



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Early preschool environments and gender: Effects of gender pedagogy in Sweden



Kristin Shutts^{a,*}, Ben Kenward^{b,c}, Helena Falk^c, Anna Ivegran^c, Christine Fawcett^c

^a Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706, USA

^b Department of Psychology, Oxford Brookes University, Oxford OX3 0BP, UK

^c Department of Psychology, Uppsala University, 751 42 Uppsala, Sweden

ARTICLE INFO

Article history:

Received 22 December 2015

Revised 17 April 2017

Keywords:

Children

Gender

Attitudes

Stereotypes

Encoding

Socialization

ABSTRACT

To test how early social environments affect children's consideration of gender, 3- to 6-year-old children ($N = 80$) enrolled in gender-neutral or typical preschool programs in the central district of a large Swedish city completed measures designed to assess their gender-based social preferences, stereotypes, and automatic encoding. Compared with children in typical preschools, a greater proportion of children in the gender-neutral school were interested in playing with unfamiliar other-gender children. In addition, children attending the gender-neutral preschool scored lower on a gender stereotyping measure than children attending typical preschools. Children at the gender-neutral school, however, were not less likely to automatically encode others' gender. The findings suggest that gender-neutral pedagogy has moderate effects on how children think and feel about people of different genders but might not affect children's tendency to spontaneously notice gender.

© 2017 Elsevier Inc. All rights reserved.

Introduction

Children are attuned to gender from an early age: infants in the United States can discriminate between male and female faces in looking time studies (Quinn et al., 2011) and look longer at faces

* Corresponding author.

E-mail address: kshutts@wisc.edu (K. Shutts).

that match the gender of their primary caregiver (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002). In addition, most U.S. children can label their own and others' gender by the time they are 2 years old (Weinraub et al., 1984; Zosuls et al., 2009). Young children not only perceive gender but also use gender to guide their social preferences and inferences about others (Dunham, Baron, & Banaji, 2016; Martin & Halverson, 1981). Preschool-age children in the United States show gender in-group favoritism (Maccoby & Jacklin, 1987; Martin, Fabes, Evans, & Wyman, 1999; Renno & Shutts, 2015; Shutts, Roben, & Spelke, 2013) and hold gender-based beliefs about others' attributes (e.g., that boys like trucks and girls like dolls; Bauer & Coyne, 1997; Kuhn, Nash, & Brucken, 1978; Martin, 1989). Although much of the research on children's consideration of gender has been conducted in the United States, studies reveal that young children in other cultures also categorize people by gender, prefer members of their own gender, and hold gender stereotypes (e.g., Brazil: de Guzman, Carlo, Ontai, Koller, & Knight, 2004; China: Knobloch, Callison, Chen, Fritzsche, & Zillmann, 2005; South Africa: Albert & Porter, 1986, and Muthukrishna & Sokoya, 2008; South Wales: Yee & Brown, 1994).

Although the prominence of gender in young children's minds is well established (see Ruble, Martin, & Berenbaum, 2006, for a thorough review), the reason for the category's preeminence is less clear (Martin, Ruble, & Szkrybalo, 2002). Some researchers have posited that gender is prominent because humans have an evolved specialized system dedicated to classifying and reasoning about others based on their gender (e.g., Cosmides, Tooby, & Kurzban, 2003; see Shutts, 2013, for a discussion). An alternative proposition in the field is that children focus on gender because their social environments highlight the importance of the category. For example, both *gender schema theory* (Bem, 1981, 1983) and *developmental intergroup theory* (Arthur, Bigler, Liben, Gelman, & Ruble, 2008; Bigler & Liben, 2007) emphasize that children receive significant input—from parents, teachers, and media—about gender categories and roles. Indeed, studies show that, at least in the United States, adults regularly use nouns and pronouns to mark gender when talking to children (Gelman, Taylor, & Nguyen, 2004; see Waxman, 2010, for a discussion). Moreover, teachers sometimes use gender to organize their classrooms (e.g., asking children to alternate by gender in seating; Bigler & Liben, 2007). Such practices may contribute to, or fully account for, children's early and robust reliance on gender as a social category.

It is difficult to determine the role of social experience in guiding children's reliance on gender categories because many young children spend significant periods of time in social environments where gender is emphasized—including preschools. Nevertheless, there are a small number of preschools that are committed to the practice of “gender-neutral” classroom environments in which teachers typically refrain from using gendered language and actively work to counteract gender stereotypes. Studying children who experience gender-neutral pedagogy (vs. more typical instruction) provides an unusual opportunity to examine the role of teachers' behaviors in guiding children's consideration of gender distinctions. The current research examined whether and how such schooling is associated with children's reliance on gender information across a range of measures.

In addition to shedding light on the theoretical proposition that adults' behavior plays a key role in children's consideration of gender, research on gender-neutral pedagogy makes a practical contribution. Many have noted the negative effects of social exclusion, gender stereotyping, and unequal treatment on young children's development (e.g., Andrews, Martin, Field, Cook, & Lee, 2016; Bian, Leslie, & Cimpian, 2017; Halpern et al., 2011), and gender-neutral pedagogy seeks to address such social problems through deemphasizing gender distinctions. Furthermore, gender-neutral pedagogy is of great international interest. For example, the first 20 results from a recent (February 2017) Google internet search for “Swedish gender neutral preschools” included relevant articles from *The New York Times*, *The Washington Post*, *Newsweek*, *The Guardian*, *The Daily Mail*, *The Independent*, and the British Broadcasting Corporation (BBC) (all published since 2010). Yet we are not aware of any research comparing children who attend gender-neutral preschools with those who attend more typical schools. Such research is needed because it bears on whether gender-neutral practices (which are effortful and require extensive teacher training) can affect children's perceptions, feelings, and thoughts about gender.

Previous research on gender pedagogy in schools

Two studies have examined how highlighting gender categories in classrooms affects children's gender attitudes and stereotyping. In one experiment (Bigler, 1995), 6- to 11-year-olds in the United States were randomly assigned to classrooms that differed in the emphasis teachers placed on gender over the course of 4 weeks. Teachers in the experimental classrooms frequently used gender noun labels to refer to students (e.g., "All the girls should be sitting down," "Jack, be a good helper for the boys") and designed classroom materials and activities that highlighted gender (e.g., a bulletin board with girls' art on one side and boys' art on the other, desks placed such that all the boys sat on one side of the room and all the girls sat on the other). Teachers in control classrooms treated their class as a unit and referred to students by their individual names only. Following the intervention, children in the experimental classrooms showed higher levels of gender stereotyping but showed similar levels of gender in-group favoritism to children in control classrooms. As Bigler (1995) noted, however, the gender in-group favoritism measures showed little variance, and most children were at ceiling. Thus, these measures might not have been sensitive enough to capture the effects of the intervention on children's social preferences.

A second, more recent experiment in the United States (Hilliard & Liben, 2010) also manipulated the salience of gender categories for children—this time in a preschool setting. Over the course of 2 weeks, 3- to 5-year-old participants in a "high-salience" condition were exposed to classroom conditions similar to those implemented in Bigler's (1995) experimental classrooms. Participants in "low-salience" classrooms experienced their usual classroom conditions, which were guided by school-level policies discouraging the use of gendered language or gendered classroom organization. From pretest to posttest, children in the high-salience condition showed increased gender stereotyping, less positive ratings of other-gender peers, and reduced play with other-gender peers, whereas children in the low-salience condition showed none of these effects. These findings, together with the gender stereotyping effects reported by Bigler (1995), are consistent with developmental intergroup theory (Bigler & Liben, 2007), which posits that when adults highlight a social distinction (e.g., through labeling and function use), it causes children to treat the highlighted distinction as a psychologically meaningful one.

Teachers in the experimental classrooms of Bigler's (1995) and Hilliard and Liben's (2010) research were asked to go to great lengths to make gender salient. Furthermore, children in the experimental conditions experienced a dramatic and sudden change of classroom practices. Although the condition differences in both studies show that extreme gender-highlighting practices intensify children's gender attitudes and stereotypes, the research cannot tell us whether subtler long-term differences in pedagogy—of the sort that distinguish practices in typical classrooms from practices in gender-neutral programs—affect how young children view gender. The answer to this question not only addresses the potency of gender socialization effects but also contributes to our understanding of whether promoting gender-neutral preschool environments is a meaningful practice during early childhood.

The current research

Cultural context and school settings

The current research focused on 3- to 6-year-old children in Sweden, a society with relatively egalitarian gender attitudes. Across the past 5 years (2011–2016), the World Economic Forum's (2015, 2016) gender gap index, for example, has consistently rated Sweden as the fourth most gender-equal society in the world based on the areas of economics, politics, education, and health. For comparison, the United States has ranged from 20th to 45th, with a median rank of 23rd. Participants came from two kinds of preschool settings, with some participants attending a preschool with several specific school policies and practices aimed at actively creating a gender-neutral environment and other participants attending more typical Swedish preschools. Throughout this article, we refer to the former schooling environment as "gender neutral" (GN) and the latter schooling environment as "typical" for ease of exposition. It is important to note, however, that all Swedish schools are required by law to aim for gender equality in classrooms; thus, the difference in gender neutrality between the

GN and typical settings in the current research was a matter of degree rather than absolute. For reviews of the almost entirely qualitative and sociological literature discussing issues of gender in Swedish schools, see [Edström \(2009\)](#), [Heikkilä \(2017\)](#), [Wahlström \(2003\)](#), and [Wernersson \(2009\)](#).

The government-mandated national curriculum for preschools in Sweden asserts that classroom practices affect preschoolers' understanding of gender and prescribes that "girls and boys in preschools should have the same opportunities to test and develop abilities and interests outside the limitations of stereotyped gender roles" and that "preschools should counteract traditional gender roles and gender patterns" ([Skolverket, 2011](#), p. 5). In practice, however, there is a great deal of variation in the implementation of government guidelines across preschools ([Eidevald, 2009](#)); even preschool teachers who intend to treat boys and girls in the same way are not always successful because treating boys and girls differently can be an unconscious habit ([Odenbring, 2010](#); [Wahlström, 2003](#)). This is one reason why a small number of Swedish preschools have taken a further step of undergoing certification offered by the Swedish Federation for Lesbian, Gay, Bisexual, and Transgender Rights (RFSL). To achieve RFSL certification, all personnel undergo comprehensive training (certification takes 6–8 months) in treating all individuals according to their individual requirements regardless of gender or sexual orientation.

In the current study, the GN preschool, but not the typical preschools, pursued and attained RFSL certification, which therefore can be taken as an independent confirmation of differences in gender pedagogy between the schools. The specific policies of the GN school included avoiding gendered language as much as possible (e.g., by using the recently adopted Swedish gender-neutral pronoun and not saying "boy" or "girl"), modifying stories and songs to counteract rather than reinforce traditional gender roles and family structures, and avoiding some behaviors traditionally directed at one gender (e.g., commenting on the attractiveness of girls' clothes). The Appendix presents information about the self-reported gender practices (and beliefs) of teachers at the GN and typical preschools in the current research.

Overview of measures and hypotheses

We tested for potential effects of gender-neutral pedagogy by presenting children at the GN and typical schools with a series of measures. Participants first completed a task designed to assess their automatic encoding of other people's gender. Participants then completed tasks meant to probe their gender attitudes and gender stereotyping. All tasks were chosen because of past research demonstrating their utility in measuring children's emerging consideration of gender during the preschool years (e.g., [LaFreniere, Strayer, & Gauthier, 1984](#); [Leinbach, Hort, & Fagot, 1997](#); [Shutts et al., 2013](#); [Weisman, Johnson, & Shutts, 2015](#)), thereby allowing us to explore possible decreased consideration of gender among children in the GN school. At the end of the session, we also tested children's ability to identify the gender of target stimuli from the tasks. The purpose of the identification task was to validate the stimuli and assess whether there were any differences between GN and typical school children's ability to identify other people's gender (which could be a confound in the study). Because our primary comparison of interest was across groups (i.e., typical vs. GN schooling), we chose to use a fixed task order so as not to increase within-group variance, which would reduce our power to detect between-group effects. The order of tasks—except for the position of the identification task—was arbitrary. The identification task was presented at the end of the session so that our use of gender labels would not cause any children to consider gender on the other tasks when they otherwise might not have done so.

Consistent with the tenets of developmental intergroup theory ([Bigler & Liben, 2007](#)), we hypothesized that children attending the GN school—where teachers deemphasize the importance of gender—would view gender as a less psychologically meaningful distinction than children in typical schools. Accordingly, we predicted that children at the GN school would be less likely to automatically take note of an individual's gender (encoding task), less likely to use gender when deciding whether to play with another individual (social partner preference tasks), and less likely to use gender when reasoning about another person's attributes (stereotyping task). Put another way, because adults in their environment do not emphasize that gender is a relevant dimension to consider, we predicted that children at the GN school would be less likely to consider other people's gender. Yet another reason to posit that children at the GN school would be less likely to use gender to reason about another

person's attributes (stereotyping task) is that teachers at the GN school report working to actively counter gender stereotypes (see the Appendix); thus, children at the GN school likely have less access to information about gender stereotypes (which should result in lower scores on the stereotyping task).

Method

Participants

Participants were recruited from four preschools (one GN and three typical) located in the central district of a large Swedish city. All the typical schools were within 1.4 km of the GN school. Participants came from three different classrooms at the GN school and from five different classrooms at the typical schools. All families with children in available classrooms (determined by the schools) received a letter inviting them to participate; parents who were interested gave written informed consent. There were 30 participants from the GN school and 50 participants from typical schools. An additional 3 children were excluded from participation because of insufficient Swedish language skills or developmental disorders.

Parents provided information about their children's age and gender on the consent form. Participants from the GN and typical schools were closely matched on these dimensions (see Table 1). The majority of families (69%) also returned an optional questionnaire that included items focused on family characteristics (e.g., parental age and education level). As displayed in Table 1, participants in the GN and typical schools came from families with similar characteristics. The questionnaire also

Table 1
Comparison of sample characteristics between the typical schools and the GN school.

	Typical schools	GN school	Comparison	<i>p</i>
<i>Children</i>				
<i>n</i>	50	30		
% Girls	62.0	63.3	$\chi^2(1) = 0.014$.905
<i>M</i> _{age} (months)	59.0	56.6	$t(53) = 0.914$.365
Girls	60.3	59.2	$t(48) = 0.335$.739
Boys	56.9	52.0	$t(28) = 1.191$.244
<i>SD</i> _{age}	10.1	12.1	$F(49, 29) = 0.693$.252
Girls	10.0	11.7	$F(30, 18) = 0.732$.437
Boys	10.1	12.0	$F(18, 10) = 0.716$.516
Minimum age	38.6	36.3		
Girls	38.6	36.3		
Boys	40.8	38.1		
Maximum age	73.4	74.3		
Girls	73.4	74.3		
Boys	70.6	73.2		
<i>Parents</i>				
Questionnaire return rate	66.0	70.0	$\chi^2 = 0.015$.902
% Same-gender parents	0	0		
Mother age	38.3	39.2	$t(49) = 0.731$.468
Father age	39.6	42.1	$t(47) = 1.417$.163
% Mothers with university degree	75.8	72.2	$t(49) = 0.272$.787
% Fathers with university degree	75.0	76.5	$t(47) = 1.176$.245
Mothers' time spent on parental leave (months)	14.7	11.9	$t(47) = 1.250$.218
Fathers' time spent on parental leave (months)	7.8	6.2	$t(45) = 0.656$.515
Mothers' current working hours (% of full-time)	79.9	89.7	$t(48) = 1.080$.286
Fathers' current working hours (% of full-time)	92.6	96.6	$t(45) = 0.623$.536
% Mothers immigrated	11.1	9.4	$t(48) = 0.193$.848
% Fathers immigrated	23.5	25.0	$t(47) = 0.112$.912

Note. Some parents neglected to provide information about one or more of the items listed in the "Parents" section of the table. GN, gender-neutral.

asked parents to indicate why they had chosen their child's school in order to determine whether parents of children attending the GN school were specifically attracted to gender-neutral pedagogy. The question was, "Why did you choose the school that your child attends?" and parents could select as many options as they wanted from the following list: (a) close to home, (b) recommended by others, (c) gender-neutral pedagogy, (d) another sibling already attends, (e) assigned to the school, and (f) other reasons. The two most common responses given by parents at both the GN school and typical schools were (a) and (b). Five families from the GN school and one family from a typical school chose (c) as one of their responses.

On average, children at the GN school had been attending for 2.2 years ($SD = 1.4$, range = 0.5–5.2), and years attending was highly correlated with age, Pearson's $r = .731$, $t(28) = 5.665$, $p < .001$. We did not collect data on when participants began attending the typical schools, but there is no reason to believe it would be significantly different from the GN school. All four schools accept children from 12 months of age; furthermore, 84% of 1- to 5-year-olds and 95% of 3- to 5-year-olds in Sweden attend preschool (Skolverket, 2014).

Materials and procedure

One of two female experimenters tested participants individually in a quiet room at their school. The tasks were presented to children in the order they are described below. The following was true of stimuli in all the tasks except for the familiar playmate preference measure: The tasks included color photographs of real children who were not familiar to participants; the photographs were obtained from internet image searches and stock photography websites (and are available from the authors on request); the children in the photographs were visible from the shoulders up and were wearing gender-neutral clothing (gray or white T-shirts); and photographs within a pair were matched for attractiveness, ethnicity (all were White), and expression (all were positive). Gender was marked by hair length (short for boys and long for girls) in all four tasks (except for the androgynous photographs; see below), but it was further marked by the provision of gender-informative proper names in the unfamiliar playmate, stereotyping, and identification tasks. Swedish wording for all measures is available from the authors on request.

Gender encoding

To measure children's automatic encoding of gender, we presented participants with a memory confusion task (Bennett & Sani, 2003; Cosmides et al., 2003; Taylor, Fiske, Etcoff, & Ruderman, 1978; Weisman et al., 2015). This task has previously shown that automatic encoding of gender is evident by at least 4 years of age in U.S. preschoolers (Weisman et al., 2015). In the task, participants saw four different target children (two boys and two girls) and 16 different animal pictures. Each of four blocks in the task consisted of a familiarization phase followed by a test phase. During familiarization, participants saw four different child–animal pairings one at a time. For each pairing (8 s total presentation time), participants saw a target child, heard what animal the child had seen (e.g., "This kid saw the horse"), and saw a picture of the animal. At test, participants were asked to match the four animals' pictures with the four targets by placing the pictures next to each other. The order in which the target photos were presented and the pairing of targets with animals were counterbalanced across participants.

For each participant, we counted the number of within-category errors (i.e., errors where the participant paired an animal with a child who matched the gender of the correct target) and the number of between-category errors (i.e., errors where the participant paired an animal with a child who did not match the gender of the correct target). We then divided the total number of between-category errors by 2 (to account for the fact that between-category errors are twice as likely to occur by chance as within-category errors in this task; see Taylor et al., 1978; Weisman et al., 2015), and subtracted the resulting number from the total number of within-category errors. This procedure generated a difference score indicating how much more likely children were to make a mistake within a gender than across genders. Scores significantly above zero indicate gender encoding.

Social partner preferences: Unfamiliar playmates

One social preference task focused on children's preferences for unfamiliar children following on a method that has previously shown that same-gender social preferences are apparent in children as young as 3 years in the United States (see [Abel & Sahinkaya, 1962](#); [Shutts et al., 2013](#)). Participants saw 14 unique pairs of photographs. Of these, 8 pairs featured a boy and a girl, 1 pair featured two boys, and 1 pair featured two girls. An additional 4 pairs included an androgynous boy or girl (with medium-length hair but a traditional gendered name) alongside a child of the same gender with typical hair length and a traditional gendered name.

On each of the 14 trials, the experimenter presented the two photographs, provided a name for each child, and asked the participant, "Which child would you like to play with? You can choose [for example] Miranda, Dennis, or both of them." The different pair types were interspersed throughout the task and presented in the same order to all participants. The boy appeared on the left for half of the boy–girl trials, and the androgynous child appeared on the left for half of the androgynous trials.

Pairs with two children of the same gender were included at the request of GN teachers (to minimize the implication that gender is the most relevant feature guiding playmate selection); participants' playmate selections on these trials did not contribute to the task score and were not analyzed. Pairs with an androgynous child were also included at the request of GN teachers (so as not to imply that boys and girls must have short and long hair, respectively). We had planned to test whether children in the different school settings might differ in their willingness to befriend androgynous children, but participants' performance on the gender identification task (described later in this section) revealed the children had difficulty in identifying androgynous children's gender. For this reason, children's playmate choices for the four pairs that included an androgynous target were not considered in the task score or analyses.

For the boy–girl trials, each same-gender choice was scored as 1, each other-gender choice was scored as 0, and each response of "both" was scored as 0.50. Individual trial scores were summed and divided by the total number of trials completed to generate a task score for each participant. Scores above 0.50 indicate gender in-group favoritism, and scores below 0.50 indicate gender out-group favoritism.

Social partner preferences: Familiar playmates

A second social preference task assessed participants' preferred play partners at school. Naturalistic observations of children indicate that same-gender play partner preferences emerge at around 2 years of age for girls and 3 years of age for boys in Canada ([LaFreniere et al., 1984](#)). In addition, children have also been shown to be able to reliably and validly nominate their preferred playmates at 3 to 5 years of age (see [Gershman & Hayes, 1983](#); [Peterson & Siegal, 2002](#)). In the current task, the experimenter presented a photograph of the participant's class and asked, "Are there any children here you usually play with?" The experimenter continued to prompt if necessary until the participant identified three children. We recorded the number of same-gender playmates choices (out of the three).

Stereotyping

The stereotyping task measured participants' thoughts about other children's behavioral propensities—in particular, the degree to which participants' beliefs matched cultural norms about what boys and girls like to do. Previous research indicates that such stereotypes are evident by around 3 years of age in U.S. children (see [Baker, Tisak, & Tisak, 2016](#); [Blakemore, 2003](#); [Kuhn et al., 1978](#); [Leinbach et al., 1997](#); [Malcolm, Defeyter, & Friedman, 2014](#); [Martin & Little, 1990](#)). In the current task, participants saw eight different trials (two with toy cars, two with toy dolls, two with a child's pair of jeans and a masculine collared shirt, and two with a child's dress). At the start of each trial, the experimenter showed participants a toy or outfit on a computer screen and provided a description (e.g., "This is a toy car"). After this, a photograph of a boy and a photograph of a girl appeared on the screen (equidistant from the toy or outfit). The experimenter said, "This is [for example] Tim, and this is Ellie. Which child do you think most wants to [for example] play with the car—Tim, Ellie, or both equally?" Participants saw one of four pseudorandomized orders that differed in object order (but not photograph pair order). The photograph of the girl appeared on the left for half of the trials within the task. Responses in line with cultural stereotypes (i.e., girls like dolls and dresses, boys like trucks and jeans/shirts)

were scored as 1, and responses counter to cultural stereotypes (i.e., girls like trucks and jeans/shirts, boys like dolls and dresses) were scored as 0. Responses of “both” were scored as 0.50 because they were neither in line with nor counter to cultural stereotypes. Individual trial scores were summed and divided by the total number of trials completed to generate a task score for each participant. Task scores above 0.50 indicate a pattern in line with cultural stereotypes, and scores below 0.50 indicate a pattern counter to cultural stereotypes.

Gender identification

Each participant saw 10 pairs of photographs taken from the unfamiliar playmate preference task and stereotyping task. All participants saw 4 boy–girl pairs, 4 pairs with an androgynous child alongside a child of the same gender (with traditional hair), 1 boy–boy pair, and 1 girl–girl pair. On each trial, the experimenter presented the pair of photographs and accompanying names and asked, for each photograph in the pair, “Do you think the child is called a girl or a boy?” The different pair types were interspersed throughout the task and were presented to participants in one of four pseudorandomized orders. Across the orders, we varied which boy–girl pairs were presented in order to ensure that all pairs used in the tasks were seen by some participants. The boy appeared on the left for half of the boy–girl trials, and the androgynous child appeared on the left for half of the androgynous trials. Participants received a score of 1 for each correct response and a score of 0 for each incorrect response. Individual trial scores were summed and divided by the total number of trials completed to generate a task score for each participant. For androgynous faces, “correct” was defined as providing a gender label that matched the gender of the target child’s proper name. However, participants had difficulty in identifying the gender of androgynous children, who had medium-length hair and gender-informative proper names ($M = .712$, $SD = .281$). For this reason, we excluded trials featuring androgynous children from all further analyses.

Results

Data analysis and presentation strategy

The majority of analyses probing for effects of school types were carried out using generalized linear regression models, which analyze data in the same way as analysis of variance (ANOVA). We used the factors of school type, gender, age, and their interactions. The interactions were included to examine whether GN pedagogy affects girls and boys or children of different ages differently. For each task, we also present descriptive information about performance by children at the GN and typical schools (see Table 2) and statistics comparing children’s performance with chance. We provide effect sizes as well as 95% confidence intervals (CIs) for these effect sizes wherever appropriate.

For each task, analyses focused on the full sample of GN and typical school children are presented first, followed by analyses focused on a more restricted sample. The restricted sample analyses address two potential sources of bias. First, as noted earlier, five families from the GN school (and one family from a typical school) selected gender-neutral education as a reason for choosing their children’s school. One possibility is that parents who desire gender-neutral education for their children differ from parents who do not have that desire (e.g., such parents may practice more gender-neutral rearing practices at home). Second, some children in our sample were not very successful in identifying others’ gender—and this ability is a prerequisite for gender-based responding in the other tasks. For example, to express a preference for same-gender individuals in the unfamiliar playmates task, one must be able to identify which individuals are girls and which are boys.

To address both confounds and probe the robustness of our school type effects, we reran all analyses for tasks where we found significant or marginally significant school type difference in the analyses focused on the full sample. For these new analyses, we excluded participants whose parents selected gender-neutral education as a reason for their school choice ($n = 1$ from a typical school and $n = 5$ from the GN school) as well as participants who scored less than 75% on the gender identification task ($n = 3$ from typical schools and $n = 6$ from the GN school). These exclusions resulted

Table 2
Descriptive statistics for all variables.

	Typical schools			GN school			Both school types		
	Girls	Boys	All	Girls	Boys	All	Girls	Boys	All
Encoding score	1.32 (2.23)	0.82 (2.13)	1.13 (2.18)	2.26 (2.44)	2.00 (2.72)	2.17 (2.50)	1.68 (2.33)	1.25 (2.39)	1.52 (2.35)
Unfamiliar playmate score	0.88 (0.15)	0.67 (0.27)	0.80 (0.23)	0.80 (0.17)	0.67 (0.20)	0.75 (0.19)	0.85 (0.16)	0.67 (0.24)	0.78 (0.21)
Unfamiliar playmate (# and % choosing all same-gender playmates)	14/31 45%	5/18 28%	19/49 39%	5/19 26%	0/11 0%	5/30 17%	19/30 63%	5/49 10%	24/79 30%
Familiar playmate (# of same-gender choices)	2.34 <i>Mdn</i> = 3	2.26 <i>Mdn</i> = 3	2.31 <i>Mdn</i> = 3	2.67 <i>Mdn</i> = 3	2.00 <i>Mdn</i> = 2	2.41 <i>Mdn</i> = 3	2.47 <i>Mdn</i> = 3	2.17 <i>Mdn</i> = 2	2.35 <i>Mdn</i> = 3
Stereotyping score	0.82 (0.17)	0.76 (0.21)	0.80 (0.18)	0.76 (0.17)	0.55 (0.15)	0.68 (0.19)	0.80 (0.17)	0.68 (0.22)	0.75 (0.19)
Identification score	0.95 (0.10)	0.91 (0.16)	0.93 (0.13)	0.88 (0.17)	0.82 (0.21)	0.86 (0.19)	0.92 (0.14)	0.88 (0.18)	0.91 (0.15)

Note. Values are based on the full sample. Means are listed first, followed by standard deviations (in parentheses), with two exceptions: (a) “Unfamiliar playmate (# choosing all same-gender choices)” lists frequencies; (b) “Familiar playmate” choices lists means and medians (*Mdn*) due to nonparametric statistics used in the analyses. Possible scores for the encoding task range from –8 to 16. Possible scores for the unfamiliar playmate task range from 0 to 1. Frequency of choosing all same-gender choices for the unfamiliar playmate task is given out of the number of possible children in that group. Possible numbers of same-gender familiar playmates range from 0 to 3. Possible scores for the stereotyping task range from 0 to 1. Possible scores for the gender identification task range from 0 to 1. GN, gender-neutral.

in a sample size of 66 for restricted analyses ($n = 46$ from typical schools and $n = 20$ from the GN school; note that 1 participant met both exclusion criteria).

Gender encoding

A generalized linear regression model with the predictors of school type, gender, age, and their interactions revealed that there was a marginally significant difference in the strength of automatic encoding across school types, $t(79) = 1.946$, $p = .055$, $\eta_p^2 = .046$, CI [0.000, 0.161], but it was in the opposite direction from our prediction, with children at the GN school scoring somewhat higher than children at the typical schools. There was no effect of participant gender or age, nor were there interactions between variables. When analyzed separately, children at both school types were above chance in their encoding of gender (typical: $t(49) = 3.661$, $p < .001$, Cohen's $d = 0.518$, CI [0.241, 0.795]; GN: $t(29) = 4.742$, $p < .001$, Cohen's $d = 0.866$, CI [0.508, 1.224]). When we repeated the regression with only the restricted sample, the marginal effect of school type on gender encoding remained, $t(64) = 1.992$, $p = .051$, $\eta_p^2 = .058$, CI [0.000, 0.193].

Social partner preferences

Unfamiliar playmates

One boy from a typical school refused to participate in this task, resulting in a sample size of 49 children at typical schools and 29 boys total. The average number of choices of same-gender unfamiliar playmates did not differ by school type, $t(78) = 1.138$, $p = .259$, $\eta_p^2 = .014$, CI [0.000, 0.109], or age, $t(78) = 0.846$, $p = .400$, $\eta_p^2 = .008$, CI [0.000, 0.091], according to a generalized linear regression model with the predictors of school type, gender, age, and their interactions. However, there was a significant effect of participant gender, with girls choosing more same-gender unfamiliar playmates than boys overall, $t(78) = 3.933$, $p < .001$, $\eta_p^2 = .167$, CI [0.042, 0.311]. Both girls and boys chose same-gender playmates at above-chance levels (girls: $t(49) = 14.935$, $p < .001$, Cohen's $d = 2.112$, CI [1.835, 2.839]; boys: $t(28) = 3.730$, $p < .001$, Cohen's $d = 0.693$, CI [0.329, 1.057]).

Given that the data overall were skewed, with 33% of children choosing same-gender playmates on every trial (i.e., a score of 1), we also compared the likelihood of this response pattern using binomial logistic regression with the same predictors as above. This analysis revealed that children from typical schools were more likely to choose no other-gender playmates than were children from the GN school, $z(75) = 1.998$, $p = .046$, odds ratio = 3.311, CI [2.945, 3.721], and choosing no other-gender playmates was also marginally more likely with increased age in months, $z(75) = 1.724$, $p = .085$, odds ratio = 1.048, CI [0.993, 1.106]. With the restricted sample, the effect of school type remained, $z(64) = 2.152$, $p = .031$, odds ratio = 4.759, CI [1.149, 19.708], as did the marginal effect of age on the rate of choosing all same-gender playmates, $z(64) = 1.921$, $p = .055$, odds ratio = 1.058, CI [0.999, 1.120].

Familiar playmates

The number of familiar playmates of the same gender (of total possible 3) chosen by participants did not vary between schools ($W = 664.5$, $p = .718$; Wilcoxon rank-sum test used because data came from an ordinal 0–3 scale with extreme skew). Children's average proportion of selected same-gender playmates was greater than chance (.50) according to a Wilcoxon signed-ranks test ($V = 2772$, $p < .001$). The lower power of nonparametric statistics precludes useful analyses of samples further divided by age and gender.

Stereotyping

A generalized linear regression model ($N = 79$ because 1 boy at a typical school refused to complete the task) with the predictors of school type, gender, age, and their interactions revealed effects of school type, gender, and age, such that children at the typical school gave more stereotyped answers than children at the GN school, $t(75) = 2.702$, $p = .009$, $\eta_p^2 = .089$, CI [0.006, 0.219], girls did so more than boys, $t(75) = 2.286$, $p = .025$, $\eta_p^2 = .065$, CI [0.000, 0.188]), and older children did so more than younger children, $t(75) = 4.580$, $p < .001$, $\eta_p^2 = .219$, CI [0.072, 0.339] (see Fig. 1A). No interactions were significant. Tests against chance revealed above-chance stereotyping scores for all groups (typical: $t(48) = 11.406$, $p < .001$, Cohen's $d = 1.629$, CI [1.349, 1.909]; GN: $t(29) = 5.170$, $p < .001$, Cohen's $d = 0.944$, CI [0.586, 1.302]; girls: $t(49) = 12.565$, $p < .001$, Cohen's $d = 1.777$, CI [1.500, 2.054]; boys: $t(28) = 4.453$, $p < .001$, Cohen's $d = 0.827$, CI [0.463, 1.191]; children above the median age: $t(38) = 13.253$, $p < .001$, Cohen's $d = 2.122$, CI [1.808, 2.436]; children below the median age: $t(37) = 5.244$, $p < .001$, Cohen's $d = 0.851$, CI [0.533, 1.169]).

For the restricted sample ($N = 65$), the analysis revealed an interaction between school type and gender, $t(64) = 2.810$, $p = .007$, $\eta_p^2 = .121$, CI [0.037, 0.416]. Thus, we separated the sample by gender to examine the effects of school type and age; girls ($n = 42$) showed only an age effect, $t(40) = 3.388$, $p = .002$, $\eta_p^2 = .223$, CI [0.035, 0.412]), whereas for boys ($n = 23$) there was a further interaction between school type and age, $t(19) = 2.102$, $p = .049$, $\eta_p^2 = .189$, CI [0.000, 0.428]. Boys were then divided at median age, and we found that older boys ($n = 11$) showed a school type effect, $t(9) = 6.794$, $p < .001$, $\eta_p^2 = .837$, CI [0.465, 0.908], with greater stereotyping at the typical schools than at the GN school, whereas younger boys ($n = 12$) showed no school type effect, $t(10) = 1.466$, $p = .174$, $\eta_p^2 = .177$, CI [0.000, 0.504]. Single-sample t tests showed that only younger boys failed to give stereotyped responses at above-chance levels, $t(11) = 1.650$, $p = .127$, Cohen's $d = 0.476$, CI [-0.089, 1.042]. Girls overall, $t(41) = 12.866$, $p < .001$, Cohen's $d = 1.985$, CI [1.683, 2.288], as well as older boys at both typical schools ($n = 8$), $t(7) = 15.601$, $p < .001$, Cohen's $d = 5.516$, CI [4.823, 6.209], and the GN school ($n = 3$), $t(2) = 5.000$, $p = .038$, Cohen's $d = 2.887$, CI [1.755, 4.018], did stereotype above chance. To summarize, the restricted sample analysis confirmed, but also qualified, the effect of school type on the stereotyping task. More specifically, for the restricted sample, we did not observe a main effect of school type; rather, only older boys showed a significant effect of school type (see Fig. 1B).

Gender identification

As noted in the Introduction, the main purpose of the gender identification task was to confirm that participants could identify the gender of target stimuli. Four participants refused to complete this task (2 girls and 2 boys; 3 from typical schools and 1 from the GN school), resulting in a sample comprising

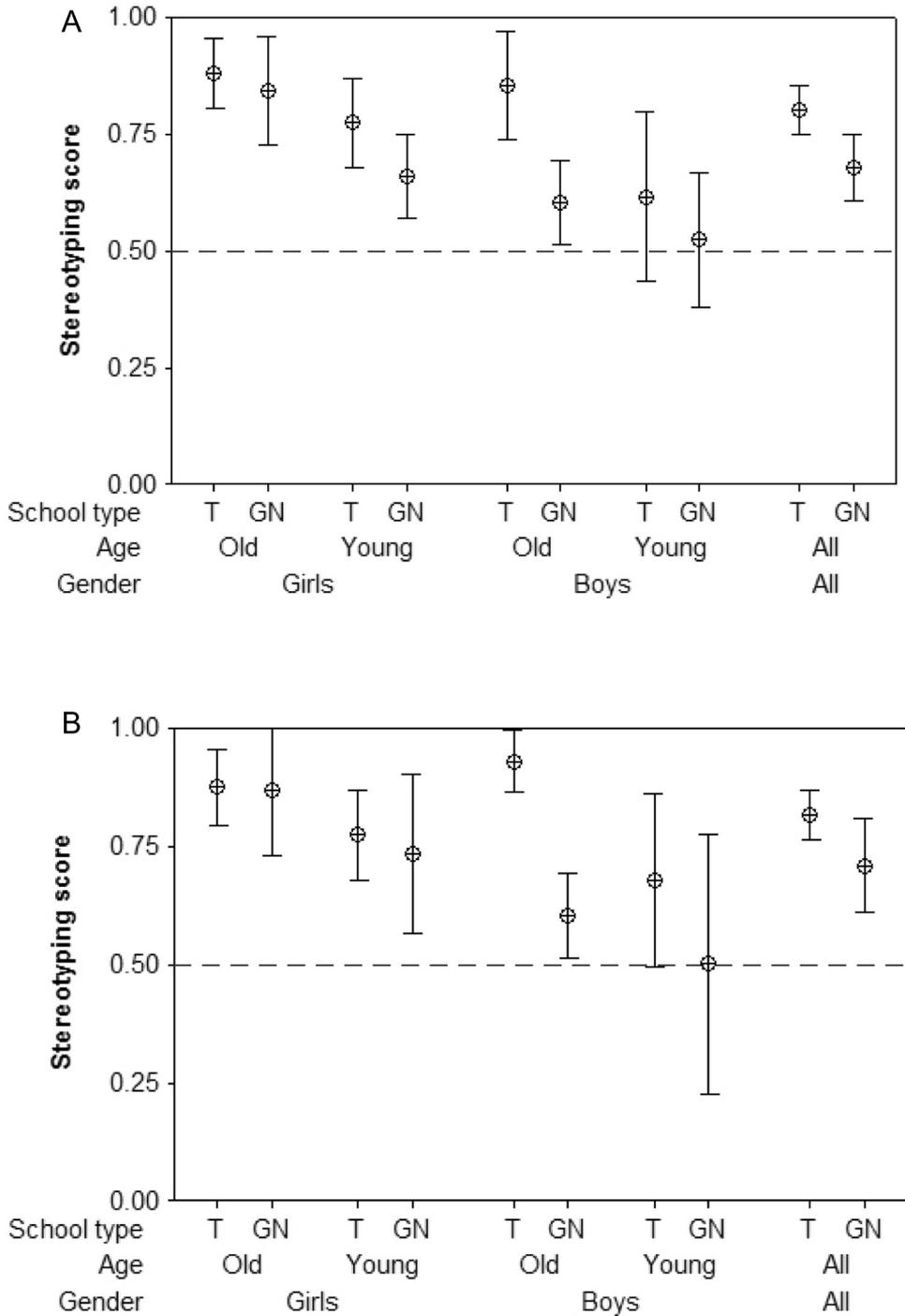


Fig. 1. Means and 95% confidence intervals for stereotyping scores from the full sample ($N = 79$) (A) and the restricted sample ($N = 65$) (B) separated by preschool type (T, typical; GN, gender neutral), age (median split), and gender. Scores greater than 0.50 indicate responses in line with cultural stereotypes.

48 girls, 28 boys, 47 typical school children, and 29 GN school children. As a group, participants performed well when asked to identify the gender of children with traditional hair (short for boys and long for girls) and gender-informative proper names (see Table 2). In a generalized linear regression model examining the factors of gender, school type, age, and their interactions, there was a significant effect of age, $t(73) = 2.593$, $p = .012$, $\eta_p^2 = .084$, CI [0.004, 0.217], with correct identifications increasing with age. There was also a marginally significant effect of school type, $t(73) = 1.695$, $p = .094$, $\eta_p^2 = .038$, CI [0.000, 0.150], with children in typical schools making marginally more correct identifications than those in the GN school. In the restricted sample, there was no longer an effect of age or school type ($ps > .10$) on the gender identification task, as would be expected given that children with the lowest scores were excluded.

Discussion

The findings from the current study provide some support for the hypothesis that differences in gender pedagogy are associated with differences in children's consideration of gender; gender-neutral pedagogy was associated both with a greater interest in playing with unfamiliar children of another gender as well as a reduced tendency to assume that unfamiliar boys' and girls' characteristics would align with cultural beliefs. These findings support the proposition that adults' behaviors affect the degree to which children use gender to guide their consideration of other children. Other tasks and analyses, however, provided no support for the hypothesis that gender-neutral pedagogy reduces children's consideration of gender. Children in the GN school did not show less robust automatic encoding of other people's gender; in fact, contrary to our hypothesis, the encoding scores of children at the GN school were somewhat higher than those of children at typical schools. However, in considering the results from the encoding task, it is worth noting the absence of theory anticipating this result as well as the marginal significance of the effect (the p value was above .05, and the confidence interval for the effect size included zero). Also contrary to our hypothesis, children in both kinds of schools were equally likely to choose same-gender peers as preferred playmates in the familiar playmates task. Finally, analyses of the restricted sample indicated that the effect of school type in the gender stereotyping task was significant only for older boys. We discuss each of these findings in turn below.

First, why did we find that exposure to gender-neutral pedagogy was associated with moderate reductions in gender in-group favoritism and application of stereotypes? One possibility is that when adults deemphasize gender groupings (e.g., "Let's have all children play together at the table" rather than "Let's have all the girls play together in the dress-up area"), children truly come to believe that gender is a less reliable basis for making judgments about other people. Such beliefs could in turn make children more open to having play partners of different genders and more likely to consider the possibility that other children's preferences and actions do not align with cultural stereotypes about boys and girls. A second possibility—which is not incompatible with the first—is that when teachers specifically work to counteract gender stereotypes in their classrooms (e.g., by modifying songs so that they do not contain gender norms and by selecting books that feature gender counter stereotypes; see the Appendix), children are less aware of gender-based cultural stereotypes. This decreased awareness could lead to a reduction in gender-based assumptions about others' toy and clothing preferences as well as a greater willingness to play with other-gender children (because children might reason that other-gender children may very well share their own toy and clothing preferences). The current research cannot distinguish between these two possibilities (see below for additional discussion).

By contrast, gender-neutralizing practices might not be robust enough to overcome a well-practiced tendency to simply take note of other people's gender. Unlike gender in-group favoritism and gender stereotyping, which appear to emerge and strengthen during early childhood (Ruble et al., 2006), children's ability to tell the difference between male and female individuals is evident during infancy (Quinn et al., 2011). Once gender categories are identified, gender encoding may be very difficult to modulate (e.g., via gender-neutral pedagogy). Indeed, studies of adults show that gender encoding, unlike race encoding, is hard—if not impossible—to suppress even when participants are given other ways to classify people (Kurzban, Tooby, & Cosmides, 2001).

Another finding that warrants discussion is the discrepancy with respect to the gender stereotyping results for the full versus restricted sample. One possibility is that the reduced sample size for the restricted analysis limited our ability to detect an overall effect of school type for the sample. Another (not incompatible) possibility is that gender pedagogy has a more pronounced effect on boys' stereotyping than on girls' stereotyping. Although older boys and girls tend to perform similarly on stereotype awareness measures (e.g., [Blakemore, 2003](#)), girls are more knowledgeable than boys about gender stereotypes very early in development. For example, one study showed that girls enacted gender-stereotypical actions with male and female dolls by 24 months of age, whereas boys did so reliably only by 31 months of age ([Poulin-Dubois, Serbin, Eichstedt, Sen, & Beissel, 2002](#)). In addition, when tested in a looking time paradigm, 18- and 24-month-old girls, but not boys, associated stereotypically masculine toys with boys' faces and associated stereotypically female toys with girls' faces ([Serbin, Poulin-Dubois, Colburne, Sen, & Eichstedt, 2001](#)). If girls are more attuned to information about gender stereotypes early in development, their beliefs may be more entrenched (and potentially less malleable) than those of boys. Nevertheless, we did not predict that participant gender would interact with school type. Further research is necessary to confirm (or disconfirm) the effect and explain it. Further research is also necessary to examine whether, and if so how, the gender of experimenters (who were female in the current study) might influence children's responses on the gender stereotyping measure as well as on other gender cognition tasks.

Finally, why is it that we detected a school type effect on the unfamiliar playmates task but not on the familiar playmates task? It seems likely that our familiar playmates measure was not extensive enough to detect a school difference given that we asked children to provide only three names. Regardless, it does appear that GN pedagogy is associated with a greater openness on the part of children when it comes to considering other-gender playmates they have just met.

Limitations and suggestions for future research

The differences we observed between children in the GN school and those in typical schools were not uniformly large across measures, and some measures revealed no differences at all. Nevertheless, it is worth mentioning three factors that could have lessened the effects of school setting (or our ability to detect them). First, children attending the GN school of course spent a considerable amount of time in other environments (e.g., their homes, informal playgroups, consuming media). Thus, it is very likely that children heard nouns and pronouns referring to gender and received exposure to gender stereotypes. Second, because Swedish law requires and emphasizes the importance of gender egalitarian practices in preschool, even children at the typical schools experienced some degree of gender neutrality in their schools. As shown in the Appendix, teachers at the GN school self-reported more rigorous gender-neutral practices, but the mean responses of teachers at typical schools did not suggest that they were prone to emphasizing gender in their classrooms. Although it is difficult to compare cultures and studies directly, it is plausible that the practices of teachers in typical Swedish schools are similar to practices of teachers in low-salience classrooms in previous studies conducted in the United States (e.g., [Bigler, 1995](#); [Hilliard & Liben, 2010](#)). Third, some lack of differences between schools could have been due to low statistical power associated with our fairly small sample sizes and wide age range. Moreover, age and time attending preschool were highly correlated, so the effect of length of attendance could not be assessed. We note that for practical reasons (difficulty of access to and rarity of GN preschools even in Sweden), increasing the sample size in this study would have been very challenging.

Although our study is unique in the psychological literature in its focus on how long-term gender-neutral pedagogy might affect children's consideration of gender, there are some methodological limitations to note. One limitation is that we assessed children's consideration of gender only in a short testing session conducted outside their classrooms. Although controlled testing conditions have some notable advantages (e.g., participants' responses cannot be influenced by those individuals around them, and all participants can experience exactly the same procedure), naturalistic observation measures (e.g., observing children's actual playmate preferences) or qualitative interviews may reveal other effects of gender-neutral pedagogy on children's thoughts and behaviors.

A second limitation is that the practices of teachers in the GN school and those in the typical schools differed in more than one way, and we were unable to conduct systematic observations of classroom practices. Thus, we cannot ascertain which practices (e.g., avoiding gender language, working to counteract gender stereotypes) may have generated the differences we observed between children in the GN school and those in the typical schools. The data from the teacher questionnaire do, however, provide some support for the idea that the classroom practices themselves—rather than transmission of teachers' general beliefs about gender—are associated with children's gender preferences and stereotyping. As shown in the Appendix, teachers in the GN school and those in the typical schools responded differently on just one of the five questions probing beliefs about gender, but they responded differently on four of the seven questions focused on classroom practices.

A third limitation is that we could not randomly assign children to their schools. The schools were located in close proximity to one another, participants at the two kinds of schools had similar family characteristics, and for the most part parents did not choose their children's school based on gender pedagogy. Nevertheless, because children were not randomly assigned to their schools, we cannot conclude with certainty that experiencing gender-neutral pedagogy caused the differences we observed between children in the two kinds of schools. Furthermore, although we collected information about family demographics, we did not have the opportunity to measure parents' gender-related beliefs or practices at home.

One way to address these limitations would be to conduct research that includes a more extensive parent questionnaire as well as a baseline assessment of children's consideration of gender. Because children generally begin preschool at 1 year of age in Sweden, such a study would have been very challenging to conduct in the current context because it would have required schools to commit to several years of onsite research. Furthermore, such research requires the development of measures of gender attitudes, stereotyping, and encoding that are (a) appropriate for children younger than 3 years, (b) reliable enough to reveal group differences, and (c) sufficiently comparable to measures that are commonly used to assess older children's gender categories.

Despite these challenges, a focus on younger children in future research will be important given that participants in the current research already showed significant levels of gender-based responding. Paralleling research conducted with children in a variety of settings in the United States (for reviews, see [Arthur et al., 2008](#), and [Ruble et al., 2006](#)), even participants in the Swedish GN school showed automatic gender encoding, tended to prefer same-gender playmates, and engaged in gender stereotyping. These findings suggest that processes or experiences that occur very early in development—or operate outside the classroom—may promote children's attention to gender in person perception and evaluation. Understanding the role of experience in guiding children's consideration of gender, therefore, may require testing infants and toddlers, including those raised in very different social environments. For example, examining correlations between parents' tendency to highlight gender in the home (e.g., by using gender labels; see [Gelman et al., 2004](#)) and children's very early consideration of gender could be useful. Additional cross-cultural research focused on young children in different cultures—for example, those raised in relatively egalitarian nations such as Sweden compared with those raised in countries with more traditional or conservative gender roles—could also shed light on how early socialization affects children's gender cognition.

Conclusions

Beyond contributing to theories of gender categorization, research focused on factors supporting young children's social preferences and stereotypes is of practical importance. Young children who favor same-gender playmates develop more extreme gender-typed interests and behaviors over time ([Martin & Fabes, 2001](#)). In addition, children become less interested in playing with particular toys once they believe that such toys are “for” the other gender ([Martin, Eisenbud, & Rose, 1995](#); see also [Bian et al., 2017](#)). These findings are of consequence given that gender-typed behaviors and interests are related to the development of skills associated with success in school ([Serbin, Zerkowitz, Doyle, Gold, & Wheaton, 1990](#)). For example, greater compliance with adult direction (a behavior more commonly demonstrated by young girls) is associated with better academic performance in elementary

school (Serbin et al., 1990), and playing with blocks (a toy stereotypically associated with boys) is associated with the development of better spatial abilities (e.g., Jirout & Newcombe, 2015). Thus, children's social partner choices and beliefs about objects are related to their social and cognitive development.

The current findings suggest that gender-neutralizing practices may reduce children's gender in-group favoritism and gender stereotyping. Thus, our findings support the hypothesis that the practices of teachers have meaningful effects on children's consideration of gender early in development. Given the consequences associated with gender segregation and stereotyping described in the previous paragraph, gender-neutral pedagogy appears to be a useful positive practice. We also note, however, that the presence of significant gender encoding, preferences, and stereotyping by young children in the GN school where we tested also indicates that other factors contribute to children's developing gender categories. This, together with the fact that most children spend significant periods of time outside their classrooms, suggests that policies and practices aimed at modifying society's consideration of gender will be important to facilitate and sustain reductions in gender attitudes and stereotypes over time.

Acknowledgment

This research was supported by grant 421-2011-1785 from the Swedish Research Council (Vetenskapsrådet).

Appendix A

We asked teachers at the GN school and the typical schools to complete a questionnaire with questions about their gender-relevant classroom behaviors ($n = 7$ questions) and beliefs about gender

Table A1

Self-reported classroom practices and beliefs of teachers in the typical preschools and those in the GN preschool.

	Typical schools		GN school		<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
<i>Classroom practices</i>						
I have used rhymes or songs that have been changed to avoid confirming gender norms [Swedish example included]	5.8	0.5	6.0	0.0	.037	0.80
When I talk to children, I avoid using words like "boy" and "girl," which specify gender	4.4	1.7	5.9	0.5	.003	1.16
I sometimes use gender-neutral personal pronouns such as "hen" [a recently invented Swedish word] and "it"	3.8	1.8	5.4	1.2	.011	1.04
Families in the stories I tell to children often have a mother and a father	3.7	1.5	4.0	1.4	.663	0.17
I specifically seek out books that have nonstereotypical gender roles and family structures	3.9	1.5	5.2	1.5	.026	0.91
I can imagine a situation in a preschool where it is appropriate to treat a girl differently than a boy	1.0	0.0	1.7	1.4	.099	0.65
I encourage children to play with toys and engage in activities that are against gender stereotypes	5.6	1.3	5.2	1.1	.565	0.26
<i>Gender beliefs</i>						
Differences between boys' and girls' behavior that depend on biology instead of culture are almost or totally nonexistent	5.3	1.2	3.7	2.0	.024	0.94
It's a big problem that children are exposed to too many normative stereotypes about gender	4.6	1.6	5.6	1.3	.071	0.72
Before the Swedish word "snippa" came into use, it was a problem that there was no neutral word for girls' genitalia	4.4	2.0	5.3	1.3	.139	0.56
It's good for children to be able to wear whatever clothes they want even if the clothes are intended for the other gender	6.0	0.0	6.0	0.0	1.00	0.00
It's only natural that boys tend to play more with cars and girls tend to play more with dolls and soft toys	1.5	0.9	1.8	1.2	.589	0.28

Note. GN, gender-neutral.

($n = 5$ questions). Teachers answered each question using a Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree). Table A1 presents all the items on the questionnaire, teachers' responses to all these items, and statistical tests comparing responses from teachers at the GN school ($n = 16$) and the typical schools ($n = 14$). We used the standard two-sample permutation test (Good, 2005) with 1 million samples to compare responses at the two kinds of schools because distributions for many items were too skewed for parametric analysis. According to teachers' responses, the majority of probed potential gender-neutral classroom practices were more frequent in the GN school than in the typical schools; none was more frequent in the typical schools. There was less clear evidence, however, that teachers in the GN school and those in the typical schools had different personal beliefs about gender.

References

- Abel, H., & Sahinkaya, R. (1962). Emergence of sex and race friendship preferences. *Child Development*, 33, 939–943.
- Albert, A. A., & Porter, J. R. (1986). Children's gender role stereotypes: A comparison of the United States and South Africa. *Journal of Cross-Cultural Psychology*, 17, 45–65.
- Andrews, N. C., Martin, C. L., Field, R. D., Cook, R. E., & Lee, J. (2016). Development of expectancies about own- and other-gender group interactions and their school-related consequences. *Child Development*, 87, 1423–1435.
- Arthur, A. E., Bigler, R. S., Liben, L. S., Gelman, S. A., & Ruble, D. N. (2008). Gender stereotyping and prejudice in young children: A developmental intergroup perspective. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 66–86). New York: Oxford University Press.
- Baker, E. R., Tisak, M. S., & Tisak, J. (2016). What can boys and girls do? Preschoolers' perspectives regarding gender roles across domains of behavior. *Social Psychology of Education*, 19, 23–29.
- Bauer, P. J., & Coyne, M. J. (1997). When the name says it all: Preschoolers' recognition and use of the gendered nature of common proper names. *Social Development*, 6, 271–291.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88, 354–364.
- Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs*, 8, 598–616.
- Bennett, M., & Sani, F. (2003). The role of target gender and race in children's encoding of category-neutral person information. *British Journal of Developmental Psychology*, 21, 99–112.
- Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355, 389–391.
- Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, 66, 1072–1087.
- 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.
- Blakemore, J. E. O. (2003). Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. *Sex Roles*, 48, 411–419.
- Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, 7, 173–179.
- de Guzman, M. R. T., Carlo, G., Ontai, L. L., Koller, S. H., & Knight, G. P. (2004). Gender and age differences in Brazilian children's friendship nominations and peer sociometric ratings. *Sex Roles*, 51, 217–225.
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science*, 19, 781–789.
- Edström, C. (2009). Preschool as an arena of gender policies: The examples of Sweden and Scotland. *European Educational Research Journal*, 8, 534–549.
- Eidevald, C. (2009). *Det finns inga tjejbästämmande: att förstå kön som position i förskolans vardagsrutiner och lek*. Doctoral dissertation, School of Education and Communication, Jönköping University. Retrieved from <<http://hj.diva-portal.org/smash/record.jsf?pid=diva2:158528>>.
- Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, 69(1, Serial No. 275).
- Gershman, E. S., & Hayes, D. S. (1983). Differential stability of reciprocal friendships and unilateral relationships among preschool children. *Merrill-Palmer Quarterly*, 29, 169–177.
- Good, P. (2005). *Permutation, parametric, and bootstrap tests of hypotheses* (3rd ed.). New York: Springer.
- Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., ... Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science*, 333, 1706–1707.
- Heikkilä, M. (2017). *Förskola och genus*. Retrieved from <<http://www.genus.se/kunskap-om-genus/fordjupning-skola/forskola/>>.
- Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development*, 81, 1787–1798.
- Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science*, 26, 302–310.
- Knobloch, S., Callison, C., Chen, L., Fritzsche, A., & Zillmann, D. (2005). Children's sex-stereotyped self-socialization through selective exposure to entertainment: Cross-cultural experiments in Germany, China, and the United States. *Journal of Communication*, 55, 122–138.
- Kuhn, D., Nash, S. C., & Bruckner, L. (1978). Sex role concepts of two- and three-year-olds. *Child Development*, 49, 445–451.
- Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 15387–15392.

- LaFreniere, P. J., Strayer, F. F., & Gauthier, R. (1984). The emergence of same-sex affiliative preferences in children's play groups: A developmental/ethological perspective. *Child Development*, 55, 1958–1965.
- Leinbach, M., Hort, B. E., & Fagot, B. I. (1997). Bears are for boys: Metaphorical associations in young children's gender stereotypes. *Cognitive Development*, 12, 107–130.
- Maccoby, E. E., & Jacklin, C. (1987). Gender segregation in childhood. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 239–287). San Diego: Academic Press.
- Malcolm, S., Defeyter, M. A., & Friedman, O. (2014). Children and adults use gender and age stereotypes in ownership judgments. *Journal of Cognition and Development*, 15, 123–135.
- Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, 25, 80–88.
- Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's gender-based reasoning about toys. *Child Development*, 66, 1453–1471.
- Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, 37, 431–446.
- Martin, C., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children's beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships*, 16, 751–771.
- Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, 52, 1119–1134.
- Martin, C. L., & Little, J. K. (1990). The relation of gender understanding to children's sex-typed preferences and gender stereotypes. *Child Development*, 61, 1427–1439.
- Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, 128, 903–933.
- Muthukrishna, N., & Sokoya, G. O. (2008). Gender differences in pretend play amongst school children in Durban, Kwazulu-Natal, South Africa. *Gender and Behaviour*, 6, 1577–1590.
- Odenbring, Y. (2010). *Kramar, kategoriseringar och hjälpfröknar: Könskonstruktioner i interaktion i förskola, förskoleklass och skolår ett*. Doctoral dissertation, University of Gothenburg. Retrieved from <<http://hdl.handle.net/2077/22144>>.
- Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology*, 20, 205–224.
- Poulin-Dubois, D., Serbin, L. A., Eichstedt, J. A., Sen, M. G., & Beissel, C. F. (2002). Men don't put on make-up: Toddlers' knowledge of the gender stereotyping of household activities. *Social Development*, 11, 166–181.
- Quinn, P. C., Anzures, G., Izard, C. E., Lee, K., Pascalis, O., Slater, A. M., & Tanaka, J. W. (2011). Looking across domains to understand infant representation of emotion. *Emotion Review*, 3, 197–206.
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, 31, 1109–1121.
- Renno, M. P., & Shutts, K. (2015). Children's social category-based giving and its correlates: Expectations and preferences. *Developmental Psychology*, 51, 533–543.
- Ruble, D. N., Martin, C., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology*, Vol. 3: *Social, emotional, and personality development* (6th ed., pp. 858–932). Hoboken, NJ: John Wiley.
- Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, 25, 7–15.
- Serbin, L. A., Zelkowitz, P., Doyle, A.-B., Gold, D., & Wheaton, B. (1990). The socialization of sex-differentiated skills and academic performance: A meditational model. *Sex Roles*, 23, 613–628.
- Shutts, K. (2013). Is gender “special”? In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: The early years* (pp. 297–300). New York: Oxford University Press.
- Shutts, K., Roben, C., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, 14, 35–62.
- Skolverket (2011). *Läroplan för förskolan* Retrieved from <http://www.skolverket.se/loroplaner-amnen-och-kurser/forskola>.
- Skolverket (2014). *Fler barn än någonsin i förskolan* Retrieved from <http://www.skolverket.se/statistik-och-utvardering/nyhetsarkiv/2.7608/Fler-barn-an-nagonsin-i-forskolan-1.170263>.
- Taylor, S. E., Fiske, S. T., Etoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36, 778–793.
- Wahlström, K. (2003). *Flickor, pojkar och pedagoger: Jämställdhetspedagogik i praktiken*. Stockholm, Sweden: Sveriges Utbildningsradio.
- Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, 40, 593–610.
- Weinraub, M., Clemens, L., Sockloff, A., Ethridge, T., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-typed toy preference, and family characteristics. *Child Development*, 55, 1493–1503.
- Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science*, 18, 1026–1043.
- Wernersson, I. (Ed.). (2009). *Genus i förskola och skola: Förändringar i policy, perspektiv och praktik*. Gothenburg, Sweden: Acta Universitatis Gothoburgensis.
- World Economic Forum (2015). *The Global Gender Gap Report 2015*. World Economic Forum: Geneva, Switzerland.
- World Economic Forum (2016). *The Global Gender Gap Report 2016*. World Economic Forum: Geneva, Switzerland.
- Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology*, 33, 183–196.
- 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.