

When do Gender Stereotypes Impair Math Performance? A Study of Stereotype Threat Among Ugandan Adolescents

Katherine Picho & Toni Schmader

Sex Roles

A Journal of Research

ISSN 0360-0025

Sex Roles

DOI 10.1007/s11199-017-0780-9



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



When do Gender Stereotypes Impair Math Performance? A Study of Stereotype Threat Among Ugandan Adolescents

Katherine Picho¹  · Toni Schmader²

© Springer Science+Business Media New York 2017

Abstract The present study examined the effect of gender-based stereotype threat (ST) on the mathematics performance of high school students in Uganda, East Africa, as moderated by students' stereotype endorsement and/or their perceptions of stereotypic expectancies by others. Participants were 190 ninth grade students (age 14–15, senior 2, in Uganda) from all-female and coed boarding schools. Only perceived stereotypic expectancies by others significantly moderated ST effects on performance. A reminder of cross-gender comparisons led both young women and young men to underperform if they assumed that the researchers expected their own gender to do worse than the other gender. Importantly, students' perceptions of the stereotypic expectancies of authority figures (i.e. researchers) mattered more for predicting their math performance than did students' own endorsement of stereotypes. Collectively, these findings support a basic assumption of ST theory— that knowledge of a cultural stereotype is a prerequisite to the ST experience. Therefore, studies conducted with younger samples and in diverse cultural contexts should establish participants' awareness of the stereotype in question. Also, regarding gender and math stereotypes, it should not be assumed that males will always be immune from stereotype-based performance deficits on quantitative tasks when tested in different cultures. Finally, results suggest that conveying an expectation that young men and young women have equal

ability and potential might be important to preventing ST among younger age groups.

Keywords Stereotype threat · Women · Mathematics · High school · Africa · Culture · Stereotypes · Adolescents

Statistics show that women participate in science and engineering at much lower rates than do men (National Science Foundation 2015). Theory and research has suggested that these disparities in participation are partly a function of gender stereotypes that restrict women to the pursuit of careers congruent with gender roles. For example, when subtly reminded of negative ability stereotypes about their group in a given domain, some individuals perform more poorly than they otherwise would (Steele and Aronson 1995). Indeed, meta-analyses (Nguyen and Ryan 2008; Picho et al. 2013) reveal that this phenomenon, stereotype threat (ST), also affects the performance of some women in quantitative domains. The deleterious impact of ST has been extensively documented (Armenta 2010; Kiefer and Sekaquaptewa 2007; McGlone and Aronson 2006; Spencer et al. 1999; Van Loo and Rydell 2014), but a key limitation of this research has been the reliance on studying college-aged adults in Western cultures.

Although individuals from Western cultural contexts constitute only 12% of the world population, a disproportionate amount of psychological research is based on samples from these societies. Henrich et al. (2010) argue that Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies differ psychologically from the rest of the world. The dearth of ST research in non-WEIRD samples raises concerns about the extent to which stereotype threat exists as a real phenomenon in other cultural contexts. Given that gender stereotypes are culturally-defined constructs that show both some consistency but also variation across cultures (Cuddy et al. 2015; Eagly 1987), there is a need for more data on this important phenomenon coming

✉ Katherine Picho
edpsychresearch@gmail.com

¹ Department of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd, Bethesda, MA 20814, USA

² Department of Psychology, University of British Columbia, Vancouver, Canada

from understudied cultural contexts. For example, in a recent meta-analysis of stereotype threat effects among the same age group we investigate, Flore and Wicherts (2015) report evidence of a small effect of stereotype threat undermining female adolescents' math performance in the published literature ($d = -.22$), but they note that 94% of the studies included in their meta-analysis were carried out in one of four developed countries: the United States, Italy, France, or Germany. To increase the published data on this phenomenon from other non-WEIRD contexts, the present study examined the effect of gender-based ST on the mathematics performance of high school students in Uganda, East Africa.

Stereotype Threat in a Non-WEIRD Culture

The theory of stereotype threat (ST) postulates that students might underperform on cognitively challenging tasks when made aware of the possibility that their performance could confirm prevalent negative stereotypes about their group (Steele 1997). As applied to gender and math, the mere suggestion that boys and men are mathematically superior to girls and women has been shown to impair women's performance on mathematical tasks (Schmader and Johns 2003; Schmader et al. 2008). The integrated process model of stereotype threat specifies that simple reminders of self-relevant negative stereotypes can activate conflicting propositional links between one's self-concept relative to one's identification to both the stereotyped group and the domain in question, that is, "My group does not have this ability, I am like my group, but I think I have this ability" (Schmader et al. 2008, p.338). The cognitive imbalance among these propositions is thought to increase physiological threat, meta-cognitive processing of one's behavior, and active efforts to suppress negative thoughts and feelings. Unfortunately, efforts allotted to these vigilance and suppression processes can result in deficits in cognitive processing and subsequent degradation in performance, which paradoxically seems to confirm the very stereotype one was so motivated to disprove (Schmader et al. 2008).

The generalizability of ST to an African cultural context is an important empirical question because Uganda is culturally distinct from the United States and other WEIRD societies where most ST research has been conducted. Compared to those countries where stereotype threat has typically been investigated, *Uganda is a relatively more collectivistic culture that emphasizes interdependence and fulfilling social roles among its members but also a less masculine culture with somewhat weaker gender-role differentiation* (Hofstede 1980; Hofstede et al. 2010; Rarick et al. 2013). Based on social role theory, enacting gender roles reinforces gender differences in behavior which reifies gender stereotypes (Wood and Eagly 2010). *Research also indicates that gender stereotypes are more pronounced in individualistic cultures*

compared to collectivist cultures where men are seen as more similar to women on the previously-mentioned individualistic and collectivist traits (Cuddy et al. 2015). Therefore, the ways in which Uganda differs from the WEIRD cultures where stereotype threat has typically been documented could raise some questions as to whether gender-based stereotype threat would be experienced by young women in this culture.

However, despite these variations on dimensions of cultural differentiation, there is some reason to suspect that stereotype threat could be experienced by young Ugandan women. First, there is evidence of a gender gap in math that favors young men. The annual performance data from Uganda's National Assessment of Progress in Education (NAPE) administered by Uganda National Examinations Board (UNEB) consistently reports gender disparities in mathematics at both elementary and high school levels (UNEB 2013). This math gender gap favoring males varies by region (UNEB 2012), and it is larger in the more rural, economically underdeveloped regions, such as Northern Uganda where our study took place.

In addition, there is evidence that the math = male stereotype equation exists in Uganda. Both survey and qualitative research conducted with students from several schools located in central Uganda reveals strong perceptions that mathematics is a subject reserved for males (Kaahwa 2012; Kakooza 2004). A longitudinal study exploring the mathematics experiences of 99 Ugandan women from secondary school through college (Kaahwa 2012) also reveals several deterrents to the pursuit of mathematics among women, like stereotype endorsement (Plante et al. 2013; Schmader et al. 2004) and solo status (Beaton et al. 2007; Sekaquaptewa and Thompson 2002, 2003), which are also well documented as stereotype threat moderators in WEIRD contexts.

Additionally, the only known published study on stereotype threat conducted in an African context found evidence of ST among 10th grade high school female adolescents in a coeducational boarding school in central Uganda (Picho and Stephens 2012). In their study, a manipulation of ST impaired performance among female adolescents attending a coed school but had no effect on young women in a same-sex (boarding) school. In the present investigation, we aimed to replicate this prior evidence of stereotype threat among young women in a coed school and also examine the potential roles of knowledge of stereotype expectancies held by others and of endorsement of gender stereotypes as potential moderators of ST among Ugandan high school students.

Stereotype Threat Among Adolescents

A key criterion of experiencing stereotype threat is the knowledge, but not necessarily the endorsement, of the stereotype that others might hold about one's group (Steele 1997).

Surprisingly, research seldom measures or analyzes variability on these variables. Rather, it is assumed that subtle primes of gender or race will bring prevalent cultural stereotypes to mind. However, subtle manipulations such as reminders of one's gender or mere mention of gender comparisons can only activate a sense of threat if girls and women have clear knowledge of the negative stereotypes about their group and an expectancy that they might be applied to them (Schmader and Johns 2003). Although such subtle manipulations might activate stereotype threat among the college-aged samples typically studied in WEIRD samples (Nguyen and Ryan 2008), when examining stereotype threat among younger samples, this assumption of consensual stereotype knowledge or expectancies cannot be taken for granted.

Developmentally, children learn stereotypes across a series of developmental stages (Martin et al. 1990). Stereotype awareness begins in early childhood (McKown and Strambler 2009), and the ability to directly infer others' stereotypes about social groups increases between the ages of 5 and 11 (McKown and Weinstein 2003). On average, by early adolescence, most individuals have developed sufficient knowledge of broad cultural stereotypes (Enesco et al. 2005; Martin et al. 1990; McKown and Strambler 2009). However, because development does not occur uniformly across individuals, there is likely to be considerable variation among adolescents in their knowledge of gender stereotypes. For example, variation in stereotype knowledge might be a result of individual differences in exposure to stereotypes at a micro-cultural level (i.e., direct interaction with family and peers at home and at school). Indeed, one study conducted with young children found a link between individual differences in cultural socialization practices among parents of the children in the study and children's knowledge of broadly held stereotypes (McKown and Strambler 2009). Furthermore, Tomasetto et al. (2011) found that elementary school-aged girls did not exhibit a typical stereotype threat effect if their mothers explicitly rejected stereotypes about gender differences in math ability.

The gender stereotypes that students learn come not only from their parents but also from the performance differences that they do or do not observe from their peers. Meta-analyses of gender differences in mathematical performance suggest that gaps in math performance and participation, to the degree that they exist at all, are not observed until late adolescence and college (Hyde et al. 1990, 2008). Thus, if children develop stereotyped expectancies based on what they directly observe among their peers, there might be considerable variability in stereotype knowledge and expectancies that are held among female and male adolescents compared to the assumptions researchers make about consensual stereotypes held among college students. Based on this reasoning, we speculated that variability in the expectation of being stereotyped could be an important moderator to adolescents' experience of stereotype threat, especially in a non-WEIRD culture with relatively less evidence of the prevalence of gender stereotypes.

The Current Research

The present study was designed to examine how a subtle reminder of gender differences in mathematical performance would affect performance among male and female adolescents in Northern Uganda on a test of mathematical ability. Because of the exploratory nature of this study in this cultural context, alternative hypotheses were tested. On the one hand, the manipulation used has led to the underperformance of women, but not of men, in other published research using college aged students in the United States (Forbes and Schmader 2010; Johns et al. 2005; Schmader and Johns 2003). Thus, one hypothesis is that a reminder of gender differences in math performance (compared to control) would impair the performance of young women, but not of young men (Hypothesis 1). However, we recognized that research in Western samples has typically assumed that students would be aware of a cultural stereotype dictating male superiority in math and, in a new cultural context and with an adolescent sample, it was not clear that this assumption would be valid.

We further examined whether students' perceptions of stereotypic expectancies held by others (i.e., the researcher in this case) would moderate ST effects on the mathematics performance of high school students in Uganda, East Africa. Thus, the alternative hypothesis is that ST would impair young women's math performance to the extent that students expect those evaluating their performance to believe that young men are better than young women are at math (Hypothesis 2). Finally, we sought to distinguish the effects of these expectancies from students' own endorsement of gender stereotypes to better isolate the role that stereotype knowledge plays in moderating effects. We thus tested a third hypothesis that students' own stereotypes (rather than the stereotypes they expect others to hold) moderate the effects of ST on math performance (Hypothesis 3). Supplementary analyses on a sample from an all-female school are also provided. Because prior research found no evidence of stereotype threat among students attending a female-only school (Picho and Stephens 2012), we did not have strong predictions for this sample.

Method

Sample

The study received ethics approval from the first author's academic university in the United States and was conducted in two schools in Northern Uganda—a region that endured a two-decade long civil war that has left it significantly impoverished both economically and educationally relative to the rest of the country. Based on 2013 reports by the Uganda National Examinations Board (UNEB), the participating school in our study ranked as performing near national averages.

Our final primary sample included a total of 128 ninth grade students (65 young men, 63 young women; age 14–15; senior 2 in Uganda) from a coed boarding school. An additional 62 female adolescents from an all-female school were also run through the same procedure and analyzed separately. A male research assistant unaffiliated with either school distributed consent forms to a total of 263 (195 from the coed school) students at the end of the school term and signed forms (participant assent and guardian/parent consent) were returned by interested parties at the beginning of the following term. Response rates were extremely high (93.3%), and attrition over the course of the 3-week study period was 20%.

Procedure

Students were assigned to control and experimental conditions using a unique six-digit index number that was randomly assigned to individual students at the beginning of the study and prior to administering the surveys. The third digit of the index number represented assignment to the control condition (1) or experimental condition (2). The last three digits of the six digit index number uniquely identified participants in each condition.

Data collection occurred in three phases, each one week apart. First, a pretest of math performance was administered by mathematics teachers one week after students had returned from their vacation and after all consent forms had been collected. This pretest was framed by teachers as an assessment to gauge their mastery of material covered in previous terms. The following week, students assigned to the control condition completed the study as a group in a separate classroom from those assigned to the threat condition. They first completed a battery of psychological measures (in the order reported in the following section), followed by the manipulation of ST and a post-test measure of math performance one week later. Students received the ST manipulation as part of the task instructions attached to the first page of the math test. The same manipulation was also read out loud by the male research assistant. The PSAT was presented as a problem-solving exercise to students in the control condition and as a math test diagnostic of ability to those in the ST condition. Additionally, participants in the ST condition were told that, previously, gender differences had consistently been shown on the math test, but the direction of this difference was left unspecified.

Priming instructions for the ST condition were as follows:

You are about to take the Math Achievement Test (MAT). The MAT is a test of one's mathematical skills, and has been reliable in predicting students' ability to excel in future advanced levels of mathematics courses. In the past, the MAT has successfully distinguished students with a natural ability to excel in mathematics from

those lacking the skills to be successful in math. The test has also consistently shown there to be differences in performance between boys and girls. In today's session we want to get a measure of your math ability using MAT. You may find some of the questions challenging, however, they are all in the range of ability for most college students. We ask that you take this test seriously and make a genuine effort so that we can collect accurate data. Your performance on this test will be used to help us establish performance norms for men and women. After the test, we will provide you with feedback about your performance and ask you some questions about the test-taking experience. Please answer the questions provided below to the best of your ability. Your performance on this exam will be compared to the performance of senior 2 boys taking the same test. Good Luck!

Students in the control condition received the following instructions:

In today's session we would like you to complete a problem solving task. This task is not diagnostic of any ability –it is just a simple exercise that allows us to study how people work at problem solving. You may find some of the questions challenging, however, they are all in the range of ability for most senior 2 students. We ask that you take this exercise seriously and make a genuine effort to solve the problems so that we can collect accurate information. Your performance will be used to help us understand the different factors that are related to problem solving processes. Afterwards, we will give you with feedback about how you did and ask you some questions about the problem solving exercise.

All students were given 35 min to complete the test. Afterwards, participants also completed a post-study questionnaire assessing stereotype endorsement and perceived researcher expectations, followed by student debriefing which lasted approximately 20 min.

Measures

All measures were administered in English, an official language in Uganda. For psychometric reasons (i.e., low psychometric properties in this cultural context), von Hippel et al.' (1997) sentence completion task was administered but not utilized in our study. We analyzed patterns of missing data for surveys in Stata 14 and found complete data for 75% of participants. For the proportion of observations that had missing data on one or more items, the percentage of missing values on individual items was quite low, ranging from 3.1%

to 9%, except for one math identification item for which 13% of responses were missing. The missing data were determined to be missing completely at random. Therefore, composite scores for multi-item scales used in our study were created by averaging items for which data were available without imputing missing values. This procedure, called available item analysis (AIA), has been shown to yield parameter estimates equivalent to other imputation methods across low levels of missing data (Parent 2013).

For the mathematics test, answers to individual items were scored 1 if correct or 0 if incorrect (or left blank). An overall test score for each participant was then computed by adding the number of items scored as correct. Because all students who participated in the experiment completed the math test, there were no missing data on the math test scores.

Math Identification

ST theory identifies domain identification as a necessary (although not sufficient) condition for ST to occur. Therefore, in line with ST theory, students' level of math identification was assessed prior to the manipulation. Math-identification was assessed as participants' average responses on the math-identification subscale of Picho and Brown's (2011) Social Identities and Attitudes Scale (SIAS; $\alpha = .82$). The five-item subscale is anchored on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) such that higher scores indicate stronger identification with mathematics. The math identification scale constituted the following items: "I value math," "Doing well in math matters to me," "Being good at math will be useful to me in my future career," "Doing well in math is critical to my future success," and "My math abilities are important to my academic success." Only 4.1% (4 young men and 1 young woman) of the sample did not identify with mathematics, and 7.6% (5 young women and 4 young men) were neutral. Therefore the study sample was mostly math-identified.

Perceived Researcher Expectations of Performance (PREP)

Students' perceptions of how the researcher expected males and females to perform was assessed by the multiple-choice item: "I think the person who administered this test expects that... (1) girls will perform much better than boys (GMB), (2) girls will perform slightly better than boys (GSB) (3) girls and boys will perform equally well, (4) boys will perform slightly better than girls (BSB), and (5) boys will perform much better than girls (BMB)." To avoid drawing attention to these beliefs prior to performance, this item was assessed in the final survey.

Stereotype Endorsement

Students' own endorsement of gender stereotypes was assessed using a three-item stereotype endorsement survey administered at post-test (Schmader et al. 2004). The survey was anchored on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*) and consisted of the following items: "It is possible that boys have more math ability than do girls," "In general, boys may be better than girls at math," and "I don't think that there are any real gender differences in math ability" (reverse scored). The stereotype endorsement subscale was created by averaging scores across items. Higher scores on the scale denoted higher levels of stereotype endorsement. As with PREP, these items were assessed in the final survey.

Math Performance

Two separate tests were used to assess math performance prior to and after the ST manipulation. The pre-test comprised select items from the 2007 Trends in International Mathematics and Science Study (TIMSS), whereas items adopted from pre-SAT (PSAT) practice tests were used as the dependent measure of performance. The pretest consisted of 18 questions (10 multiple-choice questions and 8 structured questions), and the posttest comprised 23 test questions (18 multiple-choice and 5 structured questions). Plausible test scores ranged from 0 to 18 (for the pre-test) or 23 (for the posttest).

Results

Descriptive Analysis

A core assumption of stereotype threat is that people must be aware of others' stereotypes to be threatened by the possibility of confirming them. Thus, before testing the alternative hypotheses we described previously, we first sought to analyze students' stereotyped expectancies overall and as a function of condition. Descriptive statistics of Perceived Researcher Expectations of Performance (PREP), along with stereotype endorsement and PSAT scores, are presented in Table 1. A 2 (Treatment condition: control, ST) \times 2 (Gender: males, females) ANOVA on PREP as a dependent variable revealed only a significant main effect for gender, $F(1, 121) = 15.33$, $p < .001$, $\eta^2 = .112$. Overall, only 23% of the sample endorsed the math = male stereotype by reporting a number higher than the scale mid-point, a tendency that was higher in young men (31.8%, $M = 3.54$, $SD = 1.10$) than in young women (12.1%, $M = 2.85$, $SD = .97$).

The absence of any effect of the manipulation on students' expectancies suggests that some adolescents interpreted the manipulation as implying that the researcher expected males to outperform females whereas others saw it as an expectation

Table 1 Coed school: descriptive statistics for PSAT, PREP, and stereotype endorsement

Participants' gender	Treatment group				Cohen's <i>d</i> (Control-ST)	95% CI
	Non-threat		Stereotype threat			
	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)		
Math scores on the PSAT						
Male	35	4.33 (1.91)	28	3.90 (1.93)	.22	[-.72, .28]
Female	35	3.47 (1.47)	28	3.45 (2.05)	.01	[-.51, .49]
Perceived Researcher Expectations of Performance (PREP) scores						
Male	35	3.46 (.98)	28	3.64 (1.22)	.16	[-.34, .66]
Female	35	3.00 (.85)	28	2.64 (1.10)	.37	[-.87, .13]
Stereotype endorsement scores						
Male	36	4.35 (1.44)	28	3.75 (1.86)	.36	[-.86, .14]
Female	34	4.02 (1.59)	28	3.26 (1.77)	.45	[-.95, .06]

Note. Discrepancies in sample sizes are due to missing data on the PREP and stereotype endorsement variables

that females would outperform males (the modal response was an expectation of equal performance). Given that students did not automatically assume that comparing males and females implied an expectation that males would do better on the math test, it seemed less likely that we would find support for Hypothesis 1: The expectation that ST would impair only young women's but not young men's math performance. Rather, this initial analysis of stereotype expectancies made it more essential to test Hypothesis 2: That any effect of our ST manipulation on adolescents' math performance would be moderated both by gender and students' perceptions of the testing administrator's stereotypes regarding how young men and women would perform.

Testing Hypotheses 1 and 2

Math test performance was scored as number correct on the PSAT. Overall, students' performance was low, $M = 3.82$ ($SD = 1.96$), with no one scoring higher than 8 on the 23 item test and with a non-significant, small difference in math performance between male ($M = 4.14$, $SD = 1.92$, range = 0–8) and female ($M = 3.46$, $SD = .1.99$, range = 0–8) adolescents, $t(126) = 1.96$, $p = .052$, Cohen's $d = .347$, 95% CI [-.003, .695]. To test both Hypotheses 1 and 2, we conducted a hierarchical regression using Stata 14 (see Table 2a) in which students' PSAT scores were regressed onto variables representing Gender (0 = males, 1 = females), the ST manipulation (0 = non threat,

Table 2 Coed school: moderated regression analyses for variables predicting math scores on the PSAT

Variables	Moderator					
	(a) Perceived researcher expectancies			(b) Stereotype endorsement		
	B	SEB	β	B	SEB	β
TIMSS	.20	.09	.27*	.19	.09	.25*
Treatment (ST = 1)	-2.23	1.10	-.57*	-1.41	1.10	-.35
TIMSS*treatment	.16	.13	.33	.12	.14	.23
Gender (female = 1)	-.47	.48	-.12	-.42	.49	-.11
Moderator	-.13	.31	-.07	.02	.22	.02
Gender x treatment	.58	.72	.13	.25	.72	.05
Gender x moderator	.43	.49	.15	-.09	.30	-.06
Treatment x moderator	.72	.43	.31	.23	.30	.14
Gender x treatment x moderator	-2.04	.65	-.56**	-.33	.42	-.15
R^2		.23		.14		
R^2 (adjusted)		.16		.08		
F for ΔR^2		9.74**		.62		

Note. Only final model results are shown. For the variable Treatment, 0 = control, 1 = experimental group (ST) TIMSS Trends in International Mathematics and Science Study

* $p < .05$. ** $p < .01$

1 = ST), and PREP (mean centered). Pretest performance on the TIMSS and its interaction with the ST treatment variable were also included in the model as covariates to correct for potential bias in testing the interaction between the moderator and the ST treatment variable (Yzerbyt et al. 2004). Thus, the three-block hierarchical regression model included two covariates (TIMSS, TIMSS x ST interaction, and three independent variables: gender, ST condition, and PREP), two-way interactions between the independent variables, and a three way interaction of the same, entered successively.

Results from the full model yielded a significant main effect for ST, $\beta = -.57, p = .045$, which was qualified by a significant three-way interaction among gender, ST, and PREP, $\beta = -.56, p = .002$ (see Table 2a). No other main effects or interactions were significant. Thus, given the lack of a Gender x ST interaction, there was no support for Hypothesis 1, but the significant three-way interaction yielded support for Hypothesis 2 that gender differences in math performance would be moderated by both stereotype threat and stereotype expectancies. The aforementioned moderating effect is depicted in Fig. 1.

Using procedures outlined by Aiken and West (1991), the significant three-way interaction was probed using simple slopes analysis, and in line with predictions, we explored the two-way interactions of participants' gender x Perceived Researcher Expectancies in Performance (PREP) within each level of stereotype threat (ST or not; see Fig. 1). Looking within the context of activated stereotype threat, the more young men believed that the researcher's expectancies leaned toward males doing better, the better their PSAT performance ($\beta = .52, t = 2.53, p = .01$; see Fig. 1a). The opposite pattern emerged among young women. The more young women thought that the researcher's expectancies leaned toward males doing better, the lower their performance ($\beta = -.46, t = -2.61, p = .01$). Said another way, the more strongly young women believed that the researcher's expected females to perform well, the higher was these young women's actual performance.

Turning to the context in which ST was not activated, young men's and young women's beliefs about the researcher's expectations had no influence on their actual performance ($\beta_{\text{men}} = .01, t = .07, p = .94$; $\beta_{\text{women}} = -.25, t = -1.18, p = .24$; see Fig. 1b). Thus expectancies alone were not sufficient to induce performance differences among young men and women; instead, expectancies were related to performance only within the context of stereotype threat.

Testing Hypothesis 3

We next sought to isolate the moderating role of stereotype expectancies that students had for the testing administrator as distinct from students' own stereotype beliefs. Because students' endorsement of gender stereotypes were correlated with their expectancies of being stereotyped ($r = .28, p = .001$), we

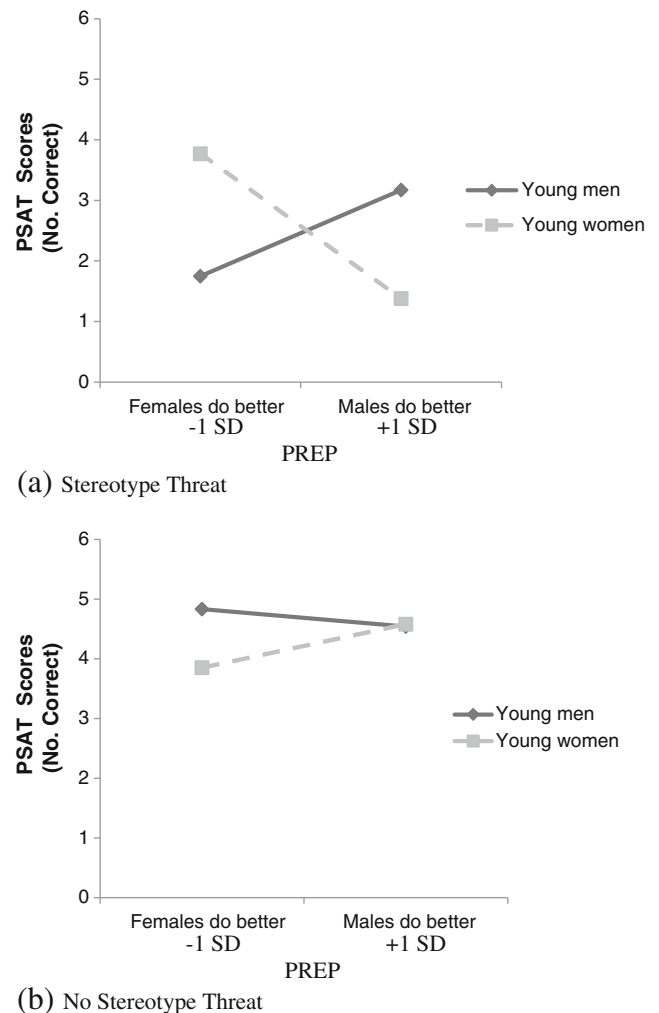


Fig. 1 Math performance as a function of students' gender, Stereotype Threat (ST), and Perceived Researcher Expectancies on Performance (PREP)

tested Hypothesis 3, which is an alternative hypothesis that ST effects on students' math performance would be moderated by their own endorsement of gender stereotypes. Our aim was to rule out the possibility that students' own endorsement of gender stereotypes explained the effects we reported previously. To assess the possible moderating effect of stereotype endorsement, we repeated the prior analysis including stereotype endorsement in place of PREP as the moderator. Results from the full model yielded no significant main or interaction effects (see Table 2b).

Supplementary Analyses: All-Female School Sample

Finally, during data collection, the same procedures, manipulation, and measures were used with a sample of 62 female adolescents recruited from an all-female school. Prior work on ST among adolescents has been mixed, with some finding evidence for (Huguet and Regner 2007, 2009) and against (Ganley et al. 2013) the phenomenon in this age group.

Also, the only known published study on adolescents in an African setting found evidence for ST among female adolescents in a coed school, but not in a same-sex school (Picho and Stephens 2012). Given that prior research suggests no effect of stereotype threat in an all-female setting (Picho and Stephens 2012), we did not have strong hypotheses that the manipulation either alone or moderated by stereotyped expectancies would affect performance. However, we summarize here parallel analyses (excluding gender as a factor given the all-female sample) with this supplemental sample.

Table 3 shows descriptive statistics of female adolescents on math performance, perceived researcher expectancies, and stereotype endorsement. As with the coed students, students' performance on the PSAT was low ($M = 2.29, SD = 1.44$), with no scores above 6 on the 23 item test. Low mean scores on the variable PREP indicated that, on average, adolescents from the all-female school perceived that the researcher expected females to perform better than males do. However, a one way ANOVA of threat condition on PREP revealed no significant differences between the threat and non-threat conditions on perceived researcher expectancies, $F(1, 60) = 1.33, p = .25$.

Second, when the same analytical model used with coed students was used to test for the effects of ST and PREP on performance, we observed no significant main or interaction effects for stereotype threat in this sample (see Table 4). In other words, in this all-female school sample, there was no evidence that young women's math performance was impaired by ST either alone or as moderated by their stereotype expectancies. Finally, although students' own endorsement of gender stereotypes was again correlated with their expectancies of being stereotyped, ($r = .30, p = .02$), there were also no significant effects of ST alone or in combination with stereotype endorsement predicting performance (see Table 4).

Discussion

The primary aim of the current study was to investigate stereotype threat among adolescents in an under-studied cultural context. Results from the present study revealed neither a

main effect of a stereotype threat manipulation on the math performance of African adolescent girls, nor any moderating effect of students' gender stereotype endorsement on ST effects. However variation in stereotypic expectancies among coed participants significantly moderated the effect of the manipulation on performance. Results from the female-only school sample revealed neither a main effect of stereotype threat nor any moderating effects of stereotypic expectancies and stereotype endorsement on stereotype threat. However, it is recommended that the null ST effects observed in the female-only school sample be interpreted with caution because the sample size of participants was small ($n = 62$) and null effects might have been a result of inadequate power.

Taken together, our results suggest that in this particular cultural context and within this age group, it is adolescents' knowledge, and not internalization of gender stereotypes, that might predict their susceptibility to experiencing ST effects. The suggestion that researchers would be conducting cross-sex comparisons led young men under stereotype threat to underperform if they assumed the researchers expected young women to do better and led young women to underperform if they assumed the researchers expected young men to do better. Thus, it should not be assumed that young men will always be immune from performance deficits on quantitative tasks.

The experience of stereotype threat assumes knowledge of a cultural stereotype, but as we have seen, this knowledge might vary greatly with younger samples (and perhaps also in novel cultural contexts). Therefore it is important to take into consideration stereotype knowledge and as well as other developmental ST factors when conducting research designed to generalize existing findings to new populations. These variables could very well account for some of the variability in ST in adolescents and children and, as such, provide a means to reconcile mixed findings in this under-studied population.

That said, it is unclear whether the variance in stereotypic expectancies observed in our study was due to cultural or developmental factors. There simply is not sufficient empirical research regarding cultural climate in Uganda as it relates to math and science education. The limited amount of research in Uganda (Kaahwa 2012; Kakooza 2004) seems to indicate

Table 3 Same-sex school: descriptive statistics for PSAT, PREP, and stereotype endorsement

Variable	Treatment group				Cohen's <i>d</i> (Control-ST)	95% CIs
	Non-threat		Threat			
	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)		
PSAT math	26	2.54 (1.65)	36	2.11 (1.25)	.30	[-.80, .21]
PREP scores	26	2.31 (1.32)	36	2.39 (1.07)	.07	[-.44, .57]
Stereotype endorsement	26	3.97 (1.61)	36	3.91 (1.65)	.04	[-.54, .47]

Note. Discrepancies in sample sizes are due to missing data on the PREP and stereotype endorsement variables
PREP Perceived Researcher Expectations of Performance

Table 4 Same-sex school: moderated regression analyses for variables predicting math scores on the PSAT

Variables	Moderator					
	Perceived researcher expectancies			Stereotype endorsement		
	B	SEB	β	B	SEB	β
TIMSS	.45	.15	.66*	.40	.15	.59**
Treatment (ST = 1)	1.66	1.21	.57	1.27	1.19	.44
TIMSS*treatment	-.32	.18	-.76	-.26	.18	-.61
Moderator	-.09	.22	-.07	-.10	.18	-.11
Treatment x moderator	.04	.31	.02	.18	.23	.15
R^2		.18		.19		
R^2 (adjusted)		.12		.11		
F for ΔR^2		3.12.*		.60		

Note. Only final model results are shown. For the variable Treatment, 0 = control, 1 = experimental group (ST) TIMSS Trends in International Mathematics and Science Study

* $p < .05$. ** $p < .01$

there might be less cultural consensus on gender stereotypes concerning math ability. However, prior test scores and survey data reviewed earlier suggests that a strong math = male bias does exist in Uganda. Another possibility is that young children and adolescents might not yet be fully aware of broader cultural stereotypes and that stereotype awareness could vary based on one's exposure (or lack thereof) of these stereotypes at the micro-cultural level (i.e. peers, and family). Indeed, previous research indicates significant variation in the beliefs of 4–8 year-old British children regarding gender differences in academics (Hartley and Sutton 2013). Also, ST in female adolescents is moderated by mothers' endorsement of gender stereotypes regarding mathematics (Tomasetto et al. 2011). A recent meta-analysis of ST in children revealed a small but reliable effect, although tests of available moderators did not explain the observed variability in effect sizes (Flore and Wicherts 2015). It appears, based on findings from our study, that stereotype knowledge might potentially explain the heterogeneity of effects especially in younger samples.

The absence of moderation by stereotype endorsement in our sample is contrary to previous research where stereotype endorsement has been linked to poorer performance outcomes among women under stereotype threat (Bonnot and Croizet 2011; Schmader et al. 2004). Either the developmental or cultural characteristics of the present sample could account for this discrepancy. We suspect that being aware of the stereotype is such a critical assumption of the phenomenon that variability along this dimension (either due to age or cultural factors) is more important than variation in personal beliefs. However, because our study was the first known study to examine stereotype endorsement in a non-WEIRD context, it is recommended that sufficient replication studies be conducted in this context to validate our finding.

Limitations and Recommendations for Future Research

Earlier we noted that performance on the math test was low, without much variability in test scores (none scored higher than 8 on the 23 item test). Therefore one limitation to our study was the restriction of range on performance that might have decreased power and attenuated bivariate relations between variables (Shadish et al. 2002). A more sensitive mathematics test with more heterogeneity might have yielded effects much larger than what was observed in our study. Future studies might benefit from using assessments that are difficult enough to elicit ST but not so difficult that floor effects arise.

Second, although the sample used in the present study was considerably large, variability on the variable PREP was uneven, with smaller samples below, and above, the scale midpoint respectively (i.e., 20.97% and 12.9% for coed females; 7.9%, and 38.1% for coed males). Thus the interpretation of findings related to the moderation of PREP on ST is tempered by the relatively small sample upon which these findings were based. We do recommend, however, that future replication studies be conducted with larger samples to validate these findings. We certainly acknowledge the need for a larger sample and attempted to do so, but obtaining samples of understudied populations in regions of the world without well-developed infrastructure for research proved to be quite challenging.

Finally, study participants came from nationally ranked low-performing schools situated in an economically impoverished region of the country. This might explain the floor effects on mathematics performance in our study, which might not be generalizable to or representative of the performance of the high school student population in Uganda. Thus the scope of our findings probably should be limited to Ugandan student subpopulations similar to that from which the study samples were derived.

Practice Implications

Findings from our study have implications for alleviating ST in authentic learning environments. Student and teacher-student interactions constitute a large part of the learning environment, and the nature of these interactions could either exacerbate or attenuate ST susceptibility among students from marginalized social groups.

Research shows that (a) members of marginalized groups rely on situational cues in the environment to assess the likelihood of experiencing ST (Murphy et al. 2007); (b) interacting with sexist males induces ST among female students (Logel et al. 2009); (c) teachers' expectations influence their behavior toward students (Good and Brophy 2000) which contributes to the ethnic achievement gap (McKown and Weinstein 2008); and (d) teachers' implicit bias negatively predicts the mathematics performance of minority students (Peterson et al. 2016). Collectively these and other studies show that stereotypic expectancies and the behaviors congruent with them create sub-optimal learning environments which can and sometimes do affect the performance of students belonging to stereotyped social groups. Therefore, the finding that students exposed to ST performed significantly worse when they believed that the researcher expected their gender to perform poorly implies that ST could be attenuated by fostering non-threatening learning environments. The process of creating intellectually non-threatening environments would, in part, require raising ST awareness among teachers as well as providing them with practical strategies to promote equitable pedagogy. Studies show that strategies such as blurring inter-group boundaries (Rosenthal and Crisp 2006), emphasizing social identities associated with positive ability stereotypes (Rydell et al. 2009), and teaching students about stereotype threat (Johns et al. 2005) might be useful in reducing stereotype threat among students susceptible to the phenomenon.

Additionally, the finding that young men exposed to stereotype threat performed less well when they believed that females were expected to perform better seems to suggest that despite positive stereotypes about their quantitative ability, young men might not necessarily be exempt from similar performance deficits that women under threat experience. Therefore, explicit efforts by educators to convey an expectation that young men and young women have equal ability and potential might also be an important means to prevent stereotype threat among younger age groups.

Finally, results showed that students' perceptions of the stereotypic expectancies of authority figures (i.e. researchers) mattered more when it came to math performance than students' own endorsement of stereotypes. This finding, which could have resulted from cultural or developmental factors (or both), highlights the importance of considering cultural and/or developmental factors that might be present in samples used to conduct ST research.

Conclusion

As noted, there has been a paucity of stereotype threat research in non-WEIRD contexts and with adolescent populations, yet diversity in research across various population groups is essential to building a unified theory of stereotype threat. The present study adds to the small literature in these populations by examining stereotype threat effects on the mathematics performance of high school students in a country that differs from the United States and other Western European countries where most ST research has been conducted, which is important for a number of reasons.

It seems likely that ST might be moderated by different factors in adolescent versus college-age groups. Thus efforts to remedy ST among susceptible individuals in the early school years might require extensive research with adolescent populations geared toward a critical understanding of moderators of stereotype threat in this age group. Accordingly, the present study contributes to our understanding of stereotype expectancies as a moderator of stereotype threat, a variable that has received surprising little attention in the prior literature although it is often assumed to be a fundamental assumption of the theory. Specifically, our study's findings show that as studies are carried out with younger samples and in diverse cultural contexts, it becomes more important to establish the basic assumption that participants have activated the stereotype in question.

Our study also improves our comprehension of ST in African cultural settings. To that end, we hope that these findings provide a platform for future research to conduct large, confirmatory replication studies in these cultural settings. This would significantly advance our understanding of the generality of the stereotype threat to other contexts culturally distinct from the West. Consequently, this research would inform future efforts to tailor culturally and age-appropriate interventions to counteract pernicious effects of the phenomenon and level the playing field for young women and young men in mathematics and science.

Acknowledgements The views expressed in the present article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

Compliance with Ethical Standards The present study was not funded. There are no conflicts of interest to report. The study was approved by the institutional research board at the first author's institution.

References

- Aiken, S. L., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks: Sage Publications.
- Armenta, B. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16*, 94–98. doi:10.1037/a0017564.

- Beaton, A., Tougas, F., Rinfret, N., & Huard, N. (2007). Strength in numbers? Women and mathematics. *European Journal of Psychology of Education, 22*, 291–306. doi:10.1007/BF03173427.
- Bonnot, V., & Croizet, J. C. (2011). Stereotype threat and stereotype endorsement: Their joint influence on women's math performance. *Revue Internationale de Psychologie Sociale, 2*(24), 105–120. Retrieved from www.cairn.info/revue-internationale-de-psychologie-sociale-2011-2-page-105.htm.
- Cuddy, A. J. C., Wolf, E. B., Glick, P., Crotty, S., Chong, J., & Norton, M. I. (2015). Men as cultural ideals: Cultural values moderate gender stereotype content. *Journal of Personality and Social Psychology, 109*(4), 622–635. doi:10.1037/pspi0000027.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale: Erlbaum.
- Enesco, I., Navarro, A., Parada, I., & Guerrero, S. (2005). Stereotypes and beliefs about different ethnic groups in Spain. A study with Spanish and Latin American children living in Madrid. *Journal of Applied Developmental Psychology, 26*, 638–659. doi:10.1016/j.appdev.2005.08.009.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*, 25–44. doi:10.1016/j.jsp.2014.10.002.
- Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology, 99*(5), 740–754. doi:10.1037/a0020971.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' math performance. *Developmental Psychology, 49*(10), 1886–1897. doi:10.1037/a0031412.
- Good, T. L., & Brophy, J. E. (2000). *Looking in classrooms* (8th ed.). New York: Longman.
- Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development, 84*(5), 1716–1733. doi:10.1111/cdev.12079.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2), 1–75. doi:10.1017/S0140525X0999152X.
- Hofstede, G. H. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: Intercultural competition and its importance for survival* (3rd ed.). New York: McGraw-Hill.
- Huguet, P., & Regner, I. (2007). Stereotype threat among school girls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99*, 545–560. doi:10.1037/0022-0663.99.3.545.
- Huguet, P., & Regner, I. (2009). Counter stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*, 1024–1027. doi:10.1016/j.jesp.2009.04.029.
- Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155. doi:10.1037/0033-2909.107.2.139.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494–495. doi:10.1126/science.1160364.
- 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.
- Kaahwa, J. (2012). The experiences of Ugandan females in mathematics. *Science Journal of Psychology, 12*, 1–12. doi:10.7237/sjpsych/103.
- Kakooza, J. (2004). *Mathematics and gender in Ugandan primary school: Influence on teachers, parents and learners*. Paper presented at the International conference to review research on science, technology, and mathematics education, Goa, India.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science, 18*(1), 13–18. doi:10.1111/j.1467-9280.2007.01841.x.
- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., Von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Attitudes and Social Cognition, 96*(6), 1089–1103. doi:10.1037/a0015703.
- Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development, 61*(6), 1891–1904. doi:10.1111/j.1467-8624.1990.tb03573.x.
- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27*, 486–493. doi:10.1016/j.appdev.2006.06.003.
- McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 74* (2), 498–515. doi:10.1111/1467-8624.7402012.
- McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social and academic consequences of stereotype consciousness in middle childhood. *Child Development, 80*(6), 1643–1159. doi:10.1111/j.1467-8624.2009.01359.x.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context and the achievement gap. *Journal of School Psychology, 46*, 235–261. doi:10.1016/j.jsp.2007.05.001.
- 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*(10), 879–885. doi:10.1111/j.1467-9280.2007.01995.x.
- National Science Foundation. (2015). Women, minorities and persons with disabilities in science and engineering. Retrieved from <http://www.nsf.gov/statistics/2015/nsf15311/digest/ssed>.
- Nguyen, 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.
- Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. *The Counseling Psychologist, 41*(4), 568–600. doi:10.1177/0011000012445176.
- Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction, 42*, 123–140. doi:10.1016/j.learninstruc.2016.01.010.
- Picho, K., & Brown, S. W. (2011). Can stereotype threat be measured? A validation of the social identities and attitudes scale (SIAS). *Journal of Advanced Academics, 22*(3), 374–411. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/1932202X1102200302>.
- Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *Journal of Social Psychology, 153*(3), 299–333. doi:10.1080/00224545.2012.737380.
- Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of single-sex and coed schools in Uganda. *Journal of Educational Research, 105*, 52–63. doi:10.1080/00220671.2010.517576.
- Plante, I., de la Sablonniere, R., Aronson, J. M., & Theoret, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*, 225–235. doi:10.1016/j.cedpsych.2013.03.004.
- Rarick, C., Winter, G., Nickerson, I., Falk, G., Barczyk, C., & Asea, P. K. (2013). An investigation of Ugandan cultural values and implications for managerial behavior. *Global Journal of Management and Business Research Administration and Management, 13*(9), 1–10.

- Retrieved from <http://journalofbusiness.org/index.php/GJMBR/article/viewFile/1056/967>.
- Rosenthal, H. E. S., & Crisp, R. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, 32(4), 501–511. doi:10.1177/0146167205281009.
- Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, 96(5), 949–966. doi:10.1037/a0014846.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 88, 934–947. doi:10.1037/0022-3514.85.3.440.
- Schmader, T., Johns, M., & Barquissau, M. (2004). The cost of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50, 835–850. doi:10.1023/B:SERS.0000029101.74557.a0.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. doi:10.1037/0033-295X.115.2.336.
- Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high and low-status groups. *Personality and Social Psychology Bulletin*, 28, 694–707. doi:10.1177/0146167202288013.
- 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.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- 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. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.370.3979&rep=rep1&type=pdf>.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629. Retrieved from http://users.nber.org/~sewp/events/2005.01.14/Bios+Links/Krieger-rec5-Steele_Threat-in-the-Air.pdf.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, 68, 797–811. doi:10.1037/0022-3514.69.5.797.
- Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology*, 47(4), 943–949. doi:10.1037/a0024047.
- Uganda National Examinations Board. (2012). *The achievement of senior two students in Uganda in mathematics, English language and biology 2008–2010*. Retrieved from http://www.uneb.ac.ug/Admin/images/Merged_report_2010-word_2003-Secondary.pdf.
- Uganda National Examinations Board. (2013). *The achievement of senior two students in Uganda in mathematics, English language and biology: 2008–2013*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.737.4513&rep=rep1&type=pdf>.
- Van Loo, K. J., & Rydell, R. J. (2014). Negative exposure: Watching another woman subjected to dominant male behavior during math interaction can induce stereotype threat. *Social Psychological and Personality Science*, 5(5), 601–607. doi:10.1177/1948550613511501.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, 33, 490–509. Retrieved from https://www.researchgate.net/profile/William_Von_Hippel/publication/232440995_The_Linguistic_Intergroup_Bias_As_an_Implicit_Indicator_of_Prejudice/links/54fec6b20cf2672e22405096/The-Linguistic-Intergroup-Bias-As-an-Implicit-Indicator-of-Prejudice.pdf.
- Wood, W., & Eagly, A. H. (2010). Gender. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., pp. 629–667). New York: Wiley. doi:10.1002/9780470561119.socpsy001017.
- Yzerbyt, V., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, 40, 424–431. doi:10.1016/j.jesp.2003.10.001.