A Stereotype Threat Account of Boys' Academic Underachievement

Bonny L Hartley and Robbie M Sutton

School of Psychology, University of Kent, Canterbury, UK

Manuscript of article published in *Child Development* (vol. 85, iss. 5, pp. 1716-1733, doi: 10.1111/cdev.12079). This may differ somewhat from published version. Copyright held by the authors and Society for Research in Child Development. Article is available from Wiley's Online Library (http://onlinelibrary.wiley.com/doi/10.1111/cdev.12079/abstract).

This research was supported by an Economic and Social Research Council 1+3 Open Competition Studentship awarded to Bonny Hartley. The authors are grateful to the head teachers, teachers, children, and parents who were involved in this research. Also, the authors thank Karen Douglas for her invaluable comments on an earlier draft of this article.

Corresponding author: Bonny L Hartley: Email: bonny.hartley@gmail.com, or Robbie Sutton: Email: r.sutton@kent.ac.uk

Abstract

Three studies examined the role of stereotype threat in boys' academic underachievement. Study 1 (children aged 4-10, *n* = 238) showed that girls from age 4 and boys from age 7 believed, and thought adults believed, that boys are academically inferior to girls. Study 2 manipulated stereotype threat, informing children aged 7-8 (*n* = 162) that boys tend to do worse than girls at school. This manipulation hindered boys' performance on a reading, writing, and math test, but did not affect girls'. Study 3 counteracted stereotype threat, informing children aged 6-9 (*n*= 184) that boys and girls were expected to perform similarly. This improved the performance of boys and did not affect that of girls.

Keywords: stereotype threat, gender, stereotypes, performance

Male advantage persists in all known societies, in many forms (Hausmann, Tyson, & Zahidi, 2010). Traditional gender stereotypes hold that men and boys are more agentic than women and girls (e.g., Rudman & Glick, 1999). Historically, boys' education has been prioritized over girls', and boys have received many more opportunities and encouragements to succeed academically (Weaver-Hightower, 2003). Yet, although there is variability across academic subjects, overall boys are underachieving in the education systems of developed countries (e.g., Department for Education and Skills [DfES], 2007; EACEA/Eurydice, 2010). This has been described as one of the most pressing educational equality challenges of current times (Equality and Human Rights Commission, 2010).

Race, class, and gender differences in educational performance cannot be taken as evidence of innate differences in ability. For example, educational disadvantage stems from inequalities of opportunity and social capital (Ogbu, 1986). However, given the widespread cultural milieu of male advantage, such inequalities would not seem to explain the phenomenon of boys' academic underachievement. Another source of educational disadvantage resides in societal expectations and stereotypes. A substantial literature shows that performance stereotypes about gender, race and socioeconomic status have self-fulfilling effects on academic attainment (see Jussim & Harber, 2005, for a review). Although traditional gender stereotypes point to male competence, it is possible that boys are at the wrong end of specifically academic stereotypes that portray them as inferior students. In this article, we examine the role of these gendered academic stereotypes in boys' academic underachievement. In particular, we focus on the role of the stereotype threat phenomenon.

Stereotype threat is one means by which stereotypes can become self-fulfilling (Steele, 1997). Stereotype threat occurs when individuals' task performance suffers as a result of their awareness that the social group they belong to is not expected to do well. In such a situation, individuals are faced with the threat of confirming a negative stereotype

about one's group and being seen or treated in terms of that stereotype. Acute exposure to stereotype threat can affect performance by causing rumination and anxiety, whereas chronic exposure can lead individuals to disengage with a performance domain to protect their self-esteem (Steele & Aronson, 1995).

Stereotype threat has been shown to degrade several aspects of performance among a range of social groups. In studies of this phenomenon, experimental participants are reminded that they are members of a group that is stereotypically expected to perform less well than other groups at the task in hand. These manipulations have been shown to negatively affect the IQ test scores of Black American undergraduate and college students (Steele & Aronson, 1995), the scholastic aptitude of students from low socio-economic backgrounds (Croizet & Claire, 1998), and the cognitive performance of elderly participants (Abrams, Eller, & Bryant, 2006). Researchers have demonstrated the scholastic aptitude of children as young as five to six years old is susceptible to stereotype threat effects regarding race and socio-economic background (Ambady, Shih, Kim & Pittinsky, 2001; Desert, Preaux, & Jund, 2009).

A large body of research has related stereotype threat specifically to women and girls' academic achievement. This makes sense considering that male advantage perseveres in society as a whole and is widely recognized (Hausmann, et al., 2010; Sutton et al., 2008). Historically, boys have been advantaged in schooling and other academic settings, especially in science, technology, engineering and mathematics (STEM) subjects (Nosek et al., 2009). Cvenek, Meltzoff, and Greenwald (2011) examined children's gender stereotypes about math, using the implicit association task (IAT). In keeping with the historical male advantage, they found that American children implicitly associated mathematics with boys more than with girls. Many published studies have also demonstrated the negative impact of stereotype threat on the performance of women in mathematics tests (e.g., Spencer, Steele, &

Quinn, 1999; Steele, 1997; for a review see Nguyen & Ryan, 2008). In addition, Ambady et al. (2001) found that girls as young as five to seven years old were affected by stereotype threat in math tests.

Given that in many respects girls and women have lower status in society, these studies illustrate how stereotype threat can impact on lower status or stigmatized groups. However, theoretically, stereotype threat can undermine the performance of anyone whose group is targeted by negative stereotypes alleging a lack of ability in a particular domain (Steele & Aronson, 1995). Indeed research has shown that even members of high status, non-stigmatized groups are susceptible to stereotype threat effects. For example, white participants underperform in sports when the task is said to reflect natural athletic ability (Stone, Lynch, Sjomeling, & Darley, 1999). Also men underperform on socio-emotional tasks under stereotype threat conditions (e.g., Leyens, Desert, Croizet, & Darcis, 2000).

Although the higher-status gender group – men and boys – may in principle be vulnerable to stereotype threat, every published study examining the role of stereotype threat in gendered academic performance has focused on female participants' stereotypical inferiority in STEM subjects. This is striking, given that in fact, there is overwhelming evidence that globally, girls outperform boys academically. One such study was conducted by the Organization for Economic Co-operation and Development (OECD, 2010a). Under the auspices of its Programme for International Student Assessment (PISA), it assessed the educational outcomes of 15 year old children in the 34 OECD countries, including the USA, Canada, Australia, the UK, France, Spain, Germany, and Italy, as well as 41 partner countries. Although boys outperformed girls in math by an average of 12 points, girls significantly outperformed boys in reading in every participating country by an average of 39 points (the equivalent of an average school year's progress).

Although the gender gap varies internationally, particularly in STEM fields (e.g., Nosek et al., 2009), in some countries, boys appear not to be advantaged even in math and science. For example, throughout elementary, middle and high school in the USA, girls obtain higher grades than boys in all major subjects, including math and science (American Association of University Women Educational Foundation, 1998; Cole, 1997). Similarly, recent statistics from England's Department for Education (DfE) reveal that girls outperformed boys on reading, writing, math and science at Key Stage 1 (age 5-7: DfE, 2010). This advantage persisted through to GCSE stage (age 16: DfE, 2012). Women also go on to outnumber men in university degree programmes. For example, women outnumber men 58:42 in UK universities (Higher Education Statistics Agency, 2011), and 57:43 in the US (Snyder & Dillow, 2011). In contrast, boys represent the majority of suspensions and expulsions from school (DfE, 2011a).

Although it is clear that boys *actually* tend to underachieve relative to girls, stereotype threat effects also require that boys are *widely perceived* to underachieve, so that they tacitly feel that they may be viewed through the lens of unfavorable stereotypes. Traditionally, gender stereotypes define men and boys as competent and dominant, but also more aggressive, less disciplined and conscientious than women and girls (Rudman & Glick, 1999). However, since the advent of feminism, sexist notions that men and boys have greater academic ability are seldom articulated explicitly (Arnot & Weiler, 1993). The remaining content of cultural portrayals of gender, as in the proverb "boys will be boys" (Foster, Kimmel, & Skelton, 2001), may hinder boys in academic settings, implying that they are too ill disciplined and inattentive to scale the same academic heights as girls.

Another source of stereotypes is likely to be the reality of individuals' experience. According to Campbell's (1967) "grain of truth" hypothesis, stereotypes partly originate from a person's experience of group differences. The male deficit in conduct and achievement is

so marked that it is unlikely to remain unnoticed by adults or the children themselves. The preponderance of girls in high ability groups (Kutnick, Blatchford, & Baines, 2002), and the greater disciplinary attention given to boys provide visible markers of academic inferiority. Indeed, research suggests that teachers regard underachieving children typically as boys and high achieving children as girls (Jones & Myhill, 2004).

Unfortunately little research directly assesses children's meta-stereotypic beliefs about gender and academic achievement. Some research suggests that children have absorbed a cultural stereotype that mathematics is traditionally masculine (Cvencek, et al., 2011). However, Martinot and Desert (2007) asked French children (aged 9 and 12), "How well do people think boys (girls) do in mathematics?" and whether they personally endorsed this stereotype. Results showed that girls of both age groups, and 12 year-old boys believed that boys do worse at mathematics than girls, and thought that others shared their belief. If boys perceive that they are viewed through the lens of such negative academic stereotypes, they become vulnerable to stereotype threat.

The Present Research

Research shows that boys underperform relative to girls, are stereotypically seen to underperform, and may believe that they are seen in these stereotypical terms. Thus, it is possible that boys' academic performance may be hindered by stereotype threat. The present research is the first empirical investigation of this possibility. Study 1 investigates whether, and at what age, children acquire the stereotype that boys are academically inferior to girls, and perceive that adults also endorse this stereotype. It therefore aims to establish whether, from whatever sources, young children receive social information that boys are not expected to do as well as girls. In Study 2, we present an experimental group of children with a direct and overt distillation of such cues. Adapting a widely-used manipulation of stereotype threat, we inform children that girls are expected to do better than boys at a test, and examine

whether this hinders boys' performance. In Study 3, we examine whether it is possible to improve boys' performance by nullifying these stereotypes, specifically by informing children that girls and boys are expected to do equally well.

Study 1

Study 1 examined whether and when children develop adverse academic stereotypes about boys, relative to girls. More specifically, we examined the age at which children develop stereotypic expectations about multiple aspects of academic ability, performance, motivation, and regulation of attention and behavior. Crucially, we also examined children's awareness that adults may endorse these gendered academic stereotypes. In the adult literature, this awareness of others' stereotypes has been referred to as meta-stereotype awareness (e.g., Vorauer, Hunter, Main, & Roy, 2000).

Consistent with theory and research on children's development of ingroup preferences, we hypothesized that in the early years of school children would display a degree of ingroup gender bias - believing that their gender is superior, and believing that adults share this view (Aboud, 1988; Yee & Brown, 1994). However, we predicted that as children progress through school they would increasingly endorse the stereotype that girls are academically superior to boys, and the meta-stereotype that adults share this view. This was based on recent findings that between the ages of 5 and 11, children become increasingly able to describe broadly held racial stereotypes held by others (McKown & Strambler, 2009), to understand perspective and establish theory of mind (Wellman, Cross, & Watson, 2001), and become increasingly aware of how they may be evaluated by others (e.g., Rutland, Cameron, Milne & McGeorge, 2005).

Children as young as 4 display ingroup bias, and for this reason we can expect very young girls to believe that they are better students than boys, and that adults share this view. Similarly, we can expect boys to be influenced by ingroup bias, even though evidence

suggests that by age 4, they are already underachieving (DfE, 2011b). Later, children's stereotypes and meta-stereotypes may be increasingly informed by environmental cues and developing stereotype consciousness. We therefore predicted that girls would endorse girls' academic superiority more strongly than boys, and that this gender difference would decline with age (Yee & Brown, 1994). We predicted that with age, boys' stereotypes and meta-stereotypes would come to agree with girls' impression that girls are, and are expected to be, better students.

Method

Participants and Design. Two hundred and thirty eight British school children (117 boys and 121 girls) from foundation stage (aged 4-5 years) through year 5 (aged 9-10 years) participated in the study (*M* age = 6.87 years). The study was therefore a 2 (gender: boy vs. girl) x 6 (year group: F/1/2/3/4/5) quasi-experimental between-groups design. Dependent measures were academic gender stereotypes and meta-stereotypes. Children were recruited from two primary schools in England, and given a sticker for their participation. Ethnicity was not recorded, but the ethnic breakdown of children in participating schools was 82% White, 8% Asian, 5% Biracial, and 2% Black (3% not recorded or refused). As an indication of economic disadvantage, the percentage of children in participating schools who were eligible for subsidized or free school meals (FSM) was 23% (this was not recorded for individual children). Parental consent was obtained prior to administration of the measures, and children gave their verbal assent to participate.

Measures and Materials.

Academic gender stereotypes. We developed a measure similar to the picture-story technique known as the Sex Stereotype Measure (Best et al., 1997; Williams, Bennett, & Best, 1975) originally designed to assess stereotypes about male and female traits. The

adapted measure was designed to assess gender stereotypes specifically about *academic* traits and behaviors.

We created two sets of 13 brief scenarios (each presented on A4 card), half of which portrayed a child (gender not specified) with good conduct and achievement, and half of which displayed a child with poor conduct and achievement. This measure reflected a range of student attributes that have been recognized in the educational literature as desirable academic criteria, such as ability, performance, motivation, and self-regulation of attention and behavior (Haahr, Nielsen, Hansen, & Jakobsen, 2005). Importantly, this measure needed to reflect attributes which are also likely to be recognizable to children themselves (a full list of scenarios is available in the online supplementary materials Appendix S1). There are many other variables relevant to academic achievement, such as self-efficacy, interpersonal confidence, dispositional optimism, implicit theories of intelligence, and locus of control. However, these are unlikely either to be known to students and teachers, or to be considered important criteria by which to assess a child's performance at school.

On the reverse of each card was a male and female silhouette represented in black against a white background. Gender was represented only by hair length and style of dress, providing no other discriminable cues. Children responded to the scenarios verbally or by pointing to the appropriate silhouette (as used by Best et al., 1977). The position of the silhouettes was counterbalanced so that in half the stereotypically 'male' scenarios the male was on the left and female on the right, and the opposite for the other half, and the same for stereotypically 'female' scenarios. When children chose the silhouette of the boy this was scored as 0, and when they chose the girl this was scored as 1.

Academic gender meta-stereotypes. Children's meta-stereotypes about adults' gendered expectations were measured by asking children "Who do grown-ups think (mostly boys or mostly girls)…" "are cleverer", "do the best at school", "want to learn the most and

try the hardest", "are better at sitting quietly and listening to the teacher", "are better at concentrating on doing their work", and "are better behaved". Again children responded verbally or by pointing to either the male or female silhouette.

Procedure. Children participated individually in a communal area of their school such as their school library or corridor, under the supervision of a female experimenter. Following some initial conversation to put the child at ease, the experimenter gave the following instructions (Best et al., 1977): "What I have here are some pictures I would like to show you and some stories that go with each one. I want you to help me by pointing to the child in each picture that the story is about. Here I'll show you what I mean". Care was taken initially to ensure that children knew that the silhouettes represented a boy and girl rather than a man and woman. The experimenter (blind to the position of the male and female silhouette for each scenario) then displayed the first silhouette card and read the first scenario. Following the child's choice, the experimenter recorded the response (0 or 1) and continued with the remaining 12 scenarios. The experimenter insisted upon a definite response to each story, and if the child hesitated the experimenter repeated the story and said "try one" or "point to one of them". Children then answered the meta-stereotypic questions about grown-ups' gender-based expectancies, were verbally debriefed, and returned to their classrooms. The total duration of the study was approximately 10 minutes.

Results

Nine children were excluded from the analysis (leaving 229). Of these excluded children, two could not choose between the boy/girl options, answering 'both' to many questions (no other children answered "both" to any questions), three were very distracted and failed to complete the task, two named individual children in their class, and two thought there was a pattern and answered boy/girl/boy/girl before hearing the scenario. There were no missing data by scale. In addition, there was insufficient sample size in year 4 (n = 16) to

yield reliable or valid findings. Combining them with year 3 or 5 did not affect the results, so they were excluded from analysis, leaving 213 children.

Academic Gender Stereotypes. The mean of the 13 items formed a single index of children's academic stereotypes (α = .80), coded such that higher scores reflected beliefs in girls' superiority. This was submitted to a 2 (gender) x 5 (year) between-participants ANOVA (see Table 1 for a breakdown of results by year group and gender). As predicted, there was a main effect of gender, such that girls' superiority was endorsed more strongly by girls (M = .82, SD = .19) than by boys (M = .64, SD = .25), $F(1, 203) = 41.50$, $p < .001$, η^2 = .170. In line with hypotheses, there was a main effect of year group, such that girls' superiority was increasingly endorsed as children progressed through school, $F(4, 203) = 10.99$, $p < .001$, $\eta^2 = .178$.

These main effects were qualified by an interaction between gender and year group, $F(4, 203) = 3.61$, $p = .007$, $\eta^2 = .066$. Simple effects showed that whereas girls' perceptions of gender differences did not change across years, $F(4, 203) = 1.19$, $p = .317$, $\eta^2 = .023$, boys increasingly endorsed perceptions of girls' academic superiority as they progressed through school, $F(4, 203) = 12.86$, $p < .001$, $\eta^2 = .202$. Post hoc comparisons using the Tukey HSD test indicated that boys' perceptions of girls' superiority jumped significantly between years 2 and 3, $p = .015$ (see also Table 1). Increases occurred between each of the other consecutive years but were not statistically significant (a figure detailing effects from this study is available in online supplementary materials Figure S1).

Academic Gender Meta-stereotypes. The mean of the six items formed a single index of children's academic meta-stereotypes (α = .71). This was submitted to a 2 (gender) x 5 (year) between participants ANOVA (see Table 1 for a breakdown of results by year group and gender). As predicted, there was a main effect for gender, such that meta-stereotypes regarding girls' superiority were perceived more strongly by girls (M = .80, SD =

.23) than by boys (M = .56, SD = .31), $F(1, 203) = 46.48$, $p < .001$, $\eta^2 = .186$. In line with hypotheses, there was a main effect of year group, such that as they progressed through school, children increasingly believed that adults saw girls as academically superior to boys, $F(4, 203) = 8.34$, $p < .001$, $\eta^2 = .141$.

These main effects were qualified by an interaction between gender and year group, $F(4, 203) = 5.01$, $p = .001$, $\eta^2 = .090$ (note that this interaction effect , and that for stereotypes, holds when we conduct hierarchical linear regression analysis, and whether we use year group or chronological age as predictors). Simple effects showed that whereas girls' perceptions of gender differences did not change across years, $F(4, 203) = 0.52$, $p = .719$, $\eta^2 = .010$, boys' beliefs that adults see girls as academically superior increased as they progressed through school, $F(4, 203) = 12.24$, $p < .001$, $\eta^2 = .194$. Post hoc comparisons using the Tukey HSD test indicated a linear increase in boys' perceptions of girls' superiority across years because increases apparent between consecutive years were not statistically significant (all ps >.05).

To test whether the relation between children's stereotypes and meta-stereotypes would weaken as children progressed through school, we computed a hierarchical linear regression. Gender, year group, gender x year group, gender x stereotypes (mean-centered), year group x stereotypes, and gender x year group x stereotypes were entered as predictors of meta-stereotypes. Consistent with hypotheses, there was a significant interaction between year group and stereotypes, $\beta = -.22$, $t(206) = -3.76$, $p < .001$, such that as children progressed through school, meta-stereotypes became increasingly independent of stereotypes (see Table 1). Further, this interaction was not qualified by children's gender, $\beta = .01$, $t(205) = 0.18$, $p = .855$. There was some negative skewness to the residuals of the stereotype and meta-stereotype variables (-.55 and-.57 respectively). Reflected inverse transformations of the original variables reduced skewness to acceptable levels (-.07 and -.15 respectively).

Performing the above analyses on the transformed stereotype and meta-stereotype variables did not affect the results.

Although all analyses presented so far were performed on the single indexes of academic stereotypes and meta-stereotypes, binomial analyses were conducted on each of the individual items. Results showed that across all stereotype items, more than half the girls (from year F) and boys (from year 3) believed that girls were superior to boys. This pattern became significant by year 5 on all items (except the 3 items measuring "cleverness", "concentration" and "sitting quietly" as rated by boys). Similarly, across all meta-stereotype items, more than half the girls (from year F) and boys (from year 5) believed that adults saw girls as superior. This pattern became significant by year 5 on all items (except "concentration" and "tries harder"). By year 3, children of both genders endorsed the specific stereotype and meta-stereotype that girls perform better than boys at school.

Discussion

Results of Study 1 support our hypothesis that as children progress through school, they increasingly endorse the stereotype that girls are academically superior to boys and the meta-stereotype that adults share this view. First, as predicted, girls endorsed girls' academic superiority from 4 years old. At this age, girls are unlikely to be aware of the stereotype that girls are academically superior to boys. Rather this finding is more likely to reflect girls' ingroup gender bias, which develops from a very young age (Aboud, 1988; Bigler & Liben, 2007; Yee & Brown, 1994).

Second, although we expected young boys (4-5 years) to display ingroup gender bias (believing that boys are academically superior), boys in foundation stage (4-5), year 1 and year 2 were reasonably equally split between believing boys or girls were superior. It seems that the normal tendency for boys to show ingroup gender bias at 4-5 years old is at odds with their (or their friends') experience of male underachievement in the classroom. With age,

boys increasingly endorse the stereotype that girls are academically superior, particularly around age 7-8 years. This is consistent with Aboud's (1988) socio-cognitive account of stereotype development which suggests that from 7 to 8 years of age children are able to base stereotypes on more complex psychological categories and traits (e.g., intelligent, naughty, friendly), in addition to simple physical categories (e.g., sex and race).

Third, in line with hypotheses, boys increasingly developed meta-stereotypical academic beliefs as they progressed through school, reporting that adults believe girls have better conduct and achievement than boys. Therefore, our findings suggest that children not only endorse gendered academic stereotypes, but believe that adults share them. The development of boys' meta-stereotype awareness is consistent with research which indicates that children become increasingly able to infer others' stereotypes between the ages of 5 and 11 (e.g., McKown & Strambler, 2009). It is also consistent with developmental increases in other meta-cognitive abilities around middle childhood, such as perspective taking, theory of mind abilities, and evaluation concerns (Rutland et al., 2005; Wellman et al., 2001).

Furthermore, the relation between children's stereotypes and meta-stereotypes weakened as children progressed through school. This suggests that older children's responses reflect stereotypic and meta-stereotypic awareness, rather than mere projections of ingroup bias and personally held beliefs. For boys, this is apparent in the changes in mean stereotypic and meta-stereotypic responses, as well as the weakening association between these responses. Although girls' mean stereotypic and meta-stereotypic responses did not vary with age, they too became increasingly dissociated with age. From this, we can tentatively conclude that girls' meta-stereotypic responses increasingly reflect bona fide awareness rather than projections of their own biases. These findings are consistent with Augostonis and Rosewarne's (2001) findings that with age, children's perceptions about what others believe become distinct from their personally held beliefs.

Study 2

Study 1 showed that from a young age, children believe that girls are expected to be better students than boys. This meta-stereotype is a necessary cultural condition for stereotype threat effects. That is, stereotype awareness at a large scale is needed for stereotype threat effects to be contributing to the real-world gender gap in school achievement. It also suggests, since children are not born with academic gender stereotypes, that children are being exposed to social cues that girls are stronger academically and expected to do better than boys. We were interested in how direct exposure to this cultural stereotype may affect children's performance. Study 2 tests whether a message conveying this meta-stereotype can influence children's performance in a self-fulfilling way. To do so, we experimentally exposed children to the stereotype that boys underperform compared to girls and then measured their performance. Specifically, following methods employed in many previous studies of stereotype threat in other intergroup contexts (Aronson et al., 1999; Leyens et al., 2000; Nguyen & Ryan, 2008), we informed an experimental group of children that boys were expected to do worse in the test they were about to take. A control group of children received no information about gender-based expectancies.

In this study, we sampled children in year 3 (7-8 years old) since by this age they have developed the meta-cognitive and perspective-taking abilities needed to be influenced by such a message (i.e., to infer that a perceiver will place them in a gender category and stereotype them accordingly). Indeed, Study 1 showed that most children by this age have developed awareness of relevant academic stereotypes. However, it is important to note that individual children need not be aware of this stereotype in order for a direct stereotype threat manipulation to affect them. For some children, a direct manipulation may act as a reminder of a stereotype that they already know. For others, it may come as new information. In

either case, the manipulation makes children situationally aware of the stereotype. Therefore, we predicted that the stereotype threat manipulation would impair boys' performance.

Hypotheses surrounding girls' subsequent performance were more exploratory. Girls' performance may be boosted after reminding them that it is meant to be better than boys' (cf. Walton & Cohen, 2003). On the other hand, positive expectations are less self-fulfilling than negative ones (Jussim & Harber, 2005), and girls may be only weakly affected by reminders of a stereotype that they already know.

Method

Participants and Design. One hundred and sixty two British school children (80 boys and 82 girls) aged 7-8 years (*M* age = 7.40) were recruited from three primary schools in England, and given a sticker for their participation. Ethnicity was not recorded, but the ethnic breakdown of children in participating schools was 89% White, 5% Asian, 4% Biracial, and 2% Black. The percentage of children in participating schools who were eligible for FSM was 20%. Children were allocated to a 2 (stereotype threat: threat vs. control) x 2 (gender: boy vs. girl) between-participants design. The dependent variable was test performance. Parental consent was obtained prior to administration of the measures, and children gave their verbal assent to participate.

Assessments and Measures. All measures were presented in an A4 question booklet. The front of the booklet provided standardized instructions about answering the questions:

> Read each question carefully, work out the answer and then write it in the space provided in the booklet. If you make a mistake you can rub/cross it out and write your new answer. Do as many of the questions as you can but if you get stuck on a question move on to the next question and then you can come back to it later. It's really important that you sit quietly, concentrate well and try really hard.

These instructions and the test items themselves were similar to those children receive in the Standard Attainment Tests (SATs) taken by all children in England toward the end of year 2, but considerably shortened for the study. SATs are a reliable age-appropriate measure of performance that children can complete independently.

Experimental manipulation. Children in the stereotype threat condition were told "We've looked at how well children do on this test and we have found that girls do better than boys. Boys don't do as well", whereas children in the control condition were told "We've looked at how well children do on this test and we just want to see how you do".

Test performance. Children's performance was measured using a combined test of math, reading and writing exercises. The order of these three exercises was counterbalanced, resulting in six versions of the question booklet.

Math. A selection of 15 math questions, varying in difficulty, was taken from two previous Key Stage 1 SATs papers (Qualifications and Curriculum Authority: QCA, 2004a, 2004b). These papers are used to assess all UK children reaching the end of year 2. The questions covered various aspects of mathematics including number (e.g., "Write the total. 25 + 32 + 13 = "), shape (e.g., "Two of these shapes have no lines of symmetry. Draw a cross [x] on them."), and measurement (e.g., "Sita has 3m 60cm of ribbon. She cuts it into 3 equal pieces. How long is each piece?"). Children received one mark for each correct answer, meaning there were 15 marks available. Final math scores represented the proportion of questions answered correctly. This score was then standardized.

Reading. This exercise was sourced from a previous Key Stage 1 SATs paper (QCA, 2001), again used for children reaching the end of year 2. The exercise consisted of two sections of text to read, accompanied by illustrations and five questions testing comprehension of the text. Three of the questions were open ended, for example, "When did Billy start to cry?" and "Where did Glenda say she was?". Children received one mark for

each question if they wrote the correct answer. Two questions were multiple choice where children picked one of four possible answers, for example, "Who came to see mum? – Aunt Bessie, Dad, Mrs Wilson, Billy". Children received one mark for each question if they ticked or crossed the correct response. If children indicated more than one response, they were not awarded the mark. There were a total of five reading marks available. Children's final reading score represented the proportion of questions answered correctly. This score was then standardized.

Writing. For the writing exercise, children were presented with a picture of an old wooden door and were asked to "Write a creative story about what you think is behind the door and what happens when you go through the door". This exercise was developed to allow children to work independently and enable the assessment of several aspects of children's writing. Children's writing was assessed on seven dimensions: story length (number of words written), spelling (percentage of words spelt correctly), punctuation (coded on a three-point scale - none, some, lots), handwriting (coded on a five-point scale – illegible, messy and difficult but possible to read, relatively neat and easy to read, neat and easy to read, very neat and easy to read and stylish), creativity (coded on a four-point scale - unclear or hard to tell, not at all, somewhat, very), detail and complexity of story (coded on a three-point scale - not at all, somewhat, very), and ability to follow instructions (coded on a four-point scale - unclear or hard to tell, not at all, somewhat or reasonably, very well).

Two researchers, blind to gender and stereotype threat condition, coded children's writing performance as above. Each component of writing performance was coded on slightly different scales and so was individually standardized. These seven standardized scores were then averaged to form a single index of writing performance for each coder. Initial inter-rater reliability was high, $r(142) = .93$, $p < .001$ and any differences between the two coders were resolved by discussion.

Procedure. Children participated individually or in small groups in a communal area of their school such as their school library or corridor, under the supervision of a female experimenter. Following some initial conversation to put the child at ease, children were read the instructions from the front of the question booklet and either the control or stereotype threat manipulation instructions. Children then completed a couple of questions to make sure they had listened to the instructions and understood what they had to do (e.g., "What is the test going to be about?") and those in the threat condition also answered an additional question, "Who does better in this test?". Children responded by ticking or circling one of three options "Boys do better", "Girls and boys do the same", or "Girls do better". Each option was coupled with a silhouette image of a boy, a boy and girl, or a girl, respectively. Children then completed the performance test. If children required help reading, questions were read aloud as necessary.

The total duration of the study was approximately 20-30 minutes, after which children were orally debriefed and given positive feedback about their performance. In order to neutralize any negative thoughts caused by the stereotype threat manipulation, children were told that actually boys and girls were both doing very well.

Results

Thirteen children were excluded from the analysis (leaving 149; 69 boys and 80 girls). Six of these were highly distracted, six answered the understanding check questions incorrectly, and one could not read or write. Boys were excluded approximately equally across conditions; of the 11 boys excluded, six were in the stereotype threat condition and five were in the control group. Children's math, reading and writing test performance was submitted to a mixed ANOVA, with task as a repeated measures factor, comparing children's performance across gender and stereotype threat. There was a significant main effect of gender, $F(1, 136) = 6.18$, $p = .014$, $\eta^2 = .043$, such that girls ($M = .13$, $SD = .60$) performed

significantly better than boys (M = -.14, SD = .75). There was also a main effect of stereotype threat, $F(1, 136) = 7.27$, $p = .008$, $\eta^2 = .051$, such that children performed significantly worse in the experimental stereotype threat condition (M = -.13, SD = .70) than in the control group (M = .14, SD = .64).

Critically, as predicted, there was a significant interaction between stereotype threat and gender, $F(1, 136) = 7.11$, $p = .009$, $\eta^2 = .050$ (see Figure 1). Boys performed significantly worse in the experimental (stereotype threat) condition (M = -.44, SD = .72) than in the control group (M = .15, SD = .67), $F(1, 136) = 13.24$, $p < .001$, $\eta^2 = .089$. In contrast, girls' overall performance was not affected by the manipulation, $F(1, 136) = 0.001$, $p = .982$, $\eta^2 = .000$ (experimental condition M = .13, SD = .59, control group M = .13, SD = .63). Further, there was no three-way interaction between gender, stereotype threat and task (math, reading, or writing), $F(2, 272) = 1.65$, $p = .193$, $\eta^2 = .012$, and no main effect of task, $F(2, 272) = 0.17$, $p = .846$, $\eta^2 = .001$.

Although we had no specific predictions about the difference between girls' and boys' performance within conditions, there was no difference between boys' and girls' performance within the control group, $F(1,136) = 0.16$, $p = .899$, $\eta^2 = .000$.

Discussion

The present findings provide the first evidence that direct messages about boys' academic inferiority can become self-fulfilling through the stereotype threat phenomenon (cf. Steele, 1997). In particular, the results highlight how boys' performance can be significantly impaired by a reminder that it is expected to be worse than girls'. Further, these findings contribute to growing evidence that children are also susceptible to stereotype threat effects in middle childhood (Ambady et al., 2001; Desert et al., 2009). In addition, they corroborate findings that traditionally non-stigmatized, high status groups (such as men and whites) are

also vulnerable to stereotype threat when stereotypes suggest that they are deficient at the task at hand (Leyens et al., 2000; Stone et al., 1999).

On the other hand, girls' performance appears unaffected when they are aware that it is expected to be better than boys'. These results suggest that stereotypes about male academic inferiority harm boys' performance rather than boosting girls'. Previous research looking at stereotype lift has only examined lift effects for dominant, high status groups such as men, whites, and members of high SES groups (Walton & Cohen, 2003). Less is known about how stereotype lift affects traditionally low-status groups in society who are subject to wider negative stereotypes operating alongside specific positive ones. Walton and Cohen (2003) argue that stereotype lift particularly benefits those who believe in the validity of negative stereotypes and the legitimacy of group-based hierarchy. Girls may not believe that their academic superiority is legitimate in the context of wider gender inequalities in society which favor men. By 7-8 years old, children believe that men have more power than women in society, and by 9-10 years old, children believe that men have greater status than women (Neff, Cooper, & Woodruff, 2007). Further, recent research suggests that by age 5, children show system-justifying tendencies which legitimize traditional social hierarchies (e.g., Baron & Banaji, 2009). As such, stereotypes about boys' academic inferiority at school may be insufficient to produce boost effects for girls (see also Cheryan & Bodenhausen, 2000).

Study 3

After showing that stereotype threat can significantly impede boys' performance, Study 3 examined whether boys' performance could be improved by counteracting their prior negative performance expectations and lowering stereotype threat. Previous research has found that the performance of stereotyped groups can be improved compared to controls, by informing individuals that there are no group-based differences in performance (Quinn & Spencer, 2001; Spencer et al., 1999). Therefore, in Study 3 we used similar instructions to

manipulate children's prior expectations about how well boys and girls are meant to perform on a test and then measured their performance. A key assumption is that pre-experimental stereotype threat is present to an extent in classroom settings and children's everyday experiences, and therefore counteracting these meta-stereotypes may improve performance.

We were interested in sampling children between the ages of six and nine (years 2 and 4), to intervene at the point when children's consciousness of gendered academic expectations is appearing. We predicted that boys' performance would improve when they believe that they are expected to perform as well as girls (counteracting any prior meta-stereotypes about performance). Given that we observed no boosting effect of the stereotype threat manipulation on girls' performance in Study 2, we predicted that girls' performance would be unaffected by this prior information. Nonetheless, previous work on stereotype lift suggests a risk that girls' performance would decline if the manipulation contradicted their previous, more positive expectations (Walton & Cohen, 2003).

Method

Participants and Design. One hundred and eighty four British school children (98 boys and 86 girls) aged 6-9 years (*M* age = 7.43) were recruited from two primary schools in England, and given a sticker for their participation. Ethnicity was not recorded, but the ethnic breakdown of children in participating schools was 86% White, 6% Asian, 4% Biracial, and 2% Black (2% not recorded or refused). The percentage of children in participating schools who were eligible for FSM was 23%. Children were allocated to a 2 (stereotype nullification: nullification vs. control) x 2 (gender: boy vs. girl) between-participants design. The dependent variable was test performance. Parental consent was obtained prior to administration of the measures, and children gave their verbal assent to participate.

Assessments and Measures. As in Study 2, all measures were presented in an A4 question booklet with the same instructions on the cover page.

Experimental manipulation. Children in the stereotype nullification condition were told, "We're looking at how well children do on this test and we expect that boys and girls will do the same", whereas children in the control condition were told, "We're looking at how well children do on this test and we just want to see how you do".

Test performance. Children's test performance was measured using a numeracy or literacy exercise. Scores were standardized to create comparable totals across the two tests.

Numeracy. Thirteen math questions, varying in difficulty, were selected from two previous Key Stage 1 SATs papers (QCA, 2004a, 2004b). Similar to Study 1, children received one mark for each correct answer. Therefore, there were a total of 13 marks available and children's final math score represented the proportion of questions answered correctly. Two versions of the exercise were created appropriate for the two age groups.

Literacy. Seventeen questions were selected from verbal reasoning practice assessment papers, specifically designed for the age groups participating (Bond, 2007a, Bond, 2007b). The questions covered various aspects of verbal reasoning, including selecting words (e.g., "Underline the two words which are the odd ones out in these groups of words. Hat, gloves, silk, coat, cotton"), code sequences (e.g., "If DEF is the code for NAP, FED is the code for..."), sorting words (e.g., "Underline the pair of words which are the most similar in meaning: alley, lane; real, false; back, forward") and logic (e.g., "Change the first word of the third pair in the same way as the other pairs to give a new word: art, part; ale, pale; ant, ...") . Each block of questions was accompanied by a completed example. Children received one mark for each correct answer, so 17 marks were available. Children's final verbal reasoning score represented the proportion of questions answered correctly.

Procedure. Children participated individually in a communal area of their school, under the supervision of a female experimenter. Following some initial conversation to put the child at ease, children were read the same standardized instructions from the front of the

question booklet as in Study 2, and either the control or nullification manipulation instructions. Children in the nullification condition then completed a question to make sure they had listened and understood the instructions: "Do we expect... Boys to do better, Girls and boys to do the same, or Girls to do better?" Each option was coupled with a silhouette image of a boy, a boy and girl, or a girl, respectively. Children responded by ticking or circling one of these options. Children then completed the numeracy or literacy test. If children required help reading, the questions were read aloud as necessary. The study took approximately 10-20 minutes, after which followed an oral debrief and positive feedback.

Results

Four children were excluded from the analysis (leaving 180, 95 boys and 85 girls); two children had very low ability and could not understand the instructions or task, one child had to go back to class, and one child did not take the task seriously and rushed through all questions. Children's numeracy and literacy scores were submitted to a two-way ANOVA comparing children's performance across gender and stereotype nullification. There was no main effect of gender, $F(1, 176) = 0.92$, $p = .340$, $\eta^2 = .005$ and a significant main effect of stereotype nullification, $F(1, 176) = 5.85$, $p = .017$, $\eta^2 = .032$, such that children performed significantly better in the stereotype nullification condition ($M = .21$, $SD = .92$) than in the control group ($M = -.15$, $SD = 1.01$).

In line with predictions, there was a significant interaction between stereotype nullification and gender, $F(1, 176) = 5.12$, $p = .025$, $\eta^2 = .028$ (a figure detailing effects from this study is available in online supplementary materials Figure S3). Boys performed significantly better in the experimental (nullification) condition ($M = .30$, $SD = .97$) than in the control group ($M = -.37$, $SD = .98$), $F(1, 176) = 11.60$, $p = .001$, $\eta^2 = .062$. In contrast, girls' overall performance was not significantly affected by the manipulation, $F(1, 176) = 0.01$, $p = .914$, $\eta^2 = .000$ (nullification condition $M = .11$, $SD = .86$, control group $M = .09$,

SD = 1.00). There was no three-way interaction between gender, stereotype nullification and task (literacy or numeracy), $F(1, 172) = 0.42$, $p = .516$, $\eta^2 = .002$, and no main effect of task, $F(1, 172) = 0.05$, $p = .825$, $\eta^2 = .000$. Likewise, year group did not moderate effects, $F(1, 172) = 0.51$, $p = .477$, $\eta^2 = .003$.

Discussion

The findings from this study support our hypothesis that boys' performance can be improved by counteracting stereotypes about gender and performance. Boys performed better when told that they were expected to perform equally as well as girls. These findings are consistent with previous research (e.g., Quinn and Spencer, 2001; Spencer et al., 1999), which has found the performance of gender groups hampered by adverse stereotypes can be improved by telling them that there are no gender differences in performance on a particular test. Girls' performance was not affected by the stereotype nullification message.

These findings have important implications for strategies to help boys fulfill their potential. Other strategies which counteract or neutralize gendered academic stereotypes may improve boys' performance and reduce inequalities in important educational outcomes. These include framing the stereotype threat as a challenge (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010), encouraging children to perceive intelligence as malleable, and providing external attributions for difficulty and anxiety (Good, Aronson, & Inzlicht, 2003).

Although our focus was on the effects of the manipulation on boys' and girls' performance, rather than simple gender differences in performance within conditions, it is worth pausing to comment on these differences. Boys did as well as girls in the control condition of Study 2, but worse than girls in the control condition of Study 3 ($F(1,176) = 5.30$, $p = .022$, $\eta^2 = .029$). A range of extraneous factors, both endogenous and exogenous to the tests themselves, will cause variability. We would expect differences to reliably emerge in reviews of multiple tests taken by many children over many situations to (e.g., DfE, 2010;

DfE, 2012), but not necessarily in any single test. For example, Steele and Aronson (1995), found no difference between Black and White participants' performance in the control condition. Croizet and Claire (1998) found no performance difference between low and high SES participants in the control condition. Good, Aronson, and Harder (2008) found no math difference between men and women in the control condition.

Several specific sources of 'noise' may account for the difference in results within control conditions of Studies 2 and 3. First, the scholastic tests in Study 3 involved more closed-ended questions than those in Study 2 and may have therefore been experienced as more formal and diagnostic of ability. Since the apparent diagnosticity of a test is sometimes employed as a stereotype threat manipulation (e.g., Croizet & Claire, 1998; Steele & Aronson, 1995), it might be that the control group of Study 3 embodied a stronger, latent stereotype threat effect than that of Study 2. Further, the attrition rate of boys was higher in Study 2. More boys dropped out of both conditions than did girls, having been highly distracted and failing to understand instructions. This might have artificially lifted boys' performance, relative to girls, in Study 2, helping render boys' educational disadvantage undetectable in the control condition. More generally, other as-yet-unidentified factors associated with schools, teachers, and cohorts may affect the extent to which boys lag, or are perceived to lag, behind. It is important for future research to explore such factors.

General Discussion

In the present studies, we systematically examined at what age children develop stereotypes surrounding boys' underachievement at school, how these stereotypes can impair boys' test performance through stereotype threat, and how to counteract them. The results provide initial evidence that stereotype threat contributes to boys' relative academic underachievement. Study 1 established that by seven to eight years old, children of both genders endorse stereotypes and meta-stereotypes regarding boys' relatively poorer school

conduct and achievement. Study 2 revealed how a direct message conveying these meta-stereotypes can become self-fulfilling, causing boys to perform significantly less well. Study 3 showed that counteracting adverse meta-stereotypes can improve boys' academic performance without negatively affecting girls'.

Our findings provide an important theoretical and empirical contribution to three fields of study – stereotype threat, gender stereotyping, and boys' underachievement. First, previous research has examined stereotype threat effects for adults in numerous groups and domains (Nguyen & Ryan, 2008), and researchers are beginning to apply stereotype threat to groups of underachieving children (Ambady et al., 2001; Desert et al., 2009; Martinot & Desert, 2007). Our paper extends this field of research by applying stereotype threat to explain boys' academic underachievement. Study 1 showed that from a young age, children have become acquainted with a cultural stereotype that boys are academically inferior to girls. Study 2 showed that an explicit reminder of this stereotype caused boys to perform less well, and Study 3 showed that an explicit nullification of the stereotype increased their performance. Together, these studies provide the first evidence as to when gendered stereotypes about male academic inferiority develop, how they may become self-fulfilling, and how they may be made the focus of remedial interventions.

Second, the present findings have important theoretical implications regarding the nature and content of gender stereotypes - specifically about male academic inferiority. Importantly, research to date reveals that gender stereotypes tend to portray women and girls as expressive and dependent, and men and boys as instrumental and agentic (e.g., Rudman & Glick, 1999). Further, in previous research, children's beliefs appeared in line with these stereotypes of male agency and female communality (e.g., Egan & Perry, 2001). Although there are negative stereotypes about men and boys, they tend to reflect traits such as increased aggression and low social skills, but never suggest deficits in academic competence. Our

research suggests that gender stereotypes extend to portray boys as academically inferior, not only in terms of behavior and self-regulation, but also crucially in terms of performance and ability. We find that these stereotypes of male academic inferiority develop in early childhood for girls, and middle childhood for boys, despite the fact that boys are members of a non-stigmatized, high status group, which remains substantially advantaged in society.

Third, our findings extend current educational research and theory regarding the potential factors that contribute to boys' educational underachievement. Previous explanations have centered on alleged biological differences, different learning styles, teacher expectations, a lack of male role models, and the feminization of the classroom (Gurian & Stevens, 2004; Mullola et al., 2011; Skelton, 2002). Our research advances understanding of the gender gap by pointing to the role of stereotype threat. Importantly, our findings have potential value in informing educational practice. Specifically, findings from Study 3 suggest that boys' performance can be improved by counteracting prior gender stereotypes and reinforcing the expectation that boys and girls perform equally well. Other interventions that similarly diffuse negative academic stereotypes could improve boys' performance.

Schools may be the most practical and effective venue to counteract gender bias and negative academic stereotypes. However, although schools aim to create a climate of respect and fairness, this may not be sufficient to counteract children's performance threatening gender stereotypes. Institutions and individuals can choose to routinely label and use some particular category within a child's environment or not. For example it is not socially acceptable to divide the class by race, yet it is by gender. Because gender bias is represented as more socially and normatively acceptable throughout society, it is widely acceptable to pitch the boys against the girls or 'harmlessly' divide the class in this way for practical ease (Thorne, 1986). Research has shown that frequent functional use of gender categorization in the classroom in this way, increases gender stereotyping (Bigler & Liben, 2007). Avoiding

these commonplace teaching strategies may reduce the salience of performance based gender stereotypes and their potential negative effects.

Mixed ability tables and classes may also help. At school pupils are often seated around tables on the basis of ability. Boys tend to be overrepresented in low ability groups (Kutnick et al., 2002). Research suggests that children are aware of the ability grouping arrangements in their school (Hallam, Ireson, & Davies, 2004). Such practice is likely to reinforce the associations children make between gender and achievement and increase the potential for stereotype threat. Moreover, assigning children to lower and higher sets during the streaming process at secondary (and in some cases primary) school could have similar consequences. Indeed, researchers from the National Foundation for Educational Research (NFER) suggest that grouping children by ability can reinforce social divisions and have a negative effect on the attitudes, motivation, and self-esteem of lower-ability groups (Sukhnandan & Lee, 1998). Their analysis of 20 research studies in the UK and USA suggests that boys, pupils from working-class families, ethnic minorities, and summer-born children are more likely to be disadvantaged by this practice.

Because the present research has important implications for education, it is particularly important to note its methodological limitations. First, the findings of each study are based on a relatively homogenous sample with respect to age, nationality, ethnicity and socio-economic status (SES). Further work is needed to establish whether adolescent boys are susceptible to stereotype threat. Future work should also examine how ethnicity, SES, and gender may interact to influence children's susceptibility to stereotypes portraying boys as academically inferior. Such factors have been shown to moderate stereotype threat affecting other target groups, such as women and math performance (e.g., see Gonzales, Blanton & Williams, 2002). Second, further work could develop more sophisticated, multidimensional scales to assess children's academic stereotypes. The present scales (Study

1) were based on methods adapted successfully to study gender and ethnic stereotypes. However they do lend themselves to factor analytic exploration of the perceived components of academic success (e.g., ability, performance, motivation, self-regulation, and conduct).

Third, the stereotype threat manipulation instructions employed in Study 2 were explicit and directional (children were told that boys are expected to perform worse on the task than girls). It was important for our purposes that the stereotype threat and nullification manipulations reflected, in a direct and unambiguous way, the message that Study 1 suggested children receive from a young age. That is, the expectation that boys are not expected to do as well as girls. It is likely that such messages represent an important source of children's developing stereotype knowledge. However, as a consequence of such a direct manipulation of stereotype threat, which creates stereotype awareness, prior stereotype awareness was not necessary at the individual level for stereotype threat effects to occur in Study 2. Despite this, children still required a somewhat complex level of social understanding and meta-knowledge in order to connect themselves to the gender category and understand that they are personally being judged accordingly.

The question now remains whether more subtle manipulations, such as merely making gender salient, for instance by using it as a functional category (see Bigler & Liben, 2007), affect boys' performance. It is likely that stereotype threat effects depend on individual awareness of cultural stereotypes, such as those in Study 1, more so when they are induced more subtly. To examine whether stereotype awareness acts as a moderator, particularly under more subtle or non-directional stereotype threat manipulations, ideally children would participate initially to assess their stereotypes and meta-stereotypes. Some weeks later, the same children could take part in a stereotype threat or nullification experiment.

Future Directions

The findings of the three studies reported here raise some larger theoretical questions. First, do children develop stereotypes of girls' academic superiority cross-culturally? International statistics show that boys are underachieving academically throughout many countries. The present research established that a sample of children in UK schools had developed gender stereotypes about conduct and achievement and were vulnerable to stereotype threat effects. Further work is required to replicate these findings across cultures. Second, how do children develop these stereotypes? The present findings document (meta)stereotypes that are deleterious to boys but do not establish where they come from. Future research should examine the impact of various implicit and explicit sources of children's academic gender stereotypes. For example, previous research suggests that stereotypes are created and maintained via parental socialization (e.g., Eccles, Jacobs, & Harold, 1990) and peer group processes (e.g., Abrams, Rutland, Cameron, & Ferrell, 2007). Other research has found that television can transmit traditional gender stereotypes portraying women and girls as passive, dependent, and emotional (Witt, 2000). Future research should examine how children's television may convey stereotypes about boys' inferior academic traits. For example, cartoons such as *The Simpsons* and *Dennis the Menace* would appear to reinforce the stereotype that boys are naughty and unscholastic. Moreover, as mentioned earlier, classroom practices which increase the salience and use of gender and ability grouping may increase children's academic gender stereotyping (Bigler & Liben, 2007).

Further, although the present studies found that boys as a group suffered the negative effects of stereotype threat, they do not reveal *which* boys are more likely to be affected by threat. Are all boys, or only a subset, affected by stereotypes surrounding academic underachievement? Among adults, identification with the domain under threat moderates stereotype threat effects (Aronson et al., 1999; Nguyen & Ryan, 2008). Therefore, boys'

susceptibility to stereotype threat may be moderated by the extent to which they identify with, and value doing well academically (and factors which are likely to influence identification with school such as prior academic performance and perceived ability). For some boys, the ongoing experience of stereotype threat may lead to disengagement and disidentification with the school (cf. Steele & Aronson, 1995).

Concluding remarks

The present findings need to be considered in the context of pervasive male advantage in society, for example the gender wage gap favoring men (AAUWEF, 2007). It is tempting to see boys' educational underachievement as one mechanism by which the scales may ultimately be balanced. However, our findings do not suggest that girls are aided by academic stereotypes, but rather that boys are hindered by them. Thus, if boys' educational underachievement does turn out to aid gender inequality, it does so by preventing children from reaching their academic potential. Further, given the evidence of inequitable treatment of women in the wage gap literature, there is also no guarantee that girls' relative advantage will translate into economic or career success.

Another consideration is that boys and girls are not locked in a zero-sum game. Research into the economic benefits of educational achievement has demonstrated that there are significant economic gains to be made if countries improve the cognitive skills of their citizens, as measured through educational outcomes (OECD, 2010b). Moreover, school attainment is positively associated with higher levels of employment, labor force-re-entry, higher wages, higher levels of productivity, and health outcomes (Council of Australian Governments, 2006). Therefore, boys' academic underachievement is likely to be deleterious to economic productivity, with negative effects for both genders.

Other research indicates that low school achievement is an important predictor of delinquency and anti-social behaviors (e.g., Lochner & Moretti, 2004). Boys'

underachievement is therefore likely to contribute to social problems, including aggression, some of which will be directed at women and girls. Indeed, if males continue to underachieve educationally in a society in which males are still advantaged and expected to be breadwinners (Hausmann et al., 2010), then there is a danger of continuing backlash against women, girls and feminism (e.g., Rudman & Glick, 1999). Indeed, "threatened masculinity" increases men's physical and social aggression as they attempt to redress the traditional social hierarchy (Vandello, Bosson, Cohen, Burnaford, & Weaver, 2008).

Thus, we suggest that boys' academic underachievement should not be seen as a compensation or correction for wider gender inequalities in society. The present research suggests that gender stereotypes have the capacity to prevent children from achieving their potential. These historically have worked against the interests of women, and are shown by the present studies to work against those of boys in the scholastic domain. It is important to note that the gender stereotypes documented in this article (Study 1) are unlikely to *produce* the gender difference in educational outcomes, which is apparent at a very young age. However, they are capable of entrenching and magnifying these gender differences (Study 2). As well as providing a partial diagnosis of the cause of boys' academic underachievement, the present studies point the way to a cure (Study 3). Specifically, the cause of educational equality may be advanced by communicating egalitarian educational expectations.

References

Abrams, D., Eller, A., & Bryant, J. (2006). An age apart: The effects of intergenerational contact and stereotype threat on performance and intergroup bias. *Psychology and Aging, 21,* 691-702. doi: 10.1037/0882-7974.21.4.691

Abrams, D., Rutland, A., Cameron, L., & Ferrell, J. (2007). Older but wilier: In-group accountability and the development of subjective group dynamics. *Developmental Psychology, 43,* 134-148. doi:10.1037/0012-1649.43.1.134

Aboud, F.E. (1988). *Children and prejudice*. Oxford, UK: Blackwell.

Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology, 46,* 166-171. doi: 10.1016/j.jesp.2009.09.014

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385-390. doi: 10.1111/1467-9280.00371

American Association of University Women Educational Foundation. (1998). *Gender gaps: Where schools still fail our children.* Washington, DC: Author.

American Association of University Women Educational Foundation. (2007). *Behind the pay gap.* Washington, DC: Author.

Arnot, M., & Weiler, K. (Eds.). (1993). *Feminism and social justice in education: International perspectives*. London: The Falmer Press.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29-46. doi:10.1006/jesp.1998.1371

Augoustinos, M., & Rosewarne, D. L. (2001). Stereotype knowledge and prejudice in children. *British Journal of Developmental Psychology, 19*, 143-156. doi: 10.1348/026151001165912

Baron, A., & Banaji, M. (2009). Evidence of system justification in young children. *Social and Personality Psychology Compass, 3*, 1-9. doi: 10.1111/j.1751-9004.2009.00214.x

Best, D. L., Williams, J. E., Cloud, J. M., Davis, S. W., Robertsson, L. S., Edwards, J. R., Giles, H., & Fowles, J. (1977). Development of sex-trait stereotypes among young children in the United States, England, and Ireland. *Child Development, 48*, 1375-1384. doi:10.2307/1128496

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science, 16*, 162-166. doi: 10.1111/j.1467-8721.2007.00496.x

Bond, J. M. (2007a). *Assessment papers: Starter papers in verbal reasoning*. Cheltenham, UK: Nelson Thornes.

Bond, J. M. (2007b).*Verbal reasoning assessment papers*. Cheltenham, UK: Nelson Thornes.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist, 22*, 817-829. doi:10.1037/h0025079

Cheryan, S. & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: Psychological hazards of "Model Minority" status. *Psychological Science, 11*, 399-402. doi:10.1111/1467-9280.00277

Cole, N. S. (1997). *The ETS gender study: How males and females perform in educational settings.* Princeton, NJ: Educational Testing Service.

Council of Australian Governments. (2006). *Human capital reform: Report by the COAG National Reform Initiative Working Group*. Retrieved from:

http://www.coag.gov.au/coag_meeting_outcomes/2006-02-10/docs/human_capital_reform_report_COAG_100206.pdf

Croizet, J., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588–594. doi: 10.1177/0146167298246003

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

Desert, M., Preaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education, 24*, 207-218. doi: 10.1007/BF03173012

DfE. (2010). *Key Stage 1 attainment by pupil characteristics in England, 2009/2010*. Retrieved from: http://www.education.gov.uk/rsgateway/DB/SFR/s000968/sfr33-2010.pdf

DfE. (2011a). *Permanent and fixed period exclusions from schools and exclusion appeals in England, 2009/2010*. Retrieved from: http://www.education.gov.uk/rsgateway/DB/SFR/s001016/sfr17-2011.pdf

DfE. (2011b). Early years foundation stage profile results in England, 2010/2011. Retrieved from: http://www.education.gov.uk/rsgateway/DB/SFR/s001033/sfr28-2011v2.pdf

DfE. (2012). *GCSE and equivalent attainment by pupil characteristics in England, 2010/2011*. Retrieved from: http://www.education.gov.uk/rsgateway/DB/SFR/s001057/sfr03-2012.pdf

DfES. (2007). *Gender and education: The evidence on pupils in England*. Retrieved from: https://www.education.gov.uk/publications/eOrderingDownload/00389-2007BKT-EN.pdf

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 42*, 183-201. doi: 10.1111/j.1540-4560.1990.tb01929.x

Egan., S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychological adjustment. *Developmental Psychology, 37*, 451-463. doi:10.1037//0012-1649.37.4.451

Equality and Human Rights Commission. (2010). *How fair is Britain? Equality, human rights and good relations in 2010. The first triennial review* (p. 664). Retrieved from: http://www.equalityhumanrights.com/uploaded_files/triennial_review/how_fair_is_britain_-_complete_report.pdf

EACEA/Eurydice. (2010). *Gender differences in educational outcomes: Study on the measures taken and the current situation in Europe*. Brussels: EACEA P9 Eurydice. doi: 10.2797/3598

Foster, V., Kimmel, M., & Skelton, C. (2001). What about the boys? An Overview of the debates. In W. Martino & B. Meyenn (Eds.), *What about the boys? Issues of masculinity in schools* (pp. 1-23). Buckingham: Open University Press.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority statues on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659-670. doi:10.1177/0146167202288010

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28. doi:10.1016/j.appdev.2007.10.004

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*, 645-662. doi: 10.1016/j.appdev.2003.09.002

Gurian, M., & Stevens, K. (2004). With boys and girls in mind. *Educational Leadership, 62*, 21-27.

Haahr, J. H., Nielsen, T. K., Hansen, M. E., & Jakobsen, S. T. (2005). *Explaining student performance: Evidence from the international PISA, TIMSS and PIRLS surveys.* Danish Technological Institute.

Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experiences of different types of grouping in school. *British Educational Research Journal, 30*, 515-534. doi: 10.1080/0141192042000237211

Hausmann, R., Tyson, L. D., & Zahidi, S. (2010). *The global gender gap report 2010*. Retrieved from World Economic Forum website: http://www3.weforum.org/docs/WEF_GenderGap_Report_2010.pdf

Higher Education Statistics Agency. (2011). *Students in higher education institutions*. Retrieved from: http://www.hesa.ac.uk/index.php/content/view/1974/278/

Jones, S., & Myhill, D. (2004). 'Troublesome boys' and 'compliant girls': Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education, 25*, 547-561. doi: 10.1080/0142569042000252044

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*, 131-155. doi: 10.1207/s15327957pspr0902_3

Kutnick, P., Blatchford, P., & Baines, E. (2002). Pupil groupings in primary school classrooms: sites for learning and social pedagogy? *British Educational Research Journal, 28*, 187-206. doi: 10.1080/0141192012012214 9

Leyens, J. P., Desert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26,* 1189–1199. doi: 10.1177/0146167200262002

Lochner, L., & Moretti, E. (2004). The effect of education on criminal activity: Evidence from prison inmates, arrests and self-reports. *American Economic Review, 94*, 155-89. doi:10.1257/000282804322970751

Martinot, D., & Desert, M. (2007). Awareness of a gender stereotype, personal beliefs and self-perceptions regarding math ability: when boys do not surpass girls. *Social Psychology of Education, 10*, 455-471. doi: 10.1007/s11218-007-9028-9

McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social academic consequences of stereotype-consciousness in middle childhood. *Child Development, 80*, 1643-1659. doi:10.1111/j.1467-8624.2009.01359.x

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen,M., Jokela, M., & Keltikangas-Jarvinin, L. (2011). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *British Journal of Educational Psychology*, no. 1-22. doi:10.1111/j.2044-8279.2010.02017.x

Neff, K. D., Cooper, C. E., & Woodruff, A. L. (2007). Children's and adolescents' developing perceptions of gender inequality. *Social Development, 16*, 682–699. doi: 10.1111/j.1467-9507.2007.00411.x

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314-1334. doi: 10.1037/a0012702.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala., ...Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex

differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*, 10593-10597. doi:10.1073/pnas.0809921106

OECD (2010a). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (Volume I)*. Retrieved from: http://www.oecd.org/dataoecd/10/61/48852548.pdf

OECD. (2010b*). The high cost of low educational performance: The long run economic impact of improving PISA outcomes*. Retrieved from: http://www.oecd.org/dataoecd/11/28/44417824.pdf

Ogbu, J. (1986). The consequences of the American caste system. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (pp. 19-56). Hillsdale, NJ: Erlbaum.

Qualifications and Curriculum Authority. (2001). *Key Stage 1 SATS*. Retrieved from: http://www.hamilton-school.co.uk/sats/ks1_sats.htm

Qualifications and Curriculum Authority. (2004a). *Key Stage 1 mathematics booklet 2004 (Level 2)*. Retrieved from: http://www.emaths.co.uk/KS1SATS/2004A.pdf

Qualifications and Curriculum Authority. (2004b). *Key Stage 1 mathematics booklet 2004 (Level 3)*. Retrieved from: http://www.emaths.co.uk/KS1SATS/2004B.pdf

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55–71. doi: 10.1111/0022-4537.00201

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology, 77*, 1004-1010. doi: 10.1037/0022-3514.77.5.1004

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development, 76*, 451-466. doi:10.1111/j.1467-8624.2005.00856.x

Skelton, C. (2002). The 'feminisation of schooling' or 're-masculinising primary education'? *International Studies in Sociology of Education, 12*, 77-96. doi:10.1080/09620210200200084

Snyder, T. D., & Dillow, S. A. (2011). *Digest of education statistics 2010*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28. doi:10.1006/jesp.1998.1373

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*, 797–811. doi: 10.1037/0022-3514.69.5.797

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629. doi: 10.1037/0003-066X.52.6.613

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology, 77*, 1213-1227. doi:10.1037//0022-3514.77.6.1213

Sukhnandan, L., & Lee, B. (1998). *Streaming, setting and grouping by ability*. Slough, UK: National Foundation for Educational Research.

Sutton, R. M., Douglas, K. M., Wilkin, K., Elder, T. J., Cole, J. M., & Stathi, S. (2008). Justice for whom, exactly? Beliefs in justice for the self and various others. *Personality and Social Psychology Bulletin, 34*, 528-541. doi:10.1177/0146167207312526

Thorne, B. (1986). Girls and boys together...but mostly apart: Gender arrangements in elementary schools. In W. Hartup & Z. Rubin (Eds.), *Relationships and Development* (pp. 167-184). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vandello, J. A., Bosson, J. K., Cohen, D., Burnaford, R. M., & Weaver, J. R. (2008). Precarious manhood. *Journal of Personality and Social Psychology, 95*, 1325-1339. doi: 10.1037/a0012453

Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology, 78*, 690-707. doi:10.1037//0022-3514.78.4.690

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467. doi:10.1016/S0022-1031(03)00019-2

Weaver-Hightower, M. (2003). The "boy turn" in research on gender and education. *Review of Educational Research, 73*, 471–498. doi:10.3102/00346543073004471

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684. doi:10.1111/1467-8624.00304

Williams, J. E., Bennett, S. M., & Best, D. L. (1975). Awareness and expression of sex stereotypes in young children. *Developmental Psychology, 11*, 635-642. doi:10.1037//0012-1649.11.5.635

Witt, S. D. (2000). The influence of television on children's gender role socialization. *Childhood Education, 76*, 322-324.

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33*, 183–196. doi:10.1111/j.2044-8309.1994.tb01017.x

Table 1

Mean (and Standard Deviation) Stereotypical and Meta-Stereotypical Beliefs About School Achievement Across Gender and Year Group, and Regression Coefficients of the Relation Between Stereotypes and Meta-Stereotypes in Study 1

		M (*SD*)		*N*	*β*
		Stereotype	Meta-stereotype		
Year F	Boys	.47 (.15)	.27 (.27)**	21	.78***
	Girls	.77 (.21)***	.75 (.27)***	23	
Year 1	Boys	.54 (.27)	.50 (.29)	19	.72***
	Girls	.78 (.21)***	.77 (.32)**	23	
Year 2	Boys	.59 (.29)	.59 (.31)	21	.45**
	Girls	.83 (.19)***	.83 (.21)***	25	
Year 3	Boys	.81 (.18)***	.69 (.21)***	21	.41**
	Girls	.87 (.16)***	.84 (.17)***	21	
Year 5	Boys	.82 (.12)***	.77 (.23)***	21	.12
	Girls	.86 (.13)***	.79 (.14)***	21	

Note. Mean values represent the proportion of stereotypical and meta-stereotypical responses (choosing girls as better than boys).

p values refer to differences from mid-point (.50) where values greater than .50 indicate the belief that girls are academically superior.

** $p < .01$. *** $p < .001$

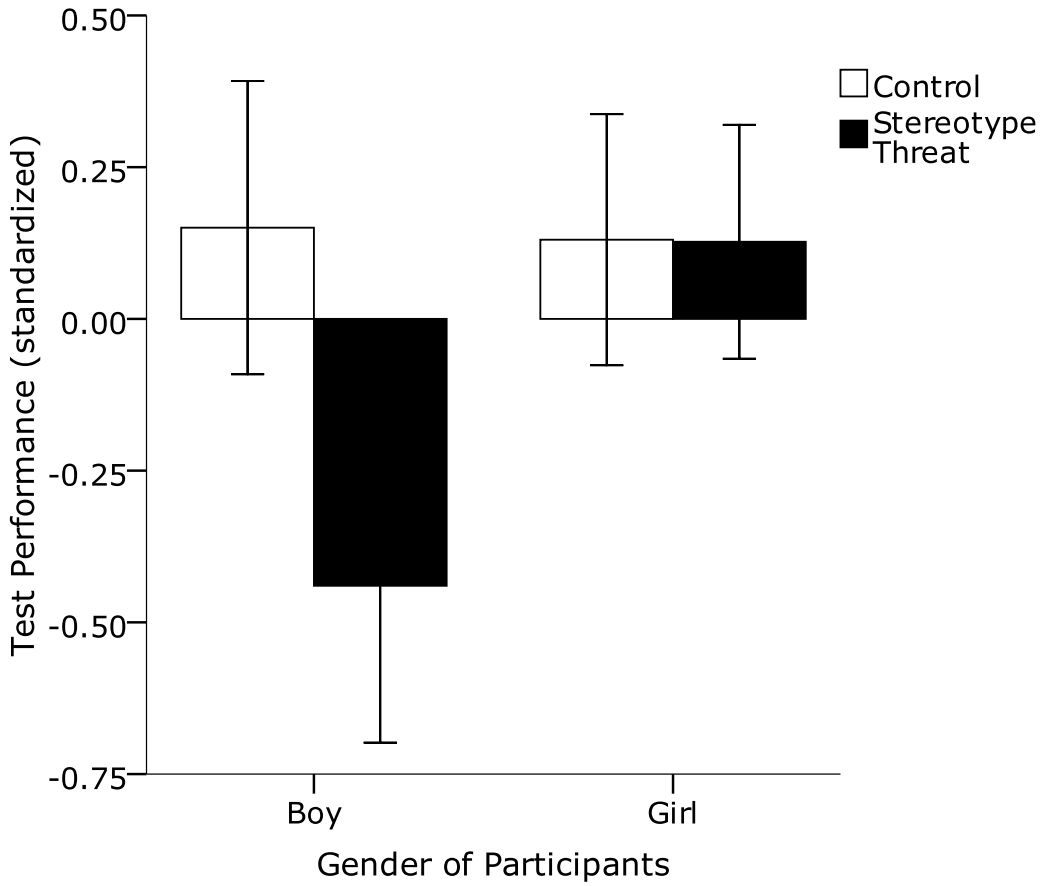

Figure 1. Girls' and Boys' Overall Test Performance as a Function of Stereotype Threat Condition in Study 2.

Note. Error bars represent 95% confidence intervals.