



Creating a critical mass eliminates the effects of stereotype threat on women's mathematical performance

Charlotte R. Pennington* and Derek Heim

Department of Psychology, Edge Hill University, Ormskirk, Lancashire, UK

Background. Women in mathematical domains may become attuned to situational cues that signal a discredited social identity, contributing to their lower achievement and underrepresentation.

Aim. This study examined whether heightened in-group representation alleviates the effects of stereotype threat on women's mathematical performance. It further investigated whether single-sex testing environments and stereotype threat influenced participants to believe that their ability was fixed (fixed mindset) rather than a trait that could be developed (growth mindset).

Sample and method. One hundred and forty-four female participants were assigned randomly to a self-as-target or group-as-target stereotype threat condition or to a control condition. They completed a modular arithmetic maths test and a mindset questionnaire either alone or in same-sex groups of 3–5 individuals.

Results. Participants solved fewer mathematical problems under self-as-target and group-as-target stereotype threat when they were tested alone, but these performance deficits were eliminated when they were tested in single-sex groups. Participants reported a weaker growth mindset when they were tested under stereotype threat and in single-sex groups. Moreover, evidence of inconsistent mediation indicated that single-sex testing environments negatively predicted mindset but positively predicted mathematical performance.

Conclusions. These findings suggest that single-sex testing environments may represent a practical intervention to alleviate stereotype threat effects but may have a paradoxical effect on mindset.

Research on stereotype threat (Steele & Aronson, 1995) indicates that women underperform relative to men when they apprehend that their mathematical performance will be evaluated in line with gender-related expectations (c.f., Spencer, Steele, & Quinn, 1999; Steele, 1997). These effects appear to be robust (Nguyen & Ryan, 2008) and extend beyond the laboratory (Good, Aronson, & Harder, 2008; Hollis-Sawyer & Sawyer, 2008; Keller, 2007). As such, researchers have turned their attention to examining the moderating factors that might heighten women's susceptibility to stereotype threat. It has been proposed that seemingly benign and subtle factors, such as the gender composition of a classroom, may undermine women's mathematical performance and further

*Correspondence should be addressed to Charlotte R. Pennington, Department of Psychology, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, UK (email: penninc@edgehill.ac.uk or charlotte.pennington@outlook.com).

contribute to their underrepresentation in this domain (Bigler & Liben, 2006, 2007; Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003).

In a direct test of this notion, Inzlicht and Ben-Zeev (2000) found that women underperformed on a mathematical test when men outnumbered them. However, these performance deficits were not observed when women completed the test in same-sex groups. Moreover, women's mathematical performance was found to decrease in proportion to the number of men in the testing environment. Extending this, Sekaquaptewa and Thompson (2003) examined the dual influence of solo status and stereotype threat on women's mathematical performance. Findings indicated that women underperformed to a greater extent when they completed the test in opposite sex (solo status) relative to same-sex groups. An interaction between solo status and stereotype threat revealed further that experiencing both of these factors simultaneously was more detrimental to performance than experiencing one of these factors alone. These findings support a wealth of research which suggests that the numerical representation of minority group members may interact with their stereotyped status to determine whether an environment will promote or attenuate academic learning, engagement, and performance (Inzlicht & Ben-Zeev, 2000, 2003; Murphy, Steele, & Gross, 2007; Sekaquaptewa & Thompson, 2003).

Research has also examined the extent to which stereotype threat effects are mitigated when women work collaboratively to solve mathematical problems. For example, Aramovich (2014) found that women were buffered from the performance-impinging effects of stereotype threat when they were tested in same-sex groups, relative to alone, because they were able to work together to detect errors. Nevertheless, the practical implications of this study may be limited because in real-life testing environments women are typically required to undertake quantitative tests independently as a measure of their personal ability. Overcoming this issue, Huguet and Régner (2007; Experiment 2) revealed that stereotype-threatened females underperformed when they worked alone or in mixed-sex classrooms on a task that ostensibly measured mathematical skills. However, these performance deficits were eliminated when females worked in single-sex groups. These findings suggest that the mere presence of other in-group members (i.e., females) can promote women's mathematical performance when they are assessed individually.

Previous work has focused largely on the potential efficacy of single-sex testing environments as a practical means to bolster women's performance against stereotype threat. Less work has examined the impact that gender-segregated classrooms may exert on attitudinal outcomes. Based on a rationale garnered from same-sex schooling (c.f., Halpern *et al.*, 2011; Pahlke, Shibley-Hyde, & Allison, 2014), the current research investigates the notion that gender-segregated environments may influence a fixed-ability mindset (Dweck, 2006, 2008). When placed in same-sex classrooms, females may question why they have been separated from their male peers and attribute this to inherent sex differences (Halpern *et al.*, 2011; Pahlke *et al.*, 2014). Such environmental cues may relay a message to women that their ability to succeed in mathematics is limited by group membership, namely their gender (Dweck, 2006, 2008; Good *et al.*, 2008). Indeed, this is an important consideration in view of research indicating that a fixed-ability mindset may have a deleterious and long-term effect on educational outcomes (Verniers & Martinot, 2015; c.f., also Martin, 2015). Although single-sex classroom initiatives may have a positive effect by alleviating women's apprehensions about confirming gender-related stereotypes in the eyes of outgroup members (Picho & Stephens, 2012; Titze, Jansen, & Heil, 2011), they may also have a paradoxical effect on mindset.

Presenting as a further issue, the majority of previous research has conceptualized stereotype threat as a singular construct (e.g., Aramovich, 2014; Huguet & Régner, 2007;

Inzlicht & Ben-Zeev, 2000). However, this overlooks the important distinction between an individual's personal and social identity (Tajfel & Turner, 1979, 1986) and, resultantly, the different impact that self- and group-relevant stereotypes may exert on performance. The multithreat framework (Shapiro & Neuberg, 2007) posits that women may experience self-as-target stereotype threat when they endorse negative gender-related stereotypes as a true representation of their personal ability. Conversely, women may experience group-as-target stereotype threat when they perceive that they will reinforce the negative reputation that their group lacks a valued ability. The existence of multiple stereotype threats is therefore a noteworthy consideration, particularly when examining the efficacy of group-based interventions to ameliorate performance deficits.

The first aim of this study was to examine whether the mere presence of other females could ameliorate the effects of self- and group-relevant stereotypes on women's mathematical performance. It was predicted that female participants would solve fewer mathematical problems under self-as-target and group-as-target stereotype threat when they were tested alone relative to those in a control condition. In this situation, women may apprehend that they are single representatives of their social group, which may exacerbate situational performance pressure (c.f., Baumeister, 1984; Huguet & Régner, 2007; Steele, 1997). It was further predicted that these performance decrements would be alleviated when females were tested in single-sex groups. At first blush, it may seem that women should be susceptible to group-as-target stereotype threat in single-sex groups because this concerns their devalued group membership in the stereotyped domain. However, in line with previous research (Inzlicht & Ben-Zeev, 2000; Murphy *et al.*, 2007; Sekaquaptewa & Thompson, 2003), the numerical representation of other females within the mathematics classroom should lessen concerns about representing positively the in-group to bolster performance. Furthermore, when tested in single-sex groups, women may be less susceptible to self-as-target stereotype threat because they strive to disconfirm the negative group stereotype as being a true representation of their personal ability.

The second aim of this study was to examine the effects of stereotype threat and group composition on mindset. Underpinned by research on single-sex schooling (c.f., Halpern *et al.*, 2011; Pahlke *et al.*, 2014), it was predicted that female participants would become more cognisant of the differences between women and men when they were tested in single-sex groups relative to alone. Under such conditions, it was predicted that they would attribute their mathematical ability to internal, fixed factors (i.e., fixed mindset) rather than a trait that could be shaped and developed (i.e., growth mindset). Moreover, given that stereotypes are fixed mindset labels (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2006, 2008), it was also predicted that females who were primed explicitly with information regarding gender differences in mathematical performance would report a weaker growth mindset compared with those in the control condition.

Method

Participants and design

One hundred and forty-four females ($M_{\text{age}} = 21.60$, $SD = 4.67$, 88.9% White British, 83.3% university students) signed up via an online participation website and arranged a time to come into the laboratory. They received £3 remuneration for their participation. In a between-participants design, they were allocated randomly to one of three experimental conditions: (1) self-as-target stereotype threat, (2) group-as-target stereotype threat, and (3) a control condition. To examine the effects of in-group representation on performance, half of the participants in each experimental condition completed the

study alone, whereas the other half was tested in groups of 3–5. The study consisted of a 3 (condition: self-as-target, group-as-target, control) \times 2 (group composition: alone, group) between-participants design, with 24 participants assigned to each condition.

Measures

Stereotype threat manipulations

We employed two distinct stereotype threat manipulations, which took the form of self-as-target or group-as-target primes (Shapiro & Neuberg, 2007). The self-as-target manipulation was drawn from previous research and influenced participants to perceive that their mathematical performance would be self-characteristic of personal ability (e.g., Steele & Aronson, 1995). Specifically, participants in this condition were presented with the following information:

There is a negative stereotype that females have less mathematical aptitude comparative to males. You are a female and this maths exam is therefore diagnostic of your **personal** mathematical ability.

Participants assigned to the group-as-target stereotype threat condition were primed that their mathematical performance would be diagnostic of gender-related ability (e.g., Aronson *et al.*, 1999). Specifically, they received the following information:

There is a negative stereotype that females have less mathematical aptitude comparative to males. This maths exam is therefore diagnostic of **females'** mathematical ability.

Both stereotype threat primes included reference to the negative gender-maths stereotype in line with research suggesting that this awareness is required for participants to be susceptible to stereotype threat (Shapiro & Neuberg, 2007). Participants in the control condition were informed that the experiment was investigating factors involved in working memory and that the mathematical test was non-diagnostic of ability (c.f., Steele & Davies, 2003).

Mathematical performance

In accordance with current research (Beilock & Carr, 2005; Beilock, Rydell, & McConnell, 2007; Rydell, Van Loo, & Boucher, 2014), we utilized modular arithmetic as a measure of mathematical performance. This computerized task required participants to judge the validity of problems such as ' $43 = 16 \pmod{3}$ ' by subtracting the middle number from the first number (e.g., $43 - 16$) and then dividing it by the number in parentheses (e.g., $27/3$). Participants were required to respond 'true' when the dividend resulted in a whole number, and 'false' when the dividend resulted in a decimal number. Problem difficulty was manipulated by function of operation and presentational format (Lee & Kang, 2002; Trbovich & LeFevre, 2003). For example, problems including larger numbers and borrow operations are more difficult to solve as they involve a longer sequence of steps and require maintenance of more intermediate products in working memory (Lee & Kang, 2002). Moreover, horizontally presented problems are suggested to be more difficult as they appear in a different format to how individuals typically solve problems (Trbovich & LeFevre, 2003). Accordingly, participants completed a total of 48 mathematical problems (16 simple, 16 moderate, 16 difficult) that were presented in a random order and remained

on the computer screen until a response had been recorded. Half of the problems were presented horizontally, and half were presented vertically. Accuracy scores were calculated by dividing the number of problems answered correctly by the total number of problems, with greater scores indicating greater accuracy (Beilock *et al.*, 2007; Rydell, Rydell, & Boucher, 2010).

Mindset

Participants' mindset was measured using a 20-item self-report questionnaire (McKenzie, 2013; adapted from Dweck, 2006). This questionnaire was modified to ensure that all questions were related to mathematical ability, rather than general intelligence. Participants responded on a 4-point Likert scale anchored between 'strongly agree' and 'strongly disagree'. Questions related to a growth mindset included 'Mathematical talent can be learned by anyone', and questions related to a fixed mindset included 'Maths is much easier to learn if you are male'. Scores were totalled out of 60, with higher scores indicative of a growth-ability mindset. The questionnaire resulted in high internal consistency in the current study, Cronbach's $\alpha = .81$.

Procedure

After being assigned randomly to one of three experimental conditions, participants completed two self-report questions: 'I am good at maths' and 'It is important to me that I am good at maths'. Responses were recorded on a 9-point Likert scale anchored between 1 (*strongly disagree*) and 9 (*strongly agree*). These questions were included to control for any differences in perceived mathematical ability and domain identification as a function of experimental condition (c.f., Keller, 2007; Steele, 1997). Upon implementing the stereotype threat prime, participants completed the mindset questionnaire and the maths test, with the order of these measures counterbalanced. Participants were then introduced to the maths test with written instructions presented on a computer. They were instructed to judge the validity of each maths problem, indicating whether the answer was true (i.e., a whole number) or false (i.e., a decimal number) using the 'Z' and 'M' buttons on a standard keyboard, respectively. Participants completed the maths test on individual computers, which had screens on either side to ensure that participants could not observe others' answers. Upon completion of the study, participants received a verbal and written debrief.

Results

Perceived mathematical ability and domain identification

A multivariate analysis of variance (MANOVA) indicated that participants in the self-as-target stereotype threat condition reported lower perceived mathematical ability ($M = 5.08$, $SD = 1.75$) compared with the control condition ($M = 6.04$, $SD = 1.49$), $F(2, 141) = 4.03$, $p < .05$, $\eta_p^2 = .05$. Moreover, participants in the self-as-target condition attributed less importance to the domain of mathematics ($M = 5.38$, $SD = 1.91$) compared with the control condition ($M = 6.27$, $SD = 1.50$), $F(2, 138) = 3.53$, $p = .03$, $\eta_p^2 = .05$. Although these responses were above average, participants' self-reported mathematical ability and domain identification were entered as covariates in all analyses to ensure that they did not influence performance. Participants in the group-as-target and control condition did not differ in their reports of mathematical ability ($p = .53$) and

domain identification ($p = 1.00$). Moreover, participants' responses to these two questions did not differ as a function of group composition (group vs. alone), $p = .96$.

Mathematical performance

Modular arithmetic accuracy was examined in a 3 (condition: self-as-target, group-as-target, control) \times 2 (group composition: alone, group) \times 3 (problem difficulty: simple, moderate, difficult) \times 2 (Problem presentation: horizontal, vertical) mixed factorial analysis of covariance (ANCOVA). Experimental condition and group composition were analysed as between-participants factors, and problem difficulty and presentation were input as within-participants factors. Main effects and interactions were elucidated using Bonferroni-corrected pairwise comparisons.

Problem difficulty and presentation

There was a main effect of problem difficulty, $F(2, 272) = 19.84$, $p < .001$, $\eta_p^2 = .13$. Pairwise comparisons revealed that participants solved fewer difficult ($M = 0.78$, $SD = 0.17$) relative to simple problems ($M = 0.93$, $SD = 0.10$), $p < .001$, $d = -1.08$. They also solved fewer moderate ($M = 0.76$, $SD = 0.17$) relative to simple problems ($M = 0.93$, $SD = 0.10$), $p < .001$, $d = -1.22$. There was a significant two-way interaction between problem difficulty and presentation, $F(2, 272) = 3.22$, $p = .04$, $\eta_p^2 = .02$. Participants solved fewer horizontally presented difficult ($M = 0.75$, $SD = 0.19$) and moderate problems ($M = 0.72$, $SD = 0.18$) compared with simple problems ($M = 0.95$, $SD = 0.07$), $p < .001$, $d = -1.40$ and -1.68 , respectively. Participants also solved fewer vertically presented difficult ($M = 0.82$, $SD = 0.19$) and moderate problems ($M = 0.80$, $SD = 0.20$) compared with simple problems ($M = 0.90$, $SD = 0.13$), $p < .001$, $d = -.49$ and $-.59$, respectively.

Stereotype threat

There was a main effect of experimental condition on maths performance, $F(2, 136) = 4.67$, $p = .01$, $\eta_p^2 = .06$. Pairwise comparisons indicated that participants assigned to the self-as-target condition solved significantly fewer problems ($M = 0.79$, $SD = 0.12$) compared with the control condition ($M = 0.86$, $SD = 0.12$), $p < .01$, $d = -.58$. There was no difference in performance between the group-as-target relative to the self-as-target stereotype threat ($p = .61$) and the control condition ($p = .21$). A three-way interaction was obtained between experimental condition, problem difficulty, and presentation, $F(4, 135) = 3.78$, $p < .01$, $\eta_p^2 = .05$. Participants in the self-as-target condition solved more difficult and moderate problems when they were presented horizontally ($M = 0.72$, $SD = 0.19$, $M = 0.64$, $SD = 0.18$) relative to vertically ($M = 0.82$, $SD = 0.19$, $M = 0.78$, $SD = 0.21$), $p < .001$, $d = -.53$ and $-.72$, respectively. They solved fewer simple problems when they were presented vertically ($M = 0.85$, $SD = 0.13$) relative to horizontally ($M = 0.94$, $SD = 0.08$), $p < .001$, $d = -.83$. Participants in the group-as-target stereotype threat condition solved fewer difficult questions when they were presented horizontally ($M = 0.74$, $SD = 0.19$) relative to vertically ($M = 0.80$, $SD = 0.19$) $p = .04$, $d = -.32$. Participants in the control condition solved fewer difficult and moderate problems when they were presented horizontally ($M = 0.79$, $SD = 0.19$, $M = 0.78$, $SD = 0.18$) relative to vertically ($M = 0.84$, $SD = 0.19$, $M = 0.85$, $SD = 0.21$), $p < .03$, $d = -.26$ and $-.36$. They solved fewer simple problems when they were presented vertically ($M = 0.93$, $SD = 0.13$) relative to horizontally ($M = 1.0$,

$SD = 0.08$), $p < .001$, $d = -.74$. Participants under self-as-target condition solved fewer horizontally presented moderate problems ($M = 0.64$, $SD = 0.18$) compared with females in the group-as-target ($M = 0.73$, $SD = 0.18$), $p = .03$, $d = -.50$, and control conditions ($M = 0.78$, $SD = 0.18$), $p = .001$, $d = -.78$. Participants in the self-as-target condition solved fewer horizontally presented simple problems ($M = 0.94$, $SD = 0.08$) compared with females in the control condition ($M = 1.0$, $SD = 0.08$), $p < .001$, $d = -.75$). They also solved fewer vertically presented simple problems ($M = 0.85$, $SD = 0.13$) compared with females in the group-as-target ($M = 0.94$, $SD = 0.13$), $p < .01$, $d = -.69$, and control conditions ($M = 0.93$, $SD = 0.13$), $p = .01$, $d = -.62$. Participants under group-as-target stereotype threat solved fewer horizontally presented simple problems ($M = 0.93$, $SD = 0.08$) compared with the control condition ($M = 1.0$, $SD = 0.08$), $p < .001$, $d = -.87$. Accuracy scores for participants in the group-as-target condition did not differ from the control condition on all other problems, $p > .05$. All other pairwise comparisons were non-significant, $p > .05$. These results suggest that females solved fewer horizontally presented simple problems under group-as-target stereotype threat compared with those in the control condition. Moreover, self-as-target stereotype threat had a greater effect on mathematical performance than group-as-target stereotype threat, with females in this condition underperforming on both horizontally and vertically presented simple problems, and horizontally presented moderate problems. See Figure 1 for interaction.

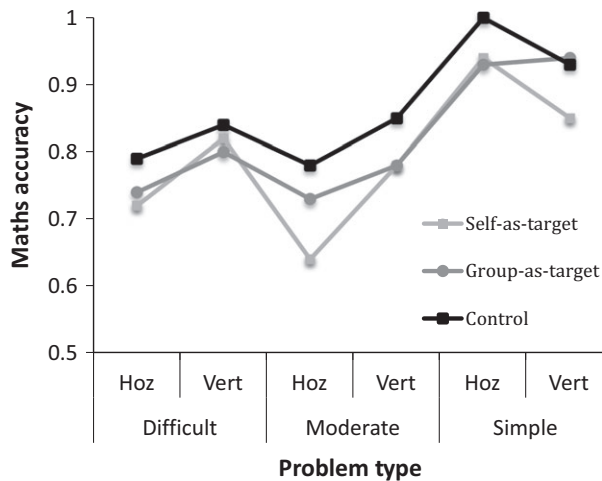


Figure 1. Three-way interaction between experimental condition, problem demand, and presentation.

Group composition

Of central importance to the aim of the current study, there was a main effect of group composition, $F(1, 136) = 3.96$, $p = .049$, $\eta_p^2 = .03$. Pairwise comparisons indicated that participants solved fewer maths problems when they were tested alone ($M = 0.81$, $SD = 0.12$) relative to in single-sex groups ($M = 0.84$, $SD = 0.12$), $p = .049$, $d = -.25$. This was qualified by a significant three-way interaction between experimental condition, group composition, and problem presentation, $F(2, 136) = 3.58$, $p = .03$, $\eta_p^2 = .05$. When tested alone, participants who were primed with a self-as-target stereotype threat solved fewer horizontally presented problems ($M = 0.73$, $SD = 0.11$) relative to participants in the control condition ($M = 0.86$, $SD = 0.11$), $p < .001$, $d = -1.18$. Participants who

were tested alone under group-as-target stereotype solved fewer horizontally presented problems ($M = 0.77$, $SD = 0.11$) compared with the control condition ($M = 0.86$, $SD = 0.11$), $p = .02$, $d = -.82$. Accuracy did not differ for vertically oriented problems, $p > .05$. Importantly, there were no performance decrements as a function of experimental condition when females were tested in groups, $p > .05$. These results suggest that the mere presence of other females bolstered participants' mathematical performance from the effects of self-as-target and group-as-target stereotype threat.

Further confirming this, females primed with a self-as-target stereotype solved fewer horizontally presented problems when they were tested alone ($M = 0.73$, $SD = 0.11$) compared with when they were tested in groups ($M = 0.80$, $SD = 0.11$), $p = .01$, $d = -.64$. They also solved fewer vertically presented problems when tested alone ($M = 0.77$, $SD = 0.14$) relative to in a group ($M = 0.85$, $SD = 0.14$), $p = .046$, $d = -.57$. There was also a trend for participants primed with a group-as-target stereotype threat to underperform on horizontally presented problems when tested alone ($M = 0.77$, $SD = 0.11$) compared with when they were tested in a group ($M = 0.83$, $SD = 0.11$), $p = .058$, $d = -.55$. Females performance in the control condition did not differ as a function of group composition, $p > .05$. When tested alone, females assigned to the self-as-target condition solved fewer horizontally presented problems ($M = 0.73$, $SD = 0.11$) relative to vertically presented problems ($M = 0.77$, $SD = 0.14$), $p = .02$, $d = -.32$. They also solved fewer horizontally presented problems ($M = 0.80$, $SD = 0.11$) compared with vertically presented problems when they were tested in a group ($M = 0.85$, $SD = 0.14$), $p = .01$, $d = -.40$. Females under group-as-target threat solved fewer horizontally presented problems ($M = 0.77$, $SD = 0.11$) compared with vertically presented problems ($M = 0.84$, $SD = 0.14$) when tested alone ($p < .001$, $d = -.56$) but not when they were tested in groups ($p = .83$). Females in the control condition solved horizontally and vertically presented problems with equivalent accuracy when tested alone and in a group, $p > .05$. Overall, these results suggest that women were susceptible to stereotype threat when they are tested individually; however, single-sex testing environments alleviated these performance deficits. See Figure 2 for interaction between experimental condition, group composition, and problem presentation. See Table 1 for descriptive statistics.

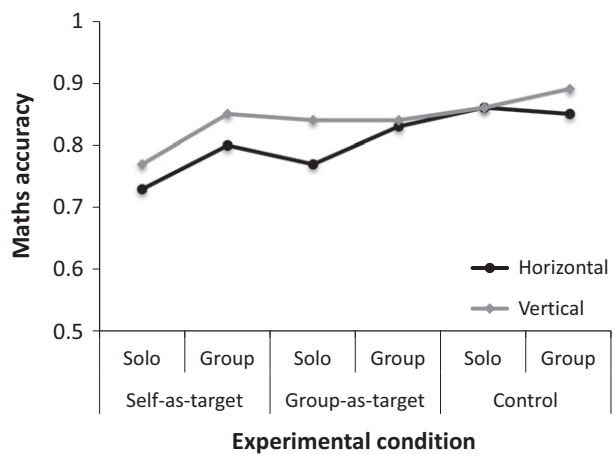


Figure 2. Three-way interaction between experimental condition, group composition, and problem presentation.

Table 1. Descriptive statistics for participants' mathematics arithmetic accuracy scores as a function of experimental condition and problem presentation

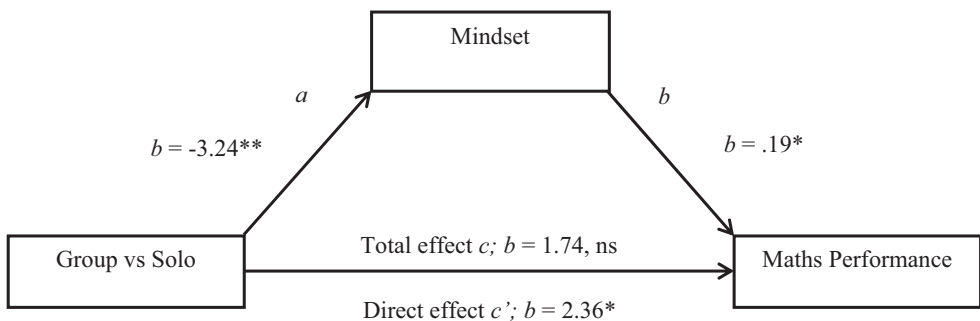
	Self-as-target		Group-as-target		Control	
	Solo	Group	Solo	Group	Solo	Group
Horizontal	.73 (0.11)	.80 (0.11)	.77 (0.11)	.83 (0.11)	.86 (0.11)	.85 (0.11)
Vertical	.77 (0.14)	.85 (0.14)	.84 (0.14)	.84 (0.14)	.86 (0.14)	.89 (0.14)

Mindset

Females' self-reported mindset did not differ dependent on whether they completed the questionnaire before or after the maths test ($p > .05$), indicating an absence of order effects. There was a significant main effect of mindset as a function of experimental condition, $F(2, 138) = 4.45$, $p = .01$, $\eta_p^2 = .06$. Participants assigned to the self-as-target stereotype threat condition ($M = 38.58$, $SD = 6.07$) reported a weaker growth mindset compared with the control condition ($M = 41.35$, $SD = 6.03$), $p = .04$, $d = -.46$. Participants in the group-as-target stereotype threat condition ($M = 38.46$, $SD = 4.45$) also reported a weaker growth mindset compared with the control condition, $p = .03$, $d = -.55$. There was a main effect of group composition, $F(1, 138) = 13.04$, $p < .001$, $\eta_p^2 = .09$. Participants who completed the test in groups reported a weaker growth mindset ($M = 37.85$, $SD = 5.32$) compared with those who completed the test alone ($M = 41.08$, $SD = 5.62$), $p < .001$, $d = -.59$. There was no interaction between stereotype threat and group composition, $p = .31$. These results suggest that negative gender-maths stereotypes, pertaining to women's personal or social identity, may hamper a growth-ability mindset. Furthermore, testing females in same-sex groups did not appear to have a positive effect on mindset.

Mediation analysis

Mediation analysis was conducted using ordinary least squares path analysis (Hayes, 2013). Here, we examined the influence that the single-sex testing environment exerted on mindset and mathematical performance. Results indicated that group composition indirectly influenced mathematical performance through its effect on mindset. Specifically, group composition negatively influenced mindset ($a = -3.24$) but positively

**Figure 3.** Mediation model indicating that mindset partially mediated the effects of group composition on women's total mathematical performance. Note. b = unstandardized co-efficient. * $p < .05$, ** $p < .001$.

predicted maths performance ($b = .19$). A bias-corrected bootstrap confidence interval for the indirect effect ($ab = -.63$) did not include zero (LLCI = -1.48 , ULCI = $-.10$). However, there was still evidence that being tested in a group influenced mathematical performance independent of its effect on mindset ($c' = 2.36$), $p = .03$. This provides evidence of partial inconsistent mediation (MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, Krull, & Lockwood, 2000), with mindset acting as a suppressor variable (MacKinnon *et al.*, 2000; Tzelgov & Henik, 1991). See Figure 3 for full mediator model.

Discussion

The current study evaluated the efficacy of single-sex testing environments as a practical means to eliminate stereotype threat effects. Moreover, it examined whether testing women in single-sex groups or under stereotype threat influenced them to perceive that their ability was a fixed trait. Results indicate that female participants underperformed when they were tested alone and were primed with a self- or group-relevant stereotype. However, these performance decrements did not emerge when they were tested in single-sex groups. These findings suggest that in-group members may function as ‘social vaccines’ who increase social belonging and inoculate fellow group members’ performance against the experience of stereotype threat (Dasgupta, 2011). Nevertheless, participants reported a weaker growth mindset when they were tested in groups relative to alone and under stereotype threat. As such, single-sex testing environments may reduce group members’ concerns about confirming a negative stereotype to bolster women’s mathematical performance but may have a paradoxical effect on mindset.

Female participants were susceptible to group-as-target stereotype threat when they were tested alone in comparison with those in a control condition. In this situation, women find themselves as single representatives of their social group, which may heighten the salience of negative stereotypes that accompany their group status. Being a minority member may result in added pressure because women apprehend that performance will confirm, and thereby reinforce, pejorative stereotypes as a true representation of their in-group (Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000). Findings also reveal that self-as-target stereotype threat had a greater negative effect than group-as-target stereotype threat. Participants may have been more vulnerable to self-as-target stereotype threat when they were tested alone because they perceived that performance would be evaluated in line with their personal ability. As such, the salience of a negative self-relevant stereotype may have interacted with the experience of being alone in the testing environment to attract a disproportionate amount of attention to one’s personal identity.

Findings indicate further that women’s mathematical performance was protected from the effects of self-as-target and group-as-target stereotype threat when they were tested in same-sex groups. This finding may be explained by distinctiveness theory (Abrams, Thomas, & Hogg, 1990; Cota & Dion, 1986), which posits that group saliency increases relative to the number of outgroup members in a particular setting (McGuire, McGuire, & Winton, 1979). Resultantly, the mere presence of other in-group members may have decreased women’s apprehensions about representing the group positively to bolster performance (Inzlicht & Ben-Zeev, 2000, 2003; Sekaquaptewa & Thompson, 2003). This finding may have practical implications in relation to gender-segregated learning environments. For example, research suggests that women may feel marginalized in mathematics classrooms when men outnumber them, with this influencing underper-

formance (Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). As such, increasing the number of women in counter-stereotypical domains, to create a critical mass, may present as a strategy to alleviate experiences of stereotype threat and encourage more women into maths-intensive fields.

Despite the positive impact that in-group representation had on performance, women who were tested in single-sex groups reported a weaker growth mindset compared with those who were tested alone. Evidence of inconsistent mediation revealed that being in a group negatively predicted mindset but positively predicted mathematical performance. When tested in single-sex groups, females may have become aware that they had been segregated from their male peers and attributed this to alleged gender differences in mathematical ability (c.f., Halpern *et al.*, 2011; Pahlke *et al.*, 2014). This may have led females to believe that gender is a fundamental characteristic of ability, thus weakening a growth-ability mindset. Participants also reported a weaker growth mindset when they were primed with a self-as-target and group-as-target stereotype relative to participants in the control condition. This is consistent with research suggesting that negative gender-maths stereotypes may influence women to believe that their mathematical ability is limited because of their group membership (Dar-Nimrod & Heine, 2006; Dweck, 2008; Good, Aronson, & Inzlicht, 2003).

Limitations and implications

A number of limitations should be borne in mind when interpreting the results of the current study. First, the study did not employ a fully matched design in that females' mathematical performance was not compared to that of males within single and mixed-sex testing environments. The rationale to only recruit female participants was underpinned by research which has suggested consistently that women's mathematical ability is hampered in the presence of men (e.g., Inzlicht & Ben-Zeev, 2000, 2003; Murphy *et al.*, 2007; Sekaquaptewa & Thompson, 2003), and by findings indicating that men are less susceptible to stereotype threat in the maths domain (Rydell *et al.*, 2014).

Participants were recruited with a wide range of demographic characteristics to ensure that the sample was not limited to university students, and sensitivity analyses indicated that maths scores and mindset did not differ as a function of participants' ethnicity or student status. However, it is viable to question whether these women encounter numerical asymmetry in terms of gender within their daily environment (e.g., educational discipline or workplace), and whether this may moderate stereotype threat effects. For example, women who are frequently outnumbered by men may be more susceptible to stereotype threat because they are conscious of their minority status in the activities they pursue (c.f., Inzlicht & Ben-Zeev, 2000). In a similar vein, participants signed up to the study via an online website, whereby they read a brief description of studies taking place and arranged a time to come into the laboratory. This purposeful sampling method may be considered a limitation because participants who were knowledgeable about negative gender-maths stereotypes or experience mathematics anxiety may have decided not to take part in the research.

The findings of the current study indicate that participants in the group-as-target and self-as-target stereotype conditions solved fewer simple and moderate problems relative to participants in the control condition. This contrasts with research indicating that stereotype threat effects are more pronounced with difficult questions (Keller, 2007). Within-participant analyses indicated that across all experimental conditions, participants

solved fewer difficult problems relative to moderate and simple problems. Resultantly, participants in the control condition may have also found these problems difficult, with this diminishing any potential differences between experimental conditions. This may particularly be the case given that a novel laboratory task was employed to ensure that participants were not familiar with the format of the test (c.f., Beilock & Carr, 2005). Future research that examines stereotype threat effects as a function of problem type and difficulty and utilizes more ecologically valid tests, such as the General Certificate in Secondary Education, is therefore recommended.

The current study may point to the possible mechanisms that underpin stereotype threat effects. Specifically, results reveal that performance decrements under stereotype threat were greatest for horizontally presented problems compared with vertically presented problems. This may reflect the difficulty of such questions as they appear in a different format to how individuals typically solve mathematical problems using pencil and paper (Trbovich & LeFevre, 2003). Moreover, research suggests that these types of questions rely more heavily on verbal working memory resources, a mechanism that has been implicated to underpin the effects of stereotype threat on women's mathematical performance (Beilock *et al.*, 2007; Rydell *et al.*, 2014). As such, negative societal stereotypes may influence intrusive performance-related thoughts that tax the verbal working memory resources required to solve mathematical problems.

Results reveal that mindset partially mediated the effects of in-group representation on women's mathematical performance. However, given that partial mediation was found, this suggests that additional (unmeasured) variables may explain further the relationship between single-sex testing environments and performance. For example, previous research suggests that single-sex testing environments may mitigate stereotype threat by decreasing anxiety (Ben-Zeev, Fein, & Inzlicht, 2005). Additional research would therefore benefit from exploring explanations for the potential efficacy of single-sex testing environments in the elimination of stereotype threat, with researchers acknowledging both the advantages and limitations of implementing such strategies.

Conclusion

The current research indicates that the salience of a negative self- or group-relevant stereotype can have a deleterious impact on women's mathematical performance, with these effects emerging after controlling for participants' perceived mathematical ability and domain identification. However, these performance deficits were reduced when women were tested in same-sex groups. These findings suggest that heightened in-group representation may bolster women's performance in counter-stereotypical domains. This finding may be particularly noteworthy when considering research which suggests that poor numerical representation may be a key determinant in women's decisions to avoid or leave maths-intensive fields, even for those who are highly skilled and domain identified (Murphy *et al.*, 2007). However, findings also reveal that females reported a weaker growth mindset when they were tested in single-sex groups. Whilst this underscores the importance of examining the potential efficacy of gender-segregated learning environments on both attitudinal and behavioural outcomes, it may lead one to question the best strategy to reduce stereotype threat effects. One approach could be to teach females in single-sex classrooms and encourage them to view mathematical ability as a malleable rather than fixed capacity (Aronson, Fried, & Good, 2002; Dweck, 2015; Good *et al.*, 2003, 2008; Spitzer & Aronson, 2015). Explaining to students that they have not been separated from males

based upon their ability but rather to foster their learning may have a positive effect on achievement, motivation, and engagement (Dweck, 2008; Elliot & Dweck, 2005). Nevertheless, research also suggests that gender-segregated educational environments may come at the cost of long-term socialization processes, particularly when females are re-integrated with males within further education and the workplace (Halpern *et al.*, 2011). An alternative strategy may therefore lie within tackling negative gender-related stereotypes within co-educational classrooms. This could be achieved by teaching students about the pervasive effects of stereotype threat and the direct influence it can exert on performance (c.f., Johns, Schmader, & Martens, 2005). Given that stereotypes about ability are fixed mindset beliefs (Blackwell *et al.*, 2007; Dweck, 2006, 2008), this strategy, in itself, may encourage students to adopt a growth mindset and increase women's participation and performance in mathematics domains.

Reference

- Abrams, D., Thomas, J., & Hogg, M. A. (1990). Numerical distinctiveness, social identity and gender salience. *British Journal of Social Psychology*, 29, 87–92. doi:10.1111/j.2044-8309.1990.tb00889.x
- Aramovich, N. P. (2014). The effect of stereotype threat on group versus individual performance. *Small Group Research*, 45, 176–197. doi:10.1177/1046496414523508
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38, 113–125. doi:10.1006/jesp.2001.1491
- 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
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality & Social Psychology*, 46, 610–620. doi:10.1037/0022-3514.46.3.610
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and 'choking under pressure' in math. *Psychological Science*, 16, 101–105. doi:10.1111/j.0956-7976.2005.00789.x
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill-over. *Journal of Experimental Psychology: General*, 136, 256–276. doi:10.1037/0096-3445.136.2.256
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, 41, 174–181. doi:10.1016/j.jesp.2003.11.007
- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behaviour*, 34, 39–89. doi:10.1016/S0065-2407(06)80004-2
- 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
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246–263. doi:10.1111/j.1467-8624.2007.00995.x
- Cota, A. A., & Dion, K. L. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality & Social Psychology*, 50, 770–776. doi:10.1037/0022-3514.50.4.770

- Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314, 435. doi:10.1126/science.1131100
- Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 22, 231–246. doi:10.1080/1047840X.2011.607313
- Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.
- Dweck, C. (2008). *Mindset: How you can fulfil your potential*. New York, NY: Random House.
- Dweck, C. S. (2015). Growth. Discussant. *British Journal of Educational Psychology*, 85, 242–245. doi:10.1111/bjep.12072
- Elliot, A. J., & Dweck, C. S. (Eds.) (2005). *Handbook of competence and motivation*. New York, NY: Guilford.
- 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
- Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., & Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science*, 333, 1706–1707. doi:10.1126/science.1205031
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guildford.
- Hollis-Sawyer, L. A., & Sawyer, T. P. (2008). Potential stereotype threat and face validity effects on cognitive-based test performance in the classroom. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 28, 291–304. doi:10.1080/01443410701532313
- Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, 99, 545–560. doi:10.1037/0022-0663.99.3.545
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11, 365–371. doi:10.1111/1467-9280.00272
- Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology*, 95, 796–805. doi:10.1037/0022-0663.95.4.796
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16, 175–179. doi:10.1111/j.0956-7976.2005.00799.x
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77, 323–338. doi:10.1348/000709906X113662
- Lee, K. M., & Kang, S. Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, 83, 63–68. doi:10.1016/S0010-0277(02)00010-0
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614. doi:10.1146/annurev.psych.58.110405.085542
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1, 173–181. doi:10.1023/A:1026595011371
- Martin, A. J. (2015). Implicit theories about intelligence and growth (personal best) goals: Exploring reciprocal relationships. *British Journal of Educational Psychology*, 85, 207–223. doi:10.1111/bjep.12038

- McGuire, W. J., McGuire, C. V., & Winton, W. (1979). Effects of household sex composition on the salience of one's gender in the spontaneous self-concept. *Journal of Experimental Social Psychology*, 15, 77–90. doi:10.1016/0022-1031(79)90020-9
- McKenzie, K. (2013). *Developing a growth mindset: The secret to improving your grades*. Academic Success Summit Program. East Stroudsburg, PA: East Stroudsburg University. Retrieved from http://www4.esu.edu/academics/enrichment_learning/documents/pdf/developing_growth_mindset.pdf
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18, 879–885. doi:10.1111/j.1467-9280.2007.01995.x
- Nguyen, H. H., & 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
- Pahlke, E., Shibley-Hyde, J., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140, 1042–1072. doi:10.1037/a0035740
- Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of young Ugandan women in coed and single-sex schools. *The Journal of Educational Research*, 105, 52–63. doi:10.1080/00220671.2010.517576
- Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality & Social Psychology*, 99, 883–896. doi:10.1037/a0021139
- Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes? *Personality and Social Psychology Bulletin*, 40, 377–390. doi:10.1177/0146167213513475
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39, 68–74. doi:10.1016/S0022-1031(02)00508-5
- Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, 11, 107–130. doi:10.1177/1088868306294790
- 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
- Spitzer, B., & Aronson, J. (2015). Minding and mending the gap: Social psychological interventions to reduce educational disparities. *British Journal of Educational Psychology*, 85, 1–18. doi:10.1111/bjep.12067
- 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
- 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., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance*, 16, 311–326. doi:10.1207/S15327043HUP1603_7
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

- Titze, C., Jansen, P., & Heil, M. (2011). Single-sex school girls outperform girls attending a co-educative school in mental rotation accuracy. *Sex Roles*, 65, 704–711. doi:10.1007/s11199-011-9947-y
- Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, 31, 738–745. doi:10.3758/BF03196112
- Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109, 524–536. doi:10.1037/0033-2909.109.3.524
- Verniers, C., & Martinot, D. (2015). Perception of students' intelligence malleability and potential for future success: Unfavourable beliefs towards girls. *British Journal of Educational Psychology*, 85, 289–299. doi:10.1111/bjep.12073

Received 17 November 2015; revised version received 13 January 2016