Math–Gender Stereotypes in Elementary School Children

Dario Cvencek, Andrew N. Meltzoff, and Anthony G. Greenwald

University of Washington

A total of 247 American children between 6 and 10 years of age (126 girls and 121 boys) completed Implicit Association Tests and explicit self-report measures assessing the association of (a) me with male (gender identity), (b) male with math (math–gender stereotype), and (c) me with math (math self-concept). Two findings emerged. First, as early as second grade, the children demonstrated the American cultural stereotype that math is for boys on both implicit and explicit measures. Second, elementary school boys identified with math more strongly than did girls on both implicit and self-report measures. The findings suggest that the math–gender stereotype is acquired early and influences emerging math self-concepts prior to ages at which there are actual differences in math achievement.

Imagine yourself an elementary school teacher. One of your female students fails to complete an arithmetic assignment and offers an excuse that “Girls don’t do math.” What might be a pretext for avoiding homework could also be the outcome of social-cognitive development. Combining cultural stereotypes (“Math is for boys”) with the knowledge about one’s own gender identity (“I am a girl”) to influence one’s self-concept (“Math is not for me”) reflects the tendency to achieve what social psychologists (Heider, 1946) call cognitive balance.

In the foregoing example, “Girls don’t do math” is a widespread cultural stereotype in the United States: Studies with both adults (Nosek et al., 2009) and children (Lummis & Stevenson, 1990) show that people in United States believe that math is stereotypically a male domain. Given such stereotypes, a tendency to keep the related concepts of self, gender, and math consistent with one another (what Heider, 1946, called cognitive balance) may play a role in why a young girl would say—and possibly believe—that math is not for her (of course, there will be individual differences).

Social knowledge can be represented as a network of interconnections among concepts (Greenwald et al., 2002). In the foregoing example, three aspects of social cognition are involved. The first is the association between math and boy or girl. If this takes a societally characteristic form (e.g., math = boy), it can be called a math–gender stereotype. The second involves gender identity, defined as the association between me and either boy or girl. The third is a math self-concept, the association between self and math.

Past research using implicit measures with adults has shown that for women, the stronger the associations of (a) self with female and (b) math with male, the weaker the association of self with math (Nosek, Banaji, & Greenwald, 2002). In children, the interplay among math–gender stereotype, gender identity, and math self-concept has been studied using self-report measures. American elementary school children often reflect the stereotypic pattern for academic self-concepts: For math, girls rate their own ability lower than boys (Fredericks & Eccles, 2002) but do not do so for reading or spelling (Herbert & Stipek, 2005; Heyman & Legare, 2004). Using self-report, this pattern is evident as early as the first grade (Entwistle, Alexander, Pallas, & Cardigan, 1987), even in the absence of differences in math achievement (Herbert & Stipek, 2005). Girls’ weaker identification with math may derive from culturally communicated messages about math being more appropriate for boys than for girls (Dweck, 2007; Eccles, 2007; Guiso, Monte, Sapienza, & Zingales, 2008; National Science Foundation, 2010).

We thank the schools, their directors, teachers, parents, and the children who participated. J. Aronson, S. Cheryan, G. Heyman, C. Kaiser, the members of the LIFE Science of Learning Center, and three anonymous referees provided insightful comments on earlier versions of this article. We also thank C. Fisher, C. Harris, and G. Owen for assistance. This research was supported by a grant from the National Science Foundation (SBE-0354453) to the LIFE Science of Learning Center. Earlier versions of this article were presented at a meeting of the Society for Research in Child Development (April 2009) and the American Psychological Association (August 2009).

Correspondence concerning this article should be addressed to Dario Cvencek, Institute for Learning & Brain Sciences, University of Washington, Box 357988, Seattle, WA 98195. Electronic mail may be sent to dario1@u.washington.edu.

doi: 10.1111/j.1467-8624.2011.01529.x
These patterns are important developmentally, because as Eccles and others have shown, children have reduced interest in future academic courses and occupations that are incompatible with their academic self-concept (Denissen, Zarrett, & Eccles, 2007; Frome, Alfeld, Eccles, & Barber, 2006; Killen, Margie, & Sinno, 2006; Liben, Bigler, & Krogh, 2001; Malcom et al., 2005; Newcombe, 2007).

Previous investigations of children's math–gender stereotype and math self-concept have focused on self-report measures (for an exception, see Ambady, Shih, Kim, & Pittinsky, 2001). The wording of self-report measures often involves asking children how good they think they are at something or how much they like it, both of which conflate self-concept (nonevaluative association of self) with self-esteem (evaluative association of self). For example, a girl who reports that she is good at math may do so because she thinks that she is good at many things (high self-esteem). Similarly, a boy who reports that he likes math may do so because he believes that liking math is a positive quality and he sees himself as having many positive qualities (high self-esteem). If the focus is children's math self-concepts, it is more informative to assess how much they self-report measures often involves asking children how good they think they are at something or how much they like it, both of which conflate self-concept (nonevaluative association of self) with self-esteem (evaluative association of self). For example, a girl who reports that she is good at math may do so because she thinks that she is good at many things (high self-esteem). Similarly, a boy who reports that he likes math may do so because he believes that liking math is a positive quality and he sees himself as having many positive qualities (high self-esteem). If the focus is children's math self-concepts, it is more informative to assess how strongly a child associates self with math (i.e., whether the child has a strong math self-concept or not).

In order to differentiate the constructs more cleanly, we adapted a test used with adults in social psychology that does not require self-report, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). We modified it so that it could be used with elementary school children. The IAT originated within social psychology but in recent years has been applied in cognitive psychology (Fazio & Olson, 2003), clinical psychology (Teachman, Gregg, & Woody, 2001), and developmental psychology (Dunham, Baron, & Banaji, 2006; Rutland, Cameron, Milne, & McGeorge, 2005; Skowronski & Lawrence, 2001). In adults, IAT measures correlate with actual math performance and real-world choices and actions (Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

The IAT is a computerized categorization task that measures relative strengths of associations among concepts. IAT’s format allows the measurement of preference for one concept (e.g., math) relative to the preference for a second concept (e.g., reading). The contrasting category is of practical importance in investigations involving academic subjects, because academic choices rarely occur without alternatives. Reading offers itself readily as a contrasting category for math because: (a) reading and math education are mandated from the first grade on, (b) sex differences in self-concepts have been demonstrated most often for math and reading, and (c) standardized tests across many countries have reading and math portions.

In addition to investigating implicit math–gender stereotype, gender identity, and math self-concept via a child IAT, we also examined explicit (self-report) counterparts in the same children. One motivation comes from research suggesting that stereotypes can be separated into two underlying processes—one automatic, unconscious, and implicit and the other controlled, conscious, and explicit (Devine, 1989; see Killen, McGlothlin, & Henning, 2008, for a review of studies using explicit and implicit measures with children). In adults, positive but weak correlations are observed between implicit and explicit measures, especially in socially sensitive domains such as stereotypes (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). One of the explanations for this dissociation involves motivational influences: Implicit measures are assumed to be less susceptible to social desirability artifacts. It has also been suggested that early developmental experiences may shape implicit more than explicit cognition (Liben & Bigler, 2002; Rudman, 2004), again suggesting the value of using both implicit and explicit measures in the same study with the same children.

The present research draws on Heider’s (1946) balance theory. Heider’s principles of cognitive balance were extended by Greenwald et al. (2002) to explain how cognitive structures involving attitudes, stereotypes, and self-concepts organize themselves to become mutually consistent, or balanced. This extended formulation has been confirmed in recent studies (e.g., Greenwald, Rudman, Nosek, & Zayas, 2006; Greenwald et al., 2002). Balance theory has often been referenced in research with older adolescents (undergraduate students) and adults, but it has not been applied in early child development with the exception of a study of disadvantaged Hispanic children (ages 5–12) and adults measuring racial identity, race attitude, and self-esteem using self-report and IAT measures (Dunham, Baron, & Banaji, 2007). Within Heider’s theoretical framework, interconnections among concepts are assumed to self-organize in ways that reflect cognitive consistency or balance. Thus, a child who strongly associates self with Hispanic, and Hispanic with good, is predicted to have higher self-esteem, as found by Dunham et al. (2007).
The three chief aims of our study were to: (a) design new measures of children’s math–gender stereotypes and math self-concepts by adapting adult work from social psychology, (b) assess children’s math–gender stereotypes and math self-concepts during elementary school years, and (c) do so using both implicit and explicit measures within the same study. We examined three hypotheses: First, the child IAT created for this study will provide evidence of gender identity, in accordance with previous research that has established gender identity using self-report measures in elementary school and younger children. Second, American elementary school children will associate math more strongly with boys than with girls on both implicit and self-report measures. Third, on both implicit and self-report measures boys should self-identify with math more strongly than girls.

Method

Participants

A total of 247 American children (126 girls, 121 boys) from Grades 1–5 were tested. All children were recruited through private and public elementary schools from the greater Seattle area. The same recruitment procedure was used for both private and public schools: Schools mailed the consent forms to the parents, and completed forms were collected by the teachers. None of the children tested had repeated a grade. We were unable to obtain dates of birth for the recruited children; however, the mean age ranges for the first five elementary school grades in the Seattle area based on the school data were as follows: The mean age for children attending Grade 1 was 6.66 years (SD = 0.33) and the mean age for children in Grade 5 was 10.68 years (SD = 0.37). The sample sizes and gender breakdown for our test sample were as follows: Grade 1, n = 50 (24 boys, 26 girls); Grade 2, n = 49 (24 boys, 25 girls); Grade 3, n = 51 (25 boys, 26 girls); Grade 4, n = 49 (24 boys, 25 girls); and Grade 5, n = 48 (24 boys, 24 girls). According to the available school data, children were predominantly from middle- to upper-class families. According to parental report (collected independently by the schools for their annual reports and provided to us at testing), the children in our sample were 83.3% White, 9.6% Asian, and 7.1% African American. After the study was completed, we provided $10 checks for each participating family to school administrators who distributed them to the participating families.

Procedure

Children were tested individually in a separate quiet room outside of his or her classroom while seated at a desk facing a computer (either a 43- or 48-cm screen). Each test session began with a 3–5 min description of the study, during which children were familiarized with the test apparatus. The children were told that they would be “asked some questions” and then “play a computer game.” They were told that they would see and hear words during the game and would have to press a button to “let the computer know which word it is.” The procedure started with the administration of the self-report measures followed by the administration of the IATs.

Math–Gender Stereotype Measures

Self-report. The self-report math–gender stereotype measure was created for this study and administered as two Likert-scale questions using images from Harter and Pike’s (1984) Pictorial Scale. For each question, children were shown two pictures of a child and responded by reporting: (a) which character (boy or girl) they themselves believed possessed an attribute (e.g., liking math) to a greater degree, and (b) whether they believed the character possessed the attribute “a little” or “a lot.” This was done by their pointing to one of two circles (1.1 and 2.3 cm in diameter). One question requested selecting the boy or girl character as “liking to do math more.” The other question requested selecting the boy or girl character as “liking to read more.” All self-report questions were memorized by the experimenter and said aloud to the children. The two scores were subtracted from one another to arrive at the explicit score with lower and upper bounds of −2 and +2; positive values indicated that the child picked the same-sex character as liking to do math more. The Appendix provides a full list of all names used in self-report measures. These self-report measures were not administered to 16 of the 247 subjects because they had not been developed yet.

Child IAT. We adapted the standard, adult IAT (Greenwald et al., 1998) for use with children. Modifications were similar to those in previous child IAT procedures (Dunham et al., 2006; Rutland et al., 2005), including an adapted computer keyboard and presenting of stimuli simultaneously as written and spoken words (see details in the next section for further adaptations used in this study). An IAT score (D; Greenwald, Nosek, & Banaji,
2003) was calculated by comparing the speed with which children categorize exemplars from four categories under two instructional conditions that vary assignments of the four categories to two computer response keys, one operated with the left hand and the other with the right hand. The measure is based on the principle that it is easier to give the same response to items from two categories if the two categories are mentally associated than if they are not. Figure 1 provides a pictorial representation of the child IAT.

During the math–gender stereotype IAT, children first practiced sorting girl and boy names. They responded to girl names (Emily, Jessica, Sarah, Ashley) by pressing a response button on the left side of the keyboard (in the position of a “D” key) and to boy names (Michael, Andrew, David, Jacob) by pressing a response button on the right side of the keyboard (in the position of a “K” key). After that, children practiced sorting math words (addition, numbers, graph, math) and reading words (read, books, story, letters) using the same two response buttons (Greenwald et al., 1998).

Following these two single discrimination tasks, children completed two combined discrimination tasks in which all four categories were used. During the combined tasks, two of the four categories were mapped onto the same response key. In one condition, math words and boy names shared one response key, with reading words and girl names sharing the other. The second condition switched the key assignments of the math and reading categories. All single discrimination tasks consisted of 16 trials and all combined tasks consisted of 24 trials. Positive scores indicated stronger association of math with own gender than with opposite gender. Greenwald et al.’s (2003) scoring algorithm constrains the resulting D measure to have bounds of −2 and +2. The implicit data were also reanalyzed separately using two alternative approaches for computing the D measure by adding penalties to error trials (Greenwald et al., 2003): D–600 ms penalty as well as the D–2SD penalty measures. For all three IATs, the D–600 and D–2SD were not statistically significant from the D–as is measure (all ps > .26). The D–as is measure is therefore used throughout the text. In addition, to rule out speed of a response as a confound, we directly compared boys’ and girls’ response times (RTs) for each of our three IATs using independent t tests. In one of the three IATs, girls had slightly faster RTs, and in the other two boys had slightly faster RTs. However, none of the t test comparisons was statistically significant (all ps > .38), suggesting that boys and girls did not differ significantly in their overall speed of response on our IAT measures.

The keyboard was furnished with two large panels to replace the computers’ “D” and “K” keys (see Figure 1). Stickers with left-pointing and right-pointing arrows on those buttons indicated their use for left and right responses. To reduce the need for reading, each stimulus word—spoken in a female voice—was synchronized with the onset of the stimulus.

![Figure 1](image_url)

**Figure 1.** For the Child Implicit Association Test, items from four categories appear one at a time on a computer and are spoken over the loudspeaker, and children respond by pressing a response button. In one task (A), math words and boy names share a response key, as do reading words and girl names (stereotype congruent). In the other task (B), these assignments are reversed—math is paired with girl (stereotype incongruent).

Note. Children with the math–gender stereotype (i.e., boy = math) should respond faster to the task (A) than (B).
the written word on the screen. The intertrial interval was 500 ms. All words used as IAT stimuli were pretested with elementary school children for familiarity and comprehension. To ensure that children understood each IAT task, error responses were followed by a red question mark appearing on the computer screen. After committing an error children could not advance to the next trial until they provided the correct response. As is standard in IAT procedures, trial latency was recorded to the correct response. The Appendix provides the list of all IAT stimuli.

Additional Measures: Gender Identity and Math Self-Concept

Self-report. Two additional self-report measures were created for this study following Harter and Pike’s (1984) two-item Likert-scale format, as described earlier. The measure of gender identity consisted of two questions. For each question, children were shown two pictures of a child and the experimenter explained each picture while pointing to it (e.g., “On the left we have a boy. His name is David” and “On the right we have a girl. Her name is Emily”). Children were asked to report: (a) which character they were more like (e.g., “Are you more like David or are you more like Emily?”) and (b) the degree to which they were like the selected character (e.g., “How much like David [Emily] are you? A little or a lot?”). The measure was scored so that positive values indicated that the child picked the boy character.

The math self-concept measure also consisted of two questions. For each question, children were shown two pictures of a child (e.g., “On the left we have a girl. Her name is Jessica. Jessica likes math.” and “On the right we have another girl. Her name is Sarah. Sarah likes to read”). Children were asked to report: (a) which character they were more like (e.g., “Are you more like Jessica or are you more like Sarah?”) and (b) the degree to which they were like the selected character (e.g., “How much like Jessica [Sarah] are you? A little or a lot?”). Reading was expected to “go with” female in the sense that girls were expected to pick the same-sex character who was reading as more like them than the same-sex character who was doing math. Positive values indicated that the child picked the same-sex character who was doing math. Selecting a reading character in one of the two questions and the math character in the other would result in a value of 0 (indicating that the child, on this measure, had an equally strong identification with math and read-

ing). For the self-report measures, the order of the math–gender stereotype, gender identity, and math self-concept measures was counterbalanced across children. The order of characters assigned to left and right sides and the names used for each character were also counterbalanced across children. Order of administering self-report measures did not influence scores (all ps > .52) and was therefore not used as a factor in analyses to be reported.

Child IAT. Two additional IAT measures were administered. For the gender identity IAT, children classified the words representing me, not-me, boy, and girl. In one instructional condition, me words and boy names shared a response key, with not-me words and girl names sharing the other response key. In the other instructional condition, two of the response assignments were reversed, such that me words and girl names shared one key whereas not-me words and boy names shared the other key. Positive scores indicated stronger association of me with boy than with girl.

For the math self-concept IAT, children classified the words representing me, not-me, math, and reading. In one instructional condition, math and me words shared a response key, as did reading and not-me words. In the other instructional condition, left versus right assignment of me and not-me words was reversed. Positive scores indicated stronger association of me with math relative to reading. For the implicit measures, there were 16 counterbalancing conditions. The gender identity IAT and math self-concept IAT were counterbalanced in the first and third position, with the math–gender stereotype IAT administered in the second position. Within each IAT, order of the two instructional conditions was counterbalanced. The spatial orientation of categories assigned to left and right was counterbalanced across participants and IATs. Order of administration did not influence scores on any implicit measures (all ps > .66) and was therefore not retained as a factor in analyses to be reported.

Internal Consistency

For implicit measures, Cronbach’s alpha was calculated from two D measures computed for matched 24-trial subsets of each IAT. Cronbach’s alpha coefficients for the math–gender stereotype, gender identity, and math self-concept IATs were $\alpha = .74$, $\alpha = .89$, and $\alpha = .78$, respectively. For the self-report measures, Cronbach’s alpha coefficients for gender identity and math self-concept were $\alpha = .93$ and $\alpha = .79$, respectively. The two items of
the self-reported math–gender stereotype scale measured two distinct constructs (gender stereotype toward math vs. gender stereotype toward reading). Thus, the expectation was for low internal consistency of the self-reported math–gender stereotype measure, which was the case, $\alpha = .03$.

**Data Reduction**

Implicit measures ($N = 247$) were analyzed after excluding participants who met any one of three exclusion criteria: (a) 10% or more of their responses faster than 300 ms, (b) error rate of 35% or greater in at least one of the three IATs, or (c) average response latency 3 SD above the mean response latency for the whole sample in at least one of the three IATs. These criteria excluded 25 (10.1%) of the participants. This was done to reduce noise in the data by excluding participants who would be identified as outliers on the basis of pre-established criteria, consistent with the usual IAT procedures with adults (Greenwald et al., 2003). Self-report data ($N = 231$) were analyzed after excluding data from 11 participants (4.7%) due to excessively slow responding (either 30 s or more to respond to three or more self-report items, or 90 s or more to respond to one or more of them). The analyses following data reduction provided increased power compared to analyses of the full sample, but the pattern of significant results and the conclusions drawn from them remained unchanged.

**Results**

**Gender Identity**

Figure 2 displays the results for both the implicit and self-report measures. As expected, boys associated *me* with *boy* more strongly than did girls on both the implicit measure, $t(220) = 15.35, p < .001$, and self-report, $t(218) = 18.81, p < .001$. The IAT and self-report measures of gender identity were strongly correlated, $r = .64, p < .001$.

**Math–Gender Stereotype**

On the implicit measure, boys associated *math* with *own gender* significantly more than did girls, $t(220) = 6.46, p < .001$. Similarly, on the self-report measure boys were more likely to pick the same gender character as ‘liking to do math more’ than were girls, $t(218) = 4.75, p < .001$. These results mean that both boys and girls indicated stronger association of math with boys than with girls—evidence for math–gender stereotype. The overall implicit-explicit correlation for the math–gender stereotype measure was positive but small, $r = .14, p < .05$.

**Math Self-Concept**

There was evidence for gender-distinctive math self-concepts. On the implicit measure, boys associated *me* with *math* more than did girls, $t(220) = 2.63, p < .01$, and on the self-report measure boys identified more with a picture of a same gender character who was solving a math problem.
than did girls, $t(218) = 3.31, p < .01$. The overall implicit–explicit correlation for the math self-concept measure was also positive but small, $r = .28, p < .001$.

**Developmental Order of Emergence**

We next examined the order of emergence of the three separate measures (gender identity, math–gender stereotype, and math self-concept). To obtain adequate statistical power for examination of developmental change, pairs of adjacent grades were combined. More specifically, to smooth the irregularities due to modest sample sizes, each grade (e.g., Grade 2) was combined with the preceding (i.e., 1–2) as well as the following school grade (i.e., 2–3) to create four 2-grade levels. Figure 3 shows that gender identity was robustly evident throughout elementary school and even at the earliest grades on both implicit and self-report measures. Math–gender stereotypes were also robustly evident throughout for both the implicit and self-report measures. The data for self-concepts are more complex, as might be expected, because this concept is hypothesized to be developmentally dependent on the other two. The detailed statistical results corresponding to the data in Figure 3 are presented next.

In line with previous findings showing early identity and constancy (e.g., Martin, Ruble, & Szekrybalo, 2002; Ruble & Martin, 1998), there was robust evidence for gender identity in the youngest age group we tested (first grade) using both implicit and explicit measures. For the implicit measure,
gender identity was measured by girls associating me with girl more strongly than boys do, and conversely boys associating me with boy more strongly than girls do. The IAT results were highly significant in Grades 1–2, \( t(83) = 10.32, p < .001 \), and similar \( t \) tests were highly significant through increasing grades (all \( ps < .001 \)). Self-report gender identity showed the same pattern: Given a paired choice of two pictures, one a girl and the other a boy, boys were more likely to identify the boy picture as being more like themselves than were girls, and girls were more likely to identify the girl picture as being more like themselves; this difference was also statistically significant in Grades 1–2, \( t(87) = 12.42, p < .001 \), and similar \( t \) tests were highly significant at later grade levels (all \( ps < .001 \)). The finding of clear evidence for gender identity on both implicit and self-report measures is useful in showing that, even at the earliest grades examined, children could follow directions for both of these measures. Moreover, while it was known that explicit self-report tests would be successful in showing gender identity (e.g., Martin et al., 2002), it was not known, prior to the current tests, whether the IAT procedure would yield similar results, because no previous child IATs of gender identity have been conducted.

Sex differences indicating the presence of math–gender stereotype were also apparent for Grades 1–2, although weaker than gender identity. On the implicit measure, boys associated math with own gender more strongly than girls did in Grades 1–2, \( t(83) = 3.91, p < .001 \), and similar \( t \) tests were significant at each adjacent two-grade level thereafter (all \( ps < .001 \)). For self-report measures, boys were more likely than girls to pick the same-sex character as liking to do math in Grades 1–2, \( t(87) = 2.66, p < .01 \), and this was stable for subsequent grade levels (all \( ps < .05 \)).

The data for math self-concepts suggest that they are weaker, less stable, and may emerge later than the other two constructs. On the implicit measure of math self-concepts, boys associated me with math more strongly than did girls in Grades 1–2, \( t(83) = 2.30, p < .05 \). This remained significant in Grades 2–3, \( t(87) = 2.85, p < .01 \), but was not significant in later grades (\( ps > .23 \)). For the self-report measure of math self-concepts, in Grades 1–2 girls were slightly more likely to pick a character who was reading (and boys to pick a character who was doing math) as being more like themselves, but this was not statistically significant (\( p > .14 \)); statistical significance emerged only in Grades 4–5, \( t(87) = 2.52, p < .05 \).

In addition, we used an analysis of variance (ANOVA) to test for changes over grade in the implicit and explicit measures of gender identity, math–gender stereotype, and math self-concept. Six 2 (sex) × 5 (school grade) ANOVAs were used to test whether the measures varied as a function of sex and grade. There was no significant main effect of grade for any measure (all \( ps > .13 \)). Similarly, there were no significant Grade × Sex interactions (all \( ps > .11 \)). We also tested for possible linear trends in the Grade × Sex interactions. Three measures (explicit gender identity, implicit math–gender stereotype, and explicit math self-concept) showed weakly increasing effects (all \( ts < 1.33 \)); the other three measures, showed weakly decreasing effects (absolute value of \( ts < 1.74 \)). None of the linear trends was significant (all \( ps > .08 \)).

Building on Heider’s (1946) balance theory, Greenwald and colleagues provided a rigorous statistical method for assessing balanced identity among the three constructs tested here (gender identity, math–gender stereotypes, and math self-concepts; Greenwald et al., 2002). This method was applied to the current data, and full details are in the supplemental material (http://ilabs.washington.edu/GenderStereotypes). The main conclusions from these supplementary analyses are that evidence for cognitive balance patterns is (a) clearly present on implicit measures, (b) more apparent on explicit measures than previously reported in studies using explicit measures in adults, and (c) stronger with increasing school grade.

**Discussion**

In this study of elementary school children, we distinguished between math–gender stereotypes and math self-concepts using both implicit and explicit measures within the same study. The findings confirm that our child IAT (and self-report) procedures are effective inasmuch as they provide the expected evidence of gender identity. These methods allowed us to uncover two new findings. First, the math–gender stereotype previously found to be pervasive in American samples of adults was found in elementary school children on both implicit and self-report measures. Second, elementary school girls showed a weaker identification with math than boys on both implicit and self-report measures (math self-concept). This suggests that the math–gender stereotype develops early and differentially influences boys’ versus girls’ self-identification with math prior to ages at which differences in math achievement emerge.
Math–Gender Stereotypes

The current demonstration of math–gender stereotypes during elementary school years extends previous work on this topic (e.g., Aronson & Good, 2003; Muzzatti & Agnoli, 2007). For example, in one study of stereotypes in elementary grades, Ambady et al. (2001) found that the activation of female identity (e.g., coloring a picture of a girl holding a doll) significantly impeded girls’ performance on a subsequent math test. The children in that study did not explicitly report awareness of the American stereotype. We provide a potentially more direct way of measuring whether children have assimilated the American cultural stereotype about girls and math: We tested children’s explicit awareness (self-report) of the stereotype that “math is for boys” and the results showed that both boys and girls explicitly subscribe to this view. Like Ambady et al., we found evidence that such stereotypes also operate at an implicit level (but we did so with a different implicit measure). Children in Ambady et al.’s study used an implicit measure that did not rely on the IAT. Their participants were told a story about a student who “was especially good at math.” During the story, the gender of the student was never mentioned. The children were then asked to repeat the story, and the experimenter recorded whether the child used “he” or “she” when referring to the student. The boys were more likely to use the word “he,” but the same was not true for girls. We here provide a different, and perhaps a more sensitive, implicit measure of whether children have assimilated the American stereotype: The child IAT measure demonstrated the math–gender stereotype for both boys and girls.

Using an implicit measure and conceptualizing, the stereotype as an association between math and boy addresses an issue raised in the child stereotype literature (Signorella, Bigler, & Liben, 1993). It has been proposed that self-reported gender stereotypes in children may indicate the mere awareness of stereotypes as opposed to personal endorsement of those stereotypes. We found that the implicit stereotype about math (i.e., boys = math) was only weakly correlated with the self-reported stereotype (i.e., “boys like to do math more”). This low correlation can be interpreted as dissociation between explicit and implicit stereotypes about math ability. It may also inform the debate between awareness and endorsement. In adults, implicit measures have been shown to predict social behavior and decision making better than explicit measures in socially sensitive domains such as stereotypes (Greenwald et al., 2009), but not necessarily in other domains (such as consumer preferences). Future studies could be designed that use child implicit measures in conjunction with self-report measures to explore the development and interrelation between implicit and explicit knowledge of stereotypes, both for more sensitive (racial preferences) or less sensitive (object preferences) domains (Greenwald & Nosek, 2008; Liben & Bigler, 2002; O’Connor, Cvencek, Nasir, Wischnia, & Meltzoff, 2010).

Math Self-Concept

The definition of math self-concept used in the current study differentiates children’s identification with math from more global beliefs about themselves such as self-esteem (Marsh, Craven, & Debus, 1991; Wigfield, Battle, Keller, & Eccles, 2002). Other researchers investigating sex differences in children’s math self-concepts also recognize the value of sharp distinctions between self-concepts and self-esteem (Wigfield et al., 2002). Self-report questions that tap an evaluative aspect (“good at”) when asking questions about the self raise issues of self-esteem rather than a math self-concept; the latter entails an identification with math without regard to evaluations either about math or about me (me = math). This distinction between self-concept and self-esteem is useful because domain-specific self-concepts (e.g., math self-concepts) may have different motivational and behavioral consequences for children’s academic performance than do children’s global evaluative feelings about the self such as those involved in self-esteem (Heyman & Dweck, 1998). Older children are more prone to make domain-specific, stable attributions than younger children (Rhodes, Newman, & Ruble, 1990; Ruble & Dweck, 1995). Our methods may be useful for uncovering conditions under which children of different ages make specific attributions about themselves and how such self-attributions interact with academic achievement and choices (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 1999; Heyman, 2008; Ruble & Dweck, 1995).

Age-Related Changes

As expected, there was robust evidence for the presence of gender identity, indeed significant evidence as early as Grades 1–2 on both the implicit and explicit measures. These findings for gender identity are consistent with previous research (see Ruble & Martin, 1998, for a review). Moreover, these findings are useful because they establish...
that, even at the youngest grades we tested, the children could understand instructions for both the implicit and self-report measures and provided interpretable data for both.

Sex differences indicated that the presence of math–gender stereotypes also emerged during Grades 1–2 (see Figure 3). Given the stronger magnitude of sex differences in Grades 1–2 for gender identity than for math–gender stereotypes (see also Figure 3), our data are consistent with the speculations that gender identity develops before Grades 1–2 and that math–gender stereotypes emerge after gender identity. This time frame would be consistent with previous research on elementary school children’s susceptibility to gender stereotypes about math (Ambady et al., 2001) and their familiarity with the stereotypes associated with social identities (Bigler, Jones, & Loblinier, 1997; Signorella et al., 1993).

Sex differences relating to math self-concepts were present, but more weakly than for the other measures. This suggests that math self-concepts emerge later than both gender identity and math–gender stereotypes (see also Figure 3). This speculation is also consistent with previous research suggesting that sex differences in math self-concepts emerge in middle to late elementary school (Herbert & Stipek, 2005; Muzzatti & Agnoli, 2007).

Relation Between Implicit and Explicit Measures

For each of the three constructs of gender identity, math–gender stereotype, and math self-concept, implicit and explicit measures were positively correlated: The implicit–explicit correlation was strong for the measures of gender identity ($r = .64$) but relatively weak for the measures of math–gender stereotype ($r = .14$) and math self-concept ($r = .28$). Moderate or low positive correlations between implicit and self-report measures are often found in socially sensitive domains such as stereotypes (Hofmann et al., 2005), with IAT measures having greater predictive validity than explicit measures (Greenwald et al., 2009). The two weak correlations (between implicit and self-report measures of math–gender stereotype and math self-concept) suggest possible differential predictive use of these measures in subsequent child development research.

Relations Among Identity, Stereotypes, and Self-Concepts

Based on the current results and previous research (Ruble, Martin, & Berenbaum, 2006), we assume that gender identity emerges before the first grade but recognize that gender identity can be measured as a multidimensional construct (Egan & Perry, 2001) or more narrowly as membership in a gender category (which emerges by 3–4 years of age; see Martin et al., 2002; Slaby & Frey, 1975). The latter category membership association of $me = \text{girl}$ or $me = \text{boy}$ is what was tapped in current tests. The interesting developmental question is how children’s gender identity measured in this way interacts with the culture’s prevailing stereotypes about math ability. Two alternatives can be offered based on the current data: (a) stereotypes may be acquired first and influence self-concepts, or (b) early self-concepts may facilitate internalization of cultural stereotypes. The first holds that children who strongly identify with their gender (strong gender identity) are more likely to internalize cultural stereotypes about their gender (math–gender stereotypes), which in turn influences their math self-concepts. Considered from the perspective of girls, this developmental sequence can be expressed as: $me = \text{girl}; \text{girls} \neq \text{math}; \therefore me \neq \text{math}$. The second alternative proposes that children with a strong gender identity and a given level of self-identification with math (math self-concept) are more likely to generalize or project their own math identification to others of their own gender (math–gender stereotype). This developmental sequence can be expressed as: $me = \text{girl}; me \neq \text{math}; \therefore \text{girls} \neq \text{math}$. We favor the first alternative because it is relatively implausible that the weaker effect produces the stronger one (see Figure 2). The data and theory of Eccles, Wigfield, Harold, and Blumenfeld (1993) also do not support the second alternative. Moreover, few gender differences exist in actual ability during the elementary school-age period (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). The current correlational data do not allow us to identify actual causal mechanisms, and further research is needed on this point.

If principles of cognitive balance operate in children similarly to the way they do in adults and college students, there are implications for children’s academic development. For example, in female college students, a balanced configuration of math–gender stereotypes, gender identity, and math self-concepts is related to negative attitudes toward mathematics and lower performance on the mathematical portion of the Scholastic Aptitude Test (Nosek et al., 2002). In elementary school, boys and girls score equally well on math achievement tests (Hyde et al., 2008) and girls receive higher math grades (Kimball, 1989). Thus, the sex differences in math self-concepts detected by our tests precede
rather than follow actual differences in math achievement and may exert a developmental influence on children’s interest and effort, which could subsequently affect achievement (Barron, 2004; Good, Aronson, & Inzlicht, 2003). Recent research in adults shows that perceiving an academic field to be at odds with one’s identity leads to a sense of “not belonging” that deters people from pursuing that field (Cheryan, Meltzoff, & Kim, 2010; Cheryan, Plaut, Davies, & Steele, 2009). Early differences in identification with math demonstrated here might contribute not only to children’s current choices but also to how children project themselves in the future and think about who they aspire to be.

**Larger Implications**

Where do children’s stereotypes about academic subjects come from—parents, school, media, peers? Future studies will profit from detailed ethnographic studies following individual children in their everyday lives to document the kind and frequency of input they encounter in the real world from different sources (Bell, Lewenstein, Shouse, & Feder, 2009). At a more global level, societies themselves provide a “natural experiment.” Research over the past two decades has shown the pervasiveness of gender stereotypes about math in the United States (Killen, Sinno, & Margie, 2007; Liben & Bigler, 2002; Ruble et al., 2006; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). A recent study revealed that national estimates of implicit gender-science stereotyping correlate with national sex differences in eighth graders’ performance on international math and science assessments, even after accounting for general indicators of societal gender equality (Nosek et al., 2009). Furthermore, Nosek et al. (2009) showed that a stereotype of associating male (more than female) with science was evident in data provided by Internet respondents (nonrandom samples) from all 34 countries from which they obtained data, suggesting that gender-related academic stereotypes are not confined to America.

The current authors are planning a cross-cultural study using the experimental methods reported here. In Singapore, girls score higher than boys on standardized math assessments in fourth and eighth grades, and both sexes score higher than age- and grade-matched American children (Gonzales et al., 2008). In such a society, there may be absence or reversal of sex differences in elementary school children’s math self-concepts on our tests.

Regardless of the results of this future work, the authors believe that the concepts and data reported here will be helpful for future studies of children’s emerging math self-concepts and how such concepts differ as a function of the sex of the child and the cultural stereotypes to which they are exposed. We have conjectured that gender identity and math–gender stereotypes interact in the formation of children’s math self-concepts. Blending the work from social and developmental psychology will increase our understanding of the influence of group membership (e.g., being a boy) on how children form attitudes toward an attribute associated with their in-group (e.g., “I like math”; Banaji, Baron, Dunham, & Olson, 2008; Killen, Kelly, Richardson, Crystal, & Ruck, 2010; Rhodes & Gelman, 2008).

**Conclusions**

In the present research, young girls showed a weaker identification with math than did their male peers. Such gender differences in children’s math self-concepts may arise from the early combination of societal influences (cultural stereotypes about gender roles) and intrapersonal cognitive factors (balanced cognitive organization). Future studies will profit from unifying the concepts and experimental tools from developmental science and social psychology (Cvencek, Greenwald, & Meltzoff, in press; Dunham & Olson, 2008; Killen et al., 2008; Meltzoff, 2007; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Olson & Dweck, 2008; Rutland et al., 2005) to explore the development of academic identity and how it contributes to children’s educational choices, success, and future aspirations.

**References**


Appendix

Words for Self-Report Measures

**Boy**: Michael, Jacob, Joshua, David, Andrew, Robert, Ryan, William.

**Girl**: Emily, Sarah, Jessica, Ashley, Lauren, Hannah, Rachel, Jennifer.

Words for Implicit Association Tests

**Me**: my, mine, I, myself.

**Not-me**: they, them, theirs, other.

**Boy**: Michael, Andrew, David, Jacob.

**Girl**: Emily, Jessica, Sarah, Ashley.

**Math**: addition, numbers, graph, math.

**Reading**: read, books, story, letters.