CrossMark

Can Math-Gender Stereotypes Be Reduced? A Theory-Based Intervention Program with Adolescent Girls

Fengqing Zhao[1] · Yiyin Zhang[2] · Valeria Alterman[3] · Baoshan Zhang[2] · Guoliang Yu[4]

Abstract Math-gender stereotypes have been prevalent with adolescent girls and negatively affect their developmental outcome. The discrepancy between these negative stereotypes and early adolescent girls' pronounced mathematical performance provides a good opportunity for intervention on negative math-gender stereotypes. This study designed a three-month intervention program based on Identity Threat Model to reduce math-gender stereotypes for middle school girls. Nine intervention sessions were held in classrooms and targeted at changing adolescent girls' collective representations, situational cues and personal characteristics. Three-wave measures were collected before the intervention (pre-test), immediately after the intervention (post-test), and three months after the end of the intervention (follow-up test). Intervention effect indicators included math-gender stereotypes, self-esteem, math scores and language-gender stereotypes. The results immediately after the intervention showed that math-gender stereotypes were significantly reduced, math scores were improved, and self-esteem and language-gender stereotypes were not significantly affected among girls in the intervention group compared with control group. Follow-up test showed no significant differences between the intervention group and the control group except that math-gender stereotypes in intervention group were still lower than control group, indicating that the intervention activities were effective to reduce adolescent girls' math-gender stereotypes. The intervention program provided theoretical and educational implications for effective intervention on adolescent girls' math-gender stereotypes in school or classroom settings.

Keywords Math-gender stereotype · Intervention · Adolescent girls · Identity threat model

Introduction

Stereotypes about female inferiority in mathematics have been persistent and pervasive in daily life (Bieg et al. 2015; Miller et al. 2015). In the short run, negative math- gender stereotypes can affect adolescent girls' learning, motivation, and performance in relevant disciplines (Beilock et al. 2010; Forbes and Schmader 2010; Lane et al. 2012); in the long run, it can disrupt girls' potential development during adolescence and negatively affect their career choice and achievements in adulthood (Appel and Kronberger 2012; Good et al. 2008; Schmader et al. 2004). Therefore, it is quite necessary to study how to reduce or eliminate the negative effects of math-gender stereotypes during adolescence.

An amount of interventions aiming at reducing females' math-gender stereotypes have been conducted in laboratory settings (Johns et al. 2005; Ramsey et al. 2013; Stout et al. 2011) and school settings (Brinkman et al. 2011; Good et al. 2003; Lamb et al. 2009). However, there are three aspects that are worth further consideration. First is about the intervention strategies. Most laboratory interventions used one particular

✉ Guoliang Yu
  yugllxl@sina.com

  Valeria Alterman
  valterman@ufl.edu

[1] Department of Psychology, Renmin University of China, No. 59 Zhongguancun Street, Haidian District, Beijing 100872, China

[2] School of Psychology, Shannxi Normal University, No. 199 South Chang'an Road, Xi'an 710062, China

[3] Department of Management, University of Florida, PO Box 117165, Gainesville, FL 32611, USA

[4] Institute of Psychology, Renmin University of China, No. 59 Zhongguancun Street, Haidian District, Beijing 100872, China

intervention strategy, such as changing implicit attitudes (Forbes and Schmader 2010) and introducing role models (Shin et al. 2016). According to *Identity Threat Model*, these strategies generally target at one of the three aspects that affect individuals' appraisal of identity threat: collective representations, situational cues and personal cues, though, these three aspects should work together to shape one's appraisal of identity threat (Major and O'Brien 2005). Thus, it necessitates an integrated and systematical intervention on girls' math-gender stereotypes that target at multiple aspects of belief and behavior. Second is about the research method. Though the interventions conducted in laboratory settings examined immediate effect of specific intervention strategy effectively, they would result in a lack of ecological validity and it was unclear whether they were effective to generalize to real life settings, like classrooms. And the experimental interventions were generally conducted once in a very short period and cannot guarantee long-lasting intervention effects. Third is about the cross-cultural generalizability of school interventions. Many laboratory and school intervention strategies have been conducted in western culture (e.g., Good et al. 2003), however, it is unclear whether they are effective in eastern background. Taken together, it is important to adopt multiple intervention strategies in real circumstances like classrooms or schools and examine the long-term intervention effects on girls' math-gender stereotypes.

The present study expected to extend previous intervention research by integrating the strategies into an intervention program based on the rationale of Identity Threat Model (Major and O'Brien 2005). The intervention activities were targeted at math-gender stereotypes among adolescent girls. Self-esteem and math performance were included as convergent indicators and language-gender stereotypes as discriminant indicator of stereotype intervention effect.

Math-Gender Stereotypes in Adolescent Girls

Math-gender stereotypes have been observed in Grade 3 or 4 of elementary school and become prevalent among early adolescence (Marsh 1989; Muzzatti and Agnoli 2007). Muzzatti and Agnoli's (2007) investigation in Italian children indicated that the onset of explicit math-gender stereotypes has been found at around 9 years of age, and the patterns observed in elementary school consolidate during middle school. Another study showed that boys and girls in grades 7 and 9 hold explicit math-gender stereotypes; girls in grade 9 reveal stronger implicit math-gender stereotypes than boys (Steffens et al. 2010). These negative stereotypes can take effect on adolescents' cognitive performance via stereotype threat (Ambady et al. 2001).

However, a body of literature suggests that girls tend to outperform boys in mathematics until they reach adolescence or high school (Lai 2010); some other studies didn't find significant differences in mathematics between boys and girls in various cultures including China (Hyde et al. 2008; Zhang and Tsang 2015). For example, one study interviewed a sample from Beijing and observed that girls outperformed boys in primary and lower secondary education on total score for Middle School Graduation Exams mathematics (Lai 2010). Though girls showed pronounced performance in math or as good as boys during adolescence, girls still hold negative math-gender stereotypes and girls' lower self-concept in math relative to boys' exceed actual performance differences (Steffens et al. 2010). This discrepancy between performance and negative stereotypes is detectable among girls who achieve better grades than boys in mathematics but are underrepresented in the Mathematics Olympiads and girls who performed less well than boys (Muzzatti and Agnoli 2007). The discrepancy between performance and stereotypes makes it more significant to intervene with early adolescent girls' math-gender stereotypes since the negative math-gender stereotypes may have long-lasting effect on girls' future motivation for math-related careers and choice of math-related domain.

Math-Gender Stereotypes in Adolescent Girls: Effect on Self-Esteem and Math Performance

Math–gender stereotypes demonstrate that boys are more suitable for mathematics than girls (Franceschini et al. 2014; Steffens and Jelenec 2011; Steffens et al. 2010). These negative stereotypes have been prevalent among adolescents and have become a problem that affects female students' self-esteem and math performance.

First, math-gender stereotypes have negative effects on target individuals' self-esteem. Self-esteem refers to a positive or negative orientation toward the self (Rosenberg 1979). Self-esteem is also operationally defined as the association between the concept of self and a valence attribute (Greenwald et al. 2002). Theoretically, the balance-congruity principle has demonstrated that female group identity (self = female) and math-gender stereotypes (male = math) can work together to predict self-esteem in math domain (self ≠ math) (Greenwald et al. 2002; Nosek et al. 2002). Thus, negative math-gender stereotypes were negatively correlated to individuals' self-esteem in math domain (Guimond and Roussel 2001; Lindberg et al. 2013; Martinot and Désert 2007). Furthermore, researches indicated that self-esteem in particular domains can contribute to the global self-esteem (Daniels and Leaper 2006). Since school experience makes up the majority of middle school students' lives and shapes their self-esteem, it is presumable that the change of self-esteem in math domain is correlated with and can spillover to global self-esteem in middle school students. That is, the higher the level of negative math-gender stereotypes is, the lower the global self-esteem is (Rydell et al. 2009).

Second, math-gender stereotypes are closely associated with math achievements. Substantial studies on college students showed that the more negative academic gender stereotype individuals hold, the lower corresponding school grades they have, and vice versa (Heyder and Kessels 2013; Smeding 2012; Tine and Gotlieb 2013). A longitudinal study indicated that negative academic-gender stereotypes in grade eight are negatively related to classroom engagement in grade eleven (Swinton et al. 2011). In line with this, Spencer et al. (1999) have revealed that negative math-gender stereotypes could account at least in part for the persistent performance deficits for females on standardized tests of math achievement in school. A recent meta-analysis also indicated that stereotype manipulation significantly deteriorates females' math performance (Doyle and Voyer 2016). In brief, math-gender stereotypes are significantly correlated with math grades.

The Intervention on Math-Gender Stereotypes in Adolescent Girls: An Intervention Program Based on *Identity Threat Model*

A large body of research on intervention strategies of math-gender stereotypes has been conducted in laboratory settings. For example, training women to hold more positive attitudes about math (Forbes and Schmader 2010), taking verbal tests before math tests (Smeding et al. 2013), suppressing stereotype related thought (Logel et al. 2009), and introducing female role models (Shin et al. 2016) are experimentally demonstrated to be effective ways to reduce negative math-gender stereotypes. All these intervention strategies have undoubtedly important implications for alleviating female students' math-gender stereotypes. However, the laboratory interventions were lack of ecological validity and cannot guarantee the effectiveness in real life settings. Consequently, it is necessary to adopt previous intervention strategies systematically to real classroom activities (Zhang et al. 2014).

Major and O'Brien (2005) put forward an Identity Threat Model of Stigma to explain how people are affected by negative stereotypes. This model assumes that individuals possessing a devalued social identity tend to experience identity threat when stereotype relevant stressors are appraised as potentially harmful to one's social identity and as exceeding one's coping resources. Collective representations, situational cues and personal characteristics are three factors that affect how individuals evaluate the potential threat of specific situation on their well-being (Major and O'Brien 2005). That is, one appraises the stigma-relevant stressor as identity threat based on the interaction between perceived cues in the immediate situation that make stigma relevant to the situation, the collective representations that individual bring to the situation, and individual characteristics including stigma sensitivity, domain identification, group identification, goals and

motivation. Thus, controlling these three influence factors can reduce the negative effect of stereotypes (Major and O'Brien 2005).

Collective representations refer to the simplified and generalized cultural views toward a subdominant group. Girls endorse the dominant cultural stereotypes of their inferiority in mathematics (Miller et al. 2015; Passolunghi et al. 2014). The interventions targeted at collective representations should teach adolescent girls to reappraise their "vulnerable" groups in more positive directions and to use schema information in more reasonable and positive ways. Specifically, the present study focused on four methods to change group representations: redefining where vulnerable groups belong to by providing external attribution for difficulties (Johns et al. 2005), providing female role models (Marx and Roman 2002; Stout et al. 2011), and teaching students to think in different ways (Cheryan et al. 2015).

Situational cues have been extensively studied to have negative effect on adolescents' math-gender stereotypes. Stereotype threat can be seen as a situational social identity threat (Steele et al. 2002). Once a person is evaluated on the basis of a threatened social identity in certain environment, he or she would experience stereotype threat. Eliminating task cues related to negative stereotype or redefining tasks can prevent girls from experiencing stereotype threat. For example, subtle cues such as describing the task as a challenge, announcing that the test is diagnostic for verbal rather than math ability, or being taught by an instructor who is an in-group member can reduce the negative effects of stereotypes (Alter et al. 2010).

Personal characteristics can influence how situations are perceived and appraised. Personal characteristics include stigma sensitivity, domain identification, group identification, goals and motivation (Major and O'Brien 2005). Changing the individual tendency of the self-evaluation and self-schema are previously studied to be effective in intervening academic gender stereotypes. We reviewed previous studies and summarized four strategies to intervene with personal characteristics: emphasizing on intellectual growth (Aronson et al. 2002), training implicit attitude changes (Forbes and Schmader 2010), encouraging self-affirmation (Martens et al. 2006), and establishing long-term justice belief (Kay and Jost 2003). Therefore, the intervention strategies systematically controlling collective representation, personal traits, and situational cues based on the rationale of Identity Threat Model were expected to be effective in reducing negative math-gender stereotypes.

The Characteristics of Math-Gender Stereotypes: Domain Dependent or Domain Independent

Math-gender stereotypes stress that boys are more suitable for mathematics than girls, whereas language-gender stereotypes suggest that girls are more suitable for literature and language

than boys (Galdi et al. 2014). Some researchers believe that different kinds of academic-gender stereotypes are correlated to each other. That is, if one type of academic gender stereotype is strong, the others academic stereotypes that the same person holds would also be strong (Beilock et al. 2007). However, much more researchers tend to believe that different kinds of academic gender stereotypes are independent from each other (Walker and Bridgeman 2008). Accordingly, in the practice of intervention on the effects of negative math-gender stereotypes, the intervention effect should not spill over to the language domain.

The Present Study

The aim of this study was to reduce negative stereotypes among middle school girls who were in a developmental stage with consolidated math-gender stereotypes. A comprehensive intervention program was designed to reduce math-gender stereotypes based on the principles suggested by Identity Threat Model. The aim was to test the intervention effect on decreasing negative math-gender stereotypes and the retention of that knowledge, specifically proposing the following hypothesis:

> Hypothesis 1: Compared to participants in control group, participants in intervention group will report less math-gender stereotypes after intervention and at follow-up.

The level of stereotype in negative stereotype domains should be closely associated with important outcomes such as self-esteem and math achievement (Major and O'Brien 2005). Therefore, self-esteem and math exam scores were included as two important outcome variables that result from the change of math-gender stereotypes. That is, self-esteem and math scores were taken as the criteria for testing the effectiveness of intervention strategies on reducing math-gender stereotype.

> Hypothesis 2: Compared to participants in control group, participants in intervention group will report higher self-esteem after intervention and at follow-up.
> Hypothesis 3: Compared to participants in control group, participants in intervention group will report higher math score after intervention and at follow-up.

The final aim was to test whether the intervention would be effective in decreasing participants' knowledge of language-gender stereotypes and the retention of that knowledge relative to control condition, specifically proposing the following:

> Hypothesis 4: The intervention will not decrease adolescent girls' language-gender stereotypes relative to baseline, and this effect will be maintained at follow-up.

Method

Participants

The intervention program was designed and conducted on courses of Mental Health Education in a junior high school in China by employing cluster sampling technique. Before intervention program begins, informed consent was obtained from all participants included in the study. Researchers explained and assured that the intervention activities were conducted purely for research purpose and that participation was voluntary. The students were free to decline participation without any negative consequences. One hundred and fifty-six junior high school students (79 girls and 77 boys) from two grade 7 and two grade 8 classes participated in the intervention activities at baseline and they indicated willingness to consider further participation. These four classes were randomly assigned into intervention group and control group, each including one class from grade 7 and one class from grade 8. All of the students completed a three-wave study (baseline, intervention, and follow-up). As the main interest was adolescent girls' math-gender stereotypes, boys' data was not included in the present study.

The final sample included 77 middle school girls except two girls who didn't finish the three-wave measures. The intervention group contained 38 participants (22 from grade 7 and 16 from grade 8) and the control group contained 39 participants (22 from grade 7 and 17 from grade 8). Pearson Chi-square test results showed that $\chi^2 (1) = .02$, $p = .90$, indicating that grade was not significantly different across two groups. The mean ages for the intervention and control group were 13.58 years ($SD = 0.58$) and 13.61 years ($SD = 0.59$), respectively. There were no significant differences in age, $t (71) = -.24$, $p = .81$.

Materials

Math-Gender Stereotype Math-gender stereotype scale was developed from previous research questionnaires (Greenwald et al. 2003; Steffens et al. 2010) and results from interviews with middle school teachers. Originally, the scales were distributed to 100 students. The students were asked to assess whether each item was clear and appropriate for their age and to make sure they could understand the meaning. After analyzing the students' feedback and further revising the scale, the final scale was formed. The final math-gender stereotype scale includes 10 items regarding male math-gender stereotypes and 10 items regarding female math-gender stereotypes, separately (e.g., "boys are good at math problem solving" and "girls are good at math problem solving"). Students were asked to assess these items on a Likert-type 5-point scale. The items for "girls are good at math" items were

negatively scored. The higher the score is, the stronger the math-gender stereotype held by the individual. The final scale was then distributed to another 100 students (50 boys, 48 girls, 2 unknown genders) to test the reliability and validity. Confirmatory factor analysis results showed that the fitting index of the scale structure was good (RMSEA = .079, χ^2 (34) = 1.97, p = .001, GFI = .92, NFI = .94, CFI = .97). The factor loadings for math-gender stereotype items in the present study range from 0.71 to 0.95, indicating that the measurements used in the present study was of reasonably good quality. In addition, the math-gender stereotype scale was significantly correlated with math score (r = −.54, p < .001), which showed a reasonable validity for the questionnaire. The Cronbach's α for the math gender stereotype scale was .80.

Language-Gender Stereotype The procedure for developing and validating the scale of language-gender stereotype was the same as that of the math-gender stereotype scale mentioned above. The final language-gender stereotype scale includes 10 items regarding male language-gender stereotypes and 10 items regarding female language-gender stereotypes, separately (e.g., "boys are good at language" and "girls are good at language"). Students were asked to assess these items on a Likert-type 5-point scale. The score for "boys are good at language" items were negatively scored. The higher the score is, the stronger the language-gender stereotype held by the individual. Confirmatory factor analysis results also showed a good fitting index of scale structure (RMSEA = .084, χ^2 (34) = 2.09, p < .01, GFI = .92, NFI = .92, CFI = .96). The factor loadings for language-gender stereotype items in the present study range from 0.69 to 0.89, indicating that the measurements used in the present study was of reasonably good quality. In addition, the language-gender stereotype scale was significantly correlated with Chinese score (r = −.53, p < .001), which showed a reasonable validity for the questionnaire. This measure had good reliability (α = .80).

Self-Esteem Rosenberg self-esteem scales (RSES) were used to measure the level of general self-esteem. The self-esteem scale includes 10 items (e.g., "I take a positive attitude toward myself"). The scale typically uses a Likert-type response format that employs scales ranging from Strongly disagree to Strongly agree. Total score ranges from 10 to 40 points; the higher score means the higher level of self-esteem. This measure had good reliability (α = .78).

Math Grades (Exam Scores) The recent test scores for the first midterm, first final term, and second midterm examinations in math were collected for each individual. To analyze the data, we transformed test scores in each grade into T scores.

Intervention Program

The three-month comprehensive intervention program was launched at a frequency of once a week or every two weeks (depending on the school schedule) to reduce negative math-gender stereotypes for adolescent girls through systematical intervention on collective representations, situational cues and personal characteristics. The whole program consisted of 9 sessions (1 opening session, 7 intervention sessions and 1 ending session), each of the seven intervention sessions focusing on one specific intervention strategy. They include: encourage the self-affirmation, external attribution training, changing the implicit attitude, establishing long-term justice goal, emphasis on intelligence growth, introducing role model and focusing on the comparison of difference thinking. Each intervention session consisted of four parts including warming up, transition, working, and ending. Each session lasted for approximately 40 min and was conducted in the classroom.

Take the intervention of providing female role models as an example. After warming up, the class intervention included three parts: brainstorm, praise a great, and topic discussion. In the brainstorm part, students were asked to write down as many names of scientists the group members know as possible in one minute and count the number of male scientists and female scientists. Then students discussed why such few female scientists' names were listed. In the "praise a great" part, group representatives were asked to introduce the glorious achievements of female scientists so that female scientists become possible role models for girls. In the "topic discussion" part, students were asked to discuss how female scientists gained glorious achievements and to discover female role models who were good at science in their lives. For detailed information on the intervention program please see Appendix.

Participants in control group received no intervention activities; instead, they had themed class meetings as usual according to the assigned books by school, the themes included ways to improve mental health, building harmonious interpersonal relationships, emotion control, etc. Participants in the intervention group received intervention activities for three months by a female teacher. Both control groups and intervention group class activities consisted of four sections, including warming up, transition, working, and ending.

Procedure

Baseline The pre-test was conducted the week before the intervention began. Students finished a series of questionnaires including math-gender stereotypes, self-esteem and language-gender stereotypes scales. Meanwhile, the midterm math exam scores were collected from all participants.

Intervention During the intervention, the nine serial sessions for reducing math-gender stereotypes were conducted by one

female researcher for the intervention group once a week or every two weeks during the period of 3 months (depending on school schedules). At the same time, participants from the control group attended a normal class meeting by theme once a week or every two weeks (depending on school schedules). At the end of the intervention program, the post-test was conducted using the same measures of math-gender stereotypes, self-esteem and language-gender stereotypes as in the pre-test. The final-term math test scores were collected.

Follow-Up Three months after the intervention program, the same measures for math-gender stereotypes, self-esteem and language-gender stereotypes were conducted and the second mid-term math test was collected. Finally, all participants were debriefed and thanked.

Data Analysis

Since cluster sampling was used in current study, students were not randomly assigned into the control and intervention groups, which cannot guarantee sample homogeneity completely (Wright and Daniel 2006). Consequently, to examine the intervention effects, analyses of covariance were employed (Huitema 2011; Wright and Daniel 2006). In data analysis, the group (intervention group vs. control group) was the independent variable, the post-test and follow-up test scores of math-gender stereotypes, self-esteem, math exam scores and language-gender stereotypes were dependent variables, with corresponding pre-test indices as covariate variables, respectively.

Results

Descriptive Statistics

Means and standard deviations of three-wave math-gender stereotypes, language- gender stereotypes, self-esteem and math exam scores were listed on Table 1.

Math-Gender Stereotype

First, participants' math-gender stereotypes favoring boys in the first place was tested. The mean score of pre-test math-gender stereotypes was $M = 33.27$, $SD = 6.13$, one sample t test showed that $t (76) = 47.66$, $p < .001$, indicating that participants held math-gender stereotypes prior to intervention.

Then an analysis of covariance was conducted using the group as the independent variable, the post-test scores of math-gender stereotypes as a dependent variable, and the

Table 1 Means and standard deviation of variable over three times of measures

	Pre-test M(SD)	Post-test M(SD)	Follow- up test M(SD)
Math-gender stereotype			
intervention	32.23(6.96)	28.47(6.21)	30.13(6.13)
control	34.30(5.07)	34.03(5.82)	33.67(6.65)
Self-esteem			
intervention	32.03(4.35)	33.66(4.21)	33.66(4.21)
control	32.28(3.10)	32.33(3.56)	32.90(3.85)
Math grades (T score)			
intervention	50.83(11.16)	52.05(9.51)	50.17(11.37)
control	48.98(8.87)	47.86(9.89)	49.68(8.69)
Language-gender stereotype			
intervention	37.61(6.33)	35.42(6.45)	35.66(5.44)
control	37.28(4.93)	35.09(5.22)	35.54(5.63)

pre-test scores of math-gender stereotypes as a covariate variable. The homogeneity of regression coefficients of pretest math-gender stereotypes on post-test math-gender stereotypes was significant, $F (1, 73) = 6.49$, $p = .01$, partial $\eta^2 = 0.08$, indicating the slopes are heterogeneous. Thus, the traditional analysis of covariance was not proper. Huitema (2011) indicated that "when the slopes are heterogeneous, an alternative to ANCOVA should be considered". Then the Johnson–Neyman (J–N) technique was used to analyze the heterogeneous regression case. PROCESS 2.15 was used to conduct the J-N analysis (Hayes 2013). The results indicated that the value was 30.13, J-N significance region is [33.77, 66.23]. That is, the post-test math-gender stereotypes were significantly different between two groups for participants scored above 30.13 in pre-test math-gender stereotypes, but the effect was not significant for participants scored lower than 30.13 in pre-test math-gender stereotypes.

Following the same strategy, another covariance analysis was conducted on the follow-up test of math-gender stereotypes with group as the independent variable and pre-test scores of math-gender stereotypes as the covariate variable. The homogeneity of regression coefficients of pretest math-gender stereotypes on follow-up math-gender stereotypes was not significant, $F (1, 73) = 3.62$, $p > .05$, partial $\eta^2 = 0.05$, indicating the slopes are homogeneous. Results showed that the covariate effect of pre-test score of math- gender stereotypes was significant, $F (1, 74) = 13.45$, $p < .001$, partial $\eta^2 = 0.15$. The main effect of group was marginally significant, $F (1, 74) = 3.84$, $p = .05$, partial $\eta^2 = 0.05$. The math-gender stereotypes of intervention group was lower than that of the control group, indicating the effect of the intervention did still exist even three months after the end of intervention.

Self-Esteem

Taking the pre-test scores of self-esteem as the covariance variable, the results showed that the homogeneity of regression coefficients of pretest self-esteem on post-test self-esteem was not significant, $F(1, 73) = 0.19$, $p = .67$, partial $\eta^2 = 0.003$, indicating the slopes are homogeneous. The covariate effect of pre-test score of self-esteem was significant, $F(1, 74) = 21.28$, $p < .001$, partial $\eta^2 = 0.22$. The main effect of group was marginally significant, $F(1, 74) = 3.40$, $p = .07$, partial $\eta^2 = 0.07$. The self-esteem of the intervention group was marginally higher than that of the control group.

Another covariance analysis on the follow-up test of self-esteem was also conducted with group as the independent variable and pre-test scores of self-esteem as the covariate variable. The homogeneity of regression coefficients of pretest self-esteem on follow-up self-esteem was not significant, $F(1, 73) = 0.38$, $p = .54$, partial $\eta^2 = 0.005$, indicating the slopes are homogeneous. Results revealed that the covariate effect of pre-test score of self-esteem was significant, $F(1, 74) = 17.55$, $p < .001$, partial $\eta^2 = 0.19$. The main effect of the group wasn't significant, $F(1, 74) = 1.11$, $p = .30$, partial $\eta^2 = 0.02$. There was no significant difference on self-esteem between the intervention group and the control group.

Math Exam Scores

Taking the pre-test score of math score as the covariate variable, covariance analysis on the post-test math score was conducted. The homogeneity of regression coefficients of pretest math score on post-test math score was not significant, $F(1, 73) = 3.25$, $p > .05$, partial $\eta^2 = 0.04$, indicating the slopes are homogeneous. Results showed that the covariate effect of pre-test score of math was significant, $F(1, 74) = 117.05$, $p < .001$, partial $\eta^2 = 0.61$. The main effect of the group was significant, $F(1, 74) = 4.02$, $p = .049$, partial $\eta^2 = 0.05$. The post-test of math score of the intervention group was higher than that of the control group, indicating that math performance was significantly improved after intervention for adolescent girls in the intervention group.

Following the same procedure, another covariance analysis on the follow-up test of math score was conducted with the pre-test scores of math as the covariate variable. The homogeneity of regression coefficients of pretest math score on follow-up math score was not significant, $F(1, 73) = 1.17$, $p = .28$, partial $\eta^2 = 0.03$, indicating the slopes are homogeneous. Results showed that the covariate effect was significant, $F(1, 74) = 143.77$, $p < .01$, partial $\eta^2 = 0.66$. The main effect of the group was not significant, $F(1, 74) = .56$, $p > .05$,

partial $\eta^2 = 0.01$. The follow-up math score of the intervention group and that of the control group were not significantly different, indicating that the math performance did not sustain the higher level three months later after the end of intervention.

Language-Gender Stereotype

First, participants' language-gender stereotypes favoring girls in the first place was examined. The mean score of pre-test language-gender stereotypes was $M = 37.44$, $SD = 5.63$, one sample t test showed that $t(76) = 58.37$, $p < .001$, indicating that participants held language-gender stereotypes prior to intervention. Then, covariance analysis on the post-test and follow-up test of language-gender stereotypes with pre-test of language-gender stereotypes as the covariate was conducted. The homogeneity of regression coefficients of pretest language-gender stereotypes on post-test language-gender stereotypes was not significant, $F(1, 73) = 1.28$, $p = .26$, partial $\eta^2 = 0.02$, indicating the slopes are homogeneous. Results showed hat the covariate effects of the pre-test score of language-gender stereotypes on post-test score was significant, $F(1, 74) = 10.33$, $p < .01$, partial $\eta^2 = 0.12$; the main effect of the group was not significant, $F(1, 74) = .03$, $p = .86$, partial $\eta^2 < 0.001$.

What's more, the covariate effect of the pre-test score of language-gender stereotypes on follow-up test score was conducted. The homogeneity of regression coefficients of pretest language-gender stereotypes on follow-up language-gender stereotypes was not significant, $F(1, 73) = 1.07$, $p = .30$, partial $\eta^2 = 0.01$, indicating the slopes are homogeneous. The covariate effects of the pre-test score of language-gender stereotypes on follow-up test score was significant, $F(1, 74) = 10.57$, $p < .01$, partial $\eta^2 = 0.12$; the main effect of the group was not significant, $F(1, 74) = .00$, $p > .05$, partial $\eta^2 < 0.001$. The differences of post-test and follow-up test language-gender stereotypes between the intervention group and the control group were not significant, indicating that the intervention sessions did not have a significant influence on the language-gender stereotypes.

Discussion

Math-gender stereotypes, acquired as early as nine years old (Cvencek et al. 2011) and consolidated during middle school age (Muzzatti and Agnoli 2007), have been potential threat to females' development. Identity threat model of stigma demonstrated that intervention on math-gender stereotypes should target at collective representations, situational cues and personal characteristics (Major and O'Brien 2005). First, consistent with hypothesis 1, the level of math-gender

stereotypes in the intervention group was significantly lower than that of control group immediately after the intervention and at follow-up test, indicating that the intervention of adolescent girls' math-gender stereotypes targeted at multiple aspects of beliefs and behavior was effective. That is, the multiple intervention activities including encouraging the self-affirmation, external attribution training, changing the implicit attitude, establishing long-term justice goal, emphasis on intelligence growth, providing role models, and teaching students to think in different ways were effective for the reduction of math-gender stereotypes among adolescent girls.

The intervention strategy of encouraging self-affirmation based on self-affirmation theory (Steele 1988) emphasizes that the primary source of human motivation is to maintain self-integrity and self-worth. This intervention strategy encourages adolescent girls to learn defensive methods to protect their self-worth (Martens et al. 2006). The external attribution training demonstrates that negative stereotypes can affect females' attribution of failure. When females are in negative conditions, they tend to attribute failure internally to their ability; while when they are in positive conditions, they would attribute their failure externally, which in turn affect their performance (Koch et al. 2008). The intervention of changing implicit attitudes has used IAT test to retrain girls to associate their gender with being good at math (Forbes and Schmader 2010). In the present intervention program, girls were asked to imagine their image being a female scientist through drawing to change their implicit attitudes. The intervention of establishing long-term justice goal stresses that gender gap between male and female mathematics have become much smaller and guides students to view mathematic gender gap in more positive perspective. The intervention of emphasis on intelligence growth argues that the way students think about intelligence may have powerful effect on their achievement and encourages students to see intelligence as a malleable capacity. Students who hold an entity view of intelligence tend to pursue "performance goals" and become disengaged when task becomes challenging; while those who hold an incremental view of intelligence tend to pursue "learning goals" and experience less anxiety, take more effort and increase engagement when the task becomes challenging (Aronson et al. 2002). The intervention of providing role models was adopted because when adolescent girls were encouraged to think about role models of their stereotyped group, their performance deficits under stereotype threat can be alleviated (Marx and Roman 2002; McIntyre et al. 2003). The intervention of teaching students to think in different ways stresses altering stereotypes by

broadening the representation of people who do math-related work, math domain itself and the environments stereotype occurs (Cheryan et al. 2015). These intervention strategies predicted students' less vulnerable to stereotype threat and more psychological engagement in math tests.

Second, the intervention of math-gender stereotypes didn't improve adolescent girls' self-esteem score over time. The post-test self-esteem of the intervention group was marginally higher than that of the control group; while no significant difference on self-esteem between the intervention group and the control group was found in follow-up test. This might be explained from following reasons. On the one hand, the intervention period is relatively too short to improve self-esteem because global self-esteem is relatively stable in a period. On the other hand, global self-esteem is a complex psychological variable influenced by many factors such as relationships with parents, parenting style, and relationships with friends. Although stable during elementary school years, self-esteem tends to decrease with the transition into junior high (Wigfield and Eccles 1994). This might also partially explain why self-esteem was not significantly improved after intervention. Therefore, future studies may use state self-esteem or math-related domain specific self-esteem as an indicator of math-gender stereotype intervention effect.

Third, the post-test math score of the intervention group was higher than that of the control group, however, the follow-up math score of the intervention group and that of the control group were not significantly different, indicating that the math performance did not sustain the higher level three months after the end of intervention. Stereotype threat has been put forward as one explanation that caused math performance gap between boys and girls. When girls are stereotyped as inferior in mathematic achievement, this social identity raises high level of anxiety during mathematic test, resulting in lower scores (Osborne 2001; Schmader 2002; Schmader et al. 2004). Considering that stereotype threat can impair math performance, the intervention on negative math-gender stereotypes is a possible way to improve adolescent girls' math performance (Good et al. 2003). However, the present study didn't support the long-lasting intervention effect on math performance. This can be explained from two aspects. From one aspect, girls tend to perform similarly or outperform boys during early adolescence in various cultures (Hyde et al. 2008; Lai 2010). This makes the improvement of math score via decreasing math-gender stereotypes difficult. From the other aspect, the present study used classroom math exam rather than laboratory math test, which may result in inconsistent conclusions with laboratory interventions. Previous studies have

indicated that classroom grades measure mastery of material explicitly taught in school and teachers are likely to take into account nonacademic factors. This can provide a different measure of mathematics performance when compared to test scores (Friedman and Frisbie 1995; Ganley et al. 2013). Though math score was not significantly improved, the intervention may have influence on math-related attitudes, motivation and future development in math-related areas (Franceschini et al. 2014).

Fourth, language-gender stereotypes was used as a discriminant criterion of effectiveness of intervention strategies because language-gender stereotypes represent a different stereotype from the target domain. Generally, females endorse more negative math-gender stereotype and more positive language-gender stereotypes (Franceschini et al. 2014; Steffens et al. 2010; Steffens and Jelenec 2011). Along with earlier work (Marsh et al. 1988; Walker and Bridgeman 2008), results indicated that the levels of language-gender stereotypes were not influenced by the intervention. In other words, the intervention program can reduce math-gender stereotypes but not necessarily language-gender stereotypes. The results of this study refined the effectiveness of the intervention program, and at the same time indicated that different types of stereotypes are independent from each other.

Limitations and Future Directions

Three obvious limitations with respect to the intervention method and research design used in this study need to be noted. First, the intervention period was not long enough. Being limited by the school teaching schedule, this research developed nine sessions that are not enough when compared to the math-gender stereotypes existing since primary school (Cvencek et al. 2011; Neuburger et al. 2012). Many other strategies should be integrated into the intervention activities, for example, activating positive group identity to redefine where vulnerable groups belong to (Hess et al. 2003). Future studies should extend the intervention time in order to improve the effectiveness of the intervention program.

Second, the current study cannot disentangle the individual effects of collective representation, personal traits, and situational cues targeted various intervention strategies on adolescent girls' math-gender stereotypes. It was difficult to ascertain the individual contribution of each type of strategy designed from the three different factors. Future work may try to develop three different intervention programs based on the three factors to disentangle or compare the individual effects on the variance of academic gender stereotypes.

Third, girls' math-gender stereotypes might be influenced by outside resources such as parents' and teachers' expectations for children's math competence (Gunderson et al. 2012). However, the present intervention program didn't take these factors into consideration. On the one hand, future studies should examine the intervention effect by controlling the possible influence of these factors; on the other hand, future interventions should stress the significant roles of parents and teachers in shaping adolescents' math-gender stereotypes.

Practical Implications

The present study included nine intervention sessions to systematically reduce math-gender stereotypes from the perspective of changing the cognitive processing of stereotype information. Specifically, we used Identity Threat Model as a solid basis for intervention program by intervening on collective representations, situational cues and personal characteristics. To ensure the ecological validity, the intervention sessions were conducted in a real school environment; the intervention strategies could be easily adapted and catered to the actual demands of the school. With these advantages in mind, the intervention program can reduce math-gender stereotypes effectively and the effectiveness could be sustained for a reasonable period of time (three months at least). Of course, the present classroom intervention is not limited to school environment and in fact could be adapted to a number of other contexts, because these activities were easy to learn and operate. Future work is necessary to assess the duration of the intervention's effectiveness at reducing adolescent girls' math-gender stereotypes. More comprehensive interventions can be designed for teachers to intervene with students' negative stereotypes.

Conclusion

In sum, the interventions based on Identity Threat Model are topic-centered activities to help students realize that their gender group has much more possibilities for better development. It is a valid activity to buffer or reduce math-gender stereotypes among junior high school girls in China.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that they have no conflict of interest.

Appendix

Table 2 Details of intervention program

Intervention sessions	Sessions content
Session 1: Greeting	Meet-and-greet; warm-up activities; introduction of the intervention program; dividing students into small groups (each group has a leader and members).
Session 2: Encouraging the self-affirmation	Story
	Youngster Dumas. By sharing the story of Dumas, the purpose was to let the group members understand that everyone has advantages over others, encourage students to rebuild more positive images about themselves so as to reduce negative math-gender stereotype.
	Activities
	Myself Analogy: Group members were instructed to describe their strengths and advantages using plants, animals, or any other things.
	Please Leave a Message on My Back: Every group member was instructed to paste a message card on their own back, and other members would in turn write down their merits on the message card quietly. Then group members would read the cards on their backs and share their feeling about the messages their partners left.
	Homework
	Today, I Am Really Cool: Record behavior that deserves praise in a small notebook every day.
Session 3: Training external attribution	Role play
	A Math Final Exam: Math teacher of Class A thinks boys' math grades are better than girls', but math teacher of class B thinks both girls and boys are good at math. The different attitude of two math teachers leads to Linlin's (girl in Class A) worse math performance in final exam than Fangfang's (girl in Class B), though the two girls' math abilities are actually the same.
	Make up a story
	After a Math Final Exam: first, share the following story with students: Xiaojing (a girl's name) thinks her math learning ability is poor because she is a girl. Xiaona (a girl's name) thinks that the anxiety resulted from her fear of performing worse than nearby boys influences her performance on the exam.
	Then students were divided into two groups, and asked to make up a story to describe what happened to Xiaojing and Xiaona after the math final exam. They were given cue words including "mood, learning attitude, learning process, and learning results". The purpose of making up a story is to teach students about external attribution.
	Summary
	Mathematics grades are influenced by many factors, and both girls and boys can learn math well.
Session 4: Changing the implicit attitude	Warm-up
	The rain variations: Let students use group members' body to perform rain variation including light rain, moderate rain, and heavy rain.
	Activities
	Lists of imagination: First, Group members were asked to imagine that they were female scientists, female engineers, or female doctors. Then they were asked to show their imagination through drawings, essays or playlets. The purpose was to guide students to change their attitudes toward females in STEM fields.
	Award ceremony
	Awarded groups in which members performed well in *Lists of imagination* to praise their active participation.
Session 5: Establishing long-term justice goal	Discussion
	The discussion part is to let group members think and discuss what justice is.
	Listening
	Group leaders were arranged to introduce the history of women's liberation and the women of *knowledge economy time* to group members. The purpose of sharing and listening the histories is to let students realize that the gap across genders today is much smaller than that of past times.
	Discussion
	Group members were instructed to discuss whether gender-equality can really encourage girls to strive for success in their school performance and life.

Table 2 (continued)

Intervention sessions	Sessions content
Session 6: Emphasis on intelligence growth	**Role play**
	"I can" and "I can't": A member plays the role of "I can", always trying to overcome difficulties and believing that she will make progress. Another member plays the role of "I can't", always losing confidence and thinking others are smarter than her. Then students in each group discuss what the differences are between the two roles and predict what different results they will get in the future.
	Activities
	Memory Challenge: Give members 1 min to remember a picture which includes the figures of a series of poker cards. And then participants are asked to answer questions about pictures. For example, how many cards are in the picture? How many "four spades" are in picture? The level of difficulty in remembering pictures are lifting by increasing the numbers of the cards inside the picture. There are 5 pictures which represent 5 difficulty levels in all.
	Share
	Share feelings and thoughts with partner. For example, how is your feeling when the pictures become harder and harder? How do you overcome difficulties? What enlightenment did this process bring to you?
	Summary
	Intelligence grows regardless of gender.
Session 7: Listing role models	**Brain storm**
	Group members were instructed to write down as many names of scientists the group members know as possible in one minute, count how many male scientists and how many female scientists were listed. Then they were instructed to discuss why such few female scientists' names were listed.
	Praise a great
	Group leader were arranged to introduce the glorious achievements of female scientists (Rosalind Franklin, Dorothycroft Hodgkin, Zehui He, Donghua Cui). The purpose was to guide students to realize that female scientists should be role models for girls.
	Discussion topics
	How do female scientists gain glorious achievements?
	Discover female role models who are good at science in your (the participants') lives.
Session 8: Focusing on the comparison of difference thinking	**Warm-up**
	Two very different pictures are shown for group members to compare and find out the differences.
	Brain storm
	A series of topics were used to train students to compare different thinking angles and to learn to think in different ways. Thinking in different ways can benefit the reduction of math-gender stereotypes.
	Topics are listed as below:
	1. The use of the paper clip. Group members were divided into several groups and and asked to think of as many uses of the paper clip as possible.
	2. Numeration competition: Group members were asked to come up with as many different solutions as possible for a math problem.
	3. The divergent thinking in life: Ask participants to come up with as many solutions to daily problems as possible (e.g. How to eat a fish?)
Session 9: Farewell	Participants were asked to share their feelings about participating in the sessions and what they have learned.
	Then students sent blessing messages to each other.
	Teachers' summary: Girls can study well and gain achievement in math domain, as long as we strive for it using the right strategies.
	Chorus: *I believe* (The name of a song).

References

Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology, 46*(1), 166–171. doi:10.1016/j.jesp.2009.09.014.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385–390. doi:10.1111/1467-9280.00371.

Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: stereotype threat prior to test taking. *Educational Psychology Review, 24*(4), 609–635. doi:10.1007/s10648-012-9200-4.

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*(2), 113–125. doi:10.1006/jesp.2001.1491.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256–276. doi:10.1037/0096-3445.136.2.256.

Beilock, S. L., Gunderson, E. A., Gerardo, R., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860–1863. doi:10.1073/pnas.0910967107.

Bieg, M., Goetz, T., Wolter, I., & Hall, N. C. (2015). Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Frontiers in Psychology, 6*(1404). doi:10.3389/fpsyg.2015.01404.

Brinkman, B. G., Jedinak, A., Rosen, L. A., & Zimmerman, T. S. (2011). Teaching children fairness: decreasing gender prejudice among children. *Analyses of Social Issues & Public Policy, 11*(1), 61–81. doi:10.1111/j.1530-2415.2010.01222.x.

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology, 6*(49). doi:10.3389/fpsyg.2015.00049.

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

Daniels, E., & Leaper, C. (2006). A longitudinal investigation of sport participation, peer acceptance, and self-esteem among adolescent girls and boys. *Sex Roles, 55*(11–12), 875–880. doi:10.1007/s11199-006-9138-4.

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: a meta-analysis. *Learning and Individual Differences, 47*, 103–116. doi:10.1016/j.lindif.2015.12.018.

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.

Franceschini, G., Galli, S., Chiesi, F., & Primi, C. (2014). Implicit gender–math stereotype and women's susceptibility to stereotype threat and stereotype lift. *Learning and Individual Differences, 32*, 273–277. doi:10.1016/j.lindif.2014.03.020.

Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55*(1), 5–26. doi:10.1177/0013164495055001001.

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

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology, 49*(10), 1886–1897. doi:10.1037/a0031412.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645–662. doi:10.1016/j.appdev.2003.09.002.

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*(1), 17–28. doi:10.1016/j.appdev.2007.10.004.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3–25. doi:10.1037//0033-295X.109.1.3.

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

Guimond, S., & Roussel, L. (2001). Bragging about one's school grades: gender stereotyping and students' perception of their abilities in science, mathematics, and language. *Social Psychology of Education, 4*(4), 275–293. doi:10.1023/A:1011332704215.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*(3–4), 153–166. doi:10.1007/s11199-011-9996-2.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. New York, NY: Guilford Press.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *Journal of Gerontology, 58*(1), 3–11. doi:10.1093/geronb/58.1.P3.

Heyder, A., & Kessels, U. (2013). Is school feminine? Implicit gender stereotyping of school as a predictor of academic achievement. *Sex Roles, 69*(11–12), 605–617. doi:10.1007/s11199-013-0309-9.

Huitema, B. (2011). *The analysis of covariance and alternatives: statistical methods for experiments, quasi-experiments, and single-case studies* (2nd edition). New Jersey: John Wiley & Sons.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. 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*(3), 175–179. doi:10.1111/j.0956-7976.2005.00799.x.

Kay, A. C., & Jost, J. T. (2003). *Complementary justice: effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive* (Vol. 85, pp. 823–837). US: American Psychological Association.

Koch, S. C., Müller, S. M., & Sieverding, M. (2008). Women and computers. Effects of stereotype threat on attribution of failure. *Computers & Education, 51*(4), 1795–1803. doi:10.1016/j.compedu.2008.05.007.

Lai, F. (2010). Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review, 29*(3), 383–399. doi:10.1016/j.econedurev.2009.07.009.

Lamb, L. M., Bigler, R. S., Liben, L. S., & Green, V. A. (2009). Teaching children to confront peers' sexist remarks: implications for theories of gender development and educational practice. *Sex Roles, 61*(5), 361–382. doi:10.1007/s11199-009-9634-4.

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles, 66*(3–4), 220–234. doi:10.1007/s11199-011-0036-z.

Lindberg, S., Linkersdörfer, J., Ehm, J., Hasselhorn, M., & Lonnemann, J. (2013). Gender differences in children's math self-concept in the first years of elementary school. *Journal of Education and Learning, 2*(3), 1–8. doi:10.5539/jel.v2n3p1.

Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2009). The perils of double consciousness: the role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology, 45*(2), 299–312. doi:10.1016/j.jesp.2008.07.016.

Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. *Annual Review of Psychology, 56*(6), 393–421. doi:10.1146/annurev.psych.56.091103.070137.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: preadolescence to early adulthood. *Journal of Educational Psychology, 81*(3), 417–430. doi:10.1037/0022-0663.81.3.417.

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*(3), 366–380. doi:10.1037/0022-0663.80.3.366.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236–243. doi:10.1016/j.jesp.2005.04.010.

Martinot, D., & Désert, M. (2007). Awareness of a gender stereotype, personal beliefs and self-perceptions regarding math ability: when boys do not surpass girls. *Social Psychology of Education, 10*(4), 455–471. doi:10.1007/s11218-007-9028-9.

Marx, D. M., & Roman, J. S. (2002). Female role models: protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*(9), 1183–1193. doi:10.1177/01461672022812004.

McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*(1), 83–90. doi:10.1016/S0022-1031(02)00513-9.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: evidence from 66 nations. *Journal of Educational Psychology, 107*(3), 631–644. doi:10.1037/edu0000005.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*(3), 747–759. doi:10.1037/0012-1649.43.3.747.

Neuburger, S., Jansen, P., Heil, M., Quaiser-Pohl, C., & Neuburger, S. (2012). A threat in the classroom: gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift für Psychologie, 22*(2), 61–69. doi:10.1027/2151-2604/a000097.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44–59. doi:10.1037//0022-3514.83.1.44.

Osborne, J. W. (2001). Testing stereotype threat: does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26*(3), 291–310. doi:10.1006/ceps.2000.1052.

Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70–76. doi:10.1016/j.lindif.2014.05.005.

Ramsey, L. R., Betz, D. E., & Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Social Psychology of Education, 16*(3), 377–397. doi:10.1007/s11218-013-9218-6.

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

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. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194–201. doi:10.1006/jesp.2001.1500.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: the role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*(11–12), 835–850. doi:10.1023/B:SERS.0000029101.74557.a0.

Shin, J. E. L., Levy, S. R., & London, B. (2016). Effects of role model exposure on STEM and non- STEM student engagement. *Journal of Applied Social Psychology, 46*(7), 410–427. doi:10.1111/jasp.12371.

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): an investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles, 67*(11–12), 617–629. doi:10.1007/s11199-012-0209-4.

Smeding, A., Dumas, F., Loose, F., & Régner, I. (2013). Order of administration of math and verbal tests: an ecological intervention to reduce stereotype threat on girls' math performance. *Journal of Educational Psychology, 105*(3), 850–860. doi:10.1037/a0032094.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28. doi:10.1006/jesp.1998.1373.

Steele, C. M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261–302. doi:10.1016/S0065-2601(08)60229-4.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: the psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379–440. doi:10.1016/S0065-2601(02)80009-0.

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles, 64*(5), 324–335. doi:10.1007/s11199-010-9924-x.

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

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 100*(2), 255–270. doi:10.1037/a0021385.

Swinton, A. D., Kurtz-Costes, B., Rowley, S. J., & Okeke-Adeyanju, N. (2011). A longitudinal examination of African American adolescents' attributions about achievement outcomes. *Child Development, 82*(5), 1486–1500. doi:10.1111/j.1467-8624.2011.01623.x.

Tine, M., & Gotlieb, R. (2013). Gender-, race-, and income-based stereotype threat: the effects of multiple stigmatized aspects of identity on math performance and working memory function. *Social Psychology of Education, 16*(3), 353–376. doi:10.1007/s11218-013-9224-8.

Walker, M. E., & Bridgeman, B. (2008). Stereotype threat spillover and SAT® scores. *Ets Research Report, 2008*(1), i–10. doi:10.1002/j.2333-8504.2008.tb02114.x.

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem change across elementary and middle school. *The Journal of Early Adolescence, 14*(2), 107–138. doi:10.1177/027243169401400203.

Wright, & Daniel, B. (2006). Comparing groups in a before-after design: when t test and ANCOVA produce different results. *British Journal of Educational Psychology, 76*(3), 663–675. doi:10.1348/000709905X52210.

Zhang, Y., & Tsang, M. (2015). Gender gap in the National College Entrance Exam performance in China: a case study of a typical Chinese municipality. *Asia Pacific Education Review, 16*(1), 27–36. doi:10.1007/s12564-014-9351-8.

Zhang, B., Yuan, F., & Xu, L. (2014). Reducing the effects of stereotype threat: intervention strategies and future directions. *Journal of Psychological Science, 37*(1), 197–204.