A school-based intervention to reduce gender-stereotyping

School Psychology International

0(0) 1–28

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01430343211009944

journals.sagepub.com/home/spi **Lauren Spinner**

University of Kent, UK

Harriet R. Tenenbaum

University of Surrey, UK

Lindsey Cameron

University of Kent, UK

Anna-Stiina Wallinheimo

University of Surrey, UK

Abstract

Children's adherence to gender stereotypes can be detrimental, yet interventions to tackle stereotyping have achieved mixed success. Few studies have examined interventions developed collaboratively by educators and researchers, and the impact of interventions that focus on increasing perceived similarities between genders has yet to be tested. The present study evaluated an intervention among 6- to 10-year-old British children (47 boys; 37 girls) compared to a control group (61 boys; 47 girls). Led by their class teacher via weekly 30-minute lessons, over a four-month period, children learned to identify and challenge stereotypes, and explored similarities between genders and diversity within gender (e.g., not all girls like pink). Key measures of gender flexibility in relation to toy play, occupations, and perceived similarity to gender groups were utilised. After controlling for baseline scores, the intervention group felt more similar to the other gender and reported that they could do a wider range of occupations in the future. Those who showed less flexibility around toy play at baseline were more flexible after the intervention. Boys in the intervention group reported that gender

Corresponding author:

Lauren Spinner, The Tavistock Centre, Tavistock and Portman NHS Foundation Trust, 120 Belsize Lane, London NW3 5BA, UK.

Email: Lspinner@tavi-port.nhs.uk

stereotypes were more unfair than did the control group. Findings are discussed in relation to theories of gender stereotyping and intervention research.

Keywords

gender, children, gender stereotype, intervention, school, prejudice

Gender stereotypes shape children's lives and can restrict opportunities (Leaper, 2015; Leman & Tenenbaum, 2014). School-based interventions can be used to reduce gender stereotyping (Aboud et al., 2012; Beelmann & Heinemann, 2014; Pfeifer et al., 2007), which is why academic researchers have concentrated their efforts on the design and evaluation of theoretically-based interventions (see Bigler, 1999, for a review of these programs). One criticism of these interventions is that they do not involve educational practitioners in the intervention design. Similarly, practical school-based interventions that suit the school environment have been created by educational practitioners but sometimes lack rigorous evaluation. As a result, the lack of collaboration between the two has hindered efforts to reduce gender stereotypes (Bigler, 1999). The current study was consequently designed to test the impact of a school-based intervention, created through collaboration between researchers and education practitioners, to challenge children's use of gender stereotypes via lessons exploring the similarities between genders and diversity within each gender. The findings demonstrate the positive impact of interventions designed in this way on children's gender flexibility in relation to perceived similarity to the other gender, occupational aspirations, flexibility about toy play, and an increased awareness of the unfairness of gender norms (among boys), without affecting children's own gender contentedness or perceived similarity to their own gender group.

Impact of gender stereotypes

Psychological approaches may explain the development of gendered stereotypes. For example, dynamic systems approach, combines gender schema and social learning theories of gender to suggest that children appropriate gender stereotypes from their everyday interactions and observations of their social worlds (Martin & Ruble, 2010). Stereotypes are a set of beliefs about the characteristics of a particular group (Biernat & Kobrynowicz, 1999), and children have been shown to learn stereotypes about gender before learning stereotypes about any other social group (Zemore et al., 2000). Children's gender stereotype knowledge rapidly increases from 3 years of age (Aubry et al., 1999), and adherence to gender stereotypes

appears to peak at 5 or 6 years, at which point children are highly rigid in their thinking about and observance of gender stereotypes (Miller et al., 2006). Psychological theories of gender development highlight the roles of children's environment, and their developing cognitive abilities, in determining children's knowledge of, understanding of and adherence to gender stereotypes (Bem, 1981; Bigler & Liben, 2007; Gerbner, 1998; Leaper, 2015).

Children's toy choices, play styles, subject choices, self-esteem, and self-efficacy are all impacted by gender stereotypes, as children conform to societal expectations of their own gender from a young age (Leaper, 2015; Leman & Tenenbaum, 2011). In this way, gender stereotypes inhibit children's learning because children are deterred from learning about activities deemed appropriate for the other gender (Leaper & Bigler, 2018; Weisgram et al., 2014). Gender stereotypes are reflected in children's aspirational occupation choices (Gottfredson, 1981; Woods & Hampson, 2010) and limit interest in future occupations that differ from traditional gender stereotypes (Weisgram et al., 2010). Gender stereotypes may also inhibit children's play with other children, limiting opportunities for learning from peer interaction (Martin et al., 2017). Finally, gender stereotypes can lead to stereotype threat (Schmader et al., 2008), in which students' performance is reduced based on negative stereotypes about the abilities of their social group. For instance, girls' performance in mathematics may be reduced because girls appropriate the gender stereotype that boys outperform girls in mathematical tasks (Galdi et al., 2014). Thus, gender stereotypes can serve as a gatekeeper, or barrier, to children's learning, future occupational decisions, and peer interactions.

Interventions to reduce gender stereotypes

School-based interventions are potentially one of the most influential means by which gender stereotypes can be challenged and gender inclusiveness facilitated. Gender stereotypes are pernicious and engrained in children's social world, therefore any intervention to counteract such messages needs to be intensive. Schools provide an opportunity for gender stereotype interventions to reach large numbers of children simultaneously and continuously over extended periods of time. Moreover, schools are uniquely positioned to implement interventions that challenge gender stereotypes across wide age ranges. School-based interventions can be delivered by trusted, trained education professionals who can successfully engage their class with the material, and challenge stereotypes in a supportive setting.

School-based interventions have been used previously to reduce gender stereotyping (Aboud et al., 2012; Beelmann & Heinemann, 2014; Fabes et al., 2019; Martin et al., 2017; Pfeifer et al., 2007). However, a critical limitation of the majority of this work is the lack of collaboration between academic researchers and practitioners. Academic researchers have concentrated their efforts on the design and evaluation of theoretically-based interventions that aim to ameliorate gender stereotyping (see Bigler, 1999, for a review of these programs). But they may lack practicality, and are therefore unlikely to be embedded in school

curricula. On the other hand, practical school-based interventions that suit the school environment have been created by educational practitioners but sometimes lack rigorous evaluation. The lack of collaboration to date between educational practitioners and academic researchers has hindered efforts to reduce gender stereotypes (Bigler, 1999). The current study was consequently designed to further our understanding of the effects of a theory-based teacher-initiated programme to reduce gender stereotyping in an elementary school, produced in collaboration with academic researchers.

In the current research, we developed and evaluated an intervention that aimed to dismantle gender stereotypes by 1) teaching children about gender stereotypes and discrimination, and 2) challenging those stereotypes and exploring similarity between genders (e.g., teaching them that a range of emotions are experienced by women and men, boys and girls, that girls and boys often like similar toys), as well as diversity within any one gender (e.g., not all girls like pink, not all boys like football) using counter-stereotypic examples. The lessons encouraged acceptance of these similarities and differences.

Previous research has demonstrated some positive impacts of interventions to reduce gender stereotyping (Bigler, 1999; Pahlke et al., 2014). One such intervention in an elementary school tested the effects of teaching children to recognise and respond to sexism in media and peer groups (Pahlke et al., 2014). Compared to children who were taught prosocial attitudes, children in the gender intervention condition were more likely to recognise sexism and respond to exclusionary behaviour in peer groups, but they were no more likely to become more egalitarian or increase their favourability toward the other gender.

Another way to dismantle gender stereotypes is through the use of counter-stereotypic information, which constitutes one of the most commonly used types of interventions in schools. However, success has been inconsistent (e.g., Bigler, 1999; Coyle & Liben, 2016; King et al., 2020; Sherman & Zurbriggen, 2014). In their review, Olsson and Martiny (2018) suggest that the use of counter-stereotypical role models may reduce gender stereotypes in children but may not necessarily decrease gendered behaviour. For example, Spinner et al. (2018) found in their intervention with young children that whilst exposure to gender counterstereotypic peer models in children's magazines made children more accepting of other children's gender stereotypic toy play, it did not affect children's own toy preferences, which remained gender-stereotypic.

Teaching children about gender stereotypes, to challenge sexism, is a logical step in an intervention. However, research on gender development has shown that an increase in gender-knowledge can be related to an increase in gender-typed behaviour. For example, observational research with infants found that those who used gender labels most frequently (indicating knowledge of gender categories) were most likely to engage in gender-typed toy play (Zosuls et al., 2009). Therefore, teaching children about gender stereotypes may sometimes have the unintended effect of increasing gender-typed behaviour. Part of the problem may be that it is

easier to amplify biases than reduce them, particularly when the stereotypes reflect a child's reality (Jussim et al., 2009).

A central aspect of gender stereotypes is the heightened perception of differences between genders (Leaper, 2015; Ruble et al., 2006). Indeed, measures of gender stereotypes typically require differentiation between the genders, and endorsing the view that children's toys, activities, or personality traits are appropriate for one gender rather than for both genders (e.g., Cvencek et al., 2011). Interventions that use counter-stereotypic examples and challenge gender stereotypes in other ways could weaken perceived differences between genders, and thereby increase children's perceived similarity to the other gender. Yet previous interventions have not focused on perceived similarity to the other gender (e.g., Fabes et al., 2019; Martin et al., 2017). Given mixed findings from previous interventions that used counter-stereotypic examples or taught children about sexism, the present intervention additionally taught children about similarities between the genders and diversity within each gender. We expected that the school-based intervention tested here would increase perceived similarity to the other gender and lead to more gender flexibility (i.e. reduced adherence to strict gender stereotypes) in relation to toy play and occupations. We also expected that the intervention would lead to less acceptance of gender stereotypes by children indicating that they are more unfair by the end of the programme.

The effectiveness of gender-stereotype interventions may also differ between girls and boys. In a study in which elementary school children were taught to either challenge sexist remarks made by their peers (practice condition) or listen to stories about other children's retorts (narrative condition), children in the practice condition challenged their peers' sexist comments more than the children in the narrative condition. However, the behavioural responses to stereotyped comments led to decreases in gender-typing of others (i.e., the children's endorsement of cultural gender stereotypes) only among girls, not boys (Lamb et al., 2009). In light of these findings, the current research also explores the impact of the intervention on boys and girls separately.

The current study

The present study focused on a school intervention that taught children to identify and challenge gender stereotypes, explore similarities between and differences within genders, (e.g., teaching them that a range of emotions are experienced by women and men, boys and girls, that girls and boys often like similar toys) and accept these differences and similarities. A group of children from one school took part in the intervention (intervention group), while another group of children in a matched school (control group) did not. To test the effectiveness of the intervention, children in the intervention group and children in the control group completed questionnaire measures before (baseline) and after the intervention (Time 2). The current study determined the impact of the intervention on key outcomes indicating the extent to which children rely on gender stereotypes, or demonstrate

gender flexibility. It is important that prejudice-reduction interventions tackle prejudice without having a negative effect on individuals' attitudes towards their own social group, because this may damage children's self-concept (Cameron & Rutland, 2016). For this reason, in the present study, we also examined the effect of the intervention on gender contentedness and perceived similarity to one's own-gender group. We focused on children 6 years and older. Children below 6 years of age experience a rapid increase in gender stereotype knowledge (Serbin & Sprafkin, 1986; Signorella et al., 1993) and have rigid definitions of how girls and boys should behave in terms of the activities they engage in and the toys they should play with (Martin & Ruble, 2004). By 6 or 7 years of age, children begin to develop more flexibility in their expectations and application of gender stereotypes (Trautner et al., 2005), which means that children in this age group may be more receptive to messages that contradict strict gender stereotypes.

The aim of our research was to determine whether, after controlling for scores at baseline, we observed differences between the groups on our key outcomes at Time 2, following the intervention. Because participants in the intervention and control groups were similar in terms of age, ethnicity, and socio-economic status (as the two schools were in the same local authority, in the same town) we did not expect to observe differences at baseline, but research indicates that baseline scores may influence outcomes (Vickers & Altman, 2001). For this reason, our analytic approach was to compare the intervention and control groups at Time 2, controlling for scores at baseline (pre-intervention) (Cook & Campbell, 1979). We hypothesized that after controlling for baseline scores, at Time 2, children in the intervention group would: 1) perceive themselves as more similar to the other gender than those in the control group; 2) believe gender stereotypes are more unfair than would children in the control group; 3) have more flexible attitudes toward who could play with masculine- and feminine-typed toys than those in the control group; and 4) would show an increase in the number of occupations that they believed they could perform in the future compared to those in the control group. We also measured gender contentedness (i.e., children's happiness with their own gender) and perceived similarity to one's own-gender group to test for any detrimental effect due to the intervention. Gender stereotype knowledge in relation to gender-typed toy play was also examined to see if this increased as a result of the intervention. Finally, we explored participant gender to examine whether the intervention was more effective for girls or boys in any of the domains described above.

Method

Participants

Two hundred four British primary school children aged between 6 and 10 years were recruited from two schools. Of this initial sample, 192 participants (47 boys and 37 girls in the intervention group compared to 61 boys and 47 girls in the

control group; $X^2(1, N = 192) = .01, p = .94$) successfully completed both the pre- and post-intervention measures and were included in the final analysis (*Intervention group* $M\ age = 7.93$ years, $SD = 1.34$; *Control group* $M\ age = 7.92$ years, $SD = 1.29$; $F(1, 188) = .00, p = .98, \eta_p^2 = .00$). The control group were recruited from a different primary school located in an adjacent neighbourhood in the same council and economic and cultural area. Both schools were located in suburban areas in predominantly White neighbourhoods, reflecting the low ethnic diversity in the area. In the intervention group, all children in years 2, 4 and 5 took part in the intervention lessons (2 classes per year group) because the teachers of these year groups were involved in designing the intervention programme. These year groups were then also selected in the control school so that similar-aged children were included in both groups. The parents/guardians of all children in years 2, 4 and 5 in both schools were sent a letter informing them of the research taking place: Opt-in consent was used, so only those children whose parents returned a consent slip took part in the study – this totalled 84 children (out of 180; 47%) in the intervention group and 108 children (out of 282; 38%) in the control group who took part at both time points.

Materials

All study measures are detailed below. Participants completed these in a fixed order, as presented:

Gender Similarity. Based on Andrews et al. (2016), across 10 questions children rated their perceived similarity to boys ($\alpha = .82$) and perceived similarity to girls ($\alpha = .66$) (e.g., how similar do you feel to X?). Scores ranged from 0 ('not at all') to 4 ('the same'). For analysis, two mean scores were calculated: one for the five items related to the same gender as the participant (e.g. for a girl, all the items indicative of how similar they feel to other girls), and another mean score for the five items related to the other gender (e.g. for a girl, all the items indicative of how similar they feel to boys). A higher score indicated greater perceived similarity. The scale was represented pictorially using stick figures of boys and girls, and circles varying distances apart: Children were told that "The closer together the circles are, the more similar you feel to other girls/boys."

Gender Contentedness. Based on Egan and Perry (2001), children answered two questions rating 1. how happy they were with their own gender (i.e. "How much do you like being a girl(boy)?") and 2. their desire to be the other gender (i.e. "How much would you like to be a boy(girl)?"), using a 5-point scale. A pictorial scale of smiley faces was included to help children choose their answers. Scores ranged from 0 ('not at all') to 4 ('definitely'). For analysis, a gender contentedness score was calculated by subtracting the score of question two from the score of question one: Calculated scores therefore ranged from -4 to 4, with a higher score indicating

greater gender contentedness. The two items were negatively correlated at baseline ($r = -.39, p < .001$) and at Time 2 ($r = -.57, p < .001$).

Gender Unfairness. Items from Egan and Perry's (2001) measure of gender contentedness were adapted to measure perceived gender unfairness. Children rated their agreement with four statements ($\alpha = .59$) (e.g., I feel annoyed that there are some things I am not supposed to do because I am a boy(girl)) on a five-point scale from 'not at all' (0) to 'definitely' (4). A mean score was calculated, with a higher score indicating greater perceived unfair treatment due to gender

Gender Flexibility about Toys: Stereotype Endorsement. To assess gender flexible attitudes around toy play, we coded children's responses to the following question for four feminine-typed toys (baby doll, tea set, pony, wand; $\alpha = .80$) and four masculine-typed toys (dumper truck, fighter jet, tool set, car; $\alpha = .79$): "Who should play with this toy?". These gender-stereotypic toys were selected from Blakemore and Centers (2005). Children selected from one of three response options, which were also paired with the corresponding gender symbols that appear on restroom signs: "Only Girls" (coded as 0), "Only Boys" (coded as 0) or "Both Girls and Boys" (coded as 1). Children could respond verbally or by pointing to the symbols of their choice (Weisgram et al., 2014). Total scores were calculated by summing the assigned codes across the eight toys. Scores could range from 0 to 8, with higher scores indicating more gender flexible attitudes toward toy play (i.e. more likely to say a toy should be played with by both boys and girls). No children indicated a counter-stereotypical endorsement, therefore all responses coded as '0' were stereotypical responses.

Gender Flexibility about Toys: Stereotype Knowledge. To assess gender knowledge around toy play, we coded children's responses for the eight toys listed above ($\alpha = .75$ for feminine-typed toys; $\alpha = .77$ for masculine-typed toys) to the following question: "Who plays with this toy?". This time, "Only Girls" and "Only Boys" responses were coded as 1, and a "Both Girls and Boys" response was coded as 0, such that higher total scores (out of 8) indicated greater rigidity in children's knowledge of gender stereotypic toy play (i.e. more likely to say that masculine-typed toys are only played with by boys and feminine-typed toys only played with by girls). Again, no children indicated a counter-stereotypical endorsement.

Occupations. Based on the Sherman and Zurbriggen (2014) 'Career Cognitions' measure, participants viewed 9 images of occupational settings: one neutral setting (a restaurant) was presented first, followed, in a randomised order, by four settings which typified female-dominated occupations (i.e., a nursery, a library, a plane aisle, and a hospital room), and four settings which typified male-dominated occupations (i.e., a building site, a fire station, a police station, and a plane cockpit). Each image was accompanied by a sentence which described the scene, e.g., "This is a restaurant where a server works"/"This is a nursery where a childcare worker works"/"This is a

building site where a builder works". The images did not include people and the sentences did not use gendered language or pronouns. After viewing each image, children were asked: "If you wanted to, could you do this job when you grow up?" (Yes = 1; No = 0). For analysis, sum scores for the female- ($\alpha = .79$) and male-stereotypic occupations ($\alpha = .80$) were calculated separately (with a maximum score of 4 for each): the higher the score, the more occupations children believed they could do. Scores for the neutral restaurant item were not included.

Procedure

Ethical approval was obtained from the Research Ethics Committee at the University of Kent and we complied with British Psychological Society guidelines for research with children. Parents provided written informed consent for their child's participation in the study.

Before the teachers began designing the intervention they met with the first author, who shared information on previous intervention studies and the theoretical bases for these. Developmental Intergroup Theory (Bigler & Liben, 2007) was used to inform some of the interventions the collaborative team discussed, with concepts such as explicit gender labelling, proportional group sizes (under-representation), classification skills, and gender essentialism being used as a basis for the intervention.

The intervention was delivered to children in years 2, 4 and 5 by their class teacher via 30-minute weekly lessons between October and January. To give an example of an intervention lesson, children in a Year 2 class were taught how to define a stereotype. They then watched a short film clip about a girl who experienced happiness, sadness, anger, fear, and disgust to teach children that both girls and boys feel the same emotions. Teachers led a lesson on the different emotions girls and boys experience ending with children discussing, first in pairs and then as a group, emotions that they had experienced. To teach a similar lesson to Year 5 children, in pairs, children referred to a large chart of listed emotions to identify emotions experienced by men and women photographed in different situations. Teachers helped children understand the multitude of emotions experienced by both men and women/boy and girls. Table 1 displays the learning outcomes across the different age groups. In another lesson, year 4 children were asked to categorise a selection of toys and activities as stereotypically belonging to boys or girls. The teacher then led a discussion about whether these stereotypes were always accurate and asked the children to survey one another using a tally chart to find out which toys and activities they liked to engage with. The children shared their findings that some of their classmates liked to do things that were not gender stereotypic, e.g., some of the girls in the class liked to play video games and/or play football. The class were then shown gender counter-stereotypic examples of boys, girls, men and women doing different activities (e.g., a man ballet dancing) and in small groups were asked to make a list of all the people they could think of who also liked to do things not typically associated with their gender. See Supplementary Material for further examples of schemes of work and example lessons.

Table 1. Gender equity milestones.

	Able to define	Awareness of gender	Aspirations	Emotional literacy
Nursery (3 to 4 years)	Boy, girl, man, woman	Discuss toys and play	Expresses what they want to be when they grow up	Knows the emotions: angry, happy, sad, frightened
Kindergarten to Year 2 (4 to 7 years)	Male, female, gender, sex, stereotypes, fair	Common stereotypes, what they like about friends of all genders	Knows that all jobs are open to all	– Knows that girls and boys feel the same emotions –Able to recognize emotions in themselves and others
Years 3 and 4 (7 to 9 years)	Cultural, biological, messages, respect, equality, expectations	Can challenge stereotypes and recognize unequal treatment	Able to identify a male and female role model	– Can discuss subtleties of emotions –Ranks emotion words by intensity –Able to identify stereotypes around emotions
Years 5 and 6 (9 to 11 years)	Diverse, counterstereotypical, equity, masculinity, femininity	Recognition of impact of stereotypes, can argue against stereotypes with reasoning	Recognizes need for diversity. Suggests ways to encourage diversity in work	– Able to accurately and clearly describe emotions –Able to deal with negative emotions constructively

Note: All 3-year-olds are entitled to free nursery starting the 1 September after they turn 3 years. Most schools, such as the intervention school, offer nursery education.

To measure the effectiveness of the intervention, participants in both the intervention and control groups completed a short questionnaire one week before the intervention began in the experimental group (at the start of the autumn term), and then again four months later (at the start of the spring term). Participants aged 6 to 7 years completed the questionnaire individually with an experimenter, while those aged 8 years and over completed the questionnaire in groups of four to eight children with an experimenter in a quiet area of the school. Those who completed the measures in small groups were asked not to discuss their answers during the testing session.

Participants were told that they were going to be shown some pictures of toys and different jobs and would be asked a few questions about them. The researchers told them that they would like to know who they thought played with the different toys and what jobs they thought they could do when they were older. Participants were reminded that there were no right or wrong answers and that their answers were private. Before data collection commenced, children provided verbal assent individually to the researcher.

Results

Data analysis

To test whether there were differences at Time 2 (after the intervention) after controlling for baseline scores, we conducted ANCOVA models for each variable (Vickers & Altman, 2001) with Time 2 scores as the DV and condition (Condition: intervention, control) and participant gender (Gender: boy, girl) as IVs. Baseline scores served as a covariate. We checked for homogeneity of slopes and report change scores and figures when these assumptions were violated. There were no significant differences in baseline scores between the control and intervention groups for any of the DVs ($p > .05$). Means and Standard deviations for all study variables can be found in Table 2 and Table 3. Correlations between participant age and all DVs can be found in Table 4 for the intervention group, and Table 5 for the control group.

An alternative way to analyse this data would be to use mixed-design ANOVAs, with Condition and Gender as between-subjects variables and baseline and Time 2 DV scores as within-subject variables. This alternative analysis is reported in Supplementary Material.

Gender similarity

In relation to children's perceived similarity to the other-gender group, the ANCOVA revealed that baseline scores significantly predicted Time 2 scores, $F(1, 187) = 47.96$, $p < .001$, $\eta_p^2 = .21$. Above and beyond baseline scores, there was a statistically significant effect of Condition, $F(1, 187) = 5.28$, $p = .02$, $\eta_p^2 = .03$,

Table 2. Means and standard deviations of study variables.

Variables							
Gender similarity		Gender contentedness		Gender unfairness		Gender stereotype endorsement	
Baseline	Time 2	Baseline	Time 2	Baseline	Time 2	Baseline	Time 2
Intervention							
Girls	1.23 (.81)	1.70 (.83)	2.35 (1.64)	2.21 (1.86)	6.16 (3.45)	7.35 (4.89)	5.51 (2.64)
Boys	.91 (.85)	.98 (.98)	2.91 (1.52)	2.62 (1.61)	4.81 (3.78)	6.32 (4.16)	4.04 (2.75)
Total	1.05 (.84)	1.29 (.91)	2.67 (1.59)	2.44 (1.72)	5.40 (3.55)	6.77 (4.50)	4.69 (2.79)
Control							
Girls	.85 (.72)	1.08 (.90)	2.64 (1.50)	2.57 (1.88)	5.62 (4.33)	7.83 (4.62)	5.04 (2.69)
Boys	.92 (.88)	.89 (.88)	3.03 (1.47)	2.80 (2.13)	5.20 (3.55)	4.66 (3.54)	3.08 (2.81)
Total	.89 (.81)	.97 (.89)	2.86 (1.49)	2.70 (2.02)	5.38 (3.90)	6.04 (4.33)	3.94 (2.92)

Note: Standard deviations are in parentheses. Gender similarity scores refer to perceived similarity to the other gender. Higher gender stereotype endorsement scores indicate greater flexibility, whereas higher gender stereotype knowledge scores indicate greater rigidity.

Table 3. Means and standard deviations of occupations.

	Variables			
	Masculine-typed occupations		Feminine-typed occupations	
	Baseline	Time 2	Baseline	Time 2
Intervention				
Boys	2.43 (1.39)	2.94 (1.48)	1.51 (1.35)	2.21 (1.68)
Girls	1.59 (1.69)	2.43 (1.59)	2.43 (1.48)	2.95 (1.31)
Total	2.06 (1.58)	2.71 (1.54)	1.92 (1.47)	2.54 (1.56)
Control				
Boys	2.30 (1.44)	2.24 (1.48)	1.47 (1.51)	1.18 (1.35)
Girls	1.60 (1.68)	1.80 (1.78)	1.98 (1.54)	2.15 (1.55)
Total	1.99 (1.58)	2.05 (1.63)	1.69 (1.54)	1.61 (1.51)

Note: Standard deviations are in parentheses.

whereby, at Time 2, children in the intervention group felt more similar to the other gender than did children in the control group. There was also a significant effect of Gender, $F(1, 187) = 11.79, p = .001, \eta_p^2 = .06$. Across both the intervention and control groups, at Time 2, girls felt more similar to the other gender than did boys. However, the slopes for gender were not homogeneous. Figure 1 shows that although girls scored higher than boys at time 2, girls increased in their similarity to the other gender with a sharper slope than did boys. There was no interaction effect of Condition by Gender, $F(1, 187) = 2.15, p = .14$.

Because the slopes were not homogenous for Condition either, we conducted analyses using change scores. Note that change scores entail a loss of power (Cook & Campbell, 1979). Change scores were computed by subtracting baseline scores from Time 2 scores. When we conducted a 2 (Condition) $\times$ 2 (Gender) ANOVA on change scores in similarity to the other gender, the findings were in the expected direction for Condition but not statistically significant, $F(1, 188) = 1.64, p = .20$. However, the slopes for Condition were homogeneous in the ANOVA. They remained statistically significant for Gender, $F(1, 188) = 6.28, p = .01, \eta_p^2 = .03$.

Regarding children’s perceived similarity to their own-gender group, baseline scores significantly predicted Time 2 scores, $F(1, 186) = 21.82, p < .001, \eta_p^2 = .11$, but there was no significant effect of Condition, $F(1, 186) = 2.24, p = .14$, or Gender, $F(1, 186) = 1.80, p = .17$, nor was there a significant interaction between Condition and Gender, $F(1, 186) = 0.40, p = .53$. Therefore, children’s perceived similarity to their own-gender group was not affected by the intervention.

Gender contentedness

The ANCOVA indicated that baseline scores on gender contentedness predicted scores at Time 2, $F(1, 187) = 86.20, p < .001, \eta_p^2 = .32$. After controlling for

Table 4. Correlations for intervention group.

Variable	Correlations													
	Baseline							Time 2						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Gender similarity	—							—						
2. Gender contentedness	-.44**	—						-.48**	—					
3. Gender unfairness	.35**	-.31**	—					.47**	-.40**	—				
4. Gender stereotype endorsement	.26*	-.34**	.26*	—				.36**	-.30**	.25*	—			
5. Gender stereotype knowledge	-.23*	.31**	-.06	-.62**	—			.08	-.11	.02	.22*	—		
6. Masculine-typed occupations	.06	-.06	.05	.19	-.15	—		-.00	.01	.14	.11	-.03	—	
7. Feminine-typed occupations	.09	-.17	.28*	.23*	-.16	.38**	—	.15	-.09	.21	.31**	.24*	.60**	—

Note: $n = 84$. Only the gender similarity scores to the other gender are included here. Higher gender stereotype endorsement scores indicate greater flexibility, whereas higher gender stereotype knowledge scores indicate greater rigidity * $p < .05$. ** $p < .01$.

Table 5. Correlations for control group.

Variable	Correlations													
	Baseline							Time 2						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Gender similarity	—							—						
2. Gender contentedness	-.49**	—						-.65**	—					
3. Gender unfairness	.12	-.33**	—					.27**	-.29**	—				
4. Gender stereotype endorsement	.14	-.22*	.08	—				.20*	-.05	.32**	—			
5. Gender stereotype knowledge	.03	.05	.04	-.46**	—			.01	-.01	.22*	.47**	—		
6. Masculine-typed occupations	-.04	.04	-.04	-.04	-.11	—		.03	-.10	.03	.16	.14	—	
7. Feminine-typed occupations	.03	-.06	-.18	.15	-.17	.57**	—	.14	-.06	.16	.38**	.15	.57**	—

Note: $n = 108$. Only the gender similarity scores to the other gender are included here. Higher gender stereotype endorsement scores indicate greater flexibility, whereas higher gender stereotype knowledge scores indicate greater rigidity. * $p < .05$. ** $p < .01$.

Figure 1. Gender similarity as a function of participant gender. This figure demonstrates that, compared to boys, girls' perceived similarity to the other gender increased with a sharper slope.

baseline scores, there were no differences in Condition, $F(1, 187) = .29$, $p = .59$ or Gender, $F(1, 187) = .001$, $p = .97$ at Time 2. Nor was there an interaction between Condition and Gender, $F(1, 187) = .003$, $p = .95$.

Gender unfairness

The ANCOVA revealed that baseline scores predicted Time 2 scores on gender unfairness, $F(1, 187) = 42.15$, $p < .001$, $\eta_p^2 = .18$. Above and beyond baseline scores, there was no statistically significant effect of Condition, $F(1, 187) = .97$, $p = .33$, but there was a significant effect of Gender, $F(1, 187) = 8.72$, $p = .004$, $\eta_p^2 = .05$, whereby even after controlling for baseline scores, girls at Time 2 judged things are more unfair than boys did. There was an interaction effect of Condition by Gender, $F(1, 187) = 5.25$, $p = .02$, $\eta_p^2 = .03$. The slopes for Gender were not homogeneous. Figure 2 shows that girls' gender unfairness views increased more than did boys. To tease apart the interaction and see whether the intervention was more effective for one gender, we conducted ANCOVA models separately by gender. For boys, baseline scores predicted Time 2 scores, $F(1, 187) = 11.82$, $p = .001$, $\eta_p^2 = .10$. Boys in the intervention group thought gender norms more unfair at Time 2 than did boys in the control group, $F(1, 105) = 6.41$, $p = .01$,

Figure 2. Gender unfairness as a function of participant gender. This figure demonstrates the statistically significant interaction of gender unfairness and participant gender. Girls have a steeper slope than do boys in their judgements of unfairness.

$\eta_p^2 = .06$. In contrast, there was no difference at Time 2 with girls, $F(1, 81) = .87$, $p = .35$.

Because of non-homogeneity of slopes, we conducted between-subjects ANOVA models on change scores. When we conducted a 2 (Condition) $\times$ 2 (Gender) ANOVA on change scores in unfairness, the findings were in the expected direction for Gender, $F(1, 188) = 3.70$, $p = .06$, $\eta_p^2 = .02$. The Condition $\times$ Gender interaction, $F(1, 188) = 5.91$, $p = .02$, $\eta_p^2 = .03$ remained statistically significant. In addition, boys in the intervention group changed more in their views that gender norms were unfair than did boys in the control group, $F(1, 106) = 5.86$, $p = .02$, $\eta_p^2 = .05$, whereas there was no effect for girls, $F(1, 82) = 1.20$, $p = .28$. Thus, the findings replicated.

Gender flexibility about toys: Stereotype endorsement

The ANCOVA revealed that baseline scores predicted Time 2 scores, $F(1, 187) = 86.30$, $p < .001$, $\eta_p^2 = .32$. Above and beyond baseline scores, there was a statistically significant effect of Condition, $F(1, 187) = 8.97$, $p = .003$, $\eta_p^2 = .05$, and Gender, $F(1, 187) = 19.24$, $p < .0001$, $\eta_p^2 = .09$ on flexibility about toys. There was no interaction effect of Condition by Gender, $F(1, 187) = .01$, $p = .99$. Children in the intervention group were more flexible than were children in the control group at Time 2. Girls were more flexible than boys at Time 2. The slopes for Condition were not homogeneous and thus, findings need to be interpreted with caution. Figure 3 shows that the effect of Condition disappeared when children showed high levels of flexibility at baseline (5 or greater at baseline). To tease

Figure 3. Gender stereotype endorsement as a function of condition. This figure indicates where the statistically significant interaction occurs. Higher gender stereotype endorsement scores indicate greater flexibility, i.e., more agreement that both girls and boys should be allowed to play with specific toys. The intervention was more effective in the intervention school when children had lower levels of flexibility (lower levels of stereotype endorsement).

apart this effect, we conducted separate ANCOVA models for children who scored above (5 or greater) and below (below 5) the median. Children in the intervention group and control group who scored above the median at baseline did not differ from each other at Time 2, $F(1, 84) = .21$, $p = .65$. In contrast, children in the intervention group who started below the median at baseline became more flexible compared to children in the control group who started below the median at baseline, $F(1, 78) = 9.61$, $p = .03$, $\eta_p^2 = .11$.

Because of non-homogeneity of slopes, we conducted between-subjects ANOVA models on change scores. When we conducted a 2 (Condition) $\times$ 2 (Gender) ANOVA on change scores in flexibility, the findings were in the expected direction for Condition, $F(1, 188) = 2.77$, $p = .10$, $\eta_p^2 = .02$, and Gender, $F(1, 188) = 2.97$, $p = .09$, $\eta_p^2 = .02$, but were not statistically significant.

Gender flexibility about toys: Stereotype knowledge

The ANCOVA indicated that baseline scores on toy knowledge predicted children's knowledge at Time 2, $F(1, 187) = 35.76$, $p < .001$, $\eta_p^2 = .16$. After controlling for baseline scores, there were no differences in toy knowledge between the Conditions, $F(1, 187) = 3.13$, $p = .08$ or Genders, $F(1, 187) = 2.79$, $p = .10$ at Time 2. Nor was there an interaction between Condition and Gender, $F(1, 187) = .31$, $p = .58$.

Masculine-stereotyped occupations

The ANCOVA revealed that baseline scores predicted Time 2 scores, $F(1, 181) = 73.94$, $p < .001$, $\eta_p^2 = .29$. Above and beyond baseline scores, there was a statistically significant effect of Condition, $F(1, 181) = 9.48$, $p = .002$, $\eta_p^2 = .05$, with children in the intervention group reporting that they could do more masculine-stereotyped occupations than children in the control group at the end of the intervention. There was no difference at Time 2 between girls and boys, $F(1, 181) = .03$, $p = .86$, nor was there a statistically significant interaction effect of Condition by Gender, $F(1, 181) = .003$, $p = .96$.

Feminine-stereotyped occupations

The ANCOVA revealed that baseline scores predicted Time 2 scores, $F(1, 185) = 80.81$, $p < .001$, $\eta_p^2 = .30$. Above and beyond baseline scores, there was a statistically significant effect of Condition, $F(1, 185) = 19.87$, $p < .001$, $\eta_p^2 = .10$, with children in the intervention group reporting that they could do more feminine-stereotyped occupations than children in the control group. There was also a statistically significant effect of Gender, $F(1, 185) = 6.88$, $p = .01$, $\eta_p^2 = .04$, with girls reporting that they could do more feminine-stereotyped occupations than could boys. The interaction between Condition and Gender was non-significant, $F(1, 185) = 2.05$, $p = .15$.

Discussion

This study tells us about the impact of a school-based intervention to reduce gender stereotyping in young children through educating children about, and challenging, gender stereotypes, and exploring perceived similarities between genders, as well as variability within gender groups. We hypothesized that at Time 2, above and beyond baseline scores, children in the intervention group would perceive themselves as more similar to the other gender, would believe gender norms are more unfair than would children in the control group, would have more flexible attitudes toward who should play with toys, and believe that they could do a wider range of occupations. To our knowledge, our study is the first to find that teaching children about similarities between genders and diversity within each gender can increase children's perception of how similar they are to the other gender. The theoretical and practical implications are explored below.

Supporting our first hypothesis, at Time 2 after controlling for baseline scores, children in the intervention group perceived themselves as more similar to the other gender group (i.e., girls perceived themselves as similar to boys) than those in the control group, suggesting that the intervention successfully taught children about the similarities between boys and girls. There was no corresponding detrimental effect on perceived similarity to one's own-gender group, as these scores were unaffected. This implies that the intervention had the intended effect of breaking down gender stereotypes by emphasising similarities across and diversity

within genders. This finding is important because a core feature of gender stereotypes is the heightened perception of difference between girls and boys, men and women (Ruble et al., 2006). The increase in observed peer similarity is encouraging in regards to peer relations: children tend to select friends from those they perceive as similar (Shin & Ryan, 2014a, 2014b). Spinner et al. (2018) found that children were more likely to choose friends based on similarity in toy preference, as opposed to gender, following exposure to peers playing with counter-stereotypical toys. Emphasising similarity across genders could encourage more gender mixed friendship groups, which is important for further erosion of gender stereotypical preferences (Meyer & Gelman, 2016).

There are other benefits of cross-gender friendships. For example, when children feel more confident interacting with other-gender peers, they like school more and are more motivated (Field et al., 2017). As Fabes et al. (2019) argue in a recent review, interacting with a diverse set of peers (including peers who are the other gender) increases the number of contacts from whom children can learn. Given the efficacy of peer interaction as means for increasing learning (Tenenbaum et al., 2020), greater peer interaction with a diverse range of peers needs to be encouraged. Thus, increasing children's perceptions of similarity across gender groups should result in positive outcomes.

Across both groups, there was also an effect of gender whereby girls saw themselves as more similar to boys than boys did to girls after controlling for baseline scores. This trend may be due to the more relaxed social norms around girls' than boys' adherence to gender stereotypes: children are aware by 9 years of age that it is more difficult for boys to contravene gender norms and engage in counter-stereotypical activities than it is for girls (Mulvey & Killen, 2015), and girls are less likely to differentiate the exclusion of potential playmates based on their activities than are boys. For example, Israeli and Taiwanese 8-year-old boys reported that they would prefer to play with boys with masculine-stereotyped interests than feminine-stereotyped interests, whereas girls did not report a preference (Lobel et al., 2001).

Our second hypothesis predicted that, at Time 2, children in the intervention group would perceive gender norms as more unfair than children in the control group. This prediction was partially met as this effect was observed only amongst the boys, not the girls, in the intervention group. However, there was also an effect of gender whereby, overall, across both groups, girls perceived gender norms as more unfair than boys. Encouragingly, the intervention appears to have raised awareness of the damaging effect of gender stereotypes among boys. This finding is unique and important because boys' views are notoriously more difficult to shift as they possess more rigid views of gender (Bussey & Perry, 1982; Mulvey & Killen, 2015).

Supporting our third hypothesis, at Time 2, children in the intervention group were more flexible in toy endorsement (i.e., they were more likely to think that stereotypic masculine and feminine toys *should be* played with by both boys *and* girls). Greater acceptance of counter-stereotypic play choices in others was a key

goal of the curriculum. This finding supports previous research which has also shown that children's attitudes about who should play with toys (i.e., what other children should do) is amenable to shifts from gender stereotype interventions (King et al., 2020; Spinner et al., 2018).

There was also an effect of gender here, whereby, as with the gender similarity measure, girls scored higher than boys, indicating greater flexibility towards toy play overall (across both the intervention and control group). Again, this is in line with explanations that girls possess less gender stereotyped views than boys (Mulvey & Killen, 2015).

Fourth, we predicted that at Time 2, children in the intervention group would believe that they could perform a greater number of occupations than those in the control group. This hypothesis was supported as those in the intervention group indicated that they could perform a greater number of both masculine- and feminine-typed occupations than those in the control group. Typically, during middle childhood, the number of occupations that children differentially associate with either males or females increases significantly (e.g., Miller et al., 2009; Ruble et al., 2006). Previous intervention research has demonstrated that girls' interest in masculine-typed occupations increases when both male and female occupational titles are used (e.g. "firewomen" and "firemen", rather than just "firemen") (Vervecken et al., 2013).

However, there was an effect of gender whereby girls indicated, across both groups, that they could perform more of the feminine-typed occupations than did boys. This mirrors our other findings in relation to gender similarity and flexibility around toy play where girls scored higher than boys. Lower scores from the boys in relation to their perceived ability to perform feminine-typed occupations reiterates the notion that gender norms are stronger for boys and boys are therefore less willing (or believe themselves unable) to cross gender boundaries (Mulvey & Killen, 2015). The same trend among girls and masculine-typed occupations was not observed, reflecting previous findings that young children often permit women to occupy masculine-typed occupations, but do not permit men to occupy feminine-typed occupations (Schuette et al., 2012).

A problematic feature of gender stereotype interventions is that they can inadvertently increase knowledge of gender stereotypes, perhaps due to the nature of the lessons which may make stereotypes more salient. Increasing gender knowledge can be of concern because of the relationship between this and increased gender-typed play (Zosuls et al., 2009), which would contradict the aim of the intervention. Gender stereotype knowledge in relation to toys was measured in the current study, but there was no effect of the intervention on this.

Finally, gender contentedness was examined to confirm that learning about gender and challenging gender stereotypes would not impact children's own gender self-concepts. The intervention appeared to have no direct effect on gender contentedness, which meant that the positive impact of the intervention on gender stereotypes was not at the expense of children's attitudes towards their own gender. Children's social identities, such as gender, are an integral part of

children's self-concept. Therefore, gender stereotype interventions (and prejudice-reduction interventions more generally) should aim to achieve their goal without having a detrimental effect on attitudes towards one's own group or identity (Cameron & Rutland, 2016).

Limitations and future research

Although we have reported some novel findings, this study is not without its limitations. First, we recognize that it would be of interest and importance in future research to examine age effects to see whether the intervention was more or less effective for children of a particular age. This information would help inform us of the optimal age to introduce gender-stereotype reduction interventions in school settings. Unfortunately, the sample size of the present study meant we did not have enough statistical power to examine age effects. It would also be of benefit in future research to follow the children up at further time points to determine whether the positive effects of the intervention were sustained beyond the immediate end of the programme.

Second, we acknowledge that the measures included in this study are attitudinal rather than behavioural. In future, it would be of interest to also include measures of children's gender-typed behaviour in relation to playmate choice and toy and activity preferences to examine whether these are changed as a result of the intervention. In other words, are children who report greater perceived similarity to the other gender also more likely to play with children from the other-gender group? We also recognize that the measures included here treat gender as binary. The importance of treating, and assessing, gender on a spectrum has been highlighted in recent years (see Oleszki et al., 2020), and we fully support the development of measures that would allow the capturing of the nuances and multi-faceted nature of gender identity and diversity in children.

The increase in flexibility related to toy play was only observed in those who presented with low flexibility to start with (below the median): It shows that the intervention successfully shifted the views of the least flexible children. Unfortunately, the intervention only changed children's gender flexibility when children started with low levels of gender flexibility because there was more room for change. Indeed, ceiling effects may reduce variability leading to a mistaken notion that an independent variable does not have an influence on the dependent variable (see Cramer & Howitt, 2004, p. 21). Future research should include a measure of gender flexibility with a greater range of scores. Finally, despite the experimental design of the study, the intervention programme was led and delivered by teachers, not experimenters: therefore, the settings in which the programme was delivered was not highly controlled. Thus, we cannot make specific causal interpretations about which elements of the intervention programme were most influential or effective. It is important that future evaluations of teacher-led intervention programmes include treatment integrity measures to examine the extent to which the intervention was delivered as planned.

Finally, this research was conducted in a Western context, and children's gender attitudes will reflect the sociocultural values of the community in which they are raised. Future research needs to examine how attitudes in schools differ across cultures to develop effective interventions to reduce gender stereotyping. Such interventions are needed because gender segregation in occupational participation may be prevalent in more than just Western countries. Across 12 different countries, Froehlich et al. (2020) found gender segregation in occupations: Following stereotypes, individuals rated male targets in masculine-stereotyped occupations as more agentic than communal, whereas female targets were rated as more communal than agentic. Thus, occupational segregation may be fuelled by stereotypes.

Implications for education settings

The findings have a number of implications for educators and practitioners striving to reduce gender stereotyping. First, the findings show that it is possible to shift children's gender stereotypes. The impact of this intervention on gender flexibility demonstrates how gender stereotypical decisions and attitudes are not inevitable at this age, and it is possible to encourage more flexible gender attitudes, through intensive school-based interventions. Therefore, it is important to invest in dismantling gender stereotypes in this young age group. Second, the research identified a number of key principles that educators and practitioners could employ to successfully tackle gender stereotypes in schools, either through especially designed classes, or through embedding these principles in everyday practice. These include identifying and challenging gender stereotypes, but also focusing on similarities between and diversity within gender groups. Indeed, to our knowledge, the current research was the first to investigate and demonstrate the benefits of interventions focusing on similarities between, and differences within, genders. Our findings suggest that by focusing on the differences within gender groups, that is, the ways in which boys differ from other boys, and girls differ from other girls, and how boys and girls are actually similar to each other, future school-based interventions have the potential to break down gender stereotypes, creating more flexible attitudes towards play, occupation aspirations, and also playmate preference. Finally, this study illustrates the benefits of close collaboration between practitioners and researchers, and highlights how such collaboration can lead to effective and practical school-based interventions, that benefit from rigorous evaluation.

Acknowledgements

The authors sincerely thank the schools involved in this research, particularly Jordan Day as the teacher leading this intervention programme. Sincere thanks also to Lydia Jamieson and all of the research assistants and placement students in the School of Psychology at the University of Surrey and the University of Kent who assisted with data collection.

Ethical approval

This research study was approved by the School of Psychology Ethics Committee at the University of Kent (Approval number: 201,815,365,684,785,115).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Lauren Spinner <https://orcid.org/0000-0002-7294-2027>

Supplemental material

Supplementary material for this article is available online.

References

- About, F. E., Tredoux, C., Tropp, L. R., Brown, C. S., Niens, U., & Noor, N. M. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review, 32*, 307–336. doi:10.1016/J.DR.2012.05.001
- Andrews, N. C., Martin, C. L., & Gallagher, A. M. (2016). On the association between self-reported own-and other-gender similarity and the use of physical and relational aggression in sixth grade children. *Archives of Sexual Behavior, 45*, 1817–1826. doi:10.1007/s10508-015-0685-z
- Aubry, S., Ruble, D. N., & Silverman, L. B. (1999). The role of gender knowledge in children's gender-typed preferences. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 363–390). Psychology Press.
- Beelmann, A., & Heinemann, K. S. (2014). Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs. *Journal of Applied Developmental Psychology, 35*, 10–24. doi:10.1016/j.appdev.2013.11.002
- Bem, S. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*, 354–364. doi:10.1037/0033-295X.88.4.354
- Biernat, M., & Kobrynowicz, D. (1999). A shifting standards perspective on the complexity of gender stereotypes and gender stereotyping. In W. B. Swann, Jr J. H. Langlois, & L. A. Gilbert (. Eds.), *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence*. (p. 75–106). American Psychological Association. doi:10.1037/10277-004
- Bigler, R. S. (1999). Psychological interventions designed to counter sexism in children: Empirical limitations and theoretical foundations. In W.B. Swann, J.H. Langlois, & L.A. Gilbert (Eds.), *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence* (pp. 129–151). American Psychological Association. doi:10.1037/10277-006

- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16, 162–166. <https://doi.org/10.1111/j.1467-8721.2007.00496.x>
- Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles*, 53, 619–633. doi:10.1007/s11199-005-7729-0
- Bussey, K., & Perry, D. G. (1982). Same-sex imitation: The avoidance of cross-sex models or the acceptance of same-sex models? *Sex Roles*, 8, 773–784. doi:10.1007/BF00287572
- Cameron, L., & Rutland, A. (2016). Researcher–practitioner partnerships in the development of intervention to reduce prejudice among children. In K. Durkin & H. R. Schaffer (Eds.), *The Wiley handbook of developmental psychology in practice: Implementation and impact* (pp. 341–365). Wiley.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.
- Coyle, E. F., & Liben, L. S. (2016). Affecting girls' activity and job interests through play: The moderating roles of personal gender salience and game characteristics. *Child Development*, 87, 414–428. doi:10.1111/cdev.12463
- Cramer, D., & Howitt, D. L. (2004). *The SAGE dictionary of statistics: A practical resource for students in the social sciences*. SAGE.
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, 82, 766–779. doi:10.1111/j.1467-8624.2010.01529.x
- Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, 37, 451–463. doi:10.1037/0012-1649.37.4.451
- Fabes, R. A., Martin, C. L., & Hanish, L. D. (2019). Gender integration and the promotion of inclusive classroom climates. *Educational Psychologist*, 54, 271–285. doi:10.1080/00461520.2019.1631826
- Field, R. D., Martin, C. L., Andrews, N. C., England, D. E., & Zosuls, K. M. (2017). The influence of gender-based relationship efficacy on attitudes toward school. *Merrill-Palmer Quarterly*, 63, 396–422. doi:10.13110/merrpalmquar1982.63.3.0396
- Froehlich, L., Olsson, M. I., Dorrough, A. R., & Martiny, S. E. (2020). Gender at work across nations: Men and women working in male-dominated and female-dominated occupations are differentially associated with agency and communion. *Journal of Social Issues*, 76, 484–511. doi:10.1111/josi.12390
- Galdi, S., Cadinu, M., & Tomasello, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, 85, 250–263. doi:10.1111/cdev.12128
- Gerbner, G. (1998). Cultivation analysis: An overview. *Mass Communication and Society*, 1, 175–194. doi:10.1080/15205436.1998.9677855
- Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology*, 28, 545–579. doi:10.1037/0022-0167.28.6.545
- Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). Psychology Press.
- King, R. A., Scott, K. E., Renno, M. P., & Shutts, K. (2020). Counterstereotyping can change children's thinking about boys' and girls' toy preferences. *Journal of Experimental Child Psychology*, 191, 104753. doi:10.1016/j.jecp.2019.104753

- Lamb, L. M., Bigler, R. S., Liben, L. S., & Green, V. A. (2009). Teaching children to confront peers' sexist remarks: Implications for theories of gender development and educational practice. *Sex Roles*, 61, 361–382. doi:10.1007/s11199-009-9634-4
- Leaper, C. (2015). Gender and social-cognitive development. In R. M. Lerner (Series Ed.), L. S. Liben, & U. Muller (Vol. Eds.), *Handbook of child psychology and developmental science* (7th ed.), Vol. 2: *Cognitive processes* (pp. 806–853). Wiley.
- Leaper, C., & Bigler, R. S. (2018). Societal causes and consequences of gender-typing of children's toys. In E. S. Weisgram & L. M. Dinella (Eds.), *Gender-typing of children's toys: How early play experiences impact development* (p. 287–308). American Psychological Association. <https://psycnet.apa.org/doi/10.1037/0000077-013>
- Leman, P. J., & Tenenbaum, H. R. (2011). Practising gender: Children's relationships and the development of gendered behaviour and beliefs. *The British Journal of Developmental Psychology*, 29, 153–157. doi:10.1111/j.2044-835X.2011.02032.x
- Leman, P. J., & Tenenbaum, H. R. (Eds.). (2014). *Current issues in developmental psychology. Gender and development*. Psychology Press.
- Lobel, T. E., Gruber, R., Govrin, N., & Mashraki-Pedhaztur, S. (2001). Children's gender-related inferences and judgments: A cross-cultural study. *Developmental Psychology*, 37, 839–846. doi:10.1037/0012-1649.37.6.839
- Martin, C. L., Fabes, R. A., Hanish, L. D., Gaertner, B., Miller, C. F., Foster, S., & Updegraff, K. A. (2017). Using an intergroup contact approach to improve gender relationships: A case study of a classroom-based intervention. *The Wiley Handbook of Group Processes in Children and Adolescents*, 435–454.
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13, 67–70. doi:10.1111/j.0963-7214.2004.00276.x
- Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology*, 61, 353–381. doi:10.1146/annurev.psych.093008.100511
- Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles*, 75, 409–421. doi:10.1007/s11199-016-0646-6
- Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*, 60, 870–881. doi:10.1007/s11199-009-9584-x
- Miller, C. F., Trautner, H. M., & Ruble, D. N. (2006). The role of gender stereotypes in children's preferences and behavior. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 293–323). Psychology Press.
- Mulvey, K. L., & Killen, M. (2015). Challenging gender stereotypes: Resistance and exclusion. *Child Development*, 86, 681–694. <https://psycnet.apa.org/doi/10.1111/cdev.12317>
- Olezeski, C. L., Pariseau, E. M., Bamatter, W. P., & Tishelman, A. C. (2020). Assessing gender in young children: Constructs and considerations. *Psychology of Sexual Orientation and Gender Diversity*, 7, 293–303. doi:10.1037/sgd0000381
- Olsson, M., & Martiny, S. E. (2018). Does exposure to counterstereotypical role models influence girls' and women's gender stereotypes and career choices? A review of social psychological research. *Frontiers in Psychology*, 9, 2264. doi:10.3389/fpsyg.2018.02264
- Pahlke, E., Bigler, R. S., & Martin, C. L. (2014). Can fostering children's ability to challenge sexism improve critical analysis, internalization, and enactment of inclusive, egalitarian peer relationships? *Journal of Social Issues*, 70, 115–133. doi:10.1111/josi.12050

- Pfeifer, J. H., Brown, C. S., & Juvonen, J. (2007). Prejudice reduction in schools teaching tolerance in schools: Lessons learned since brown v. Board of education about the development and reduction of children's prejudice. *Social Policy Report*, 21, 1–24.
- Ruble, D. N., Martin, C., & Berenbaum, S. (2006). Gender development. In N. Eisenberg (Ed.), *Handbook of child psychology: Vol. 3, personality and social development* (pp. 858–932). Wiley.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356. <https://psycnet.apa.org/doi/10.1037/0033-295X.115.2.336>
- Schuette, C. T., Ponton, M. K., & Charlton, M. L. (2012). Middle school children's career aspirations: Relationship to adult occupations and gender. *The Career Development Quarterly*, 60, 36–46. doi:10.1002/j.2161-0045.2012.00004.x.
- Serbin, L. A., & Sprafkin, C. (1986). The salience of gender and the process of sex typing in three- to seven-year-old children. *Child Development*, 57, 1188–1199. doi:10.2307/1130442
- Sherman, A. M., & Zurbriggen, E. L. (2014). “Boys can be anything”: Effect of barbie play on girls' career cognitions. *Sex Roles*, 70, 195–208. doi:10.1007/s11199-014-0347-y
- Shin, H., & Ryan, A. M. (2014a). Early adolescent friendships and academic adjustment: Examining selection and influence processes with longitudinal social network analysis. *Developmental Psychology*, 50, 2462–2472. <https://psycnet.apa.org/doi/10.1037/a0037922>
- Shin, H., & Ryan, A. M. (2014b). Friendship networks and achievement goals: An examination of selection and influence processes and variations by gender. *Journal of Youth and Adolescence*, 43, 1453–1464. doi:10.1007/s10964-014-0132-9
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, 13, 147–183. doi:10.1006/drev.1993.1007
- Spinner, L., Cameron, L., & Calogero, R. (2018). Peer toy play as a gateway to children's gender flexibility: The effect of (counter) stereotypic portrayals of peers in children's magazines. *Sex Roles*, 79, 314–328. doi:10.1007/s11199-017-0883-3
- Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2020). How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology*, 112, 1303–1319. doi:10.1037/edu0000436
- Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, 14, 365–381. doi:10.1002/icd.399
- Vervecken, D., Hannover, B., & Wolter, I. (2013). Changing (S) expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behavior*, 82, 208–220. doi:10.1016/j.jvb.2013.01.008
- Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *British Medical Journal (Clinical Research ed.)*, 323, 1123–1124. doi:10.1136/bmj.323.7321.1123
- Weisgram, E. S., Bigler, R. S., & Liben, L. S. (2010). Gender, values, and occupational interests among children, adolescents, and adults. *Child Development*, 81, 778–796. <https://psycnet.apa.org/doi/10.1111/j.1467-8624.2010.01433.x>
- Weisgram, E. S., Fulcher, M., & Dinella, L. M. (2014). Pink gives girls permission: Exploring the roles of explicit gender labels and gender-typed colors on preschool

- children's toy preferences. *Journal of Applied Developmental Psychology*, 35, 401–409. doi:10.1016/j.appdev.2014.06.004
- Woods, S. A., & Hampson, S. E. (2010). Predicting adult occupational environments from gender and childhood personality traits. *The Journal of Applied Psychology*, 95, 1045–1057. doi:10.1037/a0020600
- Zemore, S. E., Fiske, S. T., & Kim, H.-J. (2000). Gender stereotypes and the dynamics of social interaction. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 207–241). Lawrence Erlbaum Associates Publishers.
- 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. <https://psycnet.apa.org/doi/10.1037/a0014053>

Author biographies

Lauren Spinner is a Research Psychologist at the Tavistock & Portman NHS Foundation Trust. As a postdoc researcher, she is interested in children's social development and particularly the development of children's gender identity. Her research has examined the impact gendered toys have on children's attitudes, beliefs and behaviours, and the social influences which contribute to and reinforce gender norms in children, such as parents, peers, and the media.

Harriet R. Tenenbaum is a Reader in Developmental Psychology and Director of Teaching and Learning at the University of Surrey. She is interested in how children and adolescents learn about the social and physical world. Her research focuses specifically on how children develop their understanding of social issues (e.g., discrimination and rights), emotion, and scientific topics.

Lindsey Cameron is a Reader in the School of Psychology at the University of Kent. Her area of research is social developmental psychology and she is interested in how children become 'social beings': how do children develop the social knowledge and skills that we have as adults take for granted? Specifically, she is interested in how inter-group processes develop throughout childhood.

Anna-Stiina Wallinheimo is an Associate Tutor in the School of Psychology at the University of Surrey. Her research looks the role of distributed cognition in reducing the effects of stereotype threat on mental arithmetic tasks and maths anxiety. She is also interested in the role of interactivity in relation to achievement goals (mastery-approach goals and performance-approach goals).