

Consistency of Gender Identity and Preferences Across Time: An Exploration Among Cisgender and Transgender Children

Tabea Hässler^{1, 2}, Jessica J. Glazier³, and Kristina R. Olson⁴

¹ Department of Psychology, University of Zurich

² School of Psychology, University College Dublin

³ Department of Psychology, University of Washington

⁴ Department of Psychology, Princeton University

While considerable research has examined gender development in middle childhood, little longitudinal work has been conducted at this time to indicate whether, for example, youth who show more or less gender conformity at one point continue to do so later. The present study investigated the consistency of gender identity and preferences for gender-stereotypical toys, clothing, and same-gender peer preferences among groups of transgender youth ($n = 158$), their siblings ($n = 79$), and an unrelated group of cisgender youth ($n = 128$) from a mean age of 7.0 (range 3.0–10.9) to a mean age of 9.6 (range 5.1–12.0). Furthermore, 65.5% of the youth were girls, 69.7% were White, 72.8% grew up in households with an annual household income of \$75,000 or more, and 89.9% of parents had a bachelor's degree or higher. Overall, we found a small-to-medium correlation over this 2.6-year span within each group, both across the composite of measures and most measures individually. Despite the moderate stability over time, we found a decrease in the composite and individual scores over this time span for girls and for transgender participants. Together these results suggested some stability in children's gender identity and preferences in middle childhood and that this was true regardless of whether the child's gender did or did not align with their sex assignment at birth.

Keywords: gender stability, gender identity, children, cisgender, transgender

Supplemental materials: <https://doi.org/10.1037/dev0001419.supp>

Children can typically identify their gender by $2\frac{1}{2}$ to 3 years of age (Etaugh et al., 1989; Fagot et al., 1986; Thompson, 1975; Weinraub et al., 1984). Shortly thereafter, children often show preferences for toys and clothing stereotypically associated with their genders (Campbell et al., 2000; Davis & Hines, 2020; Halim et al., 2014; Jadva et al., 2010; Serbin et al., 2001) and prefer same-gender peers (Martin et al., 2011; McHale et al., 2004). Children's preference for gender stereotypical objects, activities, and same-gender peers are generally observed throughout childhood and adolescence, though the degree to which they show these

preferences varies across developmental stages (Kanka et al., 2019; Skinner & McHale, 2018; for a review see Halim et al., 2011). While most youth show these patterns, there is variation, wherein some youth show strongly gendered identity and preferences, others show weaker identity and preferences, and still others come to identify and/or have preferences more often associated with another gender (Golombok & Rust, 1993; Golombok et al., 2012; Gülgöz et al., 2019; Martin & Ruble, 2010).

Importantly, most investigations of gender preferences and identity in children have been cross-sectional, rather than longitudinal,

This article was published Online First August 11, 2022.

Tabea Hässler  <https://orcid.org/0000-0002-9371-2423>

Jessica J. Glazier  <https://orcid.org/0000-0001-7755-5651>

Primary idea was provided by Tabea Hässler and Kristina R. Olson. Conceptualization of the project, data collection, and data management were provided by Kristina R. Olson and team. Draft of preregistration was provided by Tabea Hässler and supported by Kristina R. Olson. Data preparation was provided by Jessica J. Glazier. Data analysis was provided by Tabea Hässler. Article draft was led by Tabea Hässler and Kristina R. Olson and supported by Jessica J. Glazier.

This study's hypotheses and data analytic plan were preregistered: <https://osf.io/jgdvw>.

Data and code are stored on OSF: <https://doi.org/10.17605/OSF.IO/BWU2N>.

Data collection for this project has been supported over the years from: NSF BCS-1523632, SMA-1837857, SMA-2041463, and National Institutes of Health HD092347. Tabea Hässler was supported by a grant from the Swiss National Science Foundation (PIZHP1_184553). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. We thank Natalie Gallagher for checking the syntax, Rachel Horton and Natalie Gallagher for the data preparation, and Lily Durwood, Selin Gülgöz, Raja Liechti, Shira Kahn-Samuelson, Léila Eisner, and the Human Diversity Lab for their insightful feedback.

Correspondence concerning this article should be addressed to Tabea Hässler, Department of Psychology, University of Zurich, Binzmühlestrasse 14/15, 8050 Zurich, Switzerland. Email: tabea.haessler@uzh.ch

making it unclear whether the same children who, for example, showed especially strong same-gender preferences at one age, do so at another age. Much of the longitudinal work that does exist has focused on very early childhood (i.e., infancy, toddlerhood, and preschool years) in which there is mixed evidence about stability or connection between or across gender concepts (e.g., Campbell et al., 2000, 2004; Halim et al., 2013; Kanka et al., 2019; Zosuls et al., 2009). In contrast, in the present work, we investigated the consistency of gender identity and preferences in middle childhood—a time when gendered behavior is regularly documented and central to children’s self-concept (Carver et al., 2003), yet is less often the focus of longitudinal study. Our investigation included both cisgender children and transgender children—specifically, binary socially transitioned transgender children who live as the binary gender that differs from the one assumed by their sex at birth—further allowing us to ask whether stability (or lack thereof) differs as a function of the degree to which a child identifies as the gender associated with their sex assigned at birth or not.

Stability of Gender Concepts in Middle Childhood

One of the few longitudinal studies to investigate the stability and change of gendered preferences over middle childhood assessed children at ages 3, 8, and 13 (Golombok et al., 2008, 2012). Golombok and colleagues observed that children who showed gender nonconforming behavior in early childhood were likely to remain gender nonconforming at age 8 (Golombok et al., 2008) and age 13 (Golombok et al., 2012). Interestingly, children who showed fewer gender-stereotypical preferences (i.e., more gender nonconformity) demonstrated more stability in their preferences than children who showed strong gender-stereotypical preferences. This finding is consistent with the general pattern that for gender conforming children, around the preschool years, they have often shown especially strong gender-stereotypical preferences (e.g., girls who love pink frilly dresses; Halim et al., 2014) and this tendency abates over development (i.e., they show a more moderate/less extreme level of gender-stereotypical preferences, Golombok et al., 2012).

Most of the other longitudinal studies in this age group focused on time spans of a few months to a year, but they also suggested some consistency in gender identity and preferences across time. For example, DeLay and colleagues (2018) investigated 299 children’s (M age 11.1 at the start of the study) perceptions of their similarity to their own gender and the other binary gender across a semester of sixth grade. They found these measures to be highly stable over this brief period ($r_s = .63$ and $.67$, respectively). Rae and Olson (2018) found a medium ($r = .32$) or large ($r = .56$) correlation between two administrations of an implicit gender identification measure given, a month apart, in a cohort of 107 children depending on how it was scored (as an own-group preference vs. a preference for boys). They also found a large correlation ($r = .54$ or $r = .56$, depending on scoring method) between two administrations of the same measure a year apart when tested in 97 cisgender, transgender, and gender nonconforming youth (mean age 8.7 at the start of the study). As one additional example, a study of 408 racially and ethnically diverse second and fourth graders (M ages of 7.6 and 9.5 years) found a moderate association between the number of same-gender and other-gender friends a child had at

two time points 12 months apart ($r = .33$ for same-gender friends; $r = .32$ for other gender friends; Halim et al., 2021). These results suggested that aspects of gender identity and gender-stereotyped preferences (for same-sex peers, toys, etc.) may show some consistency over time. However, given that most of these studies involved a timespan of a year or less, just how stable these might be across longer timespans, or in more or less gender diverse groups, is unknown.

Transgender Youth

While most children identify with the gender that aligns with their assigned sex (i.e., cisgender children), some children identify with the binary gender (i.e., boy or girl) different from that associated with their assigned sex. [Of course, others also identify as nonbinary, but those youth were not the focus of the present work.] A subset of these binary youth, whom we will term *transgender*, have socially transitioned or are using pronouns, first names, and wearing clothing and hairstyles more typically associated with the “opposite” sex (Aitken et al., 2015; de Graaf & Carmichael, 2019; Kaltiala et al., 2020). Initial work with binary socially transitioned transgender children suggested they perceived themselves to be more similar to children of their own gender compared to other genders, favored toys and clothing stereotypically associated with their gender, and preferred same-gender peers, at rates comparable to their cisgender peers (Gülgöz et al., 2019). This meant that a transgender boy (i.e., a child who was assigned female at birth but who lives as a boy) was likely to perceive himself to be similar to boys, favor masculine toys and clothing, and prefer to play with boys, as much as a cisgender boy.

Little is known, however, about whether the developmental pathways of socially transitioned transgender children differ from those of cisgender children (Stynes et al., 2021). Many experts working with gender diverse youth have noted that data assessing gender identity and gender-typed preferences of socially transitioned transgender children over time are lacking (Byne et al., 2012; Coleman et al., 2012; Drescher & Pula, 2014; Singh et al., 2021; Steensma & Cohen-Kettenis, 2018; Stynes et al., 2021; Turban & Keuroghlian, 2018; Zucker, 2020). In the present research, we sought to examine the continuity (and/or change) of gender identity and gender-typed preferences for both cisgender and transgender prepubertal children across a period of 2.6 years ($SD = 1.0$).

The Golombok studies (Golombok et al., 2008, 2012), which found more consistency across time among gender nonconforming youth than among gender conforming youth, might suggest that transgender youth (i.e., youth who do not conform to expectations of their sex assigned at birth) could be more consistent across time than cisgender youth. However, since the transgender youth in this study, different from the gender nonconforming youth in their work, had already socially transitioned and previous reports suggested these youth largely conform to their current gender roles at rates similar to their cisgender peers (Gülgöz et al., 2019), we had no specific prediction about differences between groups.

The Present Research

After the data had been collected, but before the proposed analyses were run, we preregistered our data analytic plan for this

paper (<https://osf.io/jgdvw>). Some cross-sectional data from this project has previously been reported (Fast & Olson, 2018; Gülgöz et al., 2019; Rae et al., 2019), however no analyses examining stability or change over time (the central questions in the present work) had been conducted with this sample.

The present analyses focused on five measures of gender identity and gender-typed preferences (i.e., gender identity, gender similarity, toy preference, clothing preference, and peer preference) reported separately and as a composite. We assessed these measures in three groups of participants: (socially transitioned) transgender children, cisgender siblings of transgender children, and unrelated cisgender children. We tested for differences between groups and examined whether boys and girls differ in their gender development over time. Because research has suggested that the strength of gender-typed preferences might vary over time (Skinner & McHale, 2018), we also explored whether the length of time between the assessments affected the stability of gender preferences.

Method

Participants

The current project included data from a large ongoing longitudinal study of binary transgender children (henceforth, transgender), their cisgender siblings (henceforth, cisgender siblings), and unrelated cisgender children (henceforth, unrelated cisgender). The research has been approved by the ethics committee of University of Washington (STUDY00001527 and STUDY00010314 Gender Development) and Princeton University (IRB# 12624 Gender Development). Data collection started in July 2013 and continued through September 2020 for the present analyses. We have been collecting data from children (who were aged 3–12 years old at their first visit) during in-person visits (or between March and September 2020, during COVID-19, through online live sessions). For the present analyses, we included all children who had participated two or more times between the ages of 3 and 12 before September 2020 ($n = 433$). As part of the longitudinal study, children over the age of 12 did not receive most of these measures and therefore data from older youth could not be included in the present analyses. Children were excluded if they completed less than 50% of the items for any measure or less than 50% of measures in the composite ($n = 68$), as proposed in our preregistration. Thus, the final sample included 365 children—158 transgender, 79 cisgender siblings, 128 unrelated cisgender children (see Table 1). A de-identified dataset (excluding some identifiable demographics) has been stored along with the code at: <https://doi.org/10.17605/OSF.IO/BWU2N>.

Most children in this sample were tested twice ($n = 312$), but some completed three or more testing sessions within the testing window ($n = 53$). In all cases, we included the first testing session and the most recent testing session. These two time points were an average of 2.6 years apart ($SD = 1.0$). We also reported the means from Time 1 for the participants in the full longitudinal study who were not included in the present research (see Table S1 in the online supplemental materials). By doing so, we could assess whether there were any differences between those children who were included in the current paper and those who were excluded

(because they were not tested a second time during this period and/or aged out before they were given these measures again).

Measures

Participants were asked to complete the following measures, which were part of a primary battery of measures given to all children participating in the larger longitudinal project for the duration of the study. Slight modifications to the stimuli were made for 3- to 7-year-old children versus 8- to 12-year-old children (i.e., toys, clothes, and age of peers depicted in the measures are different to reflect age-appropriate preferences). Participants also received additional measures for research unrelated to the current project (e.g., measures of self-esteem, mental health, gender essentialism, etc.). These other measures (aside from mental health) frequently changed during the 6 years of data collection and are outside the scope of this paper.

In line with Rae and colleagues (2019), we assessed five dimensions of gender identity (i.e., gender identity, gender similarity) and preferences (i.e., toy preference, peer preference, and clothing preference). The internal consistency of all measures is reported in Table 2.

Gender Identity

Two items assessed a child's identity now and in adulthood: "Do you feel like you are a boy, girl, or something else?" (If the child said, "something else", then they were asked, "Ok, so do you feel like neither a girl nor a boy, both a girl and a boy, it changes over time, or you do not know?") and "When you grow up, do you think you will feel like a boy, a girl, or something else?" Again, "something else" was followed by options of "neither," "both," "it changes," and "I do not know." Some participants were given an earlier version of the task in their first assessment that provided six answers (i.e., "Boy", "Girl", "Neither", "Both", "Changes", "I do not know") at once. For full description see the supplemental materials of Gülgöz and colleagues (2019). If children provided a response corresponding with the gendered nouns or pronouns they used at Time 1 (i.e., a transgender girl or cisgender girl says "girl"), they were assigned 1 point. Each response associated with the other binary gender at Time 1 was assigned -1 point (e.g., a "boy" response for a cisgender or transgender girl). All other answers were given a score of 0. If children provided two answers (rare), their scores were averaged (i.e., if a cisgender or transgender girl said "both" and "girl," they receive .5). Scores of the two gender identity questions were averaged and then recoded according to children's gender (determined by the pronouns they used) at Time 1. Higher scores for a transgender girl [assigned male] or cisgender girl [assigned female] indicated more feminine responding; higher scores for a transgender boy [assigned female] or cisgender boy [assigned male] indicated more masculine responding. Final scores ranged from 0 (gender nonconforming) to 1 (gender conforming; Cohen et al., 1999).

Gender Similarity

Children were asked five questions about how similar they are to boys and five questions about how similar they are to girls (Martin et al., 2017). The responses were given on a 5-point

Table 1
Demographic Characteristics for All Participants

Demographic characteristics	Transgender children	Cisgender siblings	Unrelated cisgender
<i>N</i>	158	79	128
Age at Time 1 (<i>SD</i>)	6.9 (1.6)	6.8 (2.0)	7.1 (1.6)
Age at Time 2 (<i>SD</i>)	9.7 (1.5)	9.3 (1.8)	9.7 (1.5)
Gender ^a	27.8% boys 72.2% girls	57.0% boys 43.0% girls	28.9% boys 71.1% girls
Race	70.3% White, non-Hispanic 8.2% White, Hispanic 0.6% Black 4.4% Asian 0.6% Native American/Alaskan Native 14.6% Multiracial 0.0% Other 1.3% NA	59.5% White, non-Hispanic 10.1% White, Hispanic 0.0% Black 2.5% Asian 1.3% Native American/Alaskan Native 17.7% Multiracial 0.0% Other 8.8% NA	67.2% White, non-Hispanic 7.0% White, Hispanic 0.0% Black 3.1% Asian 0.0% Native American/Alaskan Native 14.8% Multiracial 0.8% Other 7.0% NA
Parent education level			
Some schooling	0.0%	0.0%	0%
High school diploma	0.7%	1.4%	0.8%
Some college/Associate's degree	11.3%	12.2%	5.0%
College/Bachelor's degree			
Advanced degree (MA, MD, PHD)	31.3%	32.4%	49.2%
Other	56.7%	54.1%	43.3%
	0.0%	0.0%	1.7%
Parent political orientation			
From 1 <i>very liberal</i> to 7 <i>very conservative</i>	1.64 (0.93)	1.57 (0.84)	2.42 (1.39)
Household annual income			
Less than \$25,000	3.8%	2.7%	0.8%
\$25,001 to \$50,000	10.3%	9.3%	1.7%
\$50,001 to \$75,000	17.9%	20.0%	15.3%
\$75,001 to \$125,000	30.8%	34.7%	26.3%
More than \$125,000	37.2%	33.3%	55.9%
Urban vs. Rural			
Urban	24.5%	24.0%	48.8%
Small town	23.8%	19.2%	2.4%
Suburban	46.4%	54.5%	45.4%
Rural	4.0%	1.6%	2.9%
Other	1.3%	0.8%	0.5%
Geographic locations			
Northeast	18.8%	15.8%	0.0%
Midwest/Upper Plains	13.9%	19.2%	0.0%
Southeast	17.4%	20.0%	0.0%
Mountain West	11.1%	15.0%	0.0%
Pacific Northwest	20.8%	14.2%	100%
Pacific South	18.1%	15.8%	0.0%

Note. Percentages have been rounded and may not add up to 100. NA = not available.

^a Gender in this table was determined by pronoun use at Time 1.

pictorial scale that we coded from 1 (*very different*) to 5 (*very similar*). Following Fast and Olson (2018), we computed a difference score by subtracting the average of the five boy items from the average of the five girl items. We then reverse-scored the resulting composite for children who were living as boys at Time 1. All scores were finally rescaled to range from 0 (*gender nonconforming*) to 1 (*gender conforming*) by converting scores such that for example 1 became a .00, 2 became a .25, 3 a .50, 4 a .75, and 5 a 1.00.

Toy Preference

Toy preferences were assessed via four trials that each included an array of five toys (from Fast & Olson, 2018). The exact toys seen by 3–7-year-olds were different from the exact toys seen by 8–12-year-olds (to be age-appropriate), but both had the same format and were scored in the same way. On each trial, children could

select which of the five toys they would most like to play with. These items had previously been pilot-tested with a different group of cisgender participants to range from *very masculine* (1) to *very feminine* (5). Within each measure, responses from the four trials were averaged and for youth living as boys at Time 1, reverse-scored. Next, we recoded all scores to range from 0 (*completely gender nonconforming*) to 1 (*completely gender conforming*). Thus, higher scores for all participants indicate more gender conformity.

Peer Preference

Peer preferences were assessed on six trials in which children saw a picture of a girl and a boy and were asked whom they would prefer to be friends with (from Olson et al., 2015). We report the proportion of trials in which participants selected a same-gender (according to gender at Time 1) child, ranging from 0 to 1.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Overview of Descriptive Statistics and Stability Over Time

Measures	Trials	<i>N</i>	Range ^a	<i>M</i>	<i>SD</i>	Reliability	<i>r</i> _{T1-T2} outlier included ^b (Pearson/Spearman)
Transgender							
T ₁ Gender identity ^c	2	155	0.25–1.00	0.90	0.18	.40* ^d	
T ₂ Gender identity	2	157	0.00–1.00	0.91	0.20	.65* ^d	61.3% ^e
T ₁ Gender similarity ^f	10	146	0.22–1.00	0.76	0.18	.77	
T ₂ Gender similarity	10	152	0.10–1.00	0.71	0.19	.83	.34*/.32*
T ₁ Toy preference	4	158	0.06–1.00	0.66	0.21	.76	
T ₂ Toy preference	4	158	0.06–1.00	0.53	0.24	.71	.55*/.55*
T ₁ Peer preference ^g	6	154	0.00–1.00	0.80	0.24	—	
T ₂ Peer preference	6	149	0.00–1.00	0.70	0.24	—	.27*/.29*
T ₁ Clothing preference	4	158	0.00–1.00	0.89	0.16	.96	
T ₂ Clothing preference	4	157	0.00–1.00	0.72	0.24	.85	.16*/.19*
T ₁ Composite score ^h	5	158	0.31–1.00	0.80	0.13	.65	
T ₂ Composite score	5	158	0.18–0.98	0.71	0.17	.80	.38*/.40*
Cisgender siblings							
T ₁ Gender identity	2	78	0.00–1.00	0.87	0.23	.65* ^d	
T ₂ Gender identity	2	79	0.00–1.00	0.92	0.19	.57* ^d	69.2% ^e
T ₁ Gender similarity ^f	10	73	0.25–1.00	0.75	0.19	.79	
T ₂ Gender similarity	10	79	0.32–1.00	0.73	0.17	.78	.33*/.38*
T ₁ Toy preference	4	79	0.00–1.00	0.70	0.19	.78	
T ₂ Toy preference	4	78	0.12–1.00	0.68	0.23	.80	.24*/.27*
T ₁ Peer preference ^g	6	78	0.00–1.00	0.76	0.26	—	
T ₂ Peer preference	6	75	0.17–1.00	0.77	0.23	—	.12/.18
T ₁ Clothing preference	4	79	0.25–1.00	0.83	0.19	.91	
T ₂ Clothing preference	4	79	0.06–1.00	0.72	0.21	.84	.17/.08
T ₁ Composite score ^h	5	79	0.25–0.96	0.78	0.14	.69	
T ₂ Composite score	5	79	0.35–1.00	0.76	0.15	.78	.29*/.30*
Unrelated cisgender							
T ₁ Gender identity	2	128	0.25–1.00	0.89	0.18	.33* ^d	
T ₂ Gender identity	2	128	0.00–1.00	0.95	0.15	.47* ^d	59.4% ^e
T ₁ Gender similarity ^f	10	124	0.43–1.00	0.76	0.14	.62	
T ₂ Gender similarity	10	128	0.35–1.00	0.76	0.14	.76	.29*/.32*
T ₁ Toy preference	4	128	0.00–1.00	0.70	0.19	.80	
T ₂ Toy preference	4	128	0.12–1.00	0.64	0.21	.75	.45*/.40*
T ₁ Peer preference ^g	6	125	0.17–1.00	0.82	0.21	—	
T ₂ Peer preference	6	124	0.17–1.00	0.81	0.22	—	.18/.14
T ₁ Clothing preference	4	128	0.31–1.00	0.84	0.17	.92	
T ₂ Clothing preference	4	128	0.12–1.00	0.76	0.19	.87	.21*/.26*
T ₁ Composite score ^h	5	128	0.46–1.00	0.80	0.11	.56	
T ₂ Composite score	5	128	0.27–0.99	0.78	0.12	.66	.33*/.41*

Note. T₁ = Time 1; T₂ = Time 2. * = *p*-value below .05.

^a All measures were rescaled to a score from 0 (*completely gender nonconforming*) to 1 (*completely gender conforming*). ^b We registered using Pearson's correlation coefficient for assessing stability over time. Since the data were non-normally distributed, we also reported Spearman's correlation coefficient for the retest stability. ^c Some participants received a version where they were initially given 6 options (boy, girl, both, neither, it changes, I don't know) and some were given 3 options (boy, girl, something else) and only if they picked "something else" did they get the other options. ^d Pearson's correlation coefficient was reported for two-item scales. ^e Percentage of participants who had identical scores both times. ^f For results for the similarity measure broken down by own vs. other (binary) gender see Tables S11–S15 in the online supplemental materials. ^g Peer preferences were assessed on six trials in which children saw a picture of a boy and a girl and were asked whom they would prefer to be friends with. Thus, we reported the proportion of trials in which participants selected a same-gender child. Therefore, equivalence test and α could not be calculated. ^h The composite score was made up of all trials of all five measures of gender identity and gender-typed behaviors. Gender identity was categorical (boy, nonbinary, girl), but the recoded items can be seen as ordinal (gender nonconform to gender-conform).

Clothing Preference

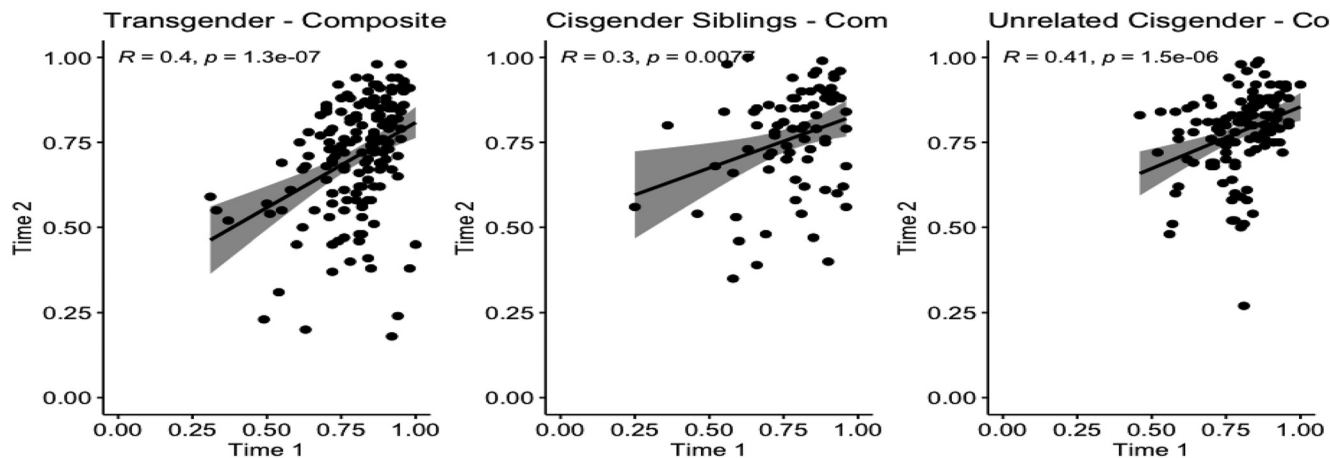
As with toy preferences, children were shown four trials, each including five images of outfits representing very masculine, slightly masculine, gender-neutral, slightly feminine, or very feminine items (from Fast & Olson, 2018). Children were asked to point to the outfit they liked best. Again, slightly different images were shown to children ages 3–7 versus 8–12 but scoring remained the same. Responses were initially coded on a 5-point-Likert scale (1 = *very masculine*, 5 = *very feminine*) before being reverse-scored for boys to mirror the other measures. Again, responses were averaged and

then rescaled to range from 0 (*completely gender nonconforming*) to 1 (*completely gender conforming*).

Gender Development Composite Score

Scores on all five measures were coded according to children's gender (determined by the pronouns they used) at Time 1 (that is, higher scores for a transgender girl [assigned male] or cisgender girl [assigned female] indicate more feminine responding; higher scores for a transgender boy [assigned female] or cisgender boy [assigned male] indicate more masculine responding). We then

Figure 1
Composite Score Over Time Among All Three Groups



rescaled measures (if they weren't already) to a score from 0 (*completely gender nonconforming*) and 1 (*completely gender conforming*). Finally, we created the gender development composite score by taking the average of all five measures.

Results

Preliminary Analyses

Before running our primary analyses, we first examined whether children tested at least twice and who had completed at least 50% of each item in the composite (i.e., those included in the primary analyses in this paper) differed from those children who were excluded. We first applied Little's MCAR test for all constructs and demographic variables listed in Table S1 in the online supplemental materials. Little's MCAR test was significant ($\chi^2[53] = 261.59, p < .001$), indicating that missingness was not completely at random. Closer examination of these data suggested only minor differences between these groups of participants (for a detailed report see Tables S1–S4 in the online supplemental materials). Thus, we decided to apply list-wise deletion for all following analyses.

Stability Over Time

Descriptive statistics of children's gender identity and preferences and the stability over time are shown in Table 2. Results indicated that both cisgender and transgender children showed a clear pattern of gender development associated with their gender at Time 1 such that those living as girls at Time 1 tended to identify as girls and prefer clothes and toys that were stereotypically feminine at both time points, and those living as boys identified as boys while preferring clothes and toys that were stereotypically masculine. These diverse measures were positively correlated with one another within time point (see Tables S5–S7 in the online supplemental materials). Most of the gender-typed preferences and the composite score showed small to moderate correlations over the 2.6-year time frame. The stability varied, however, substantially (see Figure 1 for the correlation of the composite score over

time and Tables S8–S10 in the online supplemental materials for the contingency tables of gender identity). Among transgender participants, effect sizes ranged from small-to-medium ($r_{\text{Spearman}} = .19$) to large effects ($r_{\text{Spearman}} = .55$). Among cisgender siblings effect sizes ranged from small to medium effects ($r_{\text{Spearman}} = .08$ to $r_{\text{Spearman}} = .38$). Among unrelated cisgender participants, effect sizes ranged from small to medium effects ($r_{\text{Spearman}} = .14$ to $r_{\text{Spearman}} = .41$).

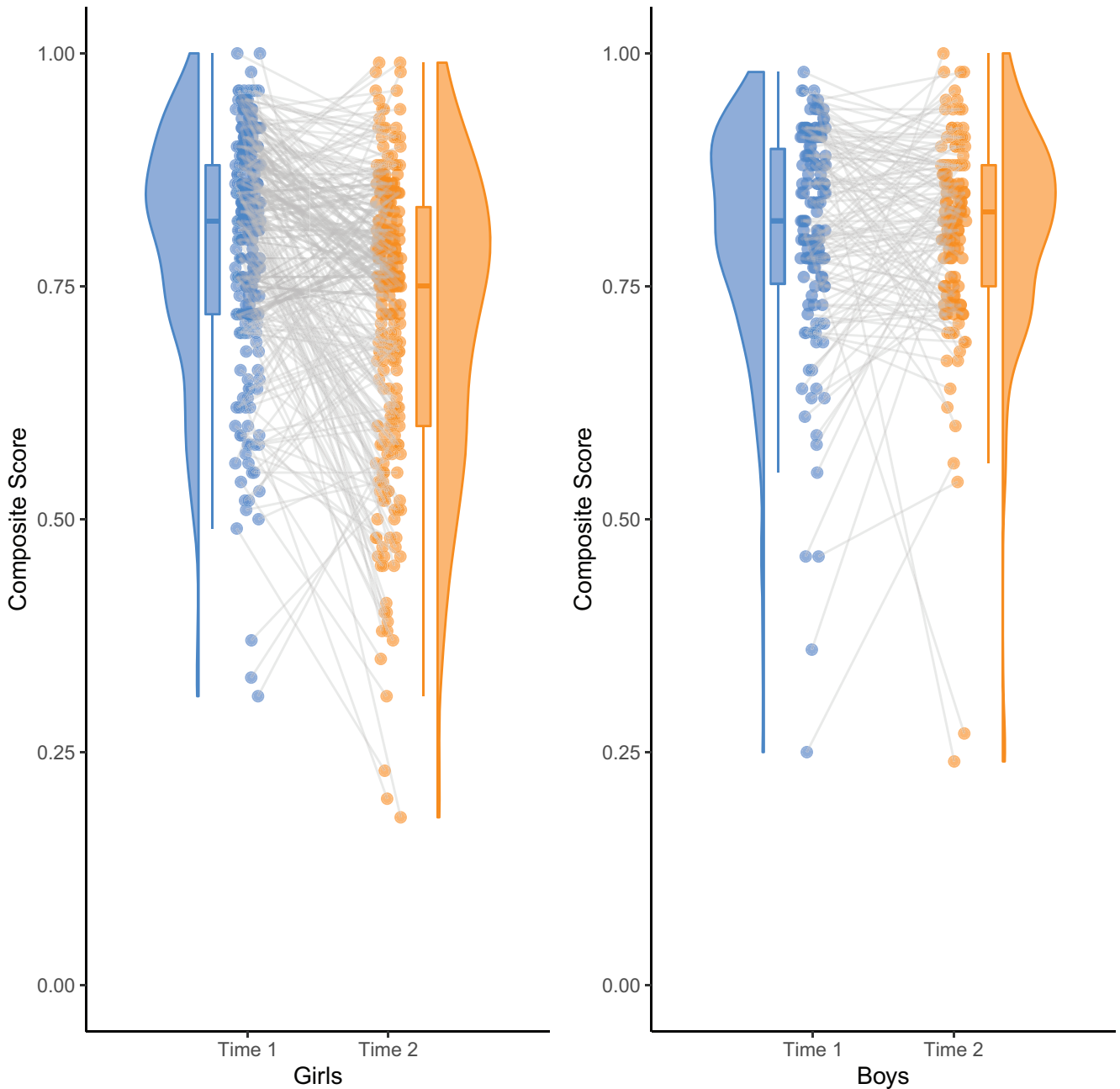
Next, we tested whether the correlations over time observed within the three groups differed significantly using Preacher's procedure (Cohen & Cohen, 1983; Preacher, 2002).¹ Results indicated that the three groups did not significantly differ in the stability of their gender identity, gender similarity, clothing preferences, or the composite score. We found, however, that cisgender siblings showed significantly lower stability of toy preferences compared to the transgender group ($r_{\text{Pearson cisgender siblings}} = .24, r_{\text{Pearson transgender}} = .55, p = .008$), while unrelated cisgender children ($r_{\text{Pearson cisgender unrelated}} = .24$) did not significantly from transgender children or cisgender siblings (see also Table 2 for the correlation coefficients). In sum, results indicated that the stability of gender preferences was generally similar across all three groups (i.e., transgender, cisgender siblings, and unrelated cisgender children).

To test for group differences in changes in the absolute mean composite score over time, we conducted a linear mixed model for repeated measures (2 Gender [i.e., boys, girls] \times 3 Groups [i.e., transgender, cisgender siblings, unrelated cisgender] \times 2 Time Points) in line with our preregistration.² There was a significant two-way interaction between gender and visit number. Follow-up analyses indicated that girls scored significantly lower on the composite of the five gender development measures during the second ($M = .71$) compared to the first visit ($M = .79$), $t(371) = 6.94, p < .001$, Cohen's $d = .73$, while we found no significant differences

¹ We preregistered using an equivalence test. However, an equivalence test for two independent correlations coefficients has not been programmed so far. Thus, we used NHST to test for the difference between the correlation coefficients.

² For an exploratory analysis among paired transgender children and their siblings see the online supplemental materials (Table S16 and Figure S2).

Figure 2
Composite Score Over Time Among Girls and Boys



Note. Descriptive means might slightly differ from the means of the general mixed model, because the latter reports estimated marginal means. See the online article for the color version of this figure.

for boys ($M_{\text{First}} = .81$, $M_{\text{Second}} = .81$, $t(371) = -.40$, $p = .688$, Cohen's $d = -.05$), see also Figure 2.

We further found a significant two-way interaction between groups [transgender versus cisgender siblings] and visit number (see Table 3). Follow-up analyses indicated that transgender participants scored significantly lower on the composite of the five gender development measures during the second ($M = .73$) compared to the first visit ($M = .81$), $t(371) = 5.58$, $p < .001$, Cohen's $d = .71$, while we found no

significant differences for cisgender siblings ($M_{\text{First}} = .78$, $M_{\text{Second}} = .75$, $t(371) = 1.55$, $p = .123$, Cohen's $d = .25$), see also Figure 3. In contrast, there was no significant two-way interaction between groups [transgender versus unrelated cisgender] and visit number (see Table 3). Nonetheless, exploratory analyses indicated no significant mean differences among unrelated cisgender children over time ($M_{\text{First}} = .80$, $M_{\text{Second}} = .79$, $t(371) = .46$, $p = .649$, Cohen's $d = .06$). Results did not reveal any other significant main or interaction effects.

Table 3
Overview of the Mixed Effect Model for the Composite Score of Gender Development

Predictor	Estimate	Std. Error	z-value	p-value
(Intercept)	0.83	0.02	41.70	<.001
Girl	-0.04	0.02	-1.63	.103
Trans vs. Unrelated cisgender	-0.03	0.03	-1.07	.285
Trans vs. Cisgender siblings	-0.04	0.03	-1.45	.146
Visit number (1 vs. 2)	-0.04	0.02	-1.68	.094
Girls × Transgender vs. Unrelated Cisgender	0.04	0.03	1.27	.204
Girls × Transgender vs. Cisgender Siblings	0.02	0.04	0.62	.535
Girls × Visit Number (1 vs. 2)	-0.08	0.03	-2.78	.005
Transgender vs. Unrelated Cisgender × Visit Number (1 vs. 2)	0.06	0.03	1.81	.070
Transgender vs. Cisgender Siblings × Visit Number (1 vs. 2)	0.07	0.03	2.18	.029
Girls × Transgender vs. Unrelated Cisgender × Visit Number (1 vs. 2)	0.02	0.04	0.39	.699
Girls × Transgender vs. Cisgender Siblings × Visit Number (1 vs. 2)	-0.04	0.04	-0.98	.329

Note. Boldface values indicate significant effects.

Exploratory analyses of mean differences among the individual dimensions and the composite score within each group also indicated weaker gender-typed preferences (but not gender identity and gender similarity) among transgender children over time. The mean differences were small to medium (see Table 4). Among both groups of cisgender children, we found a small decrease in stereotypical clothing preferences and among unrelated cisgender children also a small decrease in toy preferences. In contrast, we found a small increase in gender identity over time among the unrelated cisgender children. We did not find significant differences in any other measures of gender identity and preferences among cisgender children.

Finally, we explored whether the time between the assessments affects the stability of the gender development composite. Therefore, we created the absolute value of the difference score of the gender development composite [that is, the absolute value of the subtraction of the Time 1 composite score from the Time 2 composite score]. We correlated this difference score with the length of the time lag between the first and the second assessment. Pearson’s correlation suggested a significant decrease in the difference score over time ($r_{\text{Pearson}} = -.14, p_{\text{Pearson}} = .009$), while Spearman’s correlation was not significant ($r_{\text{Spearman}} = -.09, p_{\text{Spearman}} = .069$). Thus, there is mixed evidence about whether there was a decrease in the gender development composite score as more time passed between testing sessions, though the effect is fairly small, even if it is significant.

Discussion

The present study examined the stability of gender identity and gender-typed preferences in children during middle childhood from an average age of approximately 7 years old to age 9½, a span of years that is generally understudied. We generally found consistency—a significant positive correlation—across this 2½-year span across the composite of measures and within measures (i.e., gender identity, gender similarity, toy preference, clothing preference, and peer preference) among cisgender and transgender children. We did see some variability in the size of the correlation of different measures of gender development across time, ranging from small to large effects. For example, toy preferences for transgender children and unrelated cisgender children were particularly well-correlated over time, while clothing preferences among all

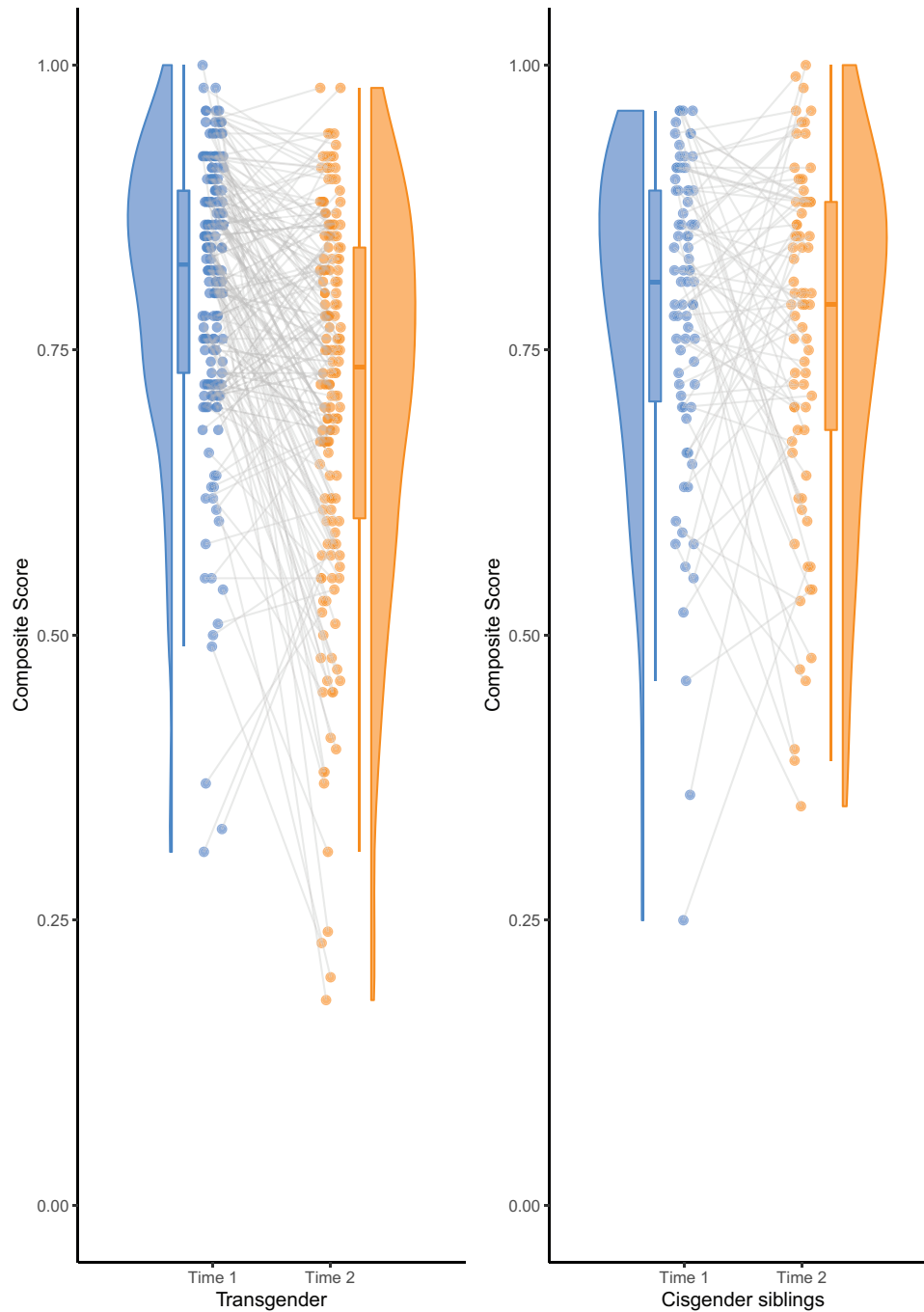
groups were not as consistent across time. Despite this variability across measures, the trajectories were highly similar across the three groups of children studied in this work (i.e., transgender, cisgender siblings, and unrelated cisgender) and did not significantly differ (except for the lower stability of toy preferences among cisgender siblings).

Another way to think about consistency is with regard to absolute mean changes of scores over time. Our data suggested a decrease in stereotypical preferences and gender identity over time among girls but not boys. Similarly, transgender children showed a decrease in their gender composite score over time as compared to cisgender siblings (the comparison to unrelated cisgender children was not statistically different). In addition, exploratory analyses of mean differences among the individual measures indicated decreases in gender-typed preferences in transgender but not in cisgender children over time.

Similarity of Trajectories in Gender-Typed Identity and Preferences Across Time

We included five different measures assessing gender identity and gender-typed preferences in the current study in line with arguments that gender consists of multiple dimensions (Ruble et al., 2006; Tate et al., 2014). Gender-typed preferences and gender identity showed mostly small to moderate correlations over the 2½-year period. In line with previous work, we found variability in the consistency across measures (Halim et al., 2013). This finding suggested that developmental patterns might slightly differ across different aspects of cognition (Kornienko et al., 2016) and behavior. One explanation might be that the different aspects of gender cognition are subject to different influences over time (Ruble et al., 2006). Another possibility is that some of the measures were just better in that they produced more variability, contained less error, or were otherwise more reliable. On the other hand, gender identity and gender-typed preferences might also assess different things: children might clearly identify as a boy or girl but prefer activities less stereotypically associated with their gender. To better understand these dynamics and to achieve a more accurate picture of the coherence of gender development, it is therefore important to assess multiple measures of gender identity and preferences simultaneously. In line with this, Rae and colleagues (2019) demonstrated that inclusion of more variables, and especially moving from just one or two to several improved predictive utility of these

Figure 3
Composite Score Over Time Among Transgender Children and Cisgender Siblings (the Groups for Whom Significant Differences Were Observed)



Note. Descriptive means might slightly differ from the means of the general mixed model, because the latter reports estimated marginal means. See the online article for the color version of this figure.

measures. In the present study, the composite tended to show one of the stronger associations across time, showing a medium effect size among transgender ($r = .40$) and unrelated cisgender youth ($r = .41$).

Although we found variability in the consistency between the separate dimensions among all three groups, the patterns of gender

development were fairly similar over time. The three groups did not significantly differ in the stability of their gender identity, gender similarity, clothing preferences, and the composite score. We found only slight differences in the stability of toy preferences in the sibling group. Together, these findings suggested that whether a child's

Table 4
Mean Differences Over Time

Dimension	Time 1			Time 2			Difference	Cohen's <i>d</i>
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	Overall statistics	
Transgender								
Gender identity	155	0.90	0.18	157	0.91	0.20	$t(154) = -0.78, p = .439$.06
Gender similarity ^a	146	0.76	0.18	152	0.71	0.19	$t(142) = 2.38, p = .019$.20
Toy preference	158	0.66	0.21	158	0.53	0.24	$t(157) = 7.49, p \leq .001$.60
Peer preference	154	0.80	0.24	149	0.70	0.24	$t(147) = 4.20, p \leq .001$.35
Clothing preference	158	0.89	0.16	157	0.72	0.24	$t(156) = 8.69, p \leq .001$.69
Composite score ^b	158	0.80	0.13	158	0.71	0.17	$t(157) = 6.95, p \leq .001$.55
Cisgender siblings								
Gender identity	78	0.87	0.23	79	0.92	0.19	$t(77) = -1.81, p = .075$.20
Gender similarity ^a	73	0.75	0.19	79	0.73	0.17	$t(72) = 0.70, p = .488$.08
Toy preference	79	0.70	0.19	78	0.68	0.23	$t(77) = 0.69, p = .489$.08
Peer preference	78	0.76	0.26	75	0.77	0.23	$t(73) = -0.26, p = .797$.03
Clothing preference	79	0.83	0.19	79	0.72	0.21	$t(78) = 3.64, p \leq .001$.41
Composite score ^b	79	0.78	0.14	79	0.76	0.15	$t(78) = 0.95, p = .345$.11
Unrelated cisgender								
Gender identity	128	0.89	0.18	128	0.95	0.15	$t(127) = -2.95, p = .004$.26
Gender similarity ^a	124	0.76	0.14	128	0.76	0.14	$t(123) = -0.14, p = .891$.01
Toy preference	128	0.70	0.19	128	0.64	0.21	$t(127) = 3.17, p = .002$.28
Peer preference	125	0.82	0.21	124	0.81	0.22	$t(120) = 0.53, p = .596$.05
Clothing preference	128	0.84	0.17	128	0.76	0.19	$t(127) = 4.22, p < .001$.37
Composite score ^b	128	0.80	0.11	128	0.78	0.12	$t(127) = 1.67, p = .097$.15

Note. We used Bonferroni-correction to adjust for the multiple testing within each group. Because we conducted six comparisons, we only rejected the null hypothesis of each comparison if it had a *p*-value less than .008.

^a For results for the similarity measure broken down by own vs. other (binary) gender see Tables S11–S15 in the online supplemental materials. ^b Descriptive means and Cohen's *d* for the composite score slightly differ from the means of the general mixed model, because the latter reports estimated marginal means. Boldface values indicate significant effect (Bonferroni-corrected).

gender aligned with their sex assigned at birth did not seem to have much impact on the stability of their gendered identity and preferences over time, at least within this unique sample of early identifying trans youth who have transitioned to live in line with their gender and are supported by their families in doing so. While not all transgender (or other gender diverse) youth come out, or transition at such early ages, this sample of youth socially transitioned on average at age 5.5 years (*SD* = 1.4 years); whether these findings generalize to trans youth who come out or transition later is currently unknown.

Differences in Reported Mean Changes Between Groups

Besides examining the consistency of gender identity and gender-typed preferences, we also examined differences in the mean gender development composite over time based on children's group, gender, and amount of time between assessments. We found a significant interaction between gender and time between assessments. Follow-up analyses suggested that girls (but not boys) scored lower on the gender development composite in the second compared to the first visit ($M_{\text{Second}} = .72, M_{\text{First}} = .79$). This indicated that girls had less stereotypical preferences and/or gender identity over time. Descriptive statistics indicated that this medium-sized decrease might be explained by greater within-gender differences at the second time point among girls compared to boys (see Figure 2). While some girls had more gender-neutral preferences during the second visit, the distribution remained relatively stable among boys. Other studies have found this overall trend toward greater rigidity in gender development among boys than girls (for a review see Leaper & Farkas, 2015). For example, a study comparing gender-typed behavior between gender-

conforming and nonconforming children over time found that within-gender differences became more pronounced over time among girls compared to boys (Golombok et al., 2008, 2012). Similarly, a study assessing the gender stability of girls and boys from preschool to elementary school years found lower gender stability over time among girls (Halim et al., 2013; for a discussion see Halim et al., 2011).

We further found a significant interaction between group (transgender versus cisgender siblings) and time of assessment. Follow-up analyses suggested that transgender (but not cisgender sibling) children scored lower (i.e., were more gender neutral than their siblings) on the gender development composite in the second compared to the first visit ($M_{\text{Second}} = .73, M_{\text{First}} = .81$). This finding indicated a decrease in stereotypical preferences and/or gender identity among transgender children over time that was of medium effect size. Follow-up analyses (see Figure S1 in the online supplemental materials) indicated that stereotypical preferences and/or gender identity decreased among transgender girls and, to a lesser extent, boys. In contrast, we found opposing trends among cisgender siblings: stereotypical preferences and/or gender identity decreased among girls but increased among boys. Further, exploratory analyses of mean differences among the individual measures indicated small to medium effect sized reductions in gender-typed preferences in transgender children over time. One reason for the decrease in gender-typed preferences in the transgender group might be that socially transitioned transgender children feel less pressure to conform to normative expectations of stereotypical "girl" or "boy" behavior over time. Another reason for the decrease in gender-typed preferences and identity might be that a minority of children—as indicated by

low values on the composite score—might start to identify with another gender (see Figure 3). It is difficult to interpret this trend at the current moment. Thus, future research should examine potential trends more closely and assess whether these patterns continue as these children mature.

Limitations

There are several limitations in the present work. For example, the current study often used relative measures meaning we could not separately determine if there were changes in children's preference for feminine- or masculine-stereotyped clothing or toys. Future studies should separately assess these constructs.

While the current study included cisgender and transgender children, the present research is still limited in its inclusion of a range of gender-diverse identities; it did not include children, or those who defy gender norms in terms of preferences but do not socially transition to live as a different gender. It would be interesting to know whether children whose gender identities are less binary, show more or less consistency in their gender identity and preferences than youth who are presenting in more binary ways (i.e., binary transgender or cisgender children).

Further, most studies on gender development, including transgender youth, have been conducted in North America or Western Europe and are largely focused on White middle- and upper-class children with relatively well-educated parents. Yet, gender socialization takes place within the context of culture, racialized experiences, religion, and social class (Brown & Mar, 2018; Crenshaw et al., 2015). For example, trans youth of color, Native trans youth, and immigrant trans youth might be confronted with specific challenges such as limited representation in the media, safe spaces, and affirmative spaces that White or nonimmigrant transgender youth are not. In addition, transgender youth of color and immigrant trans youth might encounter barriers accessing gender affirmative treatment and little awareness of the intersectional impacts of racism and transphobia. Although beyond the scope of the present research, it is important that future studies are conducted in different contexts using more diverse and representative samples to understand whether the observed effects generalize or vary by these factors.

Conclusion

The present research included a large sample of young cisgender and transgender children to assess gender identity and gender-typed behavior over time. The inclusion of transgender children allowed us to draw a more comprehensive picture of gender development among diverse children in middle childhood. We found small to moderate levels of consistency—significant positive correlations—over time for most measures of gender identity and gender-typed behavior among all children. Further, groups did not differ in terms of their overall gender development, as indicated in the consistency over time. When we examined mean changes over time, we found that girls and transgender (but not boys and cisgender sibling) children showed medium-sized decreases in their overall composite of gender identity and gender-stereotypical preferences in the second compared to the first visit. To better understand the dynamics of gender development among all children,

more longitudinal research using gender-conforming and gender-nonconforming children from different social backgrounds is needed.

References

- Aitken, M., Steensma, T. D., Blanchard, R., VanderLaan, D. P., Wood, H., Fuentes, A., Spegg, C., Wasserman, L., Ames, M., Fitzsimmons, C. L., Leef, J. H., Lishak, V., Reim, E., Takagi, A., Vinik, J., Wreford, J., Cohen-Kettenis, P. T., de Vries, A. L., Kreukels, B. P., & Zucker, K. J. (2015). Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *Journal of Sexual Medicine, 12*(3), 756–763. <https://doi.org/10.1111/jsm.12817>
- Brown, E., & Mar, K. (2018). Culturally responsive practice with children of color. In C. Keo-Meier, D. Ehrensaft (Eds.), *The Gender Affirmative Model: An interdisciplinary approach to supporting transgender and gender expansive children* (pp. 55–69). American Psychological Association.
- Byne, W., Bradley, S. J., Coleman, E., Eyler, A. E., Green, R., Menvielle, E. J., Meyer-Bahlburg, H. F., Pleak, R. R., & Tompkins, D. A. (2012). Report of the American Psychiatric Association task force on treatment of gender identity disorder. *Archives of Sexual Behavior, 41*(4), 759–796. <https://doi.org/10.1007/s10508-012-9975-x>
- Campbell, A., Shirley, L., & Candy, J. (2004). A longitudinal study of gender-related cognition and behaviour. *Developmental Science, 7*(1), 1–9. <https://doi.org/10.1111/j.1467-7687.2004.00316.x>
- Campbell, A., Shirley, L., Heywood, C., & Crook, C. (2000). Infants' visual preference for sex-congruent babies, children, toys and activities: A longitudinal study. *British Journal of Developmental Psychology, 18*(4), 479–498. <https://doi.org/10.1348/026151000165814>
- Carver, P. R., Yunger, J. L., & Perry, D. G. (2003). Gender identity and adjustment in middle childhood. *Sex Roles, 49*(3), 95–109. <https://doi.org/10.1023/A:1024423012063>
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Erlbaum.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2
- Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., Fraser, L., Green, J., Knudson, G., Meyer, W. J., Monstrey, S., Adler, R. K., Brown, G. R., Devor, A. H., Ehrbar, R., Ettner, R., Eyler, E., Garofalo, R., Karasic, D. H., . . . Zucker, K. (2012). Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *International Journal of Transgenderism, 13*(4), 165–232. <https://doi.org/10.1080/15532739.2011.700873>
- Crenshaw, K. W., Ochen, P., & Nanda, J. (2015). *Black girls matter: Pushed out, overpoliced and underprotected*. Center for Intersectionality & Social Policy.
- Davis, J. T. M., & Hines, M. (2020). How large are gender differences in toy preferences? A systematic review and meta-analysis of toy preference research. *Archives of Sexual Behavior, 49*, 373–394. <https://doi.org/10.1007/s10508-019-01624-7>
- de Graaf, N. M., & Carmichael, P. (2019). Reflections on emerging trends in clinical work with gender diverse children and adolescents. *Clinical Child Psychology and Psychiatry, 24*(2), 353–364. <https://doi.org/10.1177/1359104518812924>
- DeLay, D., Lynn Martin, C., Cook, R. E., & Hanish, L. D. (2018). The influence of peers during adolescence: Does homophobic name calling by peers change gender identity? *Journal of Youth and Adolescence, 47*(3), 636–649. <https://doi.org/10.1007/s10964-017-0749-6>
- Drescher, J., & Pula, J. (2014). Ethical issues raised by the treatment of gender-variant prepubescent children. *The Hastings Center Report, 44*(s4), S17–S22. <https://doi.org/10.1002/hast.365>

- Etaugh, C., Grinnell, K., & Etaugh, A. (1989). Development of gender labeling: Effect of age of pictured children. *Sex Roles, 21*(11), 769–773. <https://doi.org/10.1007/BF00289807>
- Fagot, B. I., Leinbach, M. D., & Hagan, R. (1986). Gender labeling and the adoption of sex-typed behaviors. *Developmental Psychology, 22*(4), 440–443. <https://doi.org/10.1037/0012-1649.22.4.440>
- Fast, A. A., & Olson, K. R. (2018). Gender development in transgender preschool children. *Child Development, 89*(2), 620–637. <https://doi.org/10.1111/cdev.12758>
- Golombok, S., & Rust, J. (1993). The measurement of gender role behaviour in pre-school children: A research note. *Journal of Child Psychology and Psychiatry, 34*(5), 805–811. <https://doi.org/10.1111/j.1469-7610.1993.tb01072.x>
- Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental trajectories of sex-typed behavior in boys and girls: A longitudinal general population study of children aged 2.5–8 years. *Child Development, 79*(5), 1583–1593. <https://doi.org/10.1111/j.1467-8624.2008.01207.x>
- Golombok, S., Rust, J., Zervoulis, K., Golding, J., & Hines, M. (2012). Continuity in sex-typed behavior from preschool to adolescence: A longitudinal population study of boys and girls aged 3–13 years. *Archives of Sexual Behavior, 41*(3), 591–597. <https://doi.org/10.1007/s10508-011-9784-7>
- Gülgöz, S., Glazier, J. J., Enright, E. A., Alonso, D. J., Durwood, L. J., Fast, A. A., Lowe, R., Ji, C., Heer, J., Martin, C. L., & Olson, K. R. (2019). Similarity in transgender and cisgender children's gender development. *Proceedings of the National Academy of Sciences of the United States of America, 116*(49), 24480–24485. <https://doi.org/10.1073/pnas.1909367116>
- Halim, M. L. D., Martin, C. L., Andrews, N. C., Zosuls, K. M., & Ruble, D. N. (2021). Enjoying each other's company: Gaining other-gender friendships promotes positive gender attitudes among ethnically diverse children. *Personality and Social Psychology Bulletin, 47*(12), 1635–1653. <https://doi.org/10.1177/0146167220984407>
- Halim, M. L., Ruble, D. N., & Amodio, D. M. (2011). From pink frilly dresses to 'one of the boys': A social-cognitive analysis of gender identity development and gender bias. *Social and Personality Psychology Compass, 5*(11), 933–949. <https://doi.org/10.1111/j.1751-9004.2011.00399.x>
- Halim, M. L., Ruble, D. N., Tamis-LeMonda, C., & Shrout, P. E. (2013). Rigidity in gender-typed behaviors in early childhood: A longitudinal study of ethnic minority children. *Child Development, 84*(4), 1269–1284. <https://doi.org/10.1111/cdev.12057>
- Halim, M. L., Ruble, D. N., Tamis-LeMonda, C. S., Zosuls, K. M., Lurye, L. E., & Greulich, F. K. (2014). Pink frilly dresses and the avoidance of all things "girly": Children's appearance rigidity and cognitive theories of gender development. *Developmental Psychology, 50*(4), 1091–1101. <https://doi.org/10.1037/a0034906>
- Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior, 39*(6), 1261–1273. <https://doi.org/10.1007/s10508-010-9618-z>
- Kaltiala, R., Bergman, H., Carmichael, P., de Graaf, N. M., Egebjerg Rischel, K., Frisé, L., Schorkopf, M., Suomalainen, L., & Waehre, A. (2020). Time trends in referrals to child and adolescent gender identity services: A study in four Nordic countries and in the U.K. *Nordic Journal of Psychiatry, 74*(1), 40–44. <https://doi.org/10.1080/08039488.2019.1667429>
- Kanka, M. H., Wagner, P., Buchmann, M., & Spiel, C. (2019). Gender-stereotyped preferences in childhood and early adolescence: A comparison of cross-sectional and longitudinal data. *European Journal of Developmental Psychology, 16*(2), 198–214. <https://doi.org/10.1080/17405629.2017.1365703>
- Kornienko, O., Santos, C. E., Martin, C. L., & Granger, K. L. (2016). Peer influence on gender identity development in adolescence. *Developmental Psychology, 52*(10), 1578–1592. <https://doi.org/10.1037/dev0000200>
- Leaper, C., & Farkas, T. (2015). The socialization of gender during childhood and adolescence. In E. Grusec, P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (2nd ed., pp. 541–565). Guilford.
- Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., & Ruble, D. N. (2017). A dual identity approach for conceptualizing and measuring children's gender identity. *Child Development, 88*(1), 167–182. <https://doi.org/10.1111/cdev.12568>
- Martin, C. L., Fabes, R. A., Hanish, L., Leonard, S., Dinella, L. M. (2011). Experienced and expected similarity to same-gender peers: Moving toward a comprehensive model of gender segregation. *Sex Roles, 65*, 421–434. <https://doi.org/10.1007/s11199-011-0029-y>
- Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology, 61*(1), 353–381. <https://doi.org/10.1146/annurev.psych.093008.100511>
- McHale, S. M., Kim, J. Y., Whiteman, S., & Crouter, A. C. (2004). Links between sex-typed time use in middle childhood and gender development in early adolescence. *Developmental Psychology, 40*(5), 868–881. <https://doi.org/10.1037/0012-1649.40.5.868>
- Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science, 26*(4), 467–474. <https://doi.org/10.1177/0956797614568156>
- Preacher, K. J. (2002). *Calculation for the test of the difference between two independent correlation coefficients* [Computer software]. <http://quantpsy.org>
- Rae, J. R., Gülgöz, S., Durwood, L., DeMeules, M., Lowe, R., Lindquist, G., & Olson, K. R. (2019). Predicting early-childhood gender transitions. *Psychological Science, 30*(5), 669–681. <https://doi.org/10.1177/0956797619830649>
- Rae, J. R., & Olson, K. R. (2018). Test-retest reliability and predictive validity of the Implicit Association Test in children. *Developmental Psychology, 54*(2), 308–330. <https://doi.org/10.1037/dev0000437>
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg (Ed.), *Handbook of child psychology* (Vol. 3, pp. 858–932). 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*(1), 7–15. <https://doi.org/10.1080/01650250042000078>
- Singh, D., Bradley, S. J., & Zucker, K. J. (2021). A follow-up study of boys with gender identity disorder. *Frontiers in Psychiatry, 12*, 632784. <https://doi.org/10.3389/fpsy.2021.632784>
- Skinner, O. D., & McHale, S. M. (2018). The development and correlates of gender role orientations in African-American youth. *Child Development, 89*(5), 1704–1719. <https://doi.org/10.1111/cdev.12828>
- Steensma, T. D., & Cohen-Kettenis, P. T. (2018). A critical commentary on "A critical commentary on follow-up studies and "desistence" theories about transgender and gender non-conforming children". *International Journal of Transgenderism, 19*(2), 225–230. <https://doi.org/10.1080/15532739.2018.1468292>
- Stynes, H., Lane, C., Pearson, B., Wright, T., Ranieri, V., Masic, U., & Kennedy, E. (2021). Gender identity development in children and young people: A systematic review of longitudinal studies. *Clinical Child Psychology and Psychiatry, 26*(3), 706–719. <https://doi.org/10.1177/13591045211002620>
- Tate, C. C., Youssef, C. P., & Bettergarcia, J. N. (2014). Integrating the study of transgender spectrum and cisgender experiences of self-categorization from a personality perspective. *Review of General Psychology, 18*(4), 302–312. <https://doi.org/10.1037/gpr0000019>

- Thompson, S. K. (1975). Gender labels and early sex role development. *Child Development, 46*(2), 339–347. <https://doi.org/10.2307/1128126>
- Turban, J. L., & Keuroghlian, A. S. (2018). Dynamic gender presentations: Understanding transition and “de-transition” among transgender youth. *Journal of the American Academy of Child & Adolescent Psychiatry, 57*(7), 451–453. <https://doi.org/10.1016/j.jaac.2018.03.016>
- Weinraub, M., Clemens, L. P., 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*(4), 1493–1503. <https://doi.org/10.2307/1130019>
- 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*(3), 688–701. <https://doi.org/10.1037/a0014053>
- Zucker, K. J. (2020). Debate: Different strokes for different folks. *Child and Adolescent Mental Health, 25*(1), 36–37. <https://doi.org/10.1111/camh.12330>

Received October 12, 2021

Revision received May 16, 2022

Accepted May 17, 2022 ■



AMERICAN PSYCHOLOGICAL ASSOCIATION

APA Journals®

ORDER INFORMATION

Subscribe to This Journal for 2023

Order Online:

Visit at.apa.org/dev-2023
for pricing and access information.

Call **800-374-2721** or **202-336-5600**

Fax **202-336-5568** | TDD/TTY **202-336-6123**

Subscription orders must be prepaid. Subscriptions are on a calendar year basis. Please allow 4-6 weeks for delivery of the first issue.

All APA journal subscriptions include Online First journal articles and access to archives. Individuals can receive online access to all of APA's scholarly journals through a subscription to APA PsycNet® or through an institutional subscription to the APA PsycArticles® database.

Visit AT.APA.ORG/CIRC2023
to browse APA's full journal collection