Journal of Educational Psychology

ISSN: 0022-0663

https://doi.org/10.1037/edu0000953

New Online Tools to Assess Children's Implicit Social Cognition: Replication and Generalization of In-Person Research

Dario Cvencek[1], Elizabeth A. Sanders[2], Craig D. Maddox[1], and Andrew N. Meltzoff[1, 3]
[1] Institute for Learning & Brain Sciences, University of Washington
[2] College of Education, University of Washington
[3] Department of Psychology, University of Washington

Reported are new online tools for testing children's implicit social cognition. These tools adapt the adult Implicit Association Test so it can be used for online, unmonitored child testing without the need for a trained experimenter. Four preregistered studies are reported. The first two involve elementary school children (Grades 2–5). Study 1 evaluated the validity of the new online procedures by investigating a well-replicated result about gender identity using both in-person ($N = 48$) and online procedures ($N = 64$). Study 2 applied the new online procedures to assess children's math–gender stereotypes ($N = 72$). The second two studies involved younger participants, preschoolers. Study 3 evaluated the validity of the new online procedures with 5-year-olds by investigating a well-replicated effect about children's preferences and attitudes (preferring flowers to insects) using both in-person and online procedures ($N = 64$ each). Study 4 then applied the new online procedures to assess gender in-group bias in 5-year-olds. Results from each of the four studies confirmed the preregistered hypotheses, and the online procedures yielded results comparable to in-person results in both direction and magnitude. Across four different studies, construct and psychometric validity of these tools were established. These new tools have the potential to: (a) be used with children who are not able to come into university-based laboratories, increasing demographic diversity in child studies, and (b) enable large-scale, national, and international collaborations investigating the development of important psychological and educational topics such as implicit biases, stereotypes, and self-concepts.

Educational Impact and Implications Statement

The presence of implicit stereotypes and attitudes in children has been shown to be related to school achievement and other outcome measures. Prior research has relied exclusively on in-person assessments of implicit cognition conducted by a trained experimenter. Here, we designed new tools for online, unmoderated testing of children and used them in four studies that included both preschool and elementary school children. These new tools involve rigorously controlled and standardized computer-based protocols. The online tools satisfied normality assumptions and demonstrated satisfactory reliability, establishing the validity of the procedures. These new tools for assessing children's implicit social cognition can now be used to assess age-related changes in constructs relevant to developmental theory and educational science, such as academic stereotypes, in-group biases, and racial biases. This work also offers new tools for broadening participation in educational and psychological science enabling testing of racially/ethnically diverse children in home and community settings, rather than requiring participants to be tested by trained testers in university laboratories, museums, or schools.

Daniel Ansari served as action editor.

Dario Cvencek https://orcid.org/0000-0002-0073-5862
Elizabeth A. Sanders https://orcid.org/0000-0002-1008-8435
Craig D. Maddox https://orcid.org/0000-0003-0237-1799
Andrew N. Meltzoff https://orcid.org/0000-0001-8683-0547

This research was supported by the Bezos Family Foundation and the Bernier McCaw Foundation. The authors thank D. Hathaway for administrative support. The authors have no known conflicts of interest to disclose. This study was registered on the Open Science Framework after receiving in-principle acceptance by the *Journal of Educational Psychology* (https://osf.io/wy423). Data from this study was shared openly following Stage 2 acceptance at the same link.

Dario Cvencek served as lead for conceptualization, data curation, investigation, methodology, and writing–original draft and contributed equally to formal analysis and supervision. Elizabeth A. Sanders served as lead for formal analysis, methodology, and validation and contributed equally to visualization. Craig D. Maddox served as lead for data curation, project administration, and software, contributed equally to investigation and writing–original draft, and served in a supporting role for writing–review and editing. Andrew N. Meltzoff served as lead for funding acquisition and supervision, contributed equally to conceptualization and writing–original draft, and served in a supporting role for formal analysis, investigation, validation, and visualization. Dario Cvencek, Elizabeth A. Sanders, and Andrew N. Meltzoff contributed equally to writing–review and editing.

The data are available at https://osf.io/w35zk/.
The experimental materials are available at https://osf.io/w35zk/.
The preregistered design is available at https://osf.io/wy423/.

Correspondence concerning this article should be addressed to Dario Cvencek, Institute for Learning & Brain Sciences, University of Washington, Room 458, Portage Bay Building, 1715 NE Columbia Road, Seattle, WA 98195, United States. Email: dario1@uw.edu

Keywords: children, implicit social cognition, online testing, stereotypes, biases

Supplemental materials: https://doi.org/10.1037/edu0000953.supp

Research on implicit bias has generated wide attention in psychology, education, business, law, and medicine (e.g., Axt, 2018; Greenwald & Banaji, 2017; Greenwald & Lai, 2020). More than 20 million adults worldwide have completed online implicit measures of attitudes and beliefs using Implicit Association Tests (IATs; Greenwald et al., 1998) and almost 100 scientific publications have already used publicly available data sets collected from the Project Implicit Demo Site (Project Implicit, 2022) using adult participants.

There is now also a growing body of work on implicit beliefs and biases in children using in-person child IATs (as well as other implicit measures such as the Draw-A-Scientist Task, Miller et al., 2018, and the Affective Misattribution Procedure, Vuletich et al., 2020). Since their initial development (Baron & Banaji, 2006) and refinement for developmental appropriateness (Cvencek, Meltzoff, & Greenwald, 2011; Cvencek et al., 2016), in-person child IATs have been used to measure racial biases (Newheiser & Olson, 2012; Steele et al., 2018), gender-linked stereotypes about math (Cvencek, Meltzoff, & Greenwald, 2011; Zhao et al., 2022), math self-concepts (Cvencek et al., 2021; Steffens et al., 2010), and self-esteem (Cvencek et al., 2016; Dunham et al., 2007). In-person child IATs have also been conducted with children across diverse cultures outside North America (e.g., Cvencek et al., 2015; Dunham et al., 2006; M. K. Qian et al., 2016; Setoh et al., 2019). A review of predictive validity of child IAT measures reported that individual differences measured by the in-person child IATs have satisfactory test–retest reliability and are predictive of children's behaviors (Rae & Olson, 2018).

All previous work with child IATs was conducted using in-person tasks, assessing children chiefly in university laboratories or schools. The present article proposes a series of four studies designed to conceptually replicate, extend, and generalize the earlier in-person research by using: (a) novel online implicit tools, as well as (b) testing across a wide age span. Collectively, the four studies reported here include elementary school children (Studies 1 and 2) and 5-year-old children (Studies 3 and 4). We report whether the data collected with these new online tools replicate the effects reported in published, in-person studies. Such results will inform the future design of longitudinal studies and may also feed into the development of age-appropriate child interventions for helping to ameliorate children's stereotypes and biases. In the following sections, we outline our rationale and proposed methods for the new online testing tools.

History of Adapting the Adult IAT for In-Person Testing of School-Aged Children

In its standard, adult form, the IAT indirectly assesses the strengths of associations among concepts (Cvencek et al., 2021; Greenwald & Banaji, 2017; for further discussion about the adult IAT and what it measures, see Greenwald & Lai, 2020; Payne et al., 2017; Schmader et al., 2022). Four adjustments of the adult IAT have already been made to make them appropriate for in-person child testing (Baron & Banaji, 2006; Cvencek, Meltzoff, & Greenwald, 2011; M. K. Qian et al., 2016). The first set of adaptations was driven by children's limited reading ability. Specifically, text stimuli were created to be age-appropriate, and text and images in the in-person child IAT work were presented simultaneously with recorded audio to allow children to hear and see stimuli at the same time (e.g., Baron & Banaji, 2006; Cvencek, Meltzoff, & Greenwald, 2011). Second, child IATs were also designed to utilize large, easily distinguishable response buttons that do not require familiarity with a standard computer keyboard (Baron & Banaji, 2006; Cvencek, Meltzoff, & Greenwald, 2011; M. K. Qian et al., 2016). Third, the overall number of trials in the standard adult IAT (180–200 trials) was reduced to 144–152 trials to limit the amount of time children needed to focus on the test (Cvencek, Meltzoff, & Greenwald, 2011). Finally, the IATs were designed for an experimenter to be present to give verbal instructions and positive feedback to ensure that the child understands the content and remains on task.

Educationally relevant research from the in-person child IAT to date has shown that: (a) elementary school children have developed implicit stereotypes and self-concepts about academic subjects, and (b) there appears to be a developmental timeline for these constructs. For example, one study of children in Grades 1–5 employed in-person IATs (Cvencek, Meltzoff, & Greenwald, 2011) to assess three constructs in the same children: (a) implicit gender identity (identifying as a boy or girl, e.g., *me* = *girl*), (b) implicit math–gender stereotypes (the belief that "math is for boys," *math* = *boys*), and (c) implicit math self-concepts (associating oneself with math, *me* = *math*). It was reported that there is a developmental sequence in which these constructs emerged. Gender identity was evident across all grades. Starting around Grades 1 and 2, boys associated math more with their own gender while girls associated math more with the opposite gender (an implicit math–gender stereotype that *math* = *boys*). Finally, math self-concepts ("I am a math person" or *me* = *math*) emerged later than gender identity and math–gender stereotypes. In this study by Cvencek, Meltzoff, and Greenwald (2011), all measures exhibited adequate reliability, Cronbach's $\alpha s \geq .74$.

History of Adapting the Adult IAT for In-Person Testing of Preschoolers

Following the successful use of the in-person child IAT, additional adaptations were made for measuring implicit attitudes in preschoolers as young as 4 years of age using a pictorial IAT (Cvencek, Greenwald, & Meltzoff, 2011; see Table 1). The pictorial version was designed to place much less demand on children's ability to read words on the computer screen. The initial in-person pictorial IAT was also designed to focus on topics of central concern to preschoolers, namely, young children's likes and dislikes for certain social groups or objects, and it yielded highly significant effects. Because good reading ability cannot be assumed, words were replaced by pictures to represent the IAT categories (e.g., a smiling face and a frowning face to represent the categories *good* and *bad*, respectively). Additionally, pictorial IAT instructions were also adapted to children's (somewhat) limited experience using computer keyboards (see below for more details). Lastly, like the text-based IAT, the pictorial IAT was developed assuming that an in-person experimenter would be present during the assessment period to provide age-appropriate verbal instructions and positive feedback. In the validation study of the in-person pictorial IAT by Cvencek, Greenwald, and Meltzoff (2011), the measure demonstrated good reliability, Cronbach's $\alpha = .85$.

Need for Online IAT Tools for Children

Across a broad swath of developmental and educational research, the need for new tools that can be used for remote, online testing is increasingly recognized (e.g., Córdova, 2020; Ibna Seraj et al., 2022). This leads to two gaps in the literature concerning children's implicit social cognition. First, trying to use the extant IAT tools online without a researcher present for delivering instructions to the children poses a unique set of challenges that require design adaptations. Second, there is a need for tools that can be used for remote testing not only for a narrow age span, but across multiple age groups in order to test a broad range of constructs that would be informative for theories of developmental and educational psychology.

Advantages and Challenges of Online Research

Even before the COVID-19 pandemic, the need for online platforms was recognized by researchers for their value in recruiting larger and more representative samples of participants into educational and developmental studies (Brummelman et al., 2022; Rhodes et al., 2020; Scott & Schulz, 2017; Sheskin & Keil, 2018). For the four studies proposed in this article, both the text-based IAT and the pictorial IAT have been redesigned so that the parent or adult caregiver need only provide consent for the child to participate. The online IAT tools introduced here do not require an experimenter to be physically present during the test or even online via synchronous screen contact. This will be of practical use because it opens up testing on weekends and after school at the convenience of the family, without coordinating schedules with university-based researchers or exiting the classroom during class time.

However, online adaptations without an experimenter are not without challenges (Gibson & Twycross, 2008; Gijbels et al., 2021). Adaptations of child studies for online use require subtle, age-specific considerations such as taking cognitive load into account or considering children's attention span. The main challenges for developmental scientists when developing online, unmoderated assessments are how best to maintain children's engagement, minimize environmental distractions, and motivate participation given varying attention spans for different age groups. In an online session, the test designers need to anticipate issues or events that might come up and adjust accordingly, whether it be troubleshooting technical problems, minimizing caregiver input, explaining instructions in detail, or providing necessary breaks when needed (Rhodes et al., 2020; Sheskin & Keil, 2018). Considerations of how tasks are structured and how stimuli are presented are also crucial to ensure a procedure's validity (Yeatman et al., 2021). In developing the materials and protocol for the proposed online studies, we paid special attention to each of these challenges and carefully designed the instructions (see Sections 5.1–5.4 in the online supplemental materials) and the tasks themselves (see below) to keep the children maximally engaged.

Table 1

Adaptations to Published In-Person IATs to Create New Online IAT Tools for Elementary School and Preschool Children

Test component	Adult IAT	Child in-person text-based IAT	Child online text-based IAT	Child in-person pictorial IAT	Child online pictorial IAT
Instructions	Text read by participant	Verbally presented by live experimenter	Text read by participant	Verbally presented by live experimenter	Text + recorded audio of experimenter
Category labels	Text	Text	Text	Pictures	Pictures
Stimuli	Text, pictures	Text + audio, pictures	Text + audio, pictures	Text + audio, pictures	Text + audio, pictures
Response buttons	Standard keyboard keys	Panels labeled with left/right arrows	Standard keyboard keys	Panels covered in solid colors	Standard keyboard keys
Blocks	7	7	7	7	7
Trials	180–200	152	152	136	136

Note. The adult IAT is virtually identical whether taken in person or online. For children, the adult IAT has been adapted to become four slightly different procedures. There is an in-person child testing procedure that requires an experimenter to administer. It has two versions, a text-based version and a pictorial version (the latter is suitable for children as young as preschoolers; see main text for details). There is also the new online child testing procedure that can be used for remote, self-administered testing of children. No involvement by a trained experimenter is necessary. This online procedure also has two versions, text-based and pictorial. The table describes the key adaptations and differences between these five test procedures (see the main text for further details). IAT = Implicit Association Test.

Developmental Implications

Previous research has demonstrated the usefulness of variants of the in-person text-based IAT with samples ranging in age between 6 and 10 years old (Baron & Banaji, 2006; Cvencek, Meltzoff, & Greenwald, 2011), and the usefulness of the pictorial IAT with samples ranging in age between 5 and 10 years old (Cvencek et al., 2018; del Río et al., 2021). (The pictorial IAT can be used starting at slightly younger ages than the text-based child IAT because it relies much more heavily on pictures for stimuli and category labels than on written words.) One of the chief goals of the proposed research is to report results using both text-based as well as pictorial online IAT tools for salient implicit psychological constructs across multiple ages.

Novel child IAT tools that can be used in a remote, unmonitored fashion across multiple age groups will have at least four uses for studying developmental continuity and change in children's implicit cognition in the future. First, it will enable the comparison of magnitudes of implicit attitudes and beliefs using an automated and standardized procedure that can be compared across age groups (and across multiple research labs). Second, the online pictorial IAT, in particular, will allow for a broad evaluation of early implicit attitudes and beliefs in children who are not yet facile readers. Third, both text-based and pictorial IAT tools use simple auditory and pictorial stimuli that could, in principle, be reliably standardized across languages and cultures to allow for rigorous cross-cultural comparisons using convenient instruments. Finally, these new tools also have the potential for investigating implicit attitudes pertaining to education, race, and self (e.g., math attitudes, race biases, self-esteem) and implicit beliefs (e.g., math self-concepts and stereotypes) at ages when implicit attitudes and beliefs are just beginning to form. This is vital because children's developing implicit social cognition may be more amenable to change during developmental windows when they are first being acquired, and somewhat less malleable and more resistant to lasting change once they have become more firmly entrenched (Baron, 2015; Cvencek et al., 2020; Halim et al., 2011; Master et al., 2021).

Rationale and Overview of the Proposed Studies

The article reports four studies using newly adapted online IATs for psychological and educational research with elementary- and preschool-aged children. One set of two studies (a) investigated elementary school children's gender identity using both in-person and online tools and (b) gathered data about math–gender stereotypes using the new online tool. Another set of two studies extended downward to test preschool children. In this young age group, we (a) investigated (repeatedly replicated) preferences that are polarized along an affective dimension (preferences for flowers over insects) using both in-person and novel online tools and (b) assessed a correlate of educational outcomes (in-group favoritism) using the novel online tool. Collectively, the four studies reported here covered a broad scope of children's cognitive (stereotypes and identities) and affective (attitudes toward social and nonsocial groups) orientations toward self, gender, and academic subjects, across an age span from preschoolers up through Grade 5.

For the online adaptation of the text-based IAT for elementary school children, we chose a socially relevant construct that has been thoroughly documented through in-person child research. Previous research has shown that children's implicit gender identity is significantly correlated with children's verbal reports of their own gender (Cvencek, Meltzoff, & Greenwald, 2011) and other theoretically expected measures (Cvencek et al., 2016). As an initial test of the validity of the new online IAT procedure, Study 1 therefore investigated gender identity using both the standard in-person procedure and the new online tool. In Study 2, the new online, text-based IAT was used in a self-administered procedure to investigate an educationally relevant construct, math–gender stereotypes, in elementary school students. Previous in-person results show that most elementary school students follow the typical adult stereotypic pattern on implicit tests and associate math with boys more than with girls (Cvencek, Meltzoff, & Greenwald, 2011; Cvencek et al., 2015).

A similar approach was used for the online adaptation of the pictorial IAT for preschool children. For the task in which a significant preference should be unequivocally expected, Study 3 examined clearly polarized categories along an affective dimension, comparing children's attitudes toward flowers versus insects using both in-person and online procedures. A preference toward flowers over insects is a phenomenon already documented in in-person testing of young children (Cvencek, Greenwald, & Meltzoff, 2011), and this predictable result lends itself well to validating the new online procedures in preschoolers. In Study 4, the new online pictorial IAT was used to investigate gender in-group biases in preschoolers. One of the earliest and strongest in-groups for children is based on gender. Previous in-person procedures using implicit measures with preschoolers have shown that both boys and girls associate their own gender more strongly with positive valence (Cvencek, Greenwald, & Meltzoff, 2011; Skowronski & Lawrence, 2001), thus showing a gender in-group bias (favoritism for their own gender).

All study procedures described here were approved by the University of Washington Human Subjects Division. Following acceptance of the Stage 1 registered report, the studies were preregistered on the Open Science Framework website (https://osf.io/wy423). Data collection took approximately 9 months. Data analysis and preparation of the article followed shortly thereafter. We estimated that the final Stage 2 article would be submitted within a year after this registered report was accepted (and this is what occurred).

Study 1—Elementary School Children: In-Person and Online Gender Identity IAT

Rationale

The online IAT created for this study assessed children's implicit gender identity. Previous in-person research on gender identity in children has established that boys tend to associate *me* = *boy*, and conversely that girls tend to associate *me* = *girl* (e.g., Cvencek, Meltzoff, & Greenwald, 2011; Cvencek et al., 2016; of course, this is at a group level, and there are individual differences among children). These findings satisfy the known group validity criterion (DeVellis, 2003). We examined gender identity in a sample of students attending Grades 2 and 3, which is an age group that has shown robust gender identity effects with an in-person procedure administered by an experimenter (e.g., Cvencek, Meltzoff, & Greenwald, 2011). In an independent group of participants within Study 1, we also assessed gender identity using the novel online version of the test.

Hypotheses

For Study 1, we tested the following four hypotheses:

> *Hypothesis 1 (H1)–Hypothesis 4 (H4): Replication of evidence of gender identity.* We hypothesized that in the in-person procedure, children would demonstrate a significant *me = own gender* identity (H1), and girls and boys separately would each show this effect (H2). We made the same predictions for the online procedure (H3 and H4). Because we cannot "prove" null results, we report 95% confidence intervals (CIs) and effect sizes for both in-person and online procedures without using null hypothesis tests to compare them analytically.

Method

Participants

Potential participants were recruited from a child human subjects registry maintained at the University of Washington. Parents were provided with an informational letter describing the study and interested families were then screened for eligibility over email or phone. To be eligible for the study, children had to be in Grade 2 or 3 and be free of developmental delays according to parental report. In this, as well as in all other studies that follow, we recruited girls and boys in approximately equal numbers.

Power Analyses. We first conducted a power analysis for the in-person procedure. Using G*Power software (Faul et al., 2007), we conducted an a priori power analysis to determine a reasonable stopping rule for data collection based on the effect size from previously published work using the in-person procedure, $d = 1.13$ (Cvencek, Meltzoff, & Greenwald, 2011). This power analysis indicated that, to detect a large effect size for the planned one-group t test (H1) assuming an α of .05 (two-tailed) and 80% power, a sample of only $N = 9$ participants would be needed. In line with recent recommendations within social psychology toward larger samples than required by power analyses (Fraley & Vazire, 2014), we chose to use a substantially larger sample of $N = 48$. We also chose this in part because of counterbalancing considerations ($n = 12$ per each of the four counterbalancing conditions; see Section 3 in the online supplemental materials for more details). Power analyses also indicated that, to conduct one-group t tests to evaluate the girls ($n = 24$) and the boys ($n = 24$) separately (H2), we would have 80% power to detect a mean difference from zero with an effect size of $d = 0.60$ (to be conservative, this effect size is below the in-person effect sizes in the published work, ds > 0.82; Cvencek, Meltzoff, & Greenwald, 2011).

For the online procedure, we conducted a power analysis based on the effect size found in a pilot study conducted using the online procedure, $d = 0.64$ (see Section 1.1 in the online supplemental materials). Using the same assumptions as above (i.e., a one-group t test with an α of .05, two-tailed, 80% power; H3), we found that a sample of $N = 22$ participants would be needed.[1] Again, for the same reasons given for the in-person power analysis, we chose to use a substantially larger sample of $N = 72$. Additionally, in order to test H4 using the $N = 72$ sample size (36 girls and 36 boys) for gender-specific one-group t tests of girls' and boys' means separately, we estimated 80% power for detecting a mean difference from zero with an effect size of $d = 0.48$ (to be conservative, this effect size is below the effect size in our online pilot study, $d = 0.64$).

Target *N* for Recruitment. To establish how many children needed to be recruited to achieve these desired *N*s for the analytic sample ($N = 48$ children for in-person procedures and $N = 72$ children for online procedures), we next took into account data exclusion and estimated that approximately 20% of data would be excluded (for further explanation, see the Acceptable Exclusions section below). We therefore planned to recruit $N = 60$ total children for the in-person procedure and $N = 90$ total children for the online procedure in order to achieve our analytic sample sizes of $N = 48$ and 72, respectively. As described below, we only needed to recruit 49 and 83 children to achieve our preregistered analytic sample sizes (see Table 2). (In this and all other studies, the stopping rule was that half of the analytic sample, after exclusions, would be girls and half would be boys.)

Analytic Samples. For the in-person procedure, of the $N = 48$ children who comprised the final analytic sample, $n = 24$ were girls, and $n = 24$ were in Grade 2. In order to achieve the preregistered $N = 48$ analytic sample size, we only needed to recruit 49 participants (i.e., only one participant who completed the test was eliminated based on IAT exclusion criteria; see Tables 2 and 3). The mean age of these 48 children was $M = 8.39$ years, $SD = 0.54$ (age range = 7.53–9.46 years old). Race for participants (Supporting Information, Appendix A in the online supplemental materials) was 72.9% White, 22.9% multiracial, and 4.2% Asian; 85.4% were non-Hispanic and 4.2% of participants were of Hispanic ethnicity (10.4% did not report whether or not they were of Hispanic ethnicity). Self-report of parental education (Supporting Information, Appendix A in the online supplemental materials) was as follows: 2.1% did not have a college degree, 43.8% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 54.2% had a graduate degree. Finally, as an indicator of socioeconomic status, an "income-to-needs" ratio (Daly et al., 2002) was computed by dividing each participant's household income by the poverty threshold based on their family size (U.S. Census Bureau, 2024). The average income-to-needs ratio was $M = 8.93$, $SD = 6.23$ (range = 2.59–33.00).

For the online procedure, a total of 83 participants were recruited to participate, of whom 79 completed the IAT (Table 3). Seven participants with completed IAT data were then excluded (8.9%) based on the IAT response criteria (Table 3), resulting in $N = 72$ participants in the analytic sample, as preregistered. Of those 72, $n = 36$ were girls, and $n = 36$ were in Grade 2. The mean age of these 72 children was $M = 8.25$ years, $SD = 0.54$ (age range = 7.18–9.40 years old). Race for participants (Supporting Information, Appendix B in the online supplemental materials) was 79.2% White, 18.1% multiracial, and 1.4% Asian (1.4% did not report race); 87.5% were non-Hispanic and 11.1% of participants were of Hispanic ethnicity (1.4% did not report whether or not they were of Hispanic ethnicity). Self-report of parental education (Supporting Information, Appendix B in the online supplemental materials) was as follows: 4.2% did not have a college degree, 43.8% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 50.7% had a graduate degree (1.4%

[1] The results of two pilot studies suggest that the online procedures may produce smaller effect sizes than the in-person procedures; hence, a slightly larger sample size was needed for the online procedures (see Sections 1.1 and 1.2 in the online supplemental materials).

Table 2
Preregistered Exclusion Criteria for Studies

Exclusion type	Criterion	Rationale
Demographics		
Missing data (Studies 1 and 2)	Missing gender, birthdate, grade level	Crucial missing data
Missing data (Studies 3 and 4)	Missing gender, birthdate	Crucial missing data
Response criteria		
Too many errors	Errors on 35% or more of combined-task trials	Preestablished criteria to reduce noise from outliers
Too fast	Responses under 300 ms on 10% or more of combined-task trials	Preestablished criteria to reduce noise from outliers
Too slow	Mean response latency across combined-task trials $\geq$3 *SD*s above the sample mean	Preestablished criteria to reduce noise from outliers
Incomplete measures	Does not finish implicit measure	Would yield missing data

Note. For the child text-based and pictorial IATs, formal exclusion criteria for in-person studies were established by Cvencek, Greenwald, and Meltzoff (2011) and Cvencek, Meltzoff, and Greenwald (2011). If a participant in any study meets any of these criteria, all of their data will be excluded from the analyses. Therefore, the analytic data set for each study will not have any missing data. IAT = Implicit Association Test.

did not report parental education). The average income-to-needs ratio was $M = 7.60$, $SD = 4.00$ (range = 1.17–22.65; 20.8% of participants did not provide all necessary information to compute the ratio; see Section 6 in the online supplemental materials for details).

Materials and Procedure

Families who participated as part of the in-person group met an experimenter at our lab. The parent signed a consent form allowing their child to participate and filled out a brief demographics survey (Supporting Information, Appendix A in the online supplemental materials). Then the researcher described the study to the child and obtained the child's assent (as per Institutional Review Board requirements). If the child said "yes" to participating, then the study began. If the child said "no," the experimenter asked one time if they were sure they did not want to participate, and if the child said "no" again, the child was not tested (all in-person participants assented). For those who gave their assent, the experimenter then sat with the child at the testing computer and the parent sat behind them (unless they chose to stay out of the room for their own reasons).

Families who participated as part of the online group used a laptop or a desktop computer with an internet connection to access an online administration platform. Upon clicking the link provided in an email, parents confirmed consent for their child to participate and completed a brief demographics survey (see Supporting Information, Appendix B in the online supplemental materials), after which they were instructed, via automated software, to ensure their child is seated in front of the computer with the mouse and keyboard at hand. Parents were encouraged to leave their child alone to complete the test procedure but they could stay and observe (as per the Institutional Review Board). Children were presented with a description of the study and asked to provide their assent to participate by using a mouse to click on a clearly marked button for "yes" or "no." If a child clicked on the "yes" button, the procedure began, and if they clicked on the "no" button, a confirmation page appeared; the application closed if they clicked "no" a second time, confirming that they do not wish to assent (all participants assented).

Table 3
Sample Sizes and Quality Checks for Studies 1–4

	Sample sizes			Quality checks					
							Method effects of counterbalancing (expected $p > .05$)		
Study and procedure	Recruited sample, *N*	Completed sample, *N*	Analytic sample, *N*	Elimination of double entries, *n* eliminated	Exclusions (up to 20% acceptable), *n* (%) excluded	Assessing reliability ($\alpha \geq .70$), α	p_{order}	p_{side}	Normalcy assumption, skewness
1. Gender identity									
In-person procedure	49	49	48	0	1 (2.0%)	.812	.148	.386	−0.139
Online procedure	83	79	72	0	7 (8.9%)	.883	.819	.107	−0.429
2. Math–gender stereotype									
Online procedure	80	78	76	1	2 (2.6%)	.781	.333	.874	0.128
3. Flower–insect attitude									
In-person procedure	71	67	64	0	3 (4.5%)	.733	.149	.255	−0.384
Online procedure	71	69	64	1	5 (7.2%)	.875	.679	.829	−0.260
4. Gender in-group bias									
Online procedure	80	72	64	0	8 (11.1%)	.868	.530	.425	−0.196

Note. Recruited sample = children who either came to the lab (in-person) or began the online study; Completed sample = children who completed the IAT; Analytic sample = children whose data were analyzed (i.e., participants who completed the IAT and who were not excluded by the preregistered exclusion criteria; see Table 2). p_{order} and p_{side} refer to p values associated with the main effects of IAT order (congruent vs. incongruent task first) and side (left–right side of screen on which the IAT categories first appear). See the Materials and Procedure section in Study 1 and the third paragraph of the Implicit Gender Identity section (below) for full details. IAT = Implicit Association Test.

The in-person and online procedures were programmed using Inquisit Version 6.6.1 (Millisecond Software, 2022) and each was completed in one session (using independent groups of participants). Families received a $10 gift card as thanks for their participation regardless of whether they assented or finished the procedure. The test session took approximately 15 min for the in-person procedure and approximately 20 for the online procedure. Each participant completed one text-based IAT. As requested during the Stage 1 review process, for exploratory purposes, children also answered an exit survey containing questions about their liking, interest, motivation, and self-efficacy regarding the IAT for both the in-person and online procedures, as well as a question about the presence of other people in the room (see Section 2 in the online supplemental materials for details).

Implicit Gender Identity. The text-based IAT is a computerized sorting task in which children sort stimuli into four categories as quickly as possible using two response buttons. As in the adult IAT, the text-based child IAT contains two critical tasks with opposite pairings of categories. It is assumed that children will sort the stimuli faster during the task in which they have a stronger mental association between the paired categories. Consider an illustrative example: If the categories *men* and *aggressive* were paired with each other, and the categories *women* and *nurturing* were paired with each other, participants would be predicted to respond quickly because those pairings may already be linked in their mind. If the pairings were reversed so that *women* and *aggressive* were paired, and *men* and *nurturing* were paired, participants would respond more slowly because those pairings are not as strongly linked in their mind (of course there may be individual differences between children depending on their experiences, which is one reason the IAT tool is so interesting). To take a simpler, more age-appropriate example, if a child verbally identifies as a girl (i.e., "I'm a girl"), then on a child IAT she would be expected to sort the stimuli faster when *me* and *girls* share a response button (called the "congruent pairing" because it reflects the sorting task that the child is expected to find easy given their reported identity) versus when *me* and *boys* share a response button (called the "incongruent pairing"). This is the logic that underlies the assessment of implicit gender identity using an IAT.

The categories (shown in italics below) and the 18 individual stimulus words (shown in parentheses) of the gender identity text-based IAT were: *me* (stimuli: "I," "me," "my," "myself"), *not me* (stimuli: "other," "theirs," "them," "they"), *boys* (stimuli: "Andrew," "David," "Jacob," "Michael," "William"), and *girls* (stimuli: "Emily," "Jennifer," "Jessica," "Rachel," "Sarah").

The length (number of trials) and block structure of the proposed in-person and online text-based procedures were identical. More specifically, Blocks 1 and 2 (16 trials each) were single-task warm-up blocks in which children sorted stimuli from only two categories per block using two response keys (e.g., sorting the *me* stimuli to the left using the left response key and sorting the *not me* stimuli to the right using the right response key, in Block 1). These "single-task" blocks were intended to help children gain familiarity with the rules of the game. Blocks 3 and 4 (24 trials each) presented the first "combined task," in which the children sorted stimuli from four categories using two response keys (e.g., sorting *me* and *girls* stimuli using the left response key and sorting *not me* and *boys* stimuli using the right response key). Block 5 (24 trials) was a single-task reversal block in which the positions of two of the categories were switched (e.g., sorting *not me* stimuli using the left response key and sorting *me* stimuli using the right response key). This allowed children to practice the new left–right orientations before completing the next round of combined-task blocks. Blocks 6 and 7 (24 trials each) were the second round of combined-task blocks which had the opposite pairings to those used in Blocks 3 and 4 (e.g., if the earlier blocks used the congruent pairing, the later combined blocks would use the incongruent pairing, and vice versa). The order of the congruent versus incongruent blocks within the text-based IAT was counterbalanced. This foregoing seven-block structure followed the established adult procedure of Greenwald et al. (1998) and is used on the Project Implicit Demo Site, but we adapted the adult test protocol in several ways to make it user-friendly for children (see Table 1). All word stimuli were recorded in a female voice because previous research comparing presentation of gender-related IAT stimuli has not found any significant differences in IAT scores when a male versus female voice was used (Cvencek et al., 2016, p. 52).

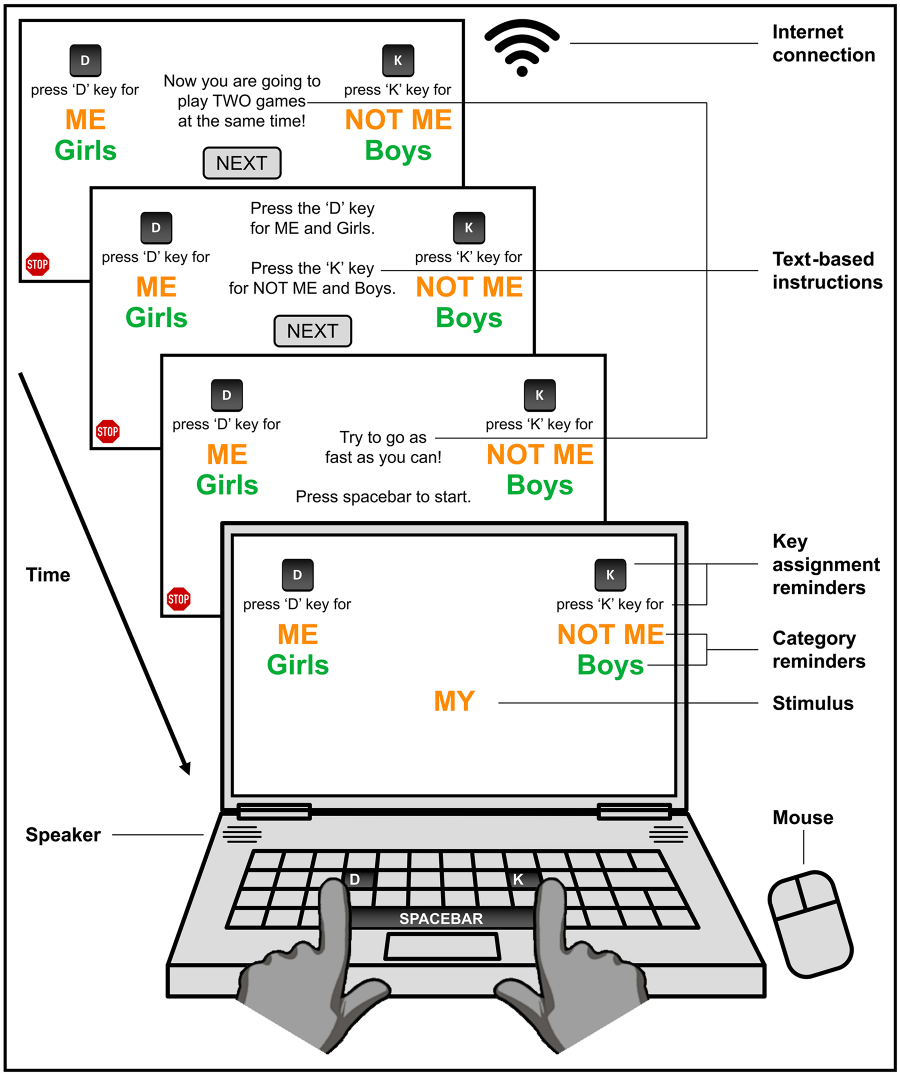
All stimuli were randomly extracted from the 18-word stimulus pool without replacement (independently for each child) until the available stimuli for a task were exhausted. At that point, the stimulus pool was replaced (with the original 18-word stimulus pool) for any additional trials that may have been needed. This resulted in different children being presented with slightly different versions of the test, which is standard best practice for IAT research with adult (Greenwald & Lai, 2020; Greenwald et al., 1998) and child (Cvencek, Meltzoff, & Greenwald, 2011; Cvencek et al., 2016) participants.

The stimuli were presented one at a time in the center of the screen. Children were provided with error feedback (in the form of a red question mark) and were asked to press the correct response key in order to proceed. In line with best practices for adult IAT research (Greenwald & Lai, 2020; Greenwald et al., 2003), a procedure that records latency to occurrence of the correct response was used: The IAT software recorded occurrence of error responses, and the trial's latency was the latency to the correct response. This follows the established adult IAT protocol (Greenwald et al., 2003).

The text-based IAT score used in all analyses was computed as the standard *D* score described in the adult literature, with a rational 0 value (Greenwald et al., 2003). The combined-task blocks (Blocks 3 and 4 and 6 and 7) were used in calculating its *D* score. The *D* score is the difference between the mean response latencies of the combined-task blocks (calculated as incongruent minus congruent) divided by the pooled standard deviation of all combined-task responses. The gender identity *D* score ranged from -2 (*me* = *opposite gender*) to $+2$ (*me* = *own gender*), with a rational 0 value indicating equal association of *me* with the child's *own gender* and *opposite gender*.

Further Specialized Adaptations to Create an Online Text-Based IAT for Gender Identity. Table 1 shows a comprehensive list of the adaptations made to the in-person text-based IAT so that it could be used for the online procedure. Figure 1 shows a visualization of the online text-based IAT gender identity apparatus. Written instructions (shown in Figure 1 as "text-based instructions") were presented to the child on screen throughout the test because there was no experimenter present. As shown in Figure 1, the presentation of each text stimulus was synchronized with the verbal pronunciation of that word through the computer speakers. Because it was not possible to provide a customized

Figure 1
Child's View of the New Online Text-Based IAT in Study 1



Note. This figure displays the online gender identity text-based IAT from Study 1; the setup remained identical in Study 2 (math–gender stereotype) except for the categories/stimuli used (see the main text). Text-based instructions appear on each page prior to the child beginning each IAT block. Children press the spacebar to start each IAT block, at which point the trials advance automatically. After pressing spacebar to begin the IAT block, the child sees one (centered) stimulus per trial, as shown in the final screen of this figure. The mouse is used to click the "NEXT" button to go through each page of instructions. IAT = Implicit Association Test. See the online article for the color version of this figure.

keyboard to each participant (as was used with the in-person procedure), the online procedure allowed the participant to classify stimuli by pressing the "D" and "K" keys on a standard keyboard, and to begin each block of the text-based IAT using the spacebar (Figure 1), as is commonly done in adult IATs. Section 5 in the online supplemental materials provides verbatim instructions for how the mapping of the "D" and "K" keys was explained to children; this procedure was successfully used with young children in a pilot study reported in Section 1.2 in the online supplemental materials.

The child's task was to classify each stimulus by pressing the response key ("D" or "K") associated with the correct category. Each child proceeded through the instructions at their own chosen pace by clicking on a button on the screen labeled "NEXT" when they were ready to move on to the next page (see Figure 1). The child had the option to stop the session by clicking on a clearly labeled "STOP" button on the screen (before each block of trials started) or pressing the escape key (during any trial). The computer was programmed to provide positive feedback to the child after each block of the test (e.g., using text reading "Good job!"). (These computerized instructions had been worked out in collaboration with parents and their children to maximize child-friendly language and ensure a test duration that children this age can manage, and many seem to enjoy.)

Quality Checks

Five preregistered criteria were used as quality checks prior to evaluating the hypotheses, including: (a) eliminating double entries, (b) implementing exclusion rules, (c) assessing measurement reliability, (d) testing for method effects, and (e) evaluating the data for nonnormality and nonindependence. These quality checks were conducted to establish the final "analytic sample" that was used to test the hypotheses with $N = 48$ total children for in-person procedure, and $N = 72$ total children for online procedure. (These quality checks apply not only for Study 1 but for all four proposed studies.)[2]

Eliminating Double Entries. The first quality check involved screening the online data file for participants who might have taken the online procedure twice by noting duplicate participant numbers and Internet protocol addresses (both captured automatically by the online administration platform). In such cases, only the data from their first complete session would be retained in our final data file. For the in-person procedure, the recruiter and experimenter ensured that each participant only took the test once.

Acceptable Exclusions. Child data were excluded from analyses based on preestablished in-person IAT exclusion criteria (see Table 2). Participants who met one or more of these criteria had all data excluded from analyses (thus there was no missing data in the analytic sample). We then tabulated the percentage of exclusions for comparisons to the "usual" exclusion rates found with in-person child IAT research. To this aim, we conducted a preliminary review of the current literature to determine the typical exclusion rates in child IAT in-person research. Our search yielded 171 published in-person child IAT articles with participants from 4 to 17 years old (the search closed on December 6, 2022). We found 98 articles (total number of participants = 26,427) that reported exclusions based on error rates and/or fast or slow responding. The mean exclusion rate was 7.2%; however, for our current studies, which are on the younger end of that age range, we took a (preregistered) exclusion rate up to 20% as showing that an acceptable number of children completed the in-person procedure as intended (e.g., not making too many errors and not responding slowly on too many trials). We also expected that the online procedure might have a higher number of participants who do not finish the task (e.g., due to computer problems such as audio not working, screen freezes), and that they might find it somewhat less fun due to lack of social support by the experimenter, which might increase the exclusion rate. Thus, an exclusion rate up to 20% was accepted for both the online and in-person procedures.

Assessing Reliability. The reliability of the gender identity text-based IAT in both the in-person and online procedures was assessed using Cronbach's α (internal consistency). Specifically, we randomly selected two sets of 48 trials from the IAT's combined-task blocks and computed a *D* score for each; the *D* values were then used to compute Cronbach's α. Cronbach's α for the gender identity in-person text-based IAT in Cvencek, Meltzoff, and Greenwald (2011) was $\alpha = .89$. We expected both the in-person and online implicit procedures to also have an acceptable alpha (i.e., $\alpha \geq .70$), which would indicate that children's responses are consistent across the items and trials used in the procedures.

Potential Effects of Counterbalancing Factors (Method Effects). We checked for possible effects of methodological counterbalancing factors, which involved IAT congruency order (congruent task first vs. incongruent task first) and the left–right side of screen on which the IAT categories first appeared, neither of which were expected to reach significance in either the in-person or online procedures (they were not significant for previous uses of the in-person procedure). For the IAT, the mean *D* score was entered as the dependent variable in a one-way analysis of variance, with congruency order (two levels: congruent task first vs. incongruent task first) and the side of screen on which the IAT category pairs appear (two levels: left vs. right) entered as between-subjects factors. At this juncture, the Stage 1 report proposed the following:

> If we find that there are method effects due to congruency order or screen location, those factors will be controlled for in the final analyses by effect-coding the condition when testing our focal hypotheses (i.e., instead of using 1- and 2-group *t*-tests, we may use linear regression).

However, no such method effects were found, and therefore these extra steps to control for them were not needed.

Evaluating Data for Statistical Assumption Violations. The fifth quality check involved checking two focal statistical assumptions that could adversely affect inferences for testing our hypotheses. First, we checked the distributional shape of the observed scores for the *D* scores (although we note that the normality assumption in statistical models applies to the residuals, not the observed scores, and the sampling distribution of most statistics with a sufficiently large sample size is normal in shape despite nonnormality of frequency distributions). At this juncture, the Stage 1 report proposed the following:

> If moderate skewness was detected, we will standardize the score values and evaluate the data for outliers exceeding three standard deviations from zero. Those values will be noted and when final analyses take place, we will analyze the data with and without the outlier(s) (our preference will be to retain all data). If we find severe skew, then we will transform the data using an inverse function and analyze the data in its original units and in its transformed units. If the results were substantively the same, we will prefer to use the original data (un-transformed). We wish to underscore, however, that given the previous in-person text-based IAT studies found that the *D* scores were fairly normally distributed, we do not believe non-normality will pose issues for our analyses.

As projected in this quoted section of the Stage 1 report, nonnormality was not a problem. This is documented in Table 3 (and in the reported results of the Quality Checks section below). Therefore, the extra steps were not needed.

Planned Analyses (H1–H4)

Analyses were conducted in SPSS. We used an α of .05, two-tailed for all statistical analyses in Study 1 and throughout the article. Analyses of the data collected in person and online were identical.

We predicted the significance of own-gender identity in both the online and in-person versions of the procedures. Specifically, in the in-person version of the procedure, we conducted a one-group *t* test for evaluating the mean level of gender identity against a null of zero

[2] The present research is not an experiment with a treatment manipulation; rather, it involves correlational studies comparing means to a neutral value on three constructs pertaining to implicit social cognition: identities, stereotypes, and attitudes. Thus, inclusion of a positive control condition was not applicable here.

(H1) as well as separately for the girl participants and the boy participants (H2). The same statistical tests were used for the online version of the procedure to evaluate H3 and H4. In addition to reporting our statistical test results, we also report Cohen's d as our effect size measure, as well as 95% CIs to describe the range of plausible values for online and in-person samples alike.

Transparency and Openness

In the Power Analyses and Quality Checks sections above, we reported how we determined our sample size, the quality checks to be applied to the data, data exclusions (if any), and all procedures in the study. All data, analysis code, stimuli, and instructions have been made available on the Open Science Framework (https://osf.io/w35zk). Data were analyzed using SPSS Version 29 (IBM Corp., 2022). All methods and procedures in this study were preregistered upon in-principle acceptance. The above transparency and openness statement applies to all Studies 1–4.

Results

Quality Checks

Table 3 shows an overview of the quality checks for Study 1, as well as each of the other three studies. As can be seen from the first line, the data from the in-person procedure passed all of the preregistered quality checks: (a) there were no double entries, (b) 2.0% of participants were excluded (which is below the acceptable 20% exclusion rate that was preregistered), (c) the internal consistency was very good, $\alpha = .812$, (d) the method effects (counterbalancing factors) were not statistically significant, both $p\text{s} > .14$, and (e) the data were not significantly skewed, skewness = −0.139. Table 3 similarly shows that the data from the online procedure passed all of the preregistered quality checks: (a) there were no double entries, (b) 8.9% of participants were excluded (below the acceptable 20% exclusions rate), (c) the internal consistency was very good, $\alpha = .883$, (d) the method effects were not statistically significant, both $p\text{s} > .10$, and (e) the data were normally distributed, skewness = −0.429.

Preregistered Hypothesis Tests

All four of the preregistered hypotheses were supported. Supporting H1, for the in-person procedure, a one-group t test showed that the mean level of gender identity was positive (i.e., *me = own gender*) and significantly different from zero, $t(47) = 6.57$, $p < .001$, $d = 0.95$, 95% CI [0.60, 1.29] (Figure 2). Supporting H2, significant results were also obtained separately for the girl participants, $t(23) = 6.85$, $p < .001$, $d = 1.40$, 95% CI [0.82, 1.96], and for the boy participants, $t(23) = 3.20$, $p = .004$, $d = 0.65$, 95% CI [0.21, 1.09] (Figure 3).

Supporting H3, for the online procedure, a one-group t test showed that the mean level of gender identity was positive (i.e., *me = own gender*) and significantly different from zero, $t(71) = 6.12$, $p < .001$, $d = 0.72$, 95% CI [0.46, 0.98] (Figure 2). Supporting H4, significant results were also obtained separately for the girl participants, $t(35) = 8.40$, $p < .001$, $d = 1.40$, 95% CI [0.93, 1.86], and for the boy participants, $t(35) = 2.03$, $p = .050$, $d = 0.34$, 95% CI [0.00, 0.67] (Figure 3).

Preregistered Exploratory Analyses

Participants' responses to the exit survey items on liking, interest, motivation, and self-efficacy did not differ between the in-person and online procedures using independent groups t tests, all $p\text{s} > .14$. As expected, children in the in-person procedure reported more people being in the room with them during the IAT than did children in the online study, $t(118) = 3.98$, $p < .001$, $d = 0.74$, 95% CI [0.36, 1.12]. This is unsurprising because there was always at least one other person with them for the in-person procedure—the experimenter—whereas the online participants could have completed the IAT without anyone else in the room.

Discussion

The goal of Study 1 was to establish the validity of the new online, unmoderated tool for assessing children's implicit cognition. Supporting all four of the preregistered hypotheses, we found a highly significant own-gender identity for both the in-person (H1 and H2) and the online, unmoderated procedures (H3 and H4). Importantly, the effect sizes for the new online procedure were comparable in magnitude and direction to those obtained with the in-person procedure. With regard to psychometric properties, the online tool exhibited high internal consistency, was resistant to method effects (counterbalancing of order and side), and the scores were normally distributed. Taken together, this provides evidence that the new online, unmoderated procedure is a valid measure of children's implicit social cognition.

Study 2—Elementary School Children: Online Math–Gender Stereotype IAT

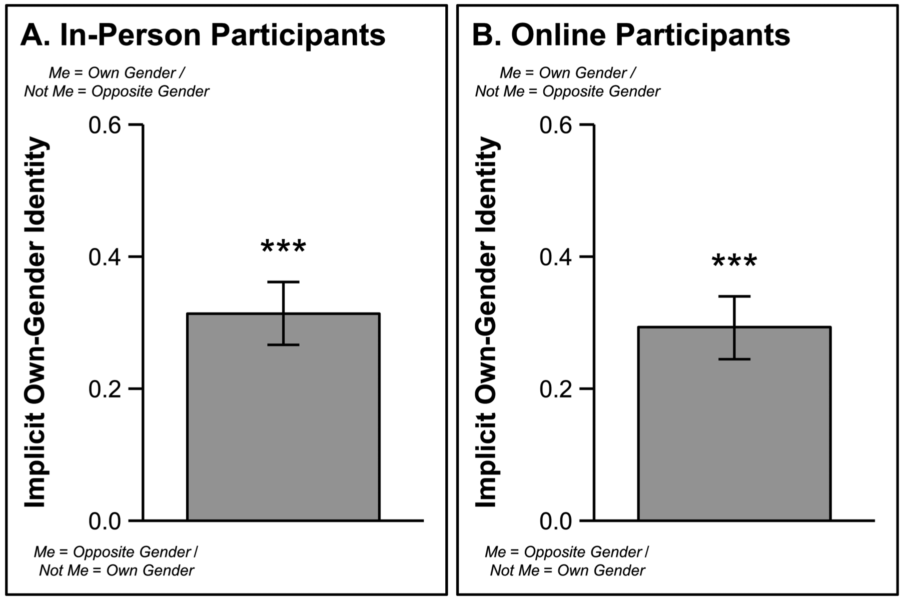
Rationale

Study 2 was an online assessment of children's implicit math–gender stereotypes based on Cvencek, Meltzoff, and Greenwald's (2011) in-person study. This study used the adaptations for the online administration procedures described in Study 1. Implicit math–gender stereotypes have been widely reported for children in Grades 2–5 in the literature, but the format of the particular child IATs used has sometimes not been age-appropriate and some studies have reported null effects, especially in the youngest samples (Hildebrand et al., 2022). Also, the extant research suggests that the strength of math–gender stereotypes tends to increase with age over the course of elementary school (Cvencek et al., 2015). Accordingly, we examined math–gender stereotypes in a sample of late elementary students (i.e., students attending Grades 4 and 5), rather than in younger age groups. The educational importance of implicit math–gender stereotypes during elementary school years is that they have previously been shown to be predictive of students' math achievement (e.g., Cvencek et al., 2015; Steffens et al., 2010). Furthermore, activating implicit math–gender stereotypes in girls has been shown to negatively influence girls' performance on standardized math test (Galdi et al., 2014). Master et al. (2021, 2025) provide detailed discussions of the educational and psychological importance of science, technology, engineering, and mathematics stereotypes in elementary school, based on self-reported responses from children.

Hypothesis

For Study 2, we tested the following hypothesis:

Figure 2
Study 1 Overall Mean Implicit Own-Gender Identity



Note. Overall means of implicit own-gender identity for in-person (Panel A) and online (Panel B) participants. Forty-eight in-person participants; 72 online participants. Error bars represent *SE*s.
*** $p < .001$.

H1: Replication of evidence of math–gender stereotypes. We hypothesized that, using the online procedure, Grade 4 and 5 children would demonstrate a significant *math* = *boy* stereotype.

Method

Participants

Recruitment procedures and participant eligibility were the same as in Study 1, except for the grade levels of the children. Participants were elementary school girls and boys attending Grades 4 and 5 and recruited in approximately equal numbers, and the data quality checks were also the same as those described in Study 1. We used the same power analysis approach for Study 2 with the same inputs for α (.05) and power (80%) and relied on the previously observed effect size from the math–gender stereotype results of the in-person study (Cvencek, Meltzoff, & Greenwald, 2011) computed by selecting Grades 4 and 5, $d = 0.44$. This power analysis indicated that to detect a medium effect size for the planned two-tailed, one-group t test, with an α of .05 and 80% power, a sample of 43 participants would be needed. Following the rationale used in Study 1, we chose to use a larger sample ($N = 76$). A sample size of $N = 76$ respondents provided us with 80% power for detecting a medium-sized mean difference, $d = 0.33$, in math–gender stereotypes using a one-group t test (H1). Similar to Study 1, we expected to recruit 20% more children than $N = 76$ in order to account for expected exclusions (see the criteria in the Acceptable Exclusions section) and smaller effect sizes for the online testing procedures, for a total of approximately $N = 96$ children needing to be recruited in order to obtain the analytic sample of $N = 76$. All of the participants of this study (and each of the four studies) only participated in a single study.

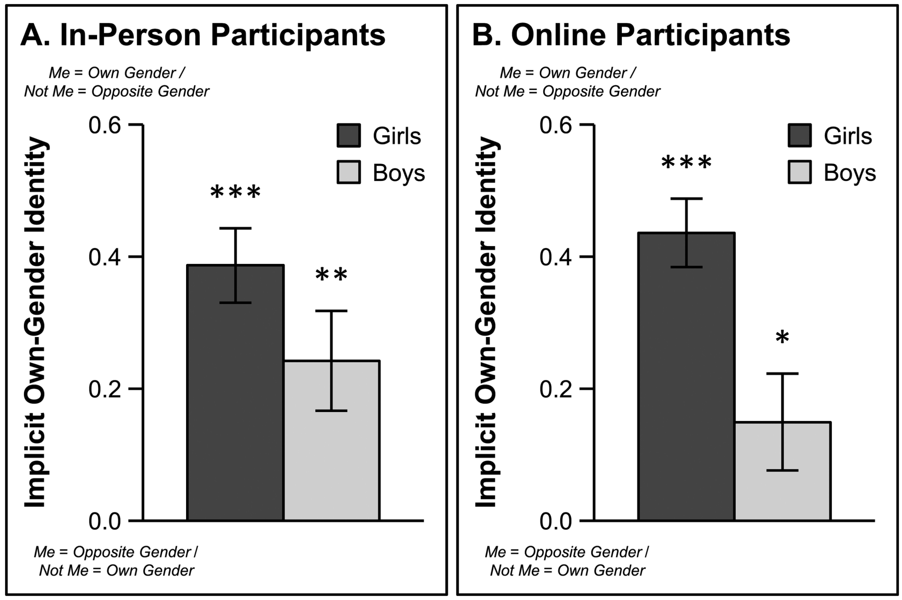
Analytic Sample. Of the $N = 76$ children who comprised the final analytic sample, $n = 38$ were girls, and $n = 38$ were in Grade 4. The mean age of these 76 children was $M = 10.19$ years, $SD = 0.54$ (age range = 9.15–11.25 years old). A total of 80 participants were recruited to participate in the online math–gender stereotype study, of whom 78 completed the IAT. Completed IAT data were excluded for $n = 2$ participants (2.6%) based on the three IAT response criteria (see Table 3). Race for participants was 78.9% White, 15.8% multiracial, 3.9% Asian, and 1.3% Black/African American, and additionally, 94.7% were non-Hispanic and 2.6% of participants were of Hispanic ethnicity (2.6% did not report whether they were of Hispanic ethnicity). For parental education, 7.2% did not have a college degree, 42.1% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 48.0% had a graduate degree (2.6% did not report parental education). The average income-to-needs ratio was $M = 6.14$, $SD = 2.83$ (range = 1.23–16.18; 15.8% of participants did not provide all necessary information to compute the ratio).

Materials and Procedure

The consent, demographics, assent, gift card, and other procedures were identical to those used with the online procedure in Study 1 unless otherwise specified below. The test session lasted approximately 15 min.

Implicit Math–Gender Stereotype. The categories in the math–gender stereotype text-based IAT were *math* (stimuli: "addition," "count," "graph," "math," "numbers"), and r*eading* (stimuli: "books," "letters," "read," "sentence," "story"), along with the same *boys* and *girls* categories as in Study 1. The math–gender stereotype D score ranged from −2 (*math* = *girls*) to +2 (*math* = *boys*), with a rational 0 value indicating equal association of *math* with *boys* and *girls*.

Figure 3
Study 1 Mean Implicit Own-Gender Identity by Gender



Note. Means of implicit own-gender identity for in-person (Panel A) and online (Panel B) participants, split by gender. Twenty-four in-person girls and 24 in-person boys; 36 online girls and 36 online boys. Error bars represent *SE*s.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Further Specialized Adaptations to Create an Online Text-Based IAT for Math–Gender Stereotypes. The same adaptations were made to the text-based IAT procedures as in Study 1 (see Table 1).

Results

Quality Checks

For Study 2, the identical five quality checks were used as in Study 1. As can be seen from the third line of Table 3, all of these preregistered quality checks were passed for the math–gender stereotype data: (a) there was one double entry (which was eliminated prior to analyses), (b) 2.6% of participants were excluded (which is under the acceptable 20% exclusion rate), (c) the internal consistency was good, $\alpha = .781$, (d) the method effects were not statistically significant, both $ps > .33$, and (e) the data were normally distributed, skewness $= 0.128$.

Preregistered Hypothesis Test

Supporting H1, the mean level of math–gender stereotype was positive (i.e., *math* = *boy*) and significantly different from zero, $t(75) = 6.85$, $p < .001$, $d = 0.79$, 95% CI [0.53, 1.04] (Figure 4).

Discussion

Study 2 extended the new online, unmoderated procedure to investigate an educationally relevant construct—implicit gender stereotypes about math and reading. Supporting our preregistered hypothesis, the online tool was sensitive to expected math–gender stereotypes: Children associated *math* with *boys* more strongly than *math* with *girls*. This result replicated previously published in-person findings with elementary school children, suggesting that gender stereotypes about math are still strongly evident among U.S. children in 2024, at least at the implicit level. Similar to the results of Study 1 (assessing implicit gender identity), the online tool exhibited good internal consistency and was resistant to method effects (counterbalancing factors).

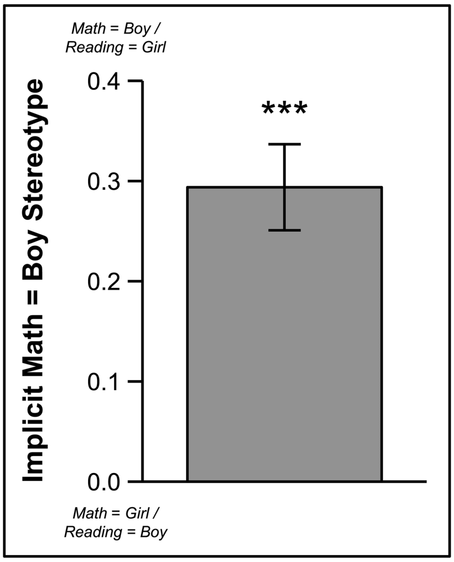
Study 3—Preschool Children: In-Person and Online Flower–Insect Attitude IAT

Rationale

Participants in Study 3 (as well as in Study 4) were 5-year-old children. We used the same pictorial IAT measuring attitudes toward flowers and insects that were used in the published in-person study (Cvencek, Greenwald, & Meltzoff, 2011). The strong preference for flowers was used by Greenwald and colleagues in an initial validation of the adult IAT (Greenwald et al., 1998). In subsequent IAT research with adults, the flower–insect IAT was repeatedly used to examine the effects of procedural changes on new variations of the adult IAT (e.g., Klauer & Mierke, 2005; Teige-Mocigemba et al., 2008), as well as to validate other IAT adaptations (Carpenter et al., 2019). Similarly, the flower–insect contrast was used to initially validate an in-person IAT for preschoolers (Cvencek, Greenwald, & Meltzoff, 2011; Thomas et al., 2007).

Study 3 included two groups of preschool participants, using a similar strategy as in Study 1: one group completed an in-person

Figure 4
Study 2 Overall Mean Implicit Math–Gender Stereotype



Note. Mean of implicit math = boy stereotype for online participants. $N = 76$. Error bar represents *SE*.
*** $p < .001$.

version with an experimenter in a laboratory setting, and the other group completed the new online version at home without an experimenter.

Hypotheses

We tested the following two hypotheses:

> *H1 and H2: Replication of evidence of flower preference.* We hypothesized that in the in-person procedure, preschool children would demonstrate significant preference for flowers (*flowers = good*; H1). We made the same prediction for the online procedure (H2).

Method

Participants

Participants were identified using the same recruitment strategy as in the previous two studies and had to meet all the same criteria as in Studies 1 and 2, except that participants in Study 3 were 5-year-olds.

We used the same power analysis approach for Study 3 as described for Study 1, which also had both an in-person and online procedure. For the in-person procedure, we used the same inputs for α (.05) and power (80%) and relied on the previously observed effect size from the flower–insect attitudes result of the in-person Cvencek, Greenwald, and Meltzoff (2011) study, $d = 0.50$. We found that a sample of 34 participants would be needed. Following the same rationale used in Studies 1 and 2, we chose to use a larger sample of $N = 64$. This provided us with 80% power for detecting a medium-sized mean difference, $d = 0.36$, in flower–insect attitudes using a one-group t test (H1).

For the online procedure, we used the same power analysis approach as in the previous two studies, setting the α at .05 and power at 80%, and used the observed effect size from the flower–insect attitudes result of our online pilot, $d = 0.65$ (see Section 1.2 in the online supplemental materials). We found that a sample of 21 participants would be needed, but we again chose to use a larger sample of $N = 64$. This provided us with 80% power for detecting a medium-sized mean difference, $d = 0.36$, in flower–insect attitudes using a one-group t test (H2). Similar to Study 1, we expected to recruit 20% more children to take into account exclusions, and therefore estimated recruiting $N = 80$ total children each for the in-person and online procedures in order to reach our stopping rule of $N = 64$ in the analytic sample for each procedure.

Analytic Samples. For the in-person study, of the $N = 64$ children who comprised the final analytic sample, $n = 32$ were girls. The mean age of these 64 children was $M = 5.19$ years, $SD = 0.14$ (age range = 5.02–5.88 years old). A total of 71 participants were recruited to participate in the in-person flower preference study, of whom 67 completed the IAT. Completed IAT data were excluded for $n = 3$ participants (4.5%) based on the three IAT response criteria (see Table 3). Race for participants was 73.4% White, 18.8% multiracial, 6.3% Asian, and 1.6% Black/African American, and additionally, 90.6% were non-Hispanic and 6.3% of participants were of Hispanic ethnicity (3.1% did not report whether they were of Hispanic ethnicity). For parental education, 7.8% did not have a college degree, 49.2% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 41.4% had a graduate degree (1.6% did not report parental education). The average income-to-needs ratio was $M = 7.70$, $SD = 5.15$ (range = 1.19–33.00; 3.1% of participants did not provide all necessary information to compute the ratio).

For the online study, of the $N = 64$ children who comprised the final analytic sample, $n = 32$ were girls. The mean age of these 64 children was $M = 5.25$ years, $SD = 0.20$ (age range = 5.03–5.94 years old). A total of 71 participants were recruited to participate in the online flower preference study, of whom 69 completed the IAT. Completed IAT data were excluded for $n = 5$ participants (7.2%) based on the three IAT response criteria (see Table 3). Race for participants was 68.8% White, 25.0% multiracial, and 6.3% Asian, and additionally, 92.2% were non-Hispanic and 6.3% of participants were of Hispanic ethnicity (1.6% did not report whether they were of Hispanic ethnicity). For parental education, 10.2% did not have a college degree, 41.4% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 44.5% had a graduate degree (3.9% did not report parental education). The average income-to-needs ratio was $M = 10.36$, $SD = 20.82$ (range = 1.51–161.81; 9.4% of participants did not provide all necessary information to compute the ratio).

Materials and Procedure

All materials and procedures were similar to Studies 1 and 2 except for those features described below. Unlike in Studies 1 and 2 with elementary-aged children, the recorded voice of an experimenter read the assent information to the child to account for 5-year-olds' limited reading ability. The session took approximately 18 min for the in-person procedure and approximately 20 min for the online procedure.

Implicit Flower–Insect Attitude. Each participant completed either an in-person pictorial IAT from Cvencek, Greenwald, and

Meltzoff (2011), or the new online version of the same test. The pictorial IAT, like the text-based IAT, is a computerized sorting task in which children sort stimuli into four categories as quickly as possible using two response buttons. The pictorial IAT follows the same principle as the text-based IAT: children are expected to sort the stimuli faster during the task in which paired categories have a stronger mental association. However, the stimuli for the pictorial IAT were either pictures presented simultaneously with an audible tone to draw the child's attention to it, or words presented simultaneously as audio and text (instead of written words alone; see below subsection titled "Further Specialized Adaptations to Create an Online Pictorial IAT for Flower–Insect Attitude").

Although the number of blocks was the same as the online text-based IAT, the online pictorial IAT reduced the total number of trials within some of the blocks to make it easier for these very young children to complete the whole test in the absence of the supportive experimenter. Blocks 1 and 2 (16 trials each) were single-task warm-up blocks. Blocks 3 (16 trials) and 4 (24 trials) were the first combined-task blocks (e.g., "congruent" blocks), Block 5 (24 trials) was the single-task reversal block, and Blocks 6 (16 trials) and 7 (24 trials) were the second combined-task blocks, which had the opposite pairings as Blocks 3 and 4 (e.g., "incongruent" blocks). The counterbalancing was the same as in Study 1.

The four categories in the flower–insect pictorial IAT were *flowers* (represented by four color images of flowers, see Figure 5), *insects* (represented by four color images of insects, see Figure 5), *good* (stimuli: "fun," "good," "happy," "nice"), and *bad* (stimuli: "bad," "mad," "mean," "yucky") represented by text and spoken word. The stimuli were presented one at a time in the center of the screen. Pictorial stimuli were synchronized with a beep tone upon presentation (as an attention-getter), and text stimuli were synchronized with the verbal presentation through the speakers.

The pictorial IAT score was computed as the standard *D* score. The *D* score ranged from −2 (*insects* = *good*) to +2 (*flowers* = *good*), with a rational 0 value indicating equal association of *flowers* and *insects* with *good*.

Further Specialized Adaptations to Create an Online Pictorial IAT for Flower–Insect Attitude. Table 1 shows a summary of adaptations made to the in-person pictorial IAT to create the online pictorial IAT. Figure 5 shows a visual illustration of the online flower–insect pictorial IAT computer setup. Verbal instructions were provided as audio recordings throughout the test by the recorded voice of an experimenter reading all text. Written instructions accompanied the verbal instructions as on-screen text (shown in Figure 5 as "recorded verbal and text-based instructions").

The child's task was to begin each block of the pictorial IAT using the spacebar and then classify each stimulus by pressing the "D" and "K" keys on their keyboard. The instructional pages advanced automatically, however the timed sorting tasks did not begin until the child was ready and pressed the spacebar. The child had the option to stop the experiment by clicking on a clearly labeled "STOP" button on the screen or pressing the escape key (see Figure 5, bottom left corners of each instructional screen). The computer provided positive feedback to the child after each block of the experiment. The feedback consisted of pictures of cartoon animals giving thumbs up while a celebratory jingle played through the speakers. (These procedures have been worked out in collaboration with families in order to ensure instructions and a test length that children could tolerate at 5 years of age.)

Results

Quality Checks

The identical quality checks described in Studies 1 and 2 were also used in Study 3. As can be seen from the fourth line of Table 3, the data from the in-person procedure passed all of the preregistered quality checks: (a) there were no double entries, (b) 4.5% of participants were excluded (which is under the acceptable 20% exclusion rate), (c) the internal consistency was good, $\alpha = .733$, (d) the effects of counterbalancing factors were not statistically significant, both $ps > .14$, and (e) the data were normally distributed, skewness = −0.384. Table 3 similarly shows that the data from the online procedure passed all of the preregistered quality checks: (a) there was one double entry (which was eliminated prior to analyses), (b) 7.2% of participants were excluded (also under the acceptable 20% exclusion rate), (c) the internal consistency was very good, $\alpha = .875$, (d) the method effects were not statistically significant, both $ps > .67$, and (e) the data were normally distributed, skewness = −0.260.

Preregistered Hypothesis Tests

Supporting H1, for the in-person procedure, the mean level of flower preference was positive (i.e., *flowers* = *good*) and significantly different from zero, $t(63) = 4.20$, $p < .001$, $d = 0.52$, 95% CI [0.26, 0.78] (Figure 6).

Supporting H2, for the online procedure, the flower preference was also positive and significantly different from zero, $t(63) = 2.52$, $p = .014$, $d = 0.32$, 95% CI [0.06, 0.57] (Figure 6).

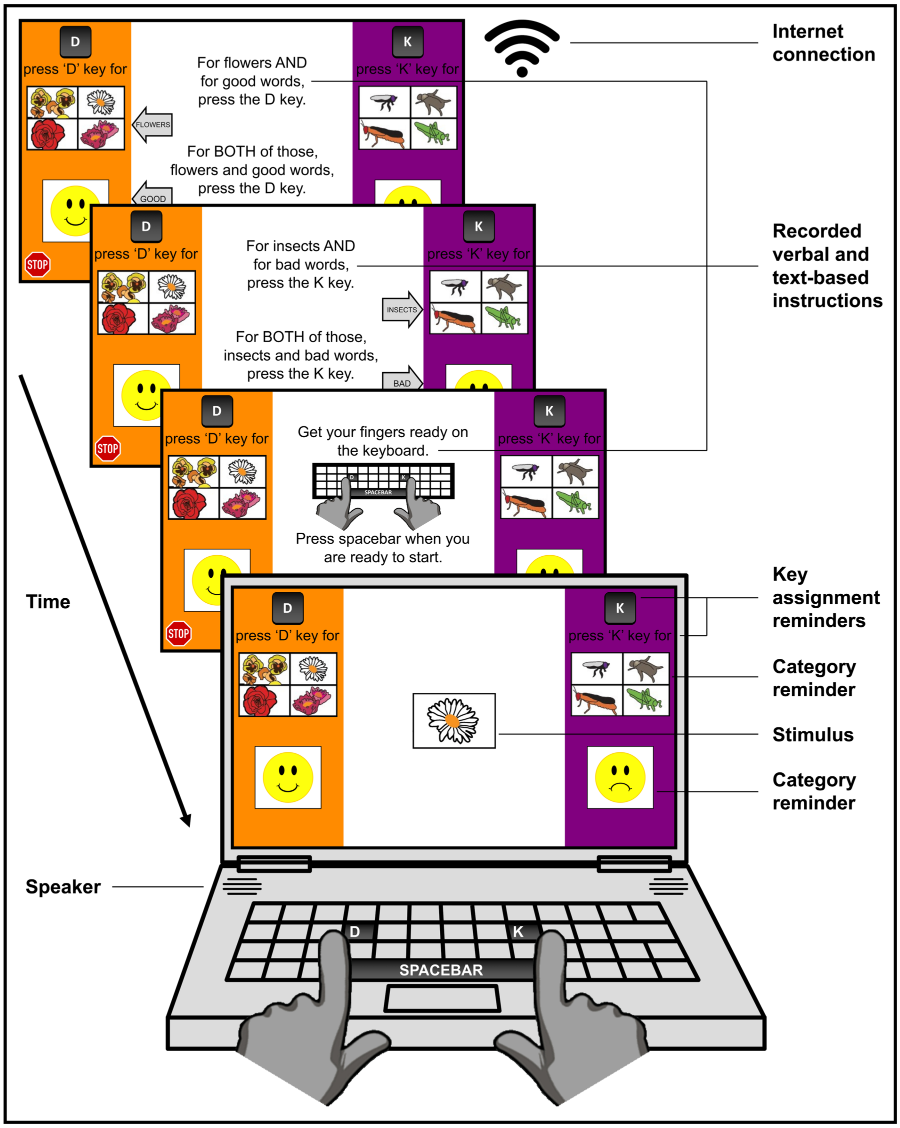
Preregistered Exploratory Analyses

Participants' responses to the exit survey item on self-efficacy regarding the IAT did not differ between the in-person and online procedures, $p = .18$. Although the responses to the items on liking, interest, and motivation also did not significantly differ between the two procedures, we note that children in the in-person procedure reported slightly more positive feelings on all three of these items than children in the online procedure, *p*s ranging from .075 to .096. Similar to the results of Study 1, children in the in-person procedure reported more people being in the room with them during the IAT than did children in the online procedure, $t(126) = 5.60$, $p < .001$, $d = 0.99$, 95% CI [0.62, 1.36].

Discussion

In Study 3 we extended the new online, unmoderated tool downward in age and tested preschool children. In order to validate this test, we used a task in which children's preferences should be unequivocally expected. In this study, we examined polarized categories along an affective dimension (*good* vs. *bad*). This was done by evaluating children's attitudes toward flowers and insects, using both in-person (H1) and online procedures (H2). The preregistered hypotheses were supported. The results showed the expected moderate preference for flowers on both online and in-person procedures ($ds \geq 0.32$). Similar to Studies 1 and 2 (with elementary school children), the new online tool exhibited good internal consistency and was resistant to method effects (counterbalancing factors), even in 5-year-old children.

Figure 5
Child's View of the New Online Pictorial IAT in Study 3



Note. This figure displays the online flower–insect pictorial IAT from Study 3; this setup remained the same for Study 4 (gender in-group bias) except for the categories/stimuli used. Children press the spacebar to start each IAT block, at which point the trials advance automatically. After pressing spacebar to begin the block, the child sees one (centered) stimulus per trial, as shown in the final screen of the figure. IAT = Implicit Association Test. See the online article for the color version of this figure.

Study 4—Preschool Children: Online Gender In-Group Bias IAT

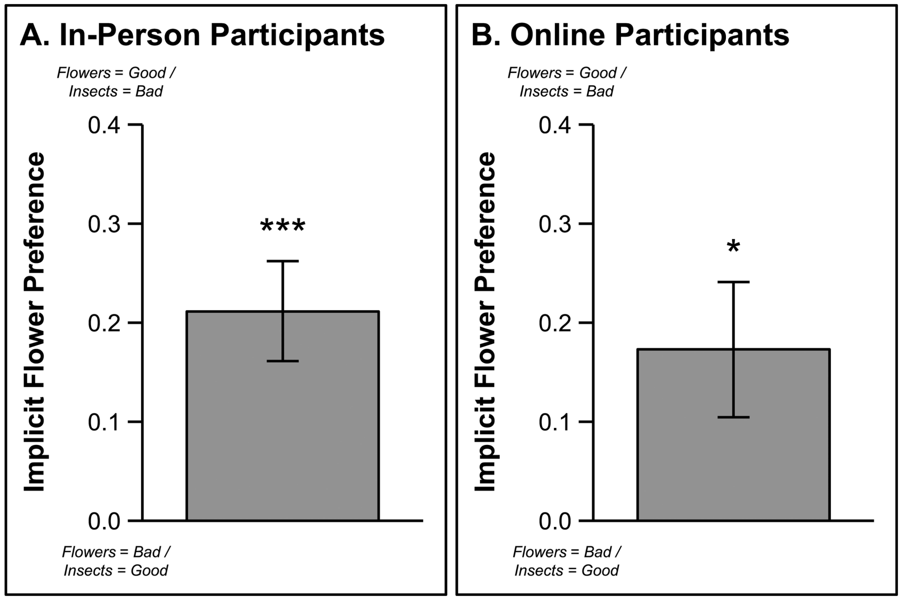
Rationale

The procedures for Study 4 were nearly identical to those in Study 3, and again involved 5-year-olds, except that Study 4 was focused on gender in-group bias (again using the new online pictorial IAT tool). In-groups are social categories that are related to the self. Prior research using implicit measures documented strong gender in-group biases in children and adults (e.g., Cvencek et al., 2016; Ebert & Steffens, 2008; Rudman & Goodwin, 2004; Skowronski & Lawrence, 2001). Gender in-group bias has been shown to be predictive of children's gendered play activities, including spatial block play and early reading, and has also been theorized to contribute to the formation of gender stereotypes and attitudes toward math in elementary school (Cvencek, Greenwald, & Meltzoff, 2011; Cvencek et al., 2021; Levine & Pantoja, 2021; Martin & Ruble, 2010).

Hypothesis

For Study 4 we tested the following hypothesis:

Figure 6
Study 3 Overall Mean Implicit Flower Preference



Note. Overall means of implicit flower preference for in-person (Panel A) and online (Panel B) participants. Sixty-four in-person participants; 64 online participants. Error bars represent *SE*s.
* $p < .05$. *** $p < .001$.

H1: Replication of evidence of gender in-group bias. We hypothesized that, using the new online procedure, preschool children would demonstrate a significant gender in-group bias (*own gender* = *good*).

Method

Participants

The recruitment procedures were the same as those in Study 3. Participants were 5-year-old girls and boys recruited in approximately equal numbers. Using the same power analysis approach as in the previous three studies with α set at .05 and power at 80%, and using the effect size from the published gender in-group results of the in-person Cvencek, Greenwald, and Meltzoff (2011) study, $d = 0.49$, we found that a sample size of 35 participants would be needed. Following the reasoning in Studies 1–3, we chose to use a larger sample of $N = 64$. A sample size of $N = 64$ respondents provided us with 80% power for detecting a medium-sized mean difference, $d = 0.36$, using a one-group t test. Similar to Study 1, we expected to recruit 20% more children in order to account for expected exclusions and smaller effect sizes for online procedures. In sum, we recruited $N = 80$ preschoolers in order to reach our stopping rule for the analytic sample of $N = 64$.

Analytic Sample. Of the $N = 64$ children who comprised the final analytic sample, $n = 32$ were girls. The mean age for these 64 children was $M = 5.29$ years, $SD = 0.20$ (range = 5.01–5.92 years old). A total of 80 participants were recruited to participate in the online gender in-group bias study, of whom 72 completed the IAT. Completed IAT data were excluded for $n = 8$ children (11.1%) based on the three IAT response criteria (see Table 3). Race for participants was 67.2% White, 25.0% multiracial, and 7.8% Asian, and additionally, 95.3% were non-Hispanic and 3.1% of participants were of Hispanic ethnicity (1.6% did not report whether they were of Hispanic ethnicity). For parental education, 8.6% did not have a college degree, 45.3% had an undergraduate (associate, technical, 2-year, or 4-year) college degree, and 44.5% had a graduate degree (1.6% did not report parental education). The average income-to-needs ratio was $M = 7.69$, $SD = 5.92$ (range = 1.19–32.36; 7.8% of participants did not provide all necessary information to compute the ratio).

Materials and Procedure

All Study 4 procedures were identical to Study 3 except those described below. The entire session took approximately 20 min. Study 4 collected data on implicit gender in-group bias (i.e., *own gender* = *good*) in 5-year-olds using the new online pictorial IAT.

Implicit Gender In-Group Bias. Two of the categories in this gender in-group bias IAT were the *good* and *bad* categories (and their respective stimuli) described in Study 3. However, in Study 4 we measured feelings toward one's in-group versus an out-group. Thus, the other two categories were *boys* and *girls*, each represented by four grayscale faces of children. The pictorial IAT score was computed as the standard *D* score. The *D* score ranged from −2 (*opposite gender* = *good*) to +2 (*own gender* = *good*), with a rational 0 value indicating equal association of *good* with the child's *own gender* and *opposite gender*.

Further Specialized Adaptations to Create an Online Pictorial IAT for Gender In-Group Bias. As in Study 3, adaptations were made from the published in-person IAT procedure to the online procedure (see Table 1).

Results

Quality Checks

The identical five quality checks described in Studies 1–3 were also used in Study 4. As can be seen from the sixth line of Table 3, the online gender in-group bias data passed all five preregistered quality checks: (a) there were no double entries, (b) 11.1% of participants were excluded (which is under the acceptable 20% exclusion rate), (c) the internal consistency was very good, $\alpha = .868$, (d) the method effects were not statistically significant, both $ps > .42$, and (e) the data were normally distributed, skewness $= -0.196$.

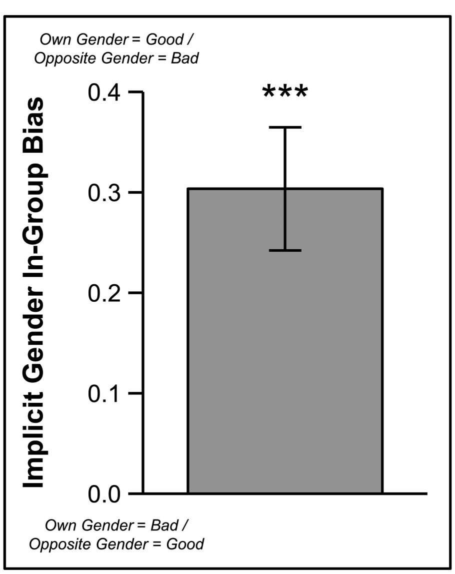
Preregistered Hypothesis Tests

Supporting H1, the mean level of gender in-group bias was positive (i.e., *own gender = good*) and significantly different from zero, $t(63) = 4.94$, $p < .001$, $d = 0.62$, 95% CI [0.35, 0.88] (Figure 7).

Discussion

Study 4 extended the new online, unmoderated tool to a social domain by investigating implicit gender in-group biases in preschoolers. The issue of children's implicit biases about social groups (e.g., gender and race) is an important topic in contemporary developmental psychology as well as in education. Supporting our preregistered hypothesis, the results using the new online tool replicated strong gender in-group bias, in line with published results with preschool children using in-person procedures. Also, the new online tool used in Study 4 exhibited high internal consistency and was resistant to method effects (counterbalancing factors). The current findings add to our knowledge of when social group biases begin to emerge in young children (e.g., Dunham et al., 2008; Meltzoff & Gilliam, 2024; Rizzo et al., 2022; Skinner-Dorkenoo et al., 2023).

Figure 7
Study 4 Overall Mean Implicit Gender In-Group Bias



Note. Mean of implicit gender in-group bias for online participants. $N = 64$. Error bar represents *SE*.
*** $p < .001$.

General Discussion

There is a need for standardized online tools that can be used to assess psychological constructs in children in an unmoderated fashion, that is, without the need for a trained experimenter. We report the development of such new tools and show that they can be used to assess implicit beliefs and attitudes toward self, social groups, and academic subjects in children ranging from preschool through elementary school. Across the four studies, eight hypotheses were preregistered, and all eight were supported. Importantly, we also report data showing the validity of these new instruments and find that the new online tools yield comparable effects to in-person laboratory testing supervised by an adult experimenter. The development of these online, unmoderated tools opens the pathway for large-scale testing of diverse samples of children who otherwise might not be able to be tested in university laboratories, schools, or museums.

Validity of the New Tools

The potential usefulness of online, unmoderated tools for children depends on their meeting the usual construct validity and psychometric standards.

Construct Validity

Construct validity of the new online, unmoderated tools for elementary and preschool children was demonstrated by finding statistically significant evidence for each of the four constructs assessed (own-gender identity, math–gender stereotypes, attitudes toward nonsocial groups, and social in-group biases). The findings obtained in the present studies were consistent with those previously reported for children on these same constructs tested in laboratories using trained experimenters (e.g., Cvencek, Greenwald, & Meltzoff, 2011; Cvencek, Meltzoff, & Greenwald, 2011). The results with preschoolers (Studies 3 and 4) are particularly noteworthy in showing that even 5-year-old children could follow the standardized directions embedded in the online, unmoderated tool.

Psychometric Validity

Internal consistencies of the new online tools were obtained by first separating all combined-task trials within a procedure into two equally sized subsets of trials. Then, a separate *D* score was computed for each subset. Finally, the correlation between the two *D* scores was computed, which revealed satisfactorily high values of Cronbach's alpha, ranging from $\alpha = .73$ to $\alpha = .88$. The fact that these online, unmoderated tools showed good internal consistencies suggests that children are not just randomly responding, but that the new online procedure was working as expected.

The resistance to method effects (counterbalancing of order and side) in the present studies further buttresses the psychometric validity (Hughes, 2018) of the new tools. The lack of order or side effects with children is noteworthy because these and other subtle procedural factors are known to influence IAT measures in adults (Greenwald et al., 2003; Nosek et al., 2005). Currently, we do not have firm knowledge about why children are not affected by these factors. One speculation is that adults may be more sensitive to procedural variations, such as whether the congruent versus incongruent task comes first, than children are. For children, these procedural nuances may not play a role (at least not until a certain age) when completing response latency measures.

Methodological Advances Embedded in the New Online Tools for Children

Over the past several years, it has become increasingly clear that online tools for testing children would be a valuable contribution to the field (e.g., Leshin et al., 2021; Lo et al., 2024; Prein et al., 2024; Rhodes et al., 2020; Rizzo et al., 2022; Scott & Schulz, 2017; Sheskin et al., 2020; Steffan et al., 2024). The existing online platforms require that the parent be present with the child during the test and that the test session be recorded via webcam and uploaded to the researcher at the end of the session.

The studies reported here advance this work in three ways. First, our online, unmoderated test procedure is entirely self-administrated by the child (not requiring the parent to be present during the test). Second, the user interface for our online test procedure involves software that automatically records and uploads data to a server, as the child responds, in real time (not requiring parents to subsequently upload a webcam video for later coding by researchers). Third, we demonstrated the validity of the online, unmoderated procedures (good psychometric properties and comparable results to in-person testing of the same age group; see also Prein et al., 2024).

The online, unmoderated tools described here have the potential for five broader impacts. First, research on implicit bias in adults has generated wide attention (more than 20 million adults completed IAT procedures through Harvard's Project Implicit platform; Charlesworth & Banaji, 2019). The child-friendly tools reported here can be used to create a new platform (a "Child Project Implicit"). The capacity to test children's beliefs and attitudes about self, social groups, and academic topics in an online, unmoderated fashion will be useful for developmental and educational research.

Second, online, unmoderated research tools that can be completed from home will make it easier to recruit participants from socioeconomically, racially, and ethnically diverse populations who otherwise do not have time or resources to participate in research projects conducted in university laboratories or schools, a point also made by the developers of other online tools (e.g., Rhodes et al., 2020; Rizzo et al., 2022; Scott & Schulz, 2017). Two of the largest obstacles for elementary schools participating in experimental research involve time demands and finding space for testing on school premises. The new tools relieve the burden of those demands, especially for schools that serve predominately Black, Indigenous, and People of Color student populations that often lack space resources for one-on-one experimental testing.

Third, the online, unmoderated tools will also facilitate cross-cultural research, which is garnering increased interest in both developmental and educational sciences (e.g., Amir & McAuliffe, 2020; Amir et al., 2023; Ansari, 2012; Cortes Barragan & Meltzoff, 2025; Cvencek et al., 2025; Lau et al., 2022; Marshall et al., 2024). They allow for a comparison of the same set of social-psychological constructs using identical, standardized procedures in different cultural contexts (of course, requiring verbal translations and culturally appropriate pictures). This will enable the research findings to be more generalizable and enhance our understanding of cultural universals and variations in children's social cognition.

Fourth, the new online, unmoderated tools should help us develop more accurate assessments of individual children's beliefs and attitudes. Work in medical science has shown that more accurate results can be obtained by taking multiple tests on the same individual (e.g., averaging blood pressure readings). Multiple assessments compensate for participants having "bad" or "good" days on a single assessment. Averaging multiple assessments has the potential advantage of being adequately diagnostic at the individual level, and not just for comparing group means. A multiassessment approach is currently underutilized in educational and developmental research because it would require taking students out of classrooms multiple times or repeat visits to the laboratory. Online tools that can be used for testing in the child's home will help address this issue.

Fifth, the new online, unmoderated tools will be useful both for research to continue during school breaks and summer months, which is useful for investigating learning loss, and also during future unplanned school closures (such as during future pandemics or natural disasters).

Replicability

Across multiple fields of psychological and educational science, concerns about replicability have been raised (e.g., Dreber et al., 2015; Nosek et al., 2022), and these concerns are now being noted with respect to online tools, too (Forsell et al., 2019; Gottfried, 2024). The studies reported here utilized a design that addresses the replicability of results in two ways.

First, each of the four studies investigated constructs for which there are previous data from in-person testing (e.g., Cvencek et al., 2016; Dunham et al., 2016; Steffens et al., 2010; Thomas et al., 2007). This provided an opportunity to evaluate whether the new online, unmoderated tools could replicate and generalize the previous results, which all involved the support of a trained experimenter. Second, two of the studies (Studies 1 and 3) also provided an opportunity for within-study replication. By including contemporaneous in-person and online tests in the same study, we were able to address the degree to which online procedures can detect the same significant effects that are regularly obtained using in-person procedures. The results were replicated, and both procedures yielded the hypothesized effects at a similar magnitude (Figures 2, 3, and 6). The new online, unmoderated procedure with children will be a useful tool in future work evaluating replicability and generalizability in developmental and educational science, using highly standardized procedures.

Contributions to Open Science

The new online, unmoderated tools make three contributions to recent calls for open science. First, we preregistered our predictions, methods, and detailed analysis plan prior to conducting the research. Second, the preregistration included important statistical information relevant to open science, such as power analyses, data exclusions, and five quality checks (e.g., assessing internal consistency and statistical assumption violations). Third, the stimuli, instructions, analysis code, and raw data have been made openly available via a trusted repository (the Open Science Framework). We hope that this level of detail will lay the groundwork for developing "best practice" recommendations for child research using both in-person and online IAT procedures. The sharing of data will also facilitate tests of alternative theoretical interpretations as well as future meta-analyses (Rhodes et al., 2020).

Limitations and Future Work

We acknowledge three limitations that warrant further research. First, the reported studies were cross-sectional. Even though we included a range of ages across the four studies, future research will add to our knowledge base by conducting studies following the same children over time. Examining individual children's continuity and change requires longitudinal work. There have been only a handful of longitudinal studies on children's implicit social cognition, and we hope that the online tools reported here will spark more.

Second, our current samples involved participants from a narrow geographical location in the Pacific Northwest. Now that these online, unmoderated procedures are available, we believe that it is a priority to use these new tools to study demographically diverse samples. Some efforts have begun to investigate child implicit cognition internationally using in-person testing (Cvencek et al., 2014, 2025; del Río et al., 2019; Dunham et al., 2006; M. Qian et al., 2021; M. K. Qian et al., 2019) and the standardized online procedures reported here will enhance this work.

Third, our studies tested only certain social-cognitive constructs, and there are others that are highly relevant for developmental and educational science. For example, in-person tests of implicit self-esteem have been linked to children's academic achievement in Grades K–2 (Cvencek et al., 2018). It would be helpful to adapt these in-person self-esteem tests so they can be done in an online, unmoderated fashion in young children. Similarly, even for the constructs tested here (e.g., own-gender identity, math–gender stereotypes), it will be interesting to empirically assess the relations between the online results and various theoretically expected behavioral correlates and outcomes (such as gender roles or school achievement).

Conclusion

Implicit stereotypes and attitudes are measurable in children and are known to relate to school achievement and other educational outcomes. This article reports the design and validation of new online, unmoderated tools for testing preschool and elementary school children using rigorously controlled and standardized computer-based protocols. The results support the preregistered hypotheses in all four studies and establish the construct and psychometric validity of the new procedures for assessing children's implicit social cognition. These new tools have the potential for: (a) broadening participation in developmental and educational research by providing tools for testing diverse children in home and community settings, and (b) enabling large-scale collaborations investigating developmental continuity and change, as well as cultural universality and variability in child development.

References

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, *41*(5), 430–444. https://doi.org/10.1016/j.evolhumbehav.2020.06.006

Amir, D., Melnikoff, D., Warneken, F., Blake, P. R., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K. L., Ross, E., Vongsachang, H., Wrangham, R., & McAuliffe, K. (2023). Computational signatures of inequity aversion in children across seven societies. *Journal of Experimental Psychology: General*, *152*(10), 2882–2896. https://doi.org/10.1037/xge0001385

Ansari, D. (2012). Culture and education: New frontiers in brain plasticity. *Trends in Cognitive Sciences*, *16*(2), 93–95. https://doi.org/10.1016/j.tics.2011.11.016

Axt, J. (2018, May 30). *Tracking the use of Project Implicit data* [Blog post]. https://implicit.harvard.edu/implicit/user/jaxt/blogposts/piblogpost020.html

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives*, *9*(1), 50–54. https://doi.org/10.1111/cdep.12105

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Brummelman, E., Grapsas, S., & van der Kooij, K. (2022). Parental praise and children's exploration: A virtual reality experiment. *Scientific Reports*, *12*(1), Article 4967. https://doi.org/10.1038/s41598-022-08226-9

Carpenter, T. P., Pogacar, R., Pullig, C., Kouril, M., Aguilar, S., LaBouff, J., Isenberg, N., & Chakroff, A. (2019). Survey-software Implicit Association Tests: A methodological and empirical analysis. *Behavior Research Methods*, *51*(5), 2194–2208. https://doi.org/10.3758/s13428-019-01293-3

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Córdova, F. (2020, March 23). *NSF letter to community regarding COVID-19: Important notice to the science and engineering community* (Important Notice No. 146). https://www.nsf.gov/pubs/issuances/in146.jsp

Cortes Barragan, R., & Meltzoff, A. N. (2025). The socialization of cultural values and the development of Latin American prosociality. *Child Development Perspectives*. Advance online publication. https://doi.org/10.1111/cdep.12545

Cvencek, D., Brečić, R., Gaćeša, D., & Meltzoff, A. N. (2021). Development of math attitudes and math self-concepts: Gender differences, implicit–explicit dissociations, and relations to math achievement. *Child Development*, *92*(5), e940–e956. https://doi.org/10.1111/cdev.13523

Cvencek, D., Fryberg, S. A., Covarrubias, R., & Meltzoff, A. N. (2018). Self-concepts, self-esteem, and academic achievement of minority and majority North American elementary school children. *Child Development*, *89*(4), 1099–1109. https://doi.org/10.1111/cdev.12802

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The Preschool Implicit Association Test. *Journal of Experimental Child Psychology*, *109*(2), 187–200. https://doi.org/10.1016/j.jecp.2010.11.002

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology*, *62*, 50–57. https://doi.org/10.1016/j.jesp.2015.09.015

Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction*, *39*, 1–10. https://doi.org/10.1016/j.learninstruc.2015.04.002

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, *117*, 73–91. https://doi.org/10.1016/j.jecp.2013.07.018

Cvencek, D., Paz-Albo, J., Master, A., Herranz Llacer, C. V., Hervas-Escobar, A., & Meltzoff, A. N. (2020). Math is for me: A field intervention to strengthen math self-concepts in Spanish-speaking 3rd grade children. *Frontiers in Psychology*, *11*, Article 593995. https://doi.org/10.3389/fpsyg.2020.593995

Cvencek, D., Sanders, E. A., del Río, M. F., Susperreguy, M. I., Strasser, K., Brečić, R., Gaćeša, D., Skala, D., Tomasetto, C., Galdi, S., Cadinu, M., Kapur, M., Passolunghi, M. C., Rueda Ferreira, T. I., Mirisola, A., Mariani, B., & Meltzoff, A. N. (2025). National disparities favoring males are reflected in girls' implicit associations about gender and academic subjects. *Developmental Psychology*, *61*(3), 579–593. https://doi.org/10.1037/dev0001797

Daly, M. C., McDonough, P., & Williams, D. R. (2002). Optimal indicators of socioeconomic status for health research. *American Journal of Public Health*, *92*(7), 1151–1157. https://doi.org/10.2105/AJPH.92.7.1151

del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., & Meltzoff, A. N. (2019). Chilean Kindergarten children's beliefs about mathematics: Family matters. *Developmental Psychology*, *55*(4), 687–702. https://doi.org/10.1037/dev0000658

del Río, M. F., Susperreguy, M. I., Strasser, K., Cvencek, D., Iturra, C., Gallardo, I., & Meltzoff, A. N. (2021). Early sources of children's math achievement in Chile: The role of parental beliefs and feelings about math. *Early Education and Development*, *32*(5), 637–652. https://doi.org/10.1080/10409289.2020.1799617

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications.

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B. A., & Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), 15343–15347. https://doi.org/10.1073/pnas.1516179112

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, *77*(5), 1268–1281. https://doi.org/10.1111/j.1467-8624.2006.00933.x

Dunham, Y., Baron, A. S., & Banaji, M. R. (2007). Children and social groups: A developmental analysis of implicit consistency in Hispanic Americans. *Self and Identity*, *6*(2–3), 238–255. https://doi.org/10.1080/15298860601115344

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(7), 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science*, *19*(5), 781–789. https://doi.org/10.1111/desc.12321

Ebert, I. D., & Steffens, M. C. (2008, June 19–23). *Generality and correlates of the implicit preference for women* [Paper presentation]. 11th Jena Workshop on Intergroup Processes: Positive Behaviour Between Social Groups—Processes, Problems, and Promises, Oppurg, Germany.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Forsell, E., Viganola, D., Pfeiffer, T., Almenberg, J., Wilson, B., Chen, Y., Nosek, B. A., Johannesson, M., & Dreber, A. (2019). Predicting replication outcomes in the Many Labs 2 study. *Journal of Economic Psychology*, *75*(Pt. A), Article 102117. https://doi.org/10.1016/j.joep.2018.10.009

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), Article e109019. https://doi.org/10.1371/journal.pone.0109019

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, *85*(1), 250–263. https://doi.org/10.1111/cdev.12128

Gibson, F., & Twycross, A. (2008). Editorial: Getting it right for children and young people's health care services. *Journal of Clinical Nursing*, *17*(23), 3081–3082. https://doi.org/10.1111/j.1365-2702.2008.02644.x

Gijbels, L., Cai, R., Donnelly, P. M., & Kuhl, P. K. (2021). Designing virtual, moderated studies of early childhood development. *Frontiers in Psychology*, *12*, Article 740290. https://doi.org/10.3389/fpsyg.2021.740290

Gottfried, J. (2024). Practices in data-quality evaluation: A large-scale review of online survey studies published in 2022. *Advances in Methods and Practices in Psychological Science*, *7*(2), Article 25152459241236414. https://doi.org/10.1177/25152459241236414

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*(1), 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

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

Hildebrand, L., Posid, T., Moss, R. C. A., Hymes, L., & Cordes, S. (2022). Does my daughter like math? Relations between parent and child math attitudes and beliefs. *Developmental Science*, *26*(1), Article e13243. https://doi.org/10.1111/desc.13243

Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (Vol. 2, pp. 751–779). Wiley. https://doi.org/10.1002/9781118489772

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29) [Computer software]. https://www.ibm.com/spss

Ibna Seraj, P. M., Chakraborty, R., Mehdi, T., & Roshid, M. M. (2022). A systematic review on pedagogical trends and assessment practices during the COVID-19 pandemic: Teachers' and students' perspectives. *Education Research International*, *2022*, Article 1534018. https://doi.org/10.1155/2022/1534018

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psychology Bulletin*, *31*(2), 208–217. https://doi.org/10.1177/0146167204271416

Lau, N. T. T., Hawes, Z., Tremblay, P., & Ansari, D. (2022). Disentangling the individual and contextual effects of math anxiety: A global perspective. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(7), Article e2115855119. https://doi.org/10.1073/pnas.2115855119

Leshin, R. A., Leslie, S.-J., & Rhodes, M. (2021). Does it matter how we speak about social kinds? A large, preregistered, online experimental study of how language shapes the development of essentialist beliefs. *Child Development*, *92*(4), e531–e547. https://doi.org/10.1111/cdev.13527

Levine, S. C., & Pantoja, N. (2021). Development of children's math attitudes: Gender differences, key socializers, and intervention approaches. *Developmental Review*, *62*, Article 100997. https://doi.org/10.1016/j.dr.2021.100997

Lo, C. H., Hermes, J., Kartushina, N., Mayor, J., & Mani, N. (2024). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *Behavior Research Methods*, *56*(5), 4530–4552. https://doi.org/10.3758/s13428-023-02200-7

Marshall, J., Mermin-Bunnell, K., Gollwitzer, A., Retelsdorf, J., & Bloom, P. (2024). Cross-cultural conceptions of third-party intervention across childhood. *Journal of Experimental Psychology: General*, *153*(9), 2216–2229. https://doi.org/10.1037/xge0001617

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

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(48), Article e2100030118. https://doi.org/10.1073/pnas.2100030118

Master, A., Meltzoff, A. N., Tang, D., & Cheryan, S. (2025). Divergence in children's gender stereotypes and motivation across STEM fields. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(18), Article e2408657122. https://doi.org/10.1073/pnas.2408657122

Meltzoff, A. N., & Gilliam, W. S. (2024). Young children & implicit racial biases. *Daedalus*, *153*(1), 65–83. https://doi.org/10.1162/daed_a_02049

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of U.S. Draw-a-Scientist studies. *Child Development*, *89*(6), 1943–1955. https://doi.org/10.1111/cdev.13039

Millisecond Software. (2022). *Inquisit* (Version 6.6.1) [Computer software]. https://www.millisecond.com

Newheiser, A.-K., & Olson, K. R. (2012). White and Black American children's implicit intergroup bias. *Journal of Experimental Social Psychology*, *48*(1), 264–270. https://doi.org/10.1016/j.jesp.2011.08.011

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Prein, J. C., Kalinke, S., Haun, D. B. M., & Bohn, M. (2024). TANGO: A reliable, open-source, browser-based task to assess individual differences in gaze understanding in 3 to 5-year-old children and adults. *Behavior Research Methods*, *56*(3), 2469–2485. https://doi.org/10.3758/s13428-023-02159-5

Project Implicit. (2022). *PI: Published manuscripts using public data*. Retrieved June 30, 2023, from https://www.projectimplicit.net/resources/publications

Qian, M., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2021). Age-related differences in implicit and explicit racial biases in Cameroonians. *Developmental Psychology*, *57*(3), 386–396. https://doi.org/10.1037/dev0001149

Qian, M. K., Heyman, G. D., Quinn, P. C., Fu, G., & Lee, K. (2019). Differential developmental courses of implicit and explicit biases for different other-race classes. *Developmental Psychology*, *55*(7), 1440–1452. https://doi.org/10.1037/dev0000730

Qian, M. K., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit racial biases in preschool children and adults from Asia and Africa. *Child Development*, *87*(1), 285–296. https://doi.org/10.1111/cdev.12442

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

Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, *21*(4), 477–493. https://doi.org/10.1080/15248372.2020.1797751

Rizzo, M. T., Britton, T. C., & Rhodes, M. (2022). Developmental origins of anti-Black bias in White children in the United States: Exposure to and beliefs about racial inequality. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(47), Article e2209129119. https://doi.org/10.1073/pnas.2209129119

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*(4), 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Schmader, T., Bareket-Shavit, C., & Baron, A. S. (2022). Beyond awareness: The many forms of implicit bias and its implications. *Psychological Inquiry*, *33*(3), 156–161. https://doi.org/10.1080/1047840X.2022.2106752

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A new online platform for developmental research. *Open Mind*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002

Setoh, P., Lee, K. J. J., Zhang, L., Qian, M. K., Quinn, P. C., Heyman, G. D., & Lee, K. (2019). Racial categorization predicts implicit racial bias in preschool children. *Child Development*, *90*(1), 162–179. https://doi.org/10.1111/cdev.12851

Sheskin, M., & Keil, F. (2018). *Thechildlab.com: A video chat platform for developmental research*. PsyArXiv. https://doi.org/10.31234/osf.io/rn7w5

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

Skinner-Dorkenoo, A. L., George, M., Wages, J. E., III, Sánchez, S., & Perry, S. P. (2023). A systemic approach to the psychology of racial bias within individuals and society. *Nature Reviews Psychology*, *2*(7), 392–406. https://doi.org/10.1038/s44159-023-00190-z

Skowronski, J. J., & Lawrence, M. A. (2001). A comparative study of the implicit and explicit gender attitudes of children and college students. *Psychology of Women Quarterly*, *25*(2), 155–165. https://doi.org/10.1111/1471-6402.00017

Steele, J. R., George, M., Williams, A., & Tay, E. (2018). A cross-cultural investigation of children's implicit attitudes toward White and Black racial outgroups. *Developmental Science*, *21*(6), Article e12673. https://doi.org/10.1111/desc.12673

Steffan, A., Zimmer, L., Arias-Trejo, N., Bohn, M., Dal Ben, R., Flores-Coronado, M. A., Franchin, L., Garbisch, I., Grosse Wiesmann, C., Hamlin, J. K., Havron, N., Hay, J. F., Hermansen, T. K., Jakobsen, K. V., Kalinke, S., Ko, E.-S., Kulke, L., Mayor, J., Meristo, M., … Schuwerk, T. (2024). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *Infancy*, *29*(1), 31–55. https://doi.org/10.1111/infa.12564

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, *102*(4), 947–963. https://doi.org/10.1037/a0019920

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A single block IAT. *European Journal of Psychological Assessment*, *24*(4), 237–245. https://doi.org/10.1027/1015-5759.24.4.237

Thomas, S., Burton Smith, R., & Ball, P. (2007). Implicit attitudes in very young children: An adaptation of the IAT. *Current Research in Social Psychology*, *13*(7), 75–85. https://crisp.org.uiowa.edu/sites/crisp.org.uiowa.edu/files/2020-04/13.7.pdf

U.S. Census Bureau. (2024). *Poverty thresholds by size of family and number of children*. https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-poverty-thresholds/thresh23.xlsx

Vuletich, H. A., Kurtz-Costes, B., Cooley, E., & Payne, B. K. (2020). Math and language gender stereotypes: Age and gender differences in implicit biases and explicit beliefs. *PLOS ONE*, *15*(9), Article e0238230. https://doi.org/10.1371/journal.pone.0238230

Yeatman, J. D., Tang, K. A., Donnelly, P. M., Yablonski, M., Ramamurthy, M., Karipidis, I. I., Caffarra, S., Takada, M. E., Kanopka, K., Ben-Shachar, M., & Domingue, B. W. (2021). Rapid online assessment of reading ability. *Scientific Reports*, *11*(1), Article 6396. https://doi.org/10.1038/s41598-021-85907-x

Zhao, S., Setoh, P., Storage, D., & Cimpian, A. (2022). The acquisition of the gender-brilliance stereotype: Age trajectory, relation to parents' stereotypes, and intersections with race/ethnicity. *Child Development*, *93*(5), e581–e597. https://doi.org/10.1111/cdev.13809

Received November 2, 2022
Revision received September 23, 2024
Accepted February 6, 2025 ■