social skills, parental attitude toward gender-typed activities is less relevant than children's actual engagement in activities that may promote those skills, such as play.

Overall, the correlations substantiated the use of masculine toys as a potential intervention tool for encouraging spatial skills at least for boys and additionally suggested that neutral play may be a useful tool for encouraging social skills for both boys and girls. The correlations (.11–.21) were small to moderate (Cohen, 1988) but are common in psychological research (Hemphill, 2003) and in prior studies relating play to spatial skills (e.g., Oostermeijer et al., 2014; Robert & Heroux, 2004). Statistically small effects should not be overlooked because they may have important real-life significance (Hemphill, 2003). Finding hypothesized correlations between play and certain skills in a Chinese sample is important, given that conceptual replication of any effect is especially strong when performed across different cultures (Crandall & Sherman, 2016).

Strengths and Limitations

The play measure improved on prior studies (e.g., Caldera & Sciaraffa, 1998; Pasterski et al., 2005; Serbin et al., 1990) by incorporating multiple methods and a range of toys. Although the toys can also be categorized by nature (e.g., spatial or non-spatial) (Jirout & Newcombe, 2015; Levine et al., 1999), categorization based on gender-typing is more appropriate given our focus on gender differences and the relations between domains of gender-typing. Also, to ensure that the masculine toys were more spatial, all had a strong spatial element and the only non-masculine toy with a clear spatial element was puzzles. Thus, post hoc analysis with play categorized by spatial nature (combining puzzles with masculine toys) did not change the main conclusions. While the current study focused on toy categories, future studies may explore the effects of specific toys, particularly in the gender-neutral category, on skills development, because even toys in the same category may differ in their dominant affordances (e.g., puzzles may be more spatial, while plush toys may be more nurturing). In addition to the frequency, future research may also explore the role of the quality of the play (Levine et al., 2012) and the presence of peers given that peer influence is strong even in preschool and same-gender peers tend to increase gender-typed behavior (Martin & Fabes, 2001).

Although we found fewer significant correlations between play and the outcomes than expected, several methodological considerations suggest that our study may offer a more

stringent test for the relations between gender-typed play and gender-typed skills than some other studies. All except one of the skill measures showed expected gender differences; play was assessed using a standardized observation that offered an objective snapshot of children's preferences in combination with parents' report of the child's daily preferences. However, the restricted age range means we cannot generalize the conclusions to other ages. It is possible that certain benefits of toys are not fully reaped until children are developmentally ready. Also, although the findings increase the confidence that current theories about early gender development apply outside the West, studies encompassing more cultures will be needed to draw a more definitive conclusion.

In an attempt to be comprehensive, this study tested multiple outcomes. The rate of Type I error increases with the number of analyses. Although most of the significant findings were as hypothesized, two findings were unexpected (feminine play correlated with lower empathy in boys and CGS-B correlated with lower targeting accuracy in girls) and should be viewed with caution. The correlations also do not offer causal evidence. The majority of past studies on the correlates of play employed cross-sectional designs and some indirectly assessed them by asking adults what they think different toys can do. We know of only one experimental study that is pertinent to causal relations between gender-typed play and skills in typically developing children (Sprafkin et al., 1983). By providing a relatively comprehensive test of the correlates of play, the current correlational findings may be useful in guiding future longitudinal and experimental studies.

Conclusion

The generalizability of existing knowledge of early gender differences has been restricted by the reliance on Western samples. The study of the correlates of gender-typing in non-Western cultures is even scarcer. Using multiple measures of social and spatial outcomes, play, and parental socialization that have been found to be gender-typed in the West, we found that the gender differences were remarkably similar in pattern, though perhaps different in magnitude, between HK Chinese children and children from prior studies. Also, the gender differences were to a large extent consistent across SES, suggesting that interventions that aim at reducing gender gaps in early childhood, at least for the outcomes assessed, are equally relevant to children from different socioeconomic backgrounds.

Inferior performance in certain abilities than the other gender may be due, in part, to insufficient experience that offers suitable practice (Levine et al., 2012; Serbin et al., 1990). For young children, a likely and enjoyable source of practice is play. Therefore, researchers advocate to narrow gender gaps in development by narrowing gender gaps in play (e.g., Casey et al., 2008; Levine et al., 2012; Sprafkin et al., 1983). Our study extended the empirical database of the literature of gender-typed play by showing that masculine play correlated with mental transformation and aggression in boys and that gender-neutral play correlated with empathy in both boys and girls as young as 5 years old. These findings suggest that much of the existing knowledge about the pattern of early gender development may be generalizable at least to other industrialized cultures. These findings also provide conceptual replication, and extension, of the hypothesized correlates of gender-typed play outside the West. Lastly, given that information on the significance and magnitude of specific developmental outcomes is often an important consideration in the study of the correlates of gender-typing, the findings will likely facilitate research on early gender development, particularly those that rely on cognate samples.

Acknowledgements This study was funded by the Early Career Scheme of the Hong Kong Research Grants Council (Grant Number 27402914). We thank Deborah Mar, Rebecca Li, Nicole Yeung, Sylvia Shi, and other assistants for their help with data collection. We also thank the families and schools for their participation in this research.

References

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Archives of Sexual Behavior, 38,* 427–433.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology, 8,* 291–322.

Auyeung, B., Wheelwright, S., Allison, C., Atkinson, M., Samarawickrema, N., & Baron-Cohen, S. (2009). The children's Empathy Quotient and Systemizing Quotient: Sex differences in typical development and in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 39,* 1509–1521.

Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal, 25,* 51–71.

Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles, 53,* 619–633.

Blakemore, J. E. O., & Hill, C. A. (2008). The Child Gender Socialization scale: A measure to compare traditional and feminist parents. *Sex Roles, 58,* 192–207.

Block, J. H. (1983). Differential premises arising from differential socialization of the sexes. *Child Development, 54,* 1335–1354.

Caldera, Y. M., McDonald Culp, A., 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.

Caldera, Y. M., & Sciaraffa, M. A. (1998). Parent-toddler play with feminine toys: Are all dolls the same? *Sex Roles, 39,* 657–668.

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

Catherine, N. L. A., & Schonert-Reichl, K. A. (2011). Children's perceptions and comforting strategies to infant crying: Relations to age, sex, and empathy-related responding. *British Journal of Developmental Psychology, 29,* 524–551.

Census and Statistics Department. (2011). *Thematic report: Household income distribution in Hong Kong*. Hong Kong Special Administrative Region.

Census and Statistics Department. (2015). *Population and household statistics analyzed by district council district: 2014*. Hong Kong Special Administrative Region.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the "Reading the Mind in the Eyes" test. *Social Neuroscience, 11,* 135–148.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes, and individual development. *Journal of Marriage and Family, 72,* 685–704.

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

Coxon, S. V. (2012). The malleability of spatial ability under treatment of a FIRST LEGO league-based robotics simulation. *Journal for the Education of the Gifted, 35,* 291–316.

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66,* 93–99.

Deming, D. K. (2015*). The growing importance of social skills in the labor market*. National Bureau of Economic Research Working Paper Series 21473.

Doering, R. W., Zucker, K. J., Bradley, S. J., & MacIntyre, R. B. (1989). Effects of neutral toys on sex-typed play in children with gender identity disorder. *Journal of Abnormal Child Psychology, 17,* 563–574.

Doyle, R., Voyer, D., & Cherney, I. D. (2012). The relation between childhood spatial activities and spatial abilities in adulthood. *Journal of Applied Developmental Psychology, 33,* 112–120.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.

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

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94,* 100–131.

Endendijk, J. J., Groeneveld, M. G., Bakermans-Kranenburg, M. J., & Mesman, J. (2016). Gender-differentiated parenting revisited: Meta-analysis reveals very few differences in parental control of boys and girls. *PLoS ONE, 11,* e0159193. https://doi.org/10.1371/journal.pone.0159193.

Escovar, E., Rosenberg-Lee, M., Uddin, L. Q., & Menon, V. (2016). The empathizing-systemizing theory, social abilities, and mathematical achievement in children. *Scientific Reports, 6,* 23011.

Etaugh, C., & Liss, M. B. (1992). Home, school, and playroom: Training grounds for adult gender roles. *Sex Roles, 26,* 129–147.

Fabes, R. A., & Eisenberg, N. (1998). Meta-analyses of age and sex differences in children's and adolescents' prosocial behavior. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (5th ed., Vol. 3, pp. 701–778). New York, NY: Wiley.

Geng, Y., Xia, D., & Qin, B. (2012). The Basic Empathy Scale: A Chinese validation of a measure of empathy in adolescents. *Child Psychiatry and Human Development, 43,* 499–510.

Gibbons, J. L. (2000). Gender development in cross-cultural perspective. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 389–416). Mahwah, NJ: Erlbaum.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320,* 1164–1165.

Heinrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83.

Hellendoorn, J., & Harinck, F. J. H. (1997). War toy play and aggression in Dutch kindergarten children. *Social Development, 6,* 340–354.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78–80.

Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences, 14,* 448–456.

Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology, 28,* 1010–2016.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592.

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing social skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26,* 302–310.

Lam, M. C. (1982). *Changing pattern of child rearing: A study of low income families in Hong Kong*. Hong Kong: The Chinese University of Hong Kong, Department of Social Work in association with UNICEF.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine, 34,* 911–924.

Leversen, I., Torsheim, T., & Samdal, O. (2012). Gendered leisure activity behavior among Norwegian adolescents across different socioeconomic status groups. *International Journal of Child, Youth and Family Studies, 4,* 355–375.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35,* 940–949.

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

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socio- economic status modifies the sex difference in spatial skill. *Psychological Science, 16,* 841–845.

Li, R. Y. H., & Wong, W. I. (2016). Gender-typed play and social abilities in boys and girls: Are they related? *Sex Roles, 74,* 399–410.

Liben, L. S., Schroeder, K. M., Borriello, G. A., & Weisgram, E. S. (2018). Cognitive consequences of gendered toy play. In E. S. Weisgram & L. M. Dinella (Eds.), *Gender typing of children's toys: How early play experiences impact development* (pp. 213–255). Washington, DC: APA Books.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior, 39,* 990–997.

Marks, J. L., Lam, C. B., & McHale, S. M. (2009). Family patterns of gender role attitudes. *Sex Roles, 61,* 221–234.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37,* 431–446.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18,* 37–45.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science, 19,* 1063–1066.

Murnen, S. K. (2018). Fashion or action? Gender-stereotyped toys and social behavior. In E. S. Weisgram & L. M. Dinella (Eds.), *Gender typing of children's toys: How early play experiences impact development* (pp. 189–211). Washington, DC: APA Books.

Nelson, A. (2005). Children's toy collections in Sweden: A less gender-typed country? *Sex Roles, 52,* 93–102.

Oakes, J. M., & Andrade, K. E. (2017). The measurement of socioeconomic status. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in social epidemiology* (2nd ed., pp. 23–42). New York, NY: Wiley.

Oostermeijer, M., Boonen, A. J. H., & Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: A mediation analysis in sixth-grade students. *Frontiers in Psychology, 17,* 782. https://doi.org/10.3389/fpsyg.2014.00782.

Owen-Anderson, A. F. H., Jenkins, J. M., Bradley, S. J., & Zucker, K. J. (2008). Empathy in boys with gender identity disorder: A comparison to externalizing clinical control boys and community control boys and girls. *Child Psychiatry and Human Development, 39,* 67–83.

Pasterski, V., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development, 76,* 264–278.

Pasterski, V., Hindmarsh, P., Geffner, M., Brook, C., Brain, C., & Hines, M. (2007). Increased aggression and activity level in 3- to 11-year-old girls with congenital adrenal hyperplasia. *Hormones and Behavior, 52,* 368–374.

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

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

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2018). *Gender achievement gaps in U.S. School Districts*. https://cepa.stanford.edu/content/gender-achievement-gaps-us-school-districts. Accessed 18 June 2018.

Robert, M., & Heroux, G. (2004). Visuo-spatial play experience: Forerunner of visuo-spatial achievement in preadolescent and adolescent boys and girls? *Infant and Child Development, 13,* 49–78.

Ruble, D. N., Martin, C. L., Berenbaum, S. A., & Lerner, R. M. (2006). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotion, and personality development* (6th ed., Vol. 3, pp. 858–932). Hoboken, NJ: Wiley.

Sakamoto, M., & Spiers, M. V. (2013). Sex and cultural differences in spatial performance between Japanese and North Americans. *Archives of Sexual Behavior, 43,* 483–491.

Sanson, A., Prior, M., Smart, D., & Oberklaid, F. (1993). Gender differences in aggression in childhood: Implications for a peaceful world. *Australian Psychologist, 28,* 86–92.

Serbin, L. A., & Connor, J. M. (1979). Sex-typing of children's play preference 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.

Sprafkin, C., Serbin, L. A., Denier, C., & Connor, J. M. (1983). Gender-differentiated play: Cognitive consequences and early interventions. In M. B. Liss (Ed.), *Social and cognitive skills: Gender roles and children's play* (pp. 167–192). New York, NY: Academic Press.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE, 8,* e57988. https://doi.org/10.1371/journal.pone.0057988.

Tenenbaum, H. R., & Leaper, C. (2002). Are parents' gender schemas related to their children's gender-related cognitions? A meta-analysis. *Developmental Psychology, 38,* 615–630.

The Women's Foundation. (2015). *Research report on gender stereotypes in the Hong Kong media: A scoping study*. Authored by Marya Saidi on behalf of The Women's Foundation.

Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin, 98,* 260–282.

Todd, B. K., Fischer, R. A., Costa, S. D., Roestorf, A., Harbour, K., Hardiman, P., & Barry, J. A. (2018). Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis. *Infant and Child Development, 27,* e2064. https://doi.org/10.1002/icd.2064.

Turner, P. J., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. *Developmental Psychology, 31,* 759–772.

Turner, P. J., Gervai, J., & Hinde, R. A. (1993). Gender-typing in young children: Preferences, behaviour and cultural differences. *British Journal of Developmental Psychology, 11,* 323–342.

van Langen, A., Bosker, R., & Dekkers, H. (2006). Exploring cross-national differences in gender gaps in education. *Educational Research and Evaluation, 12,* 155–177.

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

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.

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.

Watson, M. W., & Peng, Y. (1992). The relation between toy gun play and children's aggressive behavior. *Early Education and Development, 3,* 370–389.

Wong, W. I., & Hines, M. (2016). Interpreting digit ratio (2D:4D)–behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children. *Hormones and Behavior, 78,* 86–94.

Wood, E., Desmarais, S., & Gugula, S. (2002). The impact of parenting experience on gender stereotyped toy play of children. *Sex Roles, 47,* 39–49.

Yu, L., Winter, S., & Xie, D. (2010). The Child Play Behavior and Activity Questionnaire: A parent-report measure of childhood gender-related behavior in China. *Archives of Sexual Behavior, 39,* 807–815.

Zosuls, K. M., & Ruble, D. N. (2018). Gender-typed toy preferences among infants and toddlers. In E. S. Weisgram & L. M. Dinella (Eds.), *Gender typing of children's toys: How early play experiences impact development* (pp. 49–72). Washington, DC: APA Books.

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology, 45,* 688–701.

Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York, NY: Guilford Press.