



## OPEN ACCESS

## EDITED BY

Marisa Matias,  
University of Porto, Portugal

## REVIEWED BY

Miranda May McIntyre,  
California State University, San Bernardino,  
United States  
Sara Isabel Magalhães,  
University of Porto, Portugal

## \*CORRESPONDENCE

Carla Sanchis-Segura  
✉ csanchis@uji.es

†These authors have contributed equally to this work

RECEIVED 08 May 2023

ACCEPTED 25 July 2023

PUBLISHED 09 August 2023

## CITATION

Sebastián-Tirado A, Félix-Esbri S, Forn C and Sanchis-Segura C (2023) Are gender-science stereotypes barriers for women in science, technology, engineering, and mathematics? Exploring when, how, and to whom in an experimentally-controlled setting. *Front. Psychol.* 14:1219012. doi: 10.3389/fpsyg.2023.1219012

## COPYRIGHT

© 2023 Sebastián-Tirado, Félix-Esbri, Forn and Sanchis-Segura. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Are gender-science stereotypes barriers for women in science, technology, engineering, and mathematics? Exploring when, how, and to whom in an experimentally-controlled setting

Alba Sebastián-Tirado, Sonia Félix-Esbri, Cristina Forn<sup>†</sup> and Carla Sanchis-Segura<sup>✉\*</sup>

Departament de Psicologia Bàsica, Clínica i Psicobiologia, Universitat Jaume I, Castelló de la Plana, Spain

Based on Social Cognitive Career Theory principles, the present study sought to investigate whether stereotype threat experiences could act as a barrier and reduce the persistence of women in math-intensive activities. More specifically, we assessed whether the experimental activation of stereotypes about women's lower math capabilities affected the performance, persistence, and self-selected difficulty of engineering students in a math task which required sustained effort. We also evaluated the relationships between these effects and the participants' pre-testing gender-science stereotypes and math self-concept. A sample of 340 engineering students (175 and 165 self-identified as males and females, respectively) were recruited and randomly assigned to a control (Con) or stereotype threat (StA) condition to form four similarly sized groups. All participants rated their self-concept in four academic domains (math, chemistry, physics, and coding), completed the gender-science Implicit Association Test, and a "reading comprehension task" that served to promote the stereotype threat manipulation immediately before facing a modified version of the Math Effort Task (M-MET). We observed that, in the control condition, M-MET performance, self-selected difficulty, and persistence were similar in female and male participants, independent of their gender-science implicit stereotypes but correlated to their math self-concept scores. In contrast, the StA condition triggered opposite effects in female and male students, so they showed decreased/enhanced M-MET performance and self-selected difficulty, respectively. This experimental condition also resulted in enhanced persistence of the male students without affecting the number of trials completed by female students. These effects were correlated with the strength of the participants' gender-science implicit stereotypes but not with their math self-concept scores. In fact, as revealed by finer-grain analyses, stereotype threat only had a significant impact on individuals harboring stereotypical gender-science implicit associations. Therefore, it is concluded that: (1) stereotypes can promote differences between male and female engineering students in their performance, self-confidence, and persistence in math-related activities; (2) These effects seem to be more prominent in individuals implicitly perceiving science as a masculine domain. The relevance of these findings to explain women's enhanced abandonment rates of math-intensive studies is discussed.

## KEYWORDS

gender stereotypes, stereotype threat, math, implicit association test, STEM persistence

## 1. Introduction

Despite the notable advances toward social equity between women and men in the last decades, there is a major gender-related asymmetry in certain areas of academic and professional specialization that results in an underrepresentation of women in Science, Technology, Engineering, and Mathematics (STEM) studies and professions. Thus, although in elementary, middle, and high school, girls and boys take math and science courses in roughly equal numbers, only around 20–25% of STEM students are women, a number that declines even further in subsequent educational stages and in the workplace (Hill et al., 2010; Manske, 2010; European Commission, 2016; OECD, 2017).

The underrepresentation of women in STEM studies severely increases their risk of exclusion and/or inequality in the labor market given that STEM related careers are expected to grow several times faster than the average rate of all occupations (National Science Board, 2010) and are among the best paid jobs (National Association of Colleges and Employers, 2009; European Commission, 2016). Not only is this a problem for women, but it is a waste of talent for the STEM field (Blickenstaff, 2006; European Commission, 2014; Noland et al., 2016) and also an economic cost for society as a whole. In fact, the supply of STEM talent is not expected to meet the market demand (Hewlett et al., 2008; Fayer et al., 2017), and it has been estimated that closing the gender gap in the STEM field would increase the EU GDP *per capita* by 0.7–0.9% in 2030 and by 2.2–3.0% in 2050 (Morais Maceira, 2017). Accordingly, the gender segregation that characterizes the STEM field at the educational and professional level is seen with increasing social and institutional concern (Hill et al., 2010; European Commission, 2016).

### 1.1. Explaining the underrepresentation of women in STEM studies/careers: a social cognitive career theory perspective

Social Cognitive Career Theory (SCCT; Lent et al., 1994, 2000) has probably been one of the most widely used theoretical frameworks when attempting to explain the reduced entrance, enhanced attrition, and high drop-out rate of women in STEM studies (Rottinghaus et al., 2018). Following Bandura's general social cognitive theory (Bandura, 1977, 2000), on which it is based, SCCT hypothesizes that ability beliefs and confidence to perform well in specific cognitive/academic domains, as well as the anticipation of positive consequences derived from engaging in these activities, play a central role in the development of career and academic interests.

Ability beliefs/self-perceptions are built upon past self- and vicarious experiences, emotional reactions, and social persuasion/other people's opinions (Bandura, 1977; Bandura et al., 1999), and they are formally referred to as *self-efficacy*. Self-efficacy is not a unitary or general construct akin to self-esteem or other similar traits, but a series of domain-specific sets of perceptions and beliefs that

operate -and can be measured- at different levels of specificity (Pajares, 1996; Lent and Brown, 2006; Parsons et al., 2009). In this regard, an individual can express high self-efficacy in a cognitive/academic domain but low in others, and within that single domain, the same individual can feel confident and competent to successfully perform some tasks but not others. Self-efficacy beliefs/perceptions are not static either, and their dynamism is directly related to their level of specificity. Thus, general self-efficacy measures quite stable mental statements or representations about the self-perceived competence in a whole cognitive/academic domain (e.g., math) that are often referred to as the “*overall confidence*” (Parsons et al., 2009) or the “*self-concept*” in that domain (Pajares and Miller, 1994; Pajares, 1996). Conversely, measures of self-efficacy that are task- or even item-specific are much more malleable and can be dynamically moderated by environmental cues, so individuals may express high self-efficacy when facing a particular task in some contexts or under some circumstances but less so in others (Pajares and Miller, 1994; Pajares, 1996).

Therefore, self-efficacy -or, to better say, some of its percepts and measures- integrate the effects of past experience but also those of concurrent contextual factors. Depending on whether they facilitate or impede career-related decisions and behaviors, SCCT classifies influential contextual cues as *supports* (also referred to as *affordances*) or as *barriers*, respectively (Lent et al., 1996, 2000). Supports and barriers can be documented and also perceived as aspects of the environment but, regardless of whether or not they are grounded in reality, both have a major impact on self-efficacy, performance, and goals selection (Lent et al., 1994, 1996; Albert and Luzzo, 1999). More specifically, SCCT proposes that supports and barriers directly predict self-efficacy and that, mainly through this construct, they also predict the choice of goals and actions (Lent et al., 2002, 2003, 2015; Brown et al., 2008).

According to SCCT, gender differences in math self-efficacy could provide a reasonable explanation to the reduced number of women entering engineering and other STEM studies. In this regard, research has shown that (1) several measures of math self-efficacy predict math performance and the development of initial STEM-related interests and career goals (e.g., Hackett and Betz, 1989; Multon et al., 1991; Brown et al., 2008; Schunk and Pajares, 2009; Byars-Winston et al., 2010; Skaalvik et al., 2015); (2) despite having similar math abilities and grades (Lindberg et al., 2010; Else-Quest et al., 2013; Voyer and Voyer, 2014), boys and men tend to express higher self-efficacy than girls and women (Schunk and Pajares, 2002; Else-Quest et al., 2013; Zander et al., 2020; Mejía-Rodríguez et al., 2021); (3) women choosing STEM majors report higher levels of math self-efficacy than those opting for other studies (Hackett, 1985; Buschor et al., 2014) and, at least in some cases, similar to those exhibited by male STEM students (e.g., Brainard and Carlin, 1998; Macphree et al., 2013; Whitcomb et al., 2020). Taken together, these findings suggest that high levels of math self-efficacy are a pre-requirement to enter STEM studies and also that, despite having similar math abilities, less girls than boys complete these studies during their pre-major education years and, therefore, are less likely to choose STEM careers.

Similarly, SCCT proposes that, by determining interests and satisfaction, experience-dependent changes in self-efficacy affect academic persistence (Lent et al., 2013, 2015). Thus, different academic experiences and their effects on self-efficacy could explain why women leave STEM studies/professions at higher rates than men (Chen, 2013; Isphording and Qendrai, 2019; Clark et al., 2021; González-Pérez et al., 2022). In this regard, several studies have shown that self-efficacy is strongly related to academic persistence (Robbins et al., 2004; Wright et al., 2013) and also that math and science self-efficacy specifically predict persistence in STEM studies (e.g., Brown et al., 2008; Ackerman et al., 2013; Perez et al., 2014; Lee et al., 2015; Lent et al., 2015).

However, self-efficacy, by itself, does not suffice to adequately explain initial academic choices and persistence in their pursuit. As pointed out by Brown and Lent (1996), even those individuals with high levels of career self-efficacy, and career-aligned interests and goals may still avoid entering or pursuing a particular career if they perceive unbeatable barriers to attaining their academic/professional goals. Therefore, barriers, supports, and self-efficacy are supposed to play similarly important roles in career-related choices (Lent et al., 1994, 2000; Brown et al., 2008) and in the SCCT-based attempts to explain the underrepresentation of women in STEM studies/professions. However, research in this area has been less conclusive. Although some barriers and supports have been identified, most of these contextual factors seem to be domain-unspecific and/or do not seem to differentially affect math self-efficacy in male and female students (Lent et al., 2000; Fouad et al., 2002, 2010; Lent and Brown, 2006; Byars-Winston and Fouad, 2008; Miller M. J. et al., 2015). Accordingly, calls for identifying other possible barriers and supports and for improving their measurement have been made, as those traditionally assessed are “typically too broad to offer much precision in predicting domain-specific criteria” (Lent and Brown, 2006, p. 30) and the self-report instruments employed to evaluate them are largely dependent on the students’ awareness, ability or willingness to disclose them (Fouad et al., 2010).

## 1.2. Stereotypes as barriers for women in STEM studies

Gender-related stereotypes have been long proposed as a major factor underlying the underrepresentation of women in STEM studies and professions and, from a SCCT perspective, they can be conceptualized as distal and proximal barriers to their careers in this field. Distal barriers are environmental influences (or their perceptions) that are instrumental in determining the learning experiences through which overall self-efficacy, interests, and outcome expectations for particular academic domains are progressively forged (Lent et al., 1994). On the other hand, proximal barriers are contemporary contextual influences (or their percepts) that, by affecting more specific forms of self-efficacy, can moderate the relationships between already established interests and goals as well as between goals and behaviors, especially under adverse conditions (Lent et al., 2001).

Stereotypes about women’s lower math abilities and about the STEM field itself lead to the socially shared perception of math and “hard sciences” as “male” domains (Nosek et al., 2010; Nosek and Smyth, 2011; Cheryan et al., 2015; Makarova and Herzog, 2015).

These stereotypical views are already held by six-year-old children (Cvencek et al., 2011; Miller D. I. et al., 2015; Master et al., 2021) but also by their parents and teachers, who also bear different expectations, attributions, and evaluations of girls’ and boys’ math abilities and achievements, and tend to grant them different opportunities and support in their math-related experiences (Eccles et al., 1990; Beilock et al., 2010; Gunderson et al., 2012). This results in a reduction of girls’ math self-concept (Eccles et al., 1990; Beilock et al., 2010; Eccles, 2011; Nosek and Smyth, 2011; Steffens and Jelenec, 2011; Garriott et al., 2017; Hand et al., 2017) and, thereby, to a reduced interest, motivation, and performance in math that ultimately preclude their chances of joining STEM studies (Lent et al., 1994; Nagy et al., 2007; Nosek and Smyth, 2011; Garriott et al., 2017; Wang and Degol, 2017; Luo et al., 2021). Conversely, being favored by stereotypes and a higher math self-concept, boys become more prone to join engineering and other math-intensive studies. Accordingly, empirical evidence shows that women entering STEM majors usually have a higher math self-concept (Hackett, 1985; Buschor et al., 2014) and weaker gender-science stereotypes (Nosek and Smyth, 2011; Smeding, 2012; Smyth and Nosek, 2015; Sanchis-Segura et al., 2018) than those entering non-STEM majors, whereas the opposite seems to be true for men.

However, gender-related stereotypes do not only act as gatekeepers that limit the access of many women to engineering and other STEM majors, they also hinder the achievement and persistence of those enrolled in them. These women face a highly-stereotyped environment in which they may constantly feel at risk of being stereotypically judged and their abilities and “belongingness” are permanently under question (Marra et al., 2009; Kim et al., 2018; Casad et al., 2019).

Fears about confirming stereotypes are generically called *stereotype threat* (Steele and Aronson, 1995; Steele, 1997). Experimental research has shown that, when explicit or implicitly threatened, stereotype-targeted individuals may exhibit reduced performance in stereotyped tasks and activities, including those related to math (Nguyen and Ryan, 2008; Picho et al., 2013; Flore and Wicherts, 2015). Stereotype threat can also reduce self-efficacy, interest, and sense of belonging in the stereotyped domain while increasing negative thoughts and anxiety (Lewis and Sekaquaptewa, 2016; Spencer et al., 2016). According to the stereotype threat theory (STT; Steele, 1997; Inzlicht and Schmader, 2011), stereotype threats can be triggered by a variety of environmental and interpersonal cues, most of which are frequently encountered in STEM academic contexts (e.g., numerical gender imbalance; for a review, see Murphy and Taylor, 2012). STT also posits that both the chances of experiencing a threat and its consequences are enhanced in individuals who are highly identified with the domain being tested, as is the case of math in female STEM students (Aronson et al., 1999; Pronin et al., 2004). Therefore, women enrolled in STEM studies are likely to experience diverse and repeated stereotype threat episodes that can potentially act as proximal contextual barriers and weaken their intentions of persisting in this field (Deemer et al., 2014; Cadaret et al., 2017; Lin and Deemer, 2021).

However, not all studies have found these stereotypes’ effects and it has been suggested that the consequences of stereotype threat can be dependent on interactions with personal factors (Stoet and Geary, 2012; Flore and Wicherts, 2015; Flore et al., 2018), hence affecting some but not all the members of the negatively stereotyped groups (e.g., women). Unfortunately, stereotype threat effects are

better known than their mediators and moderators, hence limiting the chances of theoretically well-grounded interventions (Pennington et al., 2016). In this regard, only two experimental studies have assessed the possible moderating role of internalized gender stereotypes in the effects of stereotype threat on math self-efficacy (Good et al., 2008; Franceschini et al., 2014), and none of them evaluated motivational/persistence outcomes. Similarly, we are not aware of any previous experimental study assessing whether stereotype threat affects the persistence of STEM students in math-related activities through self-efficacy changes, and currently supporting evidence comes from field studies in which stereotype threat was not directly evaluated but self-reported (Deemer et al., 2014; Lin and Deemer, 2021). These voids motivated the present study.

### 1.3. Overview of the present study

The present study is developed within the frameworks provided by SCCT and STT, and was specifically designed to provide experimental evidence about the recently proposed role of stereotype threat as a possible proximal contextual barrier to women enrolled in STEM studies (Deemer et al., 2014; Cadaret et al., 2017). Specifically, we aimed to examine how experimentally activated stereotypes affect math self-efficacy, performance, and persistence of female and male engineering students in an effortful math task. Moreover, we also tested whether these effects were moderated by the strength and direction of the implicit gender-science stereotypes held by the participants before facing this task.

To pursue this goal, female and male engineering students were randomly assigned to one of two experimental conditions (control vs. stereotype threat activation), and the performance of these four experimental groups in a modified version of the Math Effort Task (Engle-Friedman et al., 2003) was compared. This task (hereafter referred to as M-MET) allowed us to obtain overall M-MET scores, but also separate measures of their principal components (total number of trials completed, percent of correctly solved problems, and level of difficulty chosen in each trial). These four dependent variables were interpreted as indexes of task performance, persistence, arithmetic accuracy, and self-confidence, respectively. In this last regard, it is worth noting that in the present study, overall math self-efficacy perceptions were also assessed by self-report before testing, hence obtaining two measures related to this construct that differed in their level of specificity (“overall” vs. task-specific and referred to as “math self-concept” and “in-task self-confidence”) and that, according to previous proposals (Pajares and Miller, 1994; Pajares, 1996), are expected to be less/more predictive of math performance and persistence, respectively.

As already mentioned, a second major goal of the present study was to assess to what extent the effects of the experimentally promoted threat were dependent on the participants’ “pre-testing stereotypes.” We use this term to refer to the stereotypical gender-science associations harbored by the participants before M-MET testing. These associations were evaluated with the implicit association test (IAT; Greenwald et al., 1998; Nosek et al., 2002a,b) and, therefore, the obtained estimates were independent of the participants’ ability and willingness to report them. These associations were expected to moderate M-MET performance in the threat, but not in the control condition.

## 2. Materials and methods

The study was approved by the Ethics Standards Committees of the Universitat Jaume I.

### 2.1. Participants

A total of 340 undergraduate engineering students from the Universitat Jaume I (Castelló, Spain) and Universitat Politècnica de València (Valencia, Spain) who volunteered in response to an invitation email. The invitation email omitted any reference to gender or gender stereotypes and presented the current study as designed to “evaluate some competences and intellectual abilities required to succeed in engineering.”

The final recruited sample had an approximately similar proportion of self-reported males (175) and females (165). Female and male participants were randomly assigned to the control (Con) or the stereotype activation (StA) experimental conditions, resulting in four similarly sized groups that did not differ in age (Table 1).

All the participants signed informed consents, and their collaboration was rewarded with €20.

### 2.2. General procedure

The study was carried out across 18 different experimental sessions, each of them involving 15–20 participants with similar proportions of females and males. Two researchers greeted the participants to the laboratory and randomly assigned them to an individual desk equipped with a personal computer. After that, participants gave their informed consent and filled out a demographic data form. Finally, they sequentially completed the programmed tasks at the pace indicated by the leading researcher.

In order of presentation, the tasks and measures included in this study were:

- *Perceived competence/self-concept in specific academic domains:* Subjects were asked to rate themselves in four academic domains (math, physics, chemistry, and coding) by answering the question “How good you are in...?” on a 10-point visual scale. This kind of question has been ordinarily used to assess overall self-efficacy beliefs (see Wigfield and Eccles, 2000) and, in the present study, it was included to obtain a measure of each participant’s math

TABLE 1 Demographic characteristics of the four participants’ groups.

	N	Median age	IQR age
M-Con	87	20	3
M-StA	88	20.5	3
F-Con	82	20	2.75
F-StA	83	20	2.5

The table reports the size of each experimental group and its median age. As reported in the main text, a two-way (gender x experimental condition) quantile-based ANOVA comparing the age medians revealed that the groups did not differ in age. IQR, Interquartile range; M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition.

self-concept (that is, his/her overall confidence in mathematics; see Parsons et al., 2009, 2011). Self-concepts in other academic domains were not of primary interest to the present study and their measurement was included to conceal the goals of the present study from the participants.

- *Implicit association test (IAT)*: The Implicit Association Test (Greenwald et al., 1998) is built upon the observation that responding to information perceived as congruent is faster than responding to information perceived as incongruent. More specifically, the IAT measures the strength of the association between concepts and attributes by calculating the reaction time difference between stereotypical non-stereotypical concept-attribute pairs (see below), and it has been widely used to assess several implicit stereotypic associations (Greenwald et al., 1998; Nosek et al., 2002a,b). In this study, the well-validated “Gender-Science” IAT was used to measure the stereotypic “males-science” and “females-humanities” association (Nosek et al., 2002a,b; Smyth and Nosek, 2015). However, to avoid biasing the participants’ responses, this task was presented to the participants as measuring two cognitive abilities: “perceptual speed” and “concepts categorization.”

The “Gender-Science” IAT was implemented using a freely available script at the Millisecond Test Library,<sup>1</sup> which we had previously adapted and translated into Spanish (see Sanchis-Segura et al., 2018). This script automatically counterbalances the order presentation of the test blocks containing stereotype-consistent trials and the blocks containing stereotype inconsistent trials, and it also automatically calculates the so-called IAT *D*-scores (Greenwald et al., 2003). The *D*-scores quantify the strength of the implicit association and are standardized deviation scores that range from +2 to -2 and whose interpretation is similar to that of the Cohen’s *d* statistic. Following the general convention, the IAT protocol used herein was arranged to provide positive *D* values for stereotype-consistent associations (e.g., “science = male/humanities = female”) and negative *D* values for stereotype-inconsistent associations.

- *Distraction task*: The Effort Expenditure for Rewards Task (EEfRT; Treadway et al., 2009) is a long (23 min) task that requires participants to select between “easy” and “hard” trials based on the probability of gaining different amounts of points by repeatedly and quickly pressing the spacebar a small/larger number of times, respectively. In this way, the EEfRT provides a measure of effort-based decision-making processes (e.g., Treadway et al., 2012). In the present study, the EEfRT was used as a mere distractor to temporally and mentally disengage the task assessing pre-existing gender-related stereotypes (IAT) from the task aimed to reactivate them (“*Reading comprehension task*,” see below). The EEfRT was implemented using the freely available script at the Millisecond Test Library.<sup>2</sup> Data from 25 individuals were lost and those of the rest of the participants were not further used.

- *“Reading comprehension task”*: This task had two phases (reading and comprehension testing) and was used to inadvertently promote the activation of gender-related stereotypes through the reading of a text.

Thus, all participants were instructed to carefully read a newspaper article that was about to be displayed on their computer screens. Participants were also informed that later on they would have to answer seven “yes/ no” questions designed to evaluate their understanding of the article’s contents. Participants could freely navigate through the two-page document for 10 min (reading phase). However, the material presented to the participants of the CON and StA groups differed. Thus, the text presented to the participants of the CON group was a real press article about the use of smartphones in primary school activities that had been published on the webpage of a national newspaper. Conversely, the text presented to the StA group had exactly the same layout as the original article, but its contents (text and illustrations) were replaced by others containing made-up psychological and neuroimaging “findings” supposedly proving that men (and their brains) are more “mathematical” and less “verbal” whereas women (and their brains) are more “verbal” and less “mathematical.” Once the pre-established reading time had elapsed, participants were instructed to start the testing phase and answer the seven upcoming questions by pressing the “b” key (masked with a green tag) or the “n” key (masked with a red tag) to provide “yes” or “no” answers, respectively. To avoid including non-complying participants, an exclusion criterion (i.e., providing two or more incorrect answers in the reading comprehension test) was set beforehand. However, no subjects needed to be actually excluded for this reason. The scores obtained in this task were not used further.

- *Modified math effort task (M-MET)*: The math effort task (MET) was originally developed by Engle-Friedman et al. (2003), and consists of 100 independent addition problems conducted after five practice trials (one per difficulty level, see below) at which performance feedback (“correct”/“false”) is provided. Each problem consists of 4 numbers that are flashed onto the screen one by one for 800 milliseconds and participants are asked to add the numbers in their heads. In each trial, the numbers for each problem are selected randomly but differ according to the participants’ self-selected level of difficulty, so the presented numbers are in the range of 0–2 (Level 1), 2–8 (Level 2), 6–13 (Level 3), 10–25 (Level 4), and 12–35 (Level 5). Once the last number is erased, a textbox is provided to collect the participants’ responses in the following 15 s, and the next trial starts without offering any performance feedback information.

In the present study, the MET was implemented through the freely available script of the Millisecond test library.<sup>3</sup> However, we modified the task in order to allow participants to choose the difficulty of each problem (instead of each 5 consecutive trials) and to have the opportunity to voluntarily abandon the task after finishing each trial. Thus, after these modifications, this task provided us with 4 different dependent variables:

1 <https://www.millisecond.com/download/library/iat/genderscienceiat>

2 <https://www.millisecond.com/download/library/effortexpenditureforrewardtask>

effortexpenditureforrewardtask

3 <https://www.millisecond.com/download/library/mathefforttask>

1. *M-MET global scores*: As in the original MET, M-MET global scores summarize the individuals' performance in this task and are calculated through the formula:  $\text{SCORE} = \Sigma[(\text{correct solutions to level 5 problems} * 1) + (\text{correct solutions to level 4 problems} * 0.8) + (\text{correct solutions to level 3 problems} * 0.6) + (\text{correct solutions to level 2 problems} * 0.4) + (\text{correct solutions to level 1 problems} * 0.2)]$ .
2. *Number of completed trials*: Unlike the original MET, this variable was under the participants' control in our task and provided an individual measure of their motivation or willingness to persist in a math task that requires sustained effort. Note that this measure is calculated in a manner that is not affected by the participants' chosen difficulty and their mental calculation abilities, hence isolating this motivational aspect.
3. *Chosen difficulty*: As mentioned in the introduction section, chosen difficulty was interpreted as a measure of in-task self-confidence. This was first assessed at the group level by comparing the relative frequency of trials completed for each selected difficulty level in each group. Of higher interest, the average difficulty level chosen by each participant (individual average chosen difficulty; I-ACD) was also calculated and these individual scores were employed to conduct more detailed group-based comparisons but also correlational analyses. Note that because they are calculated as averages or relative frequencies across all completed trials, the values of these two measures are not dependent on the number of trials completed or the participants' arithmetic accuracy.
4. *Arithmetic accuracy*: Accuracy, operationalized as the percent of correctly solved problems (over the number of problems attempted) was used as a measure of sustained arithmetic performance.

## 2.3. Statistical analyses

All statistical analyses were conducted in R (R Core Team, 2019). Statistical analyses focused on description and effect size estimation, as much as on testing statistical significance (Wasserstein and Lazar, 2016), and employed robust, non-parametric methods that simultaneously compared several location measures of the groups' distributions without making any pre-assumption about their shape or variance (Wilcox, 2016; Rousselet et al., 2017). In all analyses involving multiple comparisons, value of *ps* were adjusted with the Hochberg method and only those adjusted values are reported.

Thus, in the present study, between-group comparisons were conducted with robust equivalents of classic two-way (gender x experimental condition) ANOVAs comparing the quartile values (Q25, Q50, and Q75) instead of just the groups' means. The omnibus tests of these quantile-ANOVAs (qANOVAs) were obtained with the *q2by2* function and, when appropriate, the *Qmcp* function was employed to implement dyadic post-hoc comparisons at the quantile(s) of interest (see Wilcox, 2016; Rousselet et al., 2017). These two functions use the robust Harrell-Davis quantile estimator and a percentile bootstrap approach (2000 iterations) for effects and value of *ps* estimation and, when multiple comparisons are performed, control the probability of Type I errors with the Hochberg's method. In these comparisons, the unstandardized difference between the

estimated quantile values (denoted as  $\hat{d}$ ) and its corresponding 95% confidence intervals (calculated by the same percentile bootstrap method previously referred to) were employed to estimate the size of the observed effects.

Moreover, complementary comparisons involving the entire groups' distributions were conducted using the two-samples Kolmogorov-Smirnoff (K-S) and the Cliff's delta tests (Wilcox, 2016). The D statistic of the K-S test and the probability of superiority (PS) were employed as measures of effect size in these comparisons. The K-S D informs of the absolute value of the largest difference between the empirical cumulative distribution functions of the compared groups, whereas the PS is a probabilistic effect size index that denotes the probability that a randomly sampled member of group A will have a higher score than the score attained by a randomly sampled member of group B (Grissom and Kim, 2012).

On the other hand, some of the dependent variables considered in the present study allowed or required additional or alternative analytical approaches. More specifically:

- Given that the number of trials in the M-MET is under the participants' control but limited to a maximum value of 100, it is a time-to-event, right-censored variable that requires a specific statistical treatment (Lemeshow et al., 2011). Thus, to analyze possible group differences in the number of trials completed before abandoning or completing the M-MET, survival curves were calculated using the Kaplan-Meier's method. These curves were compared with the Long-Rank test followed by pairwise comparisons. All these analyses were conducted using different functions provided by the *survival* package (Therneau, 2022).
- Between group differences in chosen difficulty were assessed using the K-S test and the qANOVA methods previously described on the I-ACD scores, but also through contingency tables relating the frequency at which each difficulty level was chosen by the members of each group. The association between these two categorical variables was assessed with the Pearson's chi-squared test and the Pearson's residuals were inspected to identify which cells contributed the most to the overall test result (Sharpe, 2019). These analyses were conducted with the functions provided by the *vcd* package (Zeileis et al., 2007).
- Group differences in arithmetic performance were first assessed by applying the K-S test and the qANOVA methods previously described to the arithmetic accuracy scores. Given that accuracy scores are independent of the number of completed trials but not of the participants' chosen difficulty (i.e., arithmetic accuracy is expected to be higher in easier than in more difficult problems), additional between-group comparisons were conducted with a robust ANCOVA-like method. More specifically, the *ancmg1* function (see Wilcox, 2016) was employed to compare the groups' accuracy medians after removing the effects of chosen difficulty by treating I-ACD scores as a covariate of no interest.

In addition to these comparisons between groups, transversal analyses across genders and analyses assessing individual differences within each gender category were conducted.

Regarding the former, IAT D-scores were first transformed into the "IAT-influence" scores (see Sanchis-Segura et al., 2018) and then the relationships between M-MET performance (global and subcomponent scores), math self-concept, and the obtained

“IAT-influence” scores of all (male and female) participants subjected to each experimental condition were jointly assessed with correlational methods. This transformation: (1) is needed when jointly analyzing IAT D-scores of female and male participants because the same score has different implications -and it is expected to have opposite consequences in task performance- for the members of each gender; (2) just involves multiplying males’ IAT scores by 1 and those of females by  $-1$  and, therefore, does not change the strength of the IAT-measured association nor those of its possible correlations with other variables; and, (3) enriches the interpretation of these correlations, as they quantify the strength and sign of the expected impact (“influence”) of the IAT-measured association over other variables and not just their co-variation. These IAT “influence” scores were first employed in robust regression-based moderation analyses, which were carried out using the *lmrob* function of the *robustbase* package for R and that were aimed to assess if -as hypothesized- the participants’ pre-testing stereotypes did only affect M-MET performance in the participants subjected to the StA condition. Given the results obtained in these analyses, between variables’ relationships were separately explored in each experimental condition using two complementary strategies. First, the strength of the zero-order association between each pair of variables was quantified using Spearman’s rho correlation index. *The values of the obtained correlations were compared using the twopcor and twohc4cor functions (Wilcox, 2016).* Second, LASSO-regularized partial correlation networks were built to describe the whole pattern of direct associations between these variables and unbiasedly estimate their strength (Epskamp et al., 2012; Epskamp and Fried, 2018). More specifically, separate partial Spearman correlation networks for each experimental condition were first estimated with the *estimateNetwork* function (default = “EBICglasso,”  $\gamma = 0.5$ ) of the *bootnet* package (Epskamp et al., 2018) and subsequently compared with the *NCT* function of the *NetworkComparisonTest* package (Van Borkulo et al., 2022). To avoid the distortions derived from conditioning on a common effect (Pearl, 2000; Epskamp and Fried, 2018), overall M-MET scores (which are arithmetic composites of the number of completed trials, accuracy, and I-ACD scores) were not included in the construction of these networks.

On the other hand, we hypothesized that, if the opposite effects of stereotype threat on the M-MET performance of males and females are influenced by pre-existing implicit gender-science associations, changes in M-MET performance should be solely observed to be more prominent in those individuals of each gender holding stereotypical gender-science implicit associations (i.e., IAT D-scores  $> 0$ ). Conversely, stereotype threat effects should probably be minimized, absent, or even reversed in those members of the M-StA and F-StA groups lacking such stereotypes or harboring counter-stereotypic associations (i.e., IAT D-scores  $\leq 0$ ). To test this hypothesis, within-gender comparisons between participants receiving distinct (positive vs. negative) “influences” from their own implicit “gender-science” associations (i.e., IAT “influence” scores higher or lower than 0, respectively) were conducted. More specifically, the *qcomhd* function (Wilcox, 2016) was employed to compare each of these two subgroups of females and males exposed to the StA condition between them and also with their respective control groups regarding M-MET, I-ACD, and accuracy scores. To perform the same within-gender comparisons on the number of completed trials, the previously described Kaplan–Meier’s method was employed.

## 3. Results

Table 1 summarizes the composition of the experimental groups recruited for the present study. A two-way (gender x experimental condition) qANOVA comparing the age medians confirmed that participants’ age did not differ between groups (gender,  $p = 0.118$ ; experimental condition,  $p = 0.605$ ; gender x experimental condition,  $p = 0.967$ ).

### 3.1. Self-concept of males and females in different academic domains

A series of two-way (gender x experimental condition) qANOVAs comparing the participants’ overall self-efficacy beliefs yielded significant effects of gender in some, but not all, of the evaluated academic domains (Table 2). More specifically, males rated themselves higher than females in coding (Q25:  $p < 0.006$ ,  $\hat{d} = 1.31$  [0.38, 1.96]; Q50:  $p < 0.001$ ,  $\hat{d} = 1.32$  [0.51, 2.00]; Q75:  $p < 0.001$ ,  $\hat{d} = 1.54$  [0.77, 2.42]) and to a lesser extent in physics (Q25:  $p = 0.05$ ,  $\hat{d} = 0.81$  [−0.01, 1.77]; Q50:  $p = 0.027$ ,  $\hat{d} = 0.35$  [0.03, 1.06]; Q75:  $p = 0.144$ ,  $\hat{d} = 0.81$  [−0.02, 0.65]). Conversely, females and males did not significantly differ in their perceived self-concept in chemistry or math (see Table 2 for details).

When the entire distributions of these self-concept scores were compared through a series of Kolmogorov-Smirnoff tests, significant gender-based differences were also found in the participants’ self-concept in coding and physics (K-S  $D_{\text{coding}} = 0.25$ ,  $p < 0.001$ ; K-S  $D_{\text{physics}} = 0.16$ ,  $p = 0.027$ ) but not in chemistry or math (K-S  $D_{\text{chemistry}} = 0.07$ ,  $p = 0.678$ ; K-S  $D_{\text{math}} = 0.03$ ,  $p > 0.999$ ). Accordingly, the probability that a randomly picked male would judge himself as more competent than a randomly picked female in coding ( $PS_{\text{Males}} = 0.63$  [0.57, 0.68]) or physics ( $PS_{\text{Males}} = 0.58$  [0.52, 0.64]) was significantly higher than for the reverse comparisons (Cliff’s delta = 0.26 and 0.16,  $p < 0.001$  and  $p = 0.002$ , respectively).

Taken together, these results indicate that although males rated themselves higher than females in some academic domains, females and males perceived themselves as similarly competent in math. Moreover, it should be highlighted that math self-concept scores did not differ between the male nor the female participants randomly assigned to the control and StA conditions, hence ruling out any possible contribution of this variable to other possible differences between the M-Con/ M-StA and the F-Con/ F-StA groups.

### 3.2. “Gender-science” implicit association test

Figure 1 depicts the participants’ D-scores in the “Gender-Science” IAT. A two-way qANOVA yielded a statistically significant effect of the gender factor ( $p < 0.001$  in all three quartiles), but not of the experimental condition factor (Q25:  $p = 0.392$ ; Q50:  $p = 0.392$ ; Q75:  $p = 0.093$ ), nor of the between-factors’ interaction ( $p > 0.99$  in all three quartiles). Gender-based differences were similar in size at the three quartiles ( $\hat{d} = 0.50$  [0.34, 0.64],  $\hat{d} = 0.46$  [0.36, 0.55],  $\hat{d} = 0.49$  [0.38, 0.59], respectively).

In the same line, a Kolmogorov-Smirnoff test confirmed that the distribution of the “Gender-Science” IAT scores differed between

TABLE 2 Self-concept in different academic domains in the four participants' groups.

	GROUP	Q25	Q5	Q75
CODING	M-Con	5.22	7.22	8.90
	M-StA	4.52	6.49	8.03
	F-Con	3.22	5.74	7.13
	F-StA	3.87	5.46	6.92
		<b>Gender,</b> <i>p</i> = 0.006	<b>Gender,</b> <i>p</i> < 0.001	<b>Gender,</b> <i>p</i> < 0.001
		Condition, <i>p</i> = 0.812	Condition, <i>p</i> = 0.280	Condition, <i>p</i> = 0.280
		Interaction, <i>p</i> = 0.357	Interaction, <i>p</i> = 0.378	Interaction, <i>p</i> = 0.378
PHYSICS	M-Con	5.89	7.05	8.00
	M-StA	5.60	7.01	8.00
	F-Con	4.93	6.50	7.87
	F-StA	4.85	6.71	7.86
		Gender, <i>p</i> = 0.050	<b>Gender,</b> <i>p</i> < 0.027	Gender, <i>p</i> = 0.144
		Condition, <i>p</i> = 0.931	Condition, <i>p</i> = 0.931	Condition, <i>p</i> = 0.931
		Interaction, <i>p</i> = 0.946	Interaction, <i>p</i> = 0.946	Interaction, <i>p</i> = 0.946
CHEMISTRY	M-Con	3.33	5.09	7.09
	M-StA	3.47	5.40	7.41
	F-Con	4.27	6.23	7.89
	F-StA	3.27	5.09	7.43
		Gender, <i>p</i> = 0.351	Gender, <i>p</i> = 0.351	Gender, <i>p</i> = 0.351
		Condition, <i>p</i> = 0.720	Condition, <i>p</i> = 0.720	Condition, <i>p</i> = 0.810
		Interaction, <i>p</i> = 0.249	Interaction, <i>p</i> = 0.249	Interaction, <i>p</i> = 0.237
MATH	M-Con	6.19	7.17	8.34
	M-StA	6.01	7.33	8.32
	F-Con	6.12	7.67	8.66
	F-StA	6.01	7.22	8.34
		Gender, <i>p</i> = 0.945	Gender, <i>p</i> = 0.945	Gender, <i>p</i> = 0.945
		Condition, <i>p</i> = 0.691	Condition, <i>p</i> = 0.691	Condition, <i>p</i> = 0.691
		Interaction, <i>p</i> = 0.964	Interaction, <i>p</i> = 0.964	Interaction, <i>p</i> = 0.964

Displayed group scores correspond to the empirical quartile values. Statistical significance was tested using two-way (gender x condition) quantile-based ANOVAs and the reported value of *ps* correspond to the main factors and interactions of these tests. All value of *ps* were adjusted for multiple comparisons using the Hochberg method. Statistically significant effects are highlighted in bold. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition.

males and females ( $K-S D = 0.48, p < 0.001$ ). Thus, the probability that a randomly picked male would obtain a higher D-score in this IAT ( $PS_{Males} = 0.79 [0.74, 0.84]$ ) was substantially higher than for the

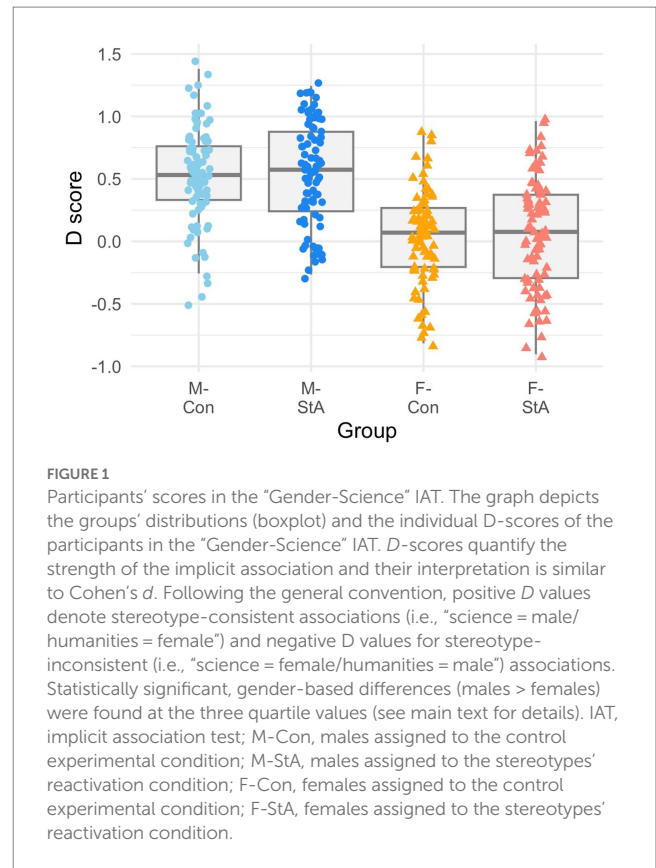


FIGURE 1 Participants' scores in the "Gender-Science" IAT. The graph depicts the groups' distributions (boxplot) and the individual D-scores of the participants in the "Gender-Science" IAT. D-scores quantify the strength of the implicit association and their interpretation is similar to Cohen's *d*. Following the general convention, positive *D* values denote stereotype-consistent associations (i.e., "science = male/humanities = female") and negative *D* values for stereotype-inconsistent (i.e., "science = female/humanities = male") associations. Statistically significant, gender-based differences (males > females) were found at the three quartile values (see main text for details). IAT, implicit association test; M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes' reactivation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotypes' reactivation condition.

reverse comparison ( $PS_{Males} = 0.21 [0.16, 0.26]$ ; Cliff delta = 0.59,  $p < 0.001$ ).

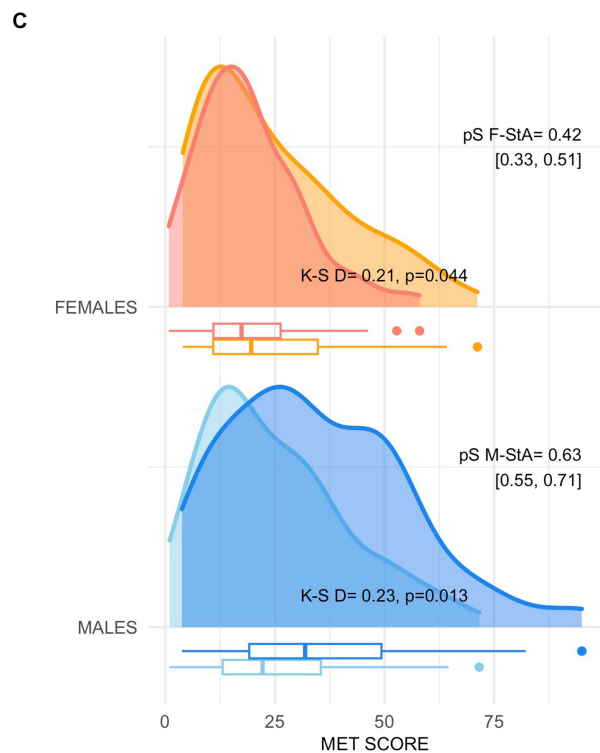
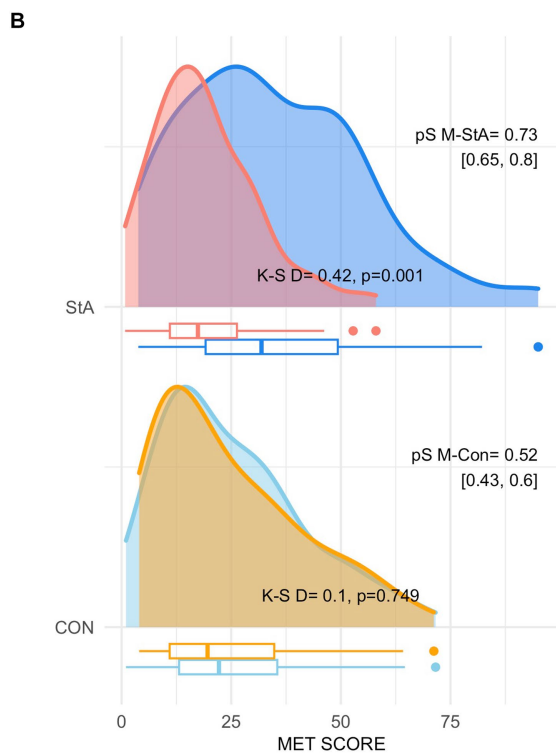
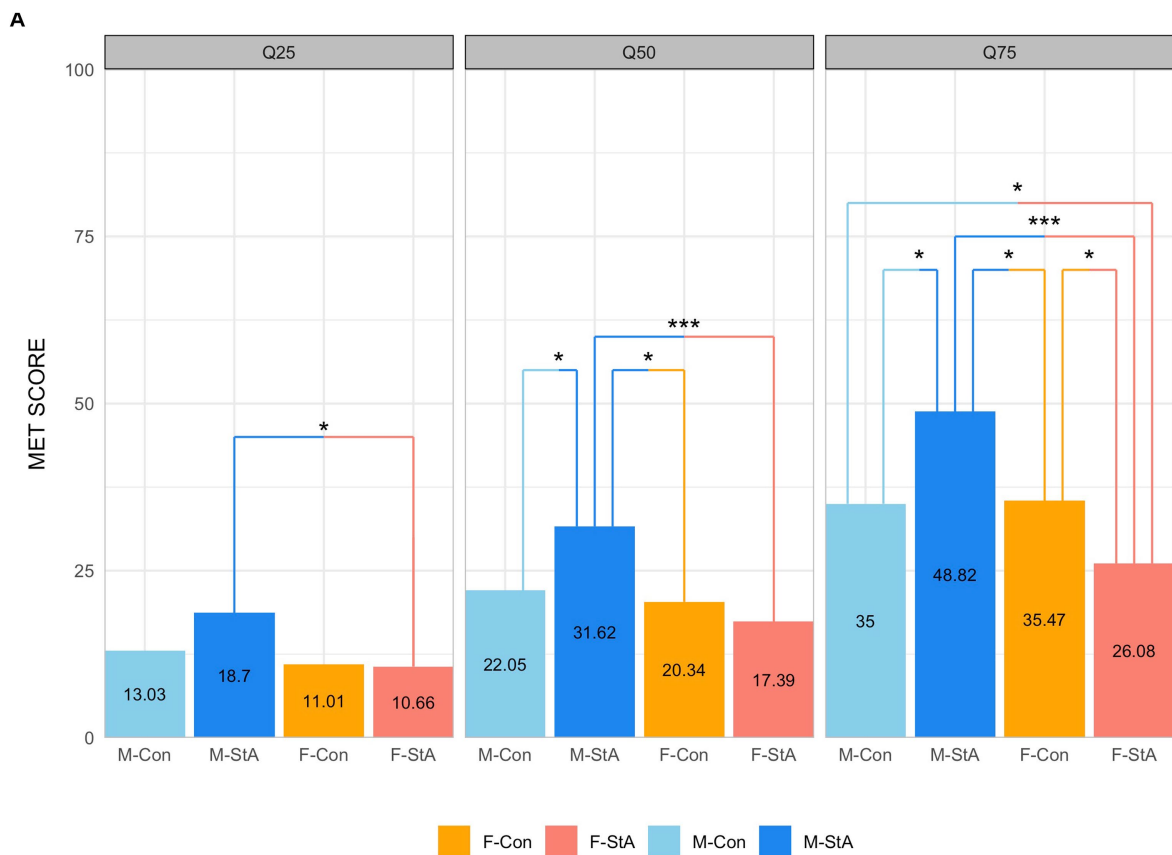
Therefore, as it could be expected from previous studies (e.g., Smyth and Nosek, 2015), we observed that male engineering students exhibit a stronger stereotypical "male-science/female-humanities" association than female engineering students. In fact, this association was harbored by the large majority of these male students (i.e., 87.43% of them had IAT D-scores >0) but only by around half (58.19%) of the female students. On the other hand, it should be noted that, as also observed for self-efficacy scores, neither the male nor the female participants randomly assigned to their respective Con and StA groups differed in their "Gender-Science" IAT. Therefore, any later difference between the M-Con/ M-StA and the F-Con/ F-StA groups should be attributed to the effects of stereotype activation and not to any pre-existing differences between these groups.

### 3.3. Modified math effort task (M-MET)

#### 3.3.1. Global M-MET scores

A two-way (gender x experimental condition) qANOVA comparing the participants' M-MET yielded statistically significant effects of the gender factor ( $p < 0.005$  in the three quartile values) and of their interactions with the experimental condition factor (Q25:  $p = 0.043$ ; Q50:  $p = 0.012$ ; Q75:  $p < 0.001$ ). As illustrated in panel A of Figure 2, post-hoc dyadic comparisons revealed that: 1) At Q25, statistically significant differences were solely observed between the F-StA and the M-StA groups; 2) At Q50, the M-StA group exhibited larger M-MET scores than the F-StA but also than the M-Con and the F-Con groups, which were not significantly different between them nor





**FIGURE 2** Participants' scores in the overall M-MET (modified math effort task). Panel (A) shows the values of the overall M-MET scores of each group at each quartile (Q25, Q50, Q75), as well as the results of their comparisons (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  after correction for multiple comparisons with the Hochberg method). Panel (B) depicts the whole distributions of the M-MET scores of the groups subjected to stereotypes activation (top) and control condition (bottom). Panel (C) illustrates the same distributions but compares them based on the gender factor (females, top; males, bottom). M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes' reactivation condition; F-Con, females

(Continued)

FIGURE 2 (Continued)  
 assigned to the control experimental condition; F-StA, females assigned to the stereotypes' reactivation condition; Con, control condition; StA, stereotypes' reactivation condition; K-S, Kogolmorov-Smirnoff test; PS, probability of superiority.

TABLE 3 Between-group differences in overall M-MET scores.

Quantile	Comparison	<i>p</i> value	$\hat{d}$	95% CI
Q25	M-StA vs. M-Con	0.170	5.67	[1.40, 13.46]
	M-StA vs. F-Con	0.053	7.7	[-0.04, 15.27]
	<b>M-StA vs. F-StA</b>	<b>0.015</b>	<b>8.04</b>	<b>[1.12, 15.60]</b>
	F-StA vs. M-Con	0.228	-2.37	[-6.35, 2.52]
	F-StA vs. F-Con	0.814	-0.34	[-5.73, 4.08]
	M-Con vs. F-Con	0.679	2.03	[-3.81, 6.08]
Q50	<b>M-StA vs. M-Con</b>	<b>0.014</b>	<b>9.58</b>	<b>[0.823, 20.82]</b>
	<b>M-StA vs. F-Con</b>	<b>0.012</b>	<b>11.29</b>	<b>[1.80, 21.40]</b>
	<b>M-StA vs. F-StA</b>	<b>&lt;0.001</b>	<b>14.23</b>	<b>[7.32, 24.29]</b>
	F-StA vs. M-Con	0.270	-4.65	[-11.95, -2.11]
	F-StA vs. F-Con	0.428	-2.95	[-11.79, 3.26]
	M-Con vs. F-Con	0.652	1.71	[-8.46, 10.20]
Q75	<b>M-StA vs. M-Con</b>	<b>0.015</b>	<b>13.82</b>	<b>[1.69, 21.53]</b>
	<b>M-StA vs. F-Con</b>	<b>0.015</b>	<b>13.35</b>	<b>[1.15, 22.34]</b>
	<b>M-StA vs. F-StA</b>	<b>&lt;0.001</b>	<b>22.74</b>	<b>[13.37, 30.28]</b>
	<b>F-StA vs. M-Con</b>	<b>0.015</b>	<b>-8.92</b>	<b>[-19.55, -1.16]</b>
	<b>F-StA vs. F-Con</b>	<b>0.017</b>	<b>-9.39</b>	<b>[-20.50, -0.05]</b>
	M-Con vs. F-Con	0.932	-0.47	[-11.52, 11.17]

The table reports the outcome (*p*-value, quantile difference, and its corresponding 95% CI) of follow-up post-hoc comparisons conducted after the identification of statistically significant gender x group interactions in quantile-based ANOVAs comparing overall M-MET scores (see main text). Robust post-hoc comparisons were carried out with the *Qmcp* function, and those reaching statistical significance after multiple comparisons correction with the Hochberg's method are highlighted in bold. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition; CI, confidence interval.

the F-StA group; and 3) At Q75, the M-StA exhibited larger M-MET scores than all the other groups, the F-StA exhibited smaller M-MET scores than all other groups whereas the M-Con and F-Con did not significantly differ between each other. The details of these comparisons can be found in Table 3. These results suggest that the stereotype activation produced opposite effects in females and males, hence promoting gender-based differences in M-MET scores in the StA condition that were not observed in the control condition. Note that these effects seemed to be more prominent for males than for females and for higher (Q75) than for lower (Q25) M-MET scores (Table 3).

In agreement with these results, the M-MET score distributions of the M-Con and F-Con groups did not significantly differ between them ( $D=0.1$ ,  $p=0.749$ ; Figure 2A, bottom), but the M-MET score distributions of the M-StA and F-StA groups did ( $D=0.42$ ,  $p=0.001$ ; Figure 2A, top). Thus, the probability that a randomly selected individual of the M-StA group would obtain a higher M-MET score than that obtained by a randomly selected member of the F-StA group (PS M-StA=0.73 [0.65, 0.80]) was significantly higher (Cliff's delta=0.46,  $p<0.001$ ) than that for the reverse comparison (PS F-StA=0.27 [0.20, 0.35]), a difference that was not observed in the

control condition (Cliff's delta=0.03,  $p=0.710$ ). As illustrated by panel B of the same figure, gender-based differences in the StA condition seem to stem from two statistically significant but opposite sign effects of stereotype activation in females and males. Thus, compared to the distributions of their respective control groups, the M-MET score distribution of the F-StA and M-StA were significantly shifted toward lower ( $D=0.21$ ,  $p=0.044$ ) and toward higher ( $D=0.23$ ,  $p=0.013$ ) values, respectively. Note that, as indicated by the obtained Cliff's delta values, these effects appeared to be slightly more prominent in males (Cliff's delta=0.26,  $p=0.003$ ) than in females (Cliff's delta=-0.16,  $p=0.08$ ), an effect that seems to align with the observation that the M-StA, but not the F-StA, differed from their respective control group at intermediate levels of M-MET performance (Q50, Table 3).

Taken together, these results confirm that the activation of stereotypes about a gender-based difference in math competence boosted the performance of males and, to a lower extent, also decreased that of females in an effortful, math-related task.

### 3.3.2. How stereotype threat affects M-MET scores?

As described in the methods section, global M-MET scores depend on three separate components: The number of trials completed, the chosen difficulty level at each trial, and arithmetic accuracy (the proportion of correctly solved problems). Therefore, we investigated which of these components were affected by the induced stereotype threat and underlie the between-group differences of the M-MET scores described in Section 3.3.1.

#### 3.3.2.1. Does stereotype threat affect the number of trials completed?

A Survival analysis (Kaplan–Meier) was used to determine whether there were between-group differences regarding the proportion of individuals persisting in the M-MET task across trials. As it can be readily observed from the survival distribution curves depicted in Figure 3, the members of the M-StA persisted longer in the M-MET than the members of all other groups. The same conclusion was obtained when analyzing the median survival times, which reported estimated values of 86.5, 60, 58, and 54 for the M-StA, M-Con, F-Con, and F-StA groups, respectively. Confirming these observations, a Long-Rank test yielded a significant group effect ( $\chi^2(3)=12.1$ ,  $p=0.007$ ) and post-hoc comparisons only yielded statistically significant differences between the survival curve of the M-Sta group and those of all other groups (M-StA vs. M-Con,  $p=0.031$ ; M-StA vs. F-Con,  $p=0.011$ ; M-StA vs. F-StA,  $p<0.001$ ). Taken together, these results suggest that stereotype activation selectively increased the persistence on the M-MET task in males.

#### 3.3.2.2. Does stereotype threat affect the chosen trials' difficulty?

A Pearson's chi-square test ( $\chi^2(12)=237.1$ ,  $p<0.001$ ) showed that chosen difficulty level differed between groups and the analysis of the residuals (Figure 4) revealed that M-StA and F-StA groups were the

ones showing a larger deviation from the chance-expected values. More specifically, the individuals of the M-StA group rarely opted for the least difficult levels (L1 and L2) and chose the most difficult ones (especially, L5) more than could be expected. Members of the F-StA exhibited the exact opposite pattern of choices, thus opting for the easiest problems and avoiding the most difficult ones (L5). Deviations from the expected values were also observed in the M-Con and F-Con groups, but they were much smaller in size and almost exclusively affecting the most difficult level (L5). These results suggest that stereotype threat promoted opposite effects in females and males, making them select less/more difficult problems, respectively.

In the same line, a two-way qANOVA (gender  $\times$  experimental condition) comparing the I-ACD scores among groups yielded statistically significant effects of the gender factor ( $p < 0.001$ ) and of its interaction ( $p < 0.005$ ) with the experimental condition factor in the three quartile values. As illustrated in panel A of Figure 5, these global effects were a result of a pattern of between group differences (M-StA  $>$  M-Con = F-Con  $>$  F-StA) that was repeated, with a similar size across the three quartile values (see Table 4 for details).

These effects were further characterized by comparing the groups' I-ACD score distributions. These distributions did not differ between males and females in the control condition (K-S  $D = 0.16$ ,  $p = 0.179$ ; Figure 5B, bottom), but did in the StA condition (K-S  $D = 0.54$ ,  $p = 0.001$ ; Figure 5B, top). In this regard, the probability that a male would have a higher I-ACD score than a female in the control condition

( $PS_{M-Con} = 0.56$  [0.48, 0.65]) was equivalent (Cliff's  $\delta = 0.13$ ,  $p = 0.150$ ) to that of a female having a higher I-ACD score than a male ( $PS_{F-Con} = 0.44$  [0.35, 0.52]). However, after stereotype activation, there was a much higher probability that a randomly selected male would have a higher I-ACD score (i.e., consistently chose more difficult problems) than a randomly selected female ( $PS_{M-StA} = 0.81$  [0.73, 0.86];  $PS_{F-StA} = 0.19$  [0.14, 0.27]; Cliff's  $\delta = 0.61$ ,  $p < 0.001$ ). As panel C of Figure 5 illustrates, these gender-based differences emerged from two statistically significant, similarly sized, but of opposite sign effects of stereotype activation in females and males. Thus, compared to the distributions of their respective control groups, the I-ACD scores' distribution of the F-StA and M-StA were significantly shifted toward lower (K-S  $D = 0.26$ ,  $p = 0.006$ ;  $PS_{F-StA} = 0.37$  [0.29, 0.45],  $PS_{F-Con} = 0.63$  [0.55, 0.71], Cliff's  $\delta = -0.26$ ,  $p = 0.004$ ) and higher values (K-S  $D = 0.27$ ,  $p = 0.003$ ;  $PS_{M-StA} = 0.64$  [0.56, 0.72],  $PS_{M-Con} = 0.36$  [0.28, 0.44], Cliff's  $\delta = 0.29$ ,  $p = 0.001$ ), respectively. Taken together, the results obtained indicate that the activation of stereotypes about a gender-based difference in math ability made males choose more difficult math problems, whereas it promoted the opposite effect in females.

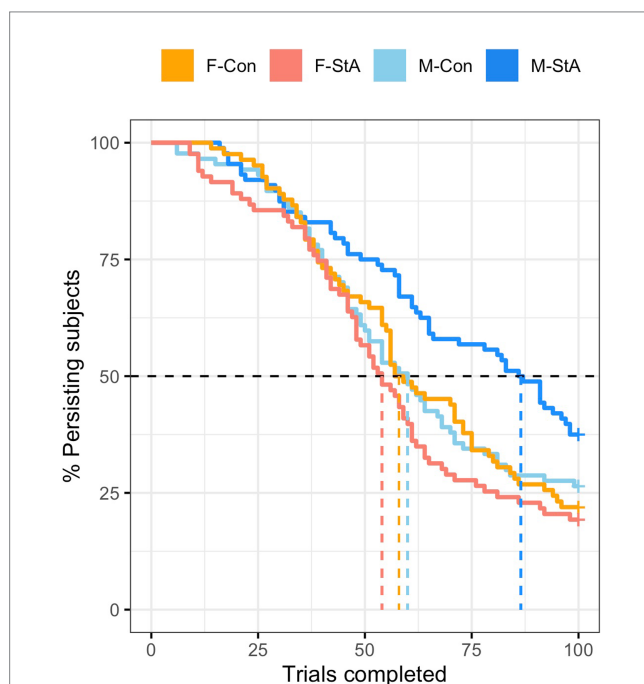
### 3.3.2.3. Does stereotype threat affect arithmetic accuracy?

Stereotypes' activation did not seem to affect arithmetic accuracy. Thus, a two-way qANOVA (gender  $\times$  experimental condition) comparing accuracy scores did not yield any statistically significant effect for any quartile (Table 5). No statistically significant differences were found when comparing the entire accuracy score distributions with Kolmogorov–Smirnov or Cliff's delta tests (Table 6). Finally, no statistically significant differences were observed when accuracy scores were compared while controlling the possible effects of chosen difficulty by introducing the I-ACD scores as a covariate of no interest (Table 7).

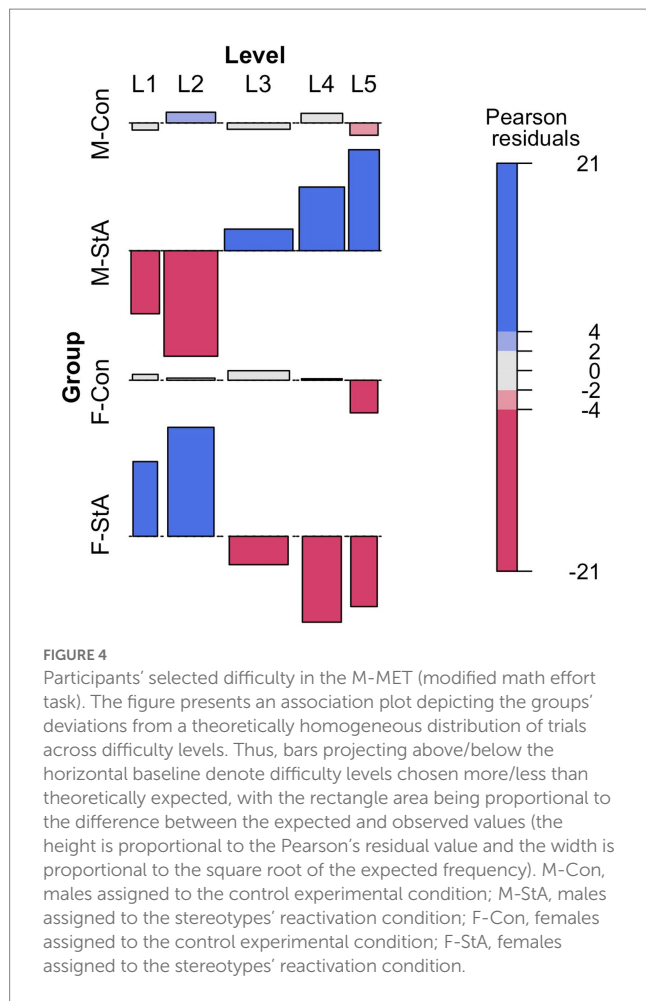
## 3.4. Relationship between math self-concept, the “influence” of pre-testing gender-related stereotypes, and M-MET performance

Robust moderation regression-based analyses revealed that the pre-testing science-gender stereotypes did only influence specific aspects of their M-MET performance in those individuals which pre-testing stereotypes were experimentally re-activated (StA condition). More specifically, it was observed that only the interaction between the imposed experimental condition and the IAT “influence” scores significantly predicted the overall MET scores (condition estimate = 1.17,  $p = 0.526$ , IAT estimate = 0.20,  $p = 0.939$ , interaction estimate = 14.50,  $p < 0.001$ ). Given these results, the relationships between the M-MET components, IAT-“influence” scores and math self-concept scores were further and separately explored for each experimental condition with correlational methods.

Panels A and B of Figure 6 depict the patterns of zero-order Spearman correlations between math self-concept, the IAT-“influence” scores, and the different measures obtained from the M-MET task in the control and the StA activation condition, respectively. As can be readily observed, the correlations between M-MET derived measures were very similar in both cases, whereas their correlations



**FIGURE 3**  
Participants' persistence in the M-MET (modified math effort task). This figure depicts the percentage of subjects of each group that persisted in the task across trials. The colored stepped lines correspond to the Kaplan–Meier estimates of the survival curves of each group, whereas the dashed vertical lines indicate the median survival time in each group. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes' reactivation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotypes' reactivation condition.



with the pre-testing math self-concept and IAT-“influence” scores clearly differed between experimental conditions. Thus, in the control condition (panel A), the overall M-MET, the I-ACD, and the accuracy scores were significantly correlated to math self-concept, but appeared to be unrelated to the IAT-“influence” scores. Conversely, in the StA condition (panel B), the number of completed trials as well as the overall M-MET and the I-ACD scores were significantly correlated with the IAT “influence,” but not with the math self-concept scores. Confirming and extending the results of our initial moderation analysis, the values of the correlations between the IAT “influence” scores and the M-MET scores (difference = 0.38 [0.21, 0.63],  $p < 0.001$ ), I-ACD scores (difference = 0.49 [0.37, 0.74],  $p < 0.001$ ), and the number of trials completed (difference = 0.28 [0.10, 0.54],  $p = 0.006$ ) were significantly larger in the StA than in the control condition.

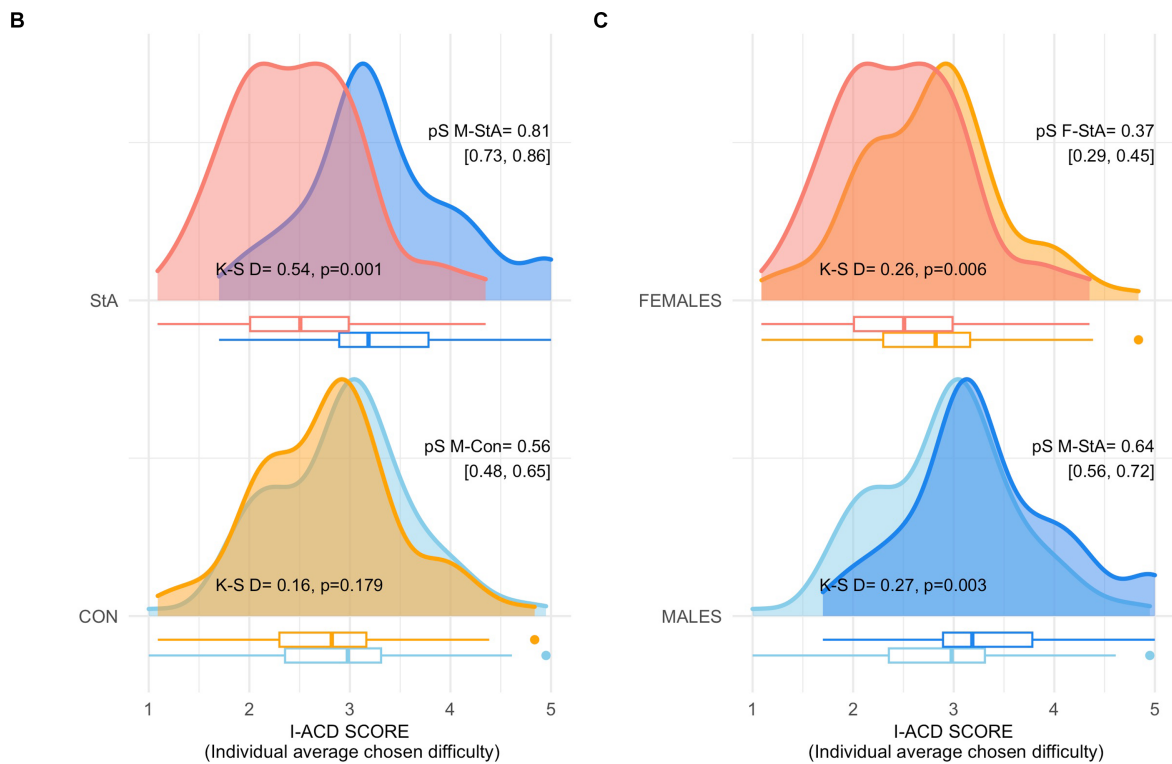
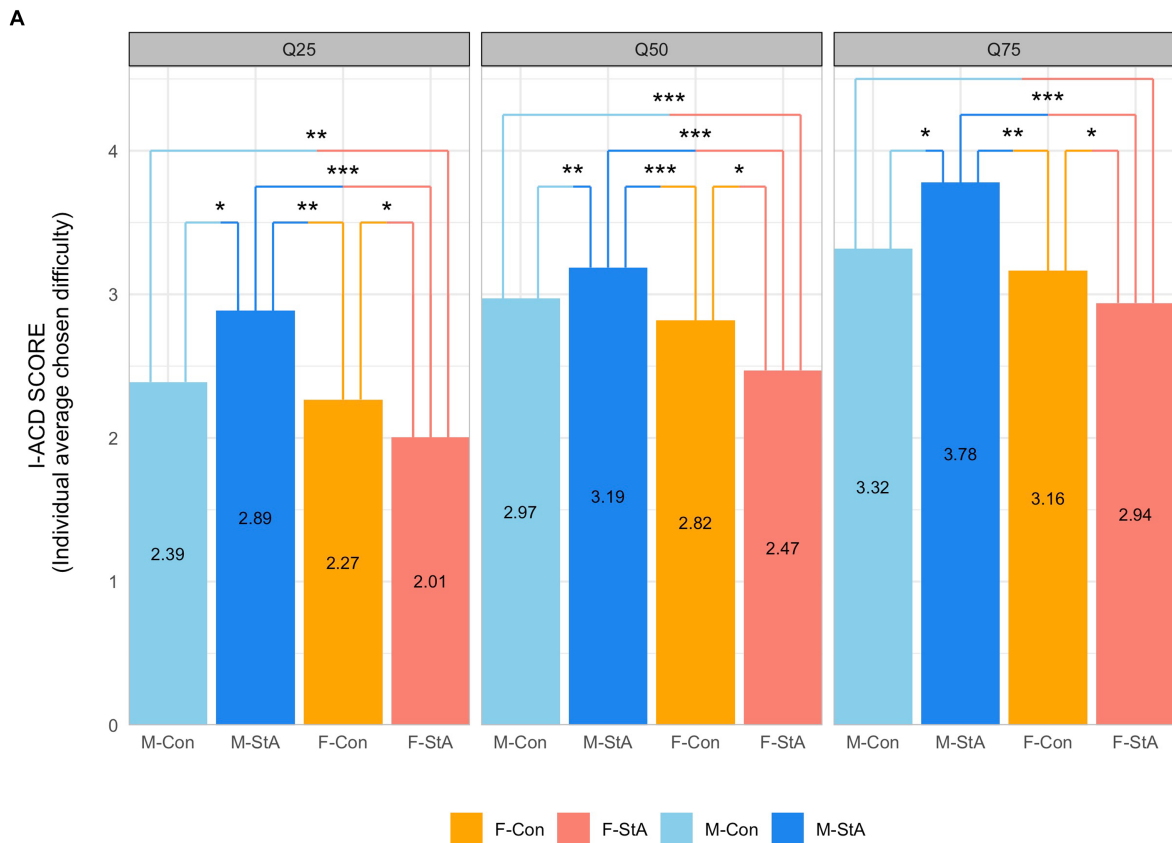
To assess the relationship between each possible pair of variables in each experimental condition while controlling the possible influence of all others and removing spurious correlations, LASSO-regularized partial correlation networks were estimated (Figures 6C,D). These networks exhibited a good fit to the data (control condition: CFI = 0.99, TLI = 0.97, RMSEA = 0.032; StA condition: CFI = 0.98, TLI = 0.96, RMSEA = 0.045) and confirmed that, although the relationships between the different M-MET components were very similar in both models, their regulation by pre-testing math self-concept and IAT-“influence” scores was very different in each experimental condition. More specifically, in the control condition

(panel C), math self-concept was the only relevant predictor, and was directly associated with accuracy and I-ACD scores but, through these two mediators, also indirectly related to the number of completed trials. By contrast, in the StA condition (panel D), IAT-“influence” scores (but not math self-concept scores) were directly associated with the number of completed trials as well as with I-ACD scores, and, through both of them, also with accuracy scores. These networks proved to be significantly different ( $M = 0.443$ ,  $p < 0.001$ ) and statistically significant differences were observed at the edges, linking IAT “influence” to I-ACD scores ( $p < 0.001$ ), the IAT “influence” scores to total trials ( $p = 0.022$ ), and math self-concept to I-ACD scores ( $p = 0.055$ ), but not in any of the edges linking the M-MET components between them ( $p > 0.2$  in all cases). Taken together, these results suggest that, under ordinary circumstances, high/low math self-concept is associated with high/low self-selected difficulty (I-ACD scores) and arithmetic accuracy that result in high/low M-MET scores, respectively. In contrast, after a forced exposure to gender-related stereotypes, M-MET performance seems to get largely disengaged from the participants' math self-concept, so their motivation to persist, self-selected difficulty, and performance in this task become dependent on the contents and strength of their pre-testing gender-science implicit associations.

In this regard, additional intra-group analyses confirmed that only those individuals of M-StA and F-StA groups harboring stereotypical associations linking “science” to “male” -and, therefore, experiencing positive and negative “influences” from these associations, respectively- showed I-ACD and M-MET (but not accuracy) scores that were significantly higher/lower than their respective control groups, respectively (Figure 7). The within-gender differences in M-MET scores appeared to be more prominent in males than in females, probably because the number of completed trials was only significantly affected in males (Figures 7B,F). On the other hand, it is worth noting that these within-gender differences cannot be explained by any pre-existing difference in math self-concept between these male/female subgroups (Table 8).

## 4. Discussion

The present study was designed to assess how experimentally activated stereotypes affect the performance and persistence of female and male engineering students in math tasks that require sustained effort (M-MET) and to which extent the effects of the induced threat are mediated by self-efficacy changes and moderated by the participants' pre-testing stereotypes. Under the control condition (Con) both genders were found to exhibit similar M-MET performances. These between-gender similarities were apparent when attending to their overall M-MET scores (medians: 22.05 vs. 20.34) as well as to each of its subcomponents (medians completed trials: 60 vs. 58, medians average difficulty: 2.97 vs. 2.82, and medians accuracy: 71 vs. 69.8%, respectively). The same level of similarity was observed for lower and higher levels of performance (Q25 and Q75) and also when considering the entire score distributions of each of these variables (Figures 2, 3, 5; Tables 3–7). Given that this is the first time that the M-MET has been used, these findings cannot be directly compared to those of any other study, although it is worth noting that they are similar to those reported for the original MET in which no gender-based differences were found (Engle-Friedman et al., 2003).



**FIGURE 5** Participants' scores in the I-ACD scores (individual average chosen difficulty) in the M-MET. Panel (A) shows the values of the I-ACD scores of each group at each quartile value (Q25, Q50, Q75), as well as the results of their comparisons (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  after correction for multiple comparisons with the Hochberg method). Panel (B) depicts the whole distributions of the I-ACD scores of the groups subjected to stereotypes' reactivation (top) and control condition (bottom). Panel (C) illustrates the same distributions but compares them based on the gender factor (females, top; males, bottom). M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes' reactivation condition; (Continued)

FIGURE 5 (Continued)

F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotypes' reactivation condition; Con, control condition; StA, stereotypes' reactivation condition; K-S, Kogolmorov-Smirnoff test; PS, probability of superiority.

TABLE 4 Between-group differences in I-ACD scores.

Quantile	Comparison	<i>p</i> value	$\hat{d}$	95% CI
Q25	M-StA vs. M-Con	<b>0.019</b>	<b>0.50</b>	<b>[0.03, 0.84]</b>
	M-StA vs. F-Con	<b>0.002</b>	<b>0.62</b>	<b>[0.17, 0.91]</b>
	M-StA vs. F-StA	<b>&lt;0.001</b>	<b>0.88</b>	<b>[0.50, 1.17]</b>
	F-StA vs. M-Con	<b>0.002</b>	<b>-0.38</b>	<b>[-0.84, -0.06]</b>
	F-StA vs. F-Con	<b>0.027</b>	<b>-0.26</b>	<b>[-0.62, -0.03]</b>
	M-Con vs. F-Con	0.414	0.12	[-0.29, 0.58]
Q50	M-StA vs. M-Con	<b>0.006</b>	<b>0.21</b>	<b>[0.04, 0.48]</b>
	M-StA vs. F-Con	<b>&lt;0.001</b>	<b>0.37</b>	<b>[0.17, 0.65]</b>
	M-StA vs. F-StA	<b>&lt;0.001</b>	<b>0.72</b>	<b>[0.47, 1.09]</b>
	F-StA vs. M-Con	<b>&lt;0.001</b>	<b>-0.50</b>	<b>[-0.81, -0.23]</b>
	F-StA vs. F-Con	<b>0.012</b>	<b>-0.35</b>	<b>[-0.67, -0.02]</b>
	M-Con vs. F-Con	0.066	0.15	[-0.08, 0.44]
Q75	M-StA vs. M-Con	<b>0.032</b>	<b>0.46</b>	<b>[0.01, 0.89]</b>
	M-StA vs. F-Con	<b>0.006</b>	<b>0.61</b>	<b>[0.14, 1.03]</b>
	M-StA vs. F-StA	<b>&lt;0.001</b>	<b>0.84</b>	<b>[0.41, 1.32]</b>
	F-StA vs. M-Con	<b>&lt;0.001</b>	<b>-0.38</b>	<b>[-0.75, -0.13]</b>
	F-StA vs. F-Con	<b>0.032</b>	<b>-0.22</b>	<b>[-0.59, -0.02]</b>
	M-Con vs. F-Con	0.199	0.15	[-0.18, 0.48]

The table reports the outcome (*p*-value, quantile difference, and its corresponding 95% CI) of dyadic post-hoc comparisons conducted after the identification of statistically significant gender x group interactions in quantile-based ANOVAs comparing I-ACD scores. Robust post-hoc comparisons were carried out with the *Qmcp* function, and those reaching statistical significance after multiple comparisons correction with the Hochberg's method are highlighted in bold. I-ACD, individual average chosen difficulty; M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition; CI, confidence interval.

By contrast, between-gender differences emerged when participants were exposed to a stereotype threat (StA condition). In this case, females showed lower overall M-MET scores than males (Figure 2; Table 3) and these differences were increasingly larger across quartiles (Q75 > Q50 > Q25; Table 3). Moreover, the M-StA group exhibited significantly higher M-MET scores than the M-Con and F-Con groups at intermediate (Q50) and high (Q75) levels of performance, whereas the M-MET scores of the F-StA were significantly lower than those of both control groups at Q75 (Figure 2; Table 3). Taken together these results suggest that stereotype activation promoted opposite effects in the M-MET performance of male and female students, promoting gender-based differences that were not observed in the control condition. These effects can be suitably accounted for by the principles of the Stereotype Threat Theory (STT). More specifically, the reduced M-MET performance observed in the F-StA group can be understood as the result of a stereotype threat, whereas the increased performance in the M-StA group could potentially arise from a stereotype lift, a stereotype boost, or a combination of both kinds of effects. In this regard, it has been repeatedly observed that pre-exposure to explicit or implicit

TABLE 5 Quartile-based comparisons for arithmetic accuracy scores.

GROUP	Q25	Q50	Q75
M-Con	60.2	71.0	79.3
M-StA	61.4	72.1	80.4
F-Con	60.4	69.8	78.3
F-StA	60.1	70.9	78.0
	Gender, <i>p</i> = 0.881	Gender, <i>p</i> = 0.881	Gender, <i>p</i> = 0.417
	Condition, <i>p</i> = 0.856	Condition, <i>p</i> = 0.856	Condition, <i>p</i> = 0.856
	Interaction, <i>p</i> = 0.970	Interaction, <i>p</i> = 0.970	Interaction, <i>p</i> = 0.970

The table reports the outcome (*p*-value, quantile difference, and its corresponding 95% CI) of dyadic post-hoc comparisons conducted after having identified statistically significant gender x group interactions in quantile-based ANOVAs comparing arithmetic accuracy scores. Robust post-hoc comparisons were carried out with the *Qmcp* function and those reaching statistical significance after multiple comparisons correction with the Hochberg's method are highlighted in bold. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition.

TABLE 6 Whole distribution between-groups comparisons for arithmetic accuracy scores.

	K-S D	<i>p</i> -value	Cliff's delta	<i>p</i> -value
M-StA vs. F-StA	0.11	0.653	0.07	0.470
M-Con vs. F-Con	0.08	0.931	0.04	0.643
M-Con vs. M-StA	0.10	0.686	0.04	0.620
F-Con vs. F-StA	0.06	0.983	0.02	0.831

The table summarizes the obtained statistics (D and delta), as well as the associated *p*-values for between-group comparisons on the whole distributions of accuracy scores conducted with the two-samples Kolmogorov-Smirnoff and the Cliff's delta tests. All *value of ps* are corrected for multiple comparisons with the Hochberg's method. K-S, two-samples Kolmogorov-Smirnoff test; M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotype activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotype activation condition.

suggestions of women having lower abilities than men in math as well as cues indicating a gender-biased evaluation are able to promote a threat to female students that reduces their performance in math-related tasks (e.g., Spencer et al., 1999; Adams et al., 2006; Good et al., 2008; Picho and Schmader, 2018; for meta-analyses, see Nguyen and Ryan, 2008; Picho et al., 2013; Flore and Wicherts, 2015; Doyle and Voyer, 2016). On the other hand, although far less studied, it has also been shown that conditions triggering a threat for women can simultaneously promote a stereotype lift effect in men, indirectly enhancing their math performance (e.g., Spencer et al., 1999; Eriksson and Lindholm, 2007; Good et al., 2008; Picho and Schmader, 2018; for

TABLE 7 Between-group median differences in accuracy scores after accounting for chosen difficulty variation.

Covariate value	Comparison	p-value	$\hat{d}$	95% CI
2	M-StA vs. M-Con	0.883	-1.63	[-11.90, -9.97]
	M-StA vs. F-Con	0.883	5.57	[-7.21, 15.27]
	M-StA vs. F-StA	0.883	5.20	[-6.89, 13.16]
	F-StA vs. M-Con	0.212	6.84	[-1.67, 12.78]
	F-StA vs. F-Con	0.883	0.37	[-8.49, 10.39]
	M-Con vs. F-Con	0.282	7.20	[-2.01, 15.94]
3	M-StA vs. M-Con	0.727	4.30	[-5.04, 15.69]
	M-StA vs. F-Con	0.727	2.18	[-6.20, 13.51]
	M-StA vs. F-StA	0.727	1.40	[-7.70, 16.51]
	F-StA vs. M-Con	0.727	-2.89	[-13.10, 11.47]
	F-StA vs. F-Con	0.727	-1.39	[-16.51, 7.70]
	M-Con vs. F-Con	0.727	-2.12	[-11.47, 9.31]
4	M-StA vs. M-Con	0.962	1.79	[-14.77, 11.43]
	M-StA vs. F-Con	0.962	-0.51	[-19.04, 13.91]
	M-StA vs. F-StA	0.962	4.91	[-14.39, 37.44]
	F-StA vs. M-Con	0.962	0.02	[-11.36, 35.71]
	F-StA vs. F-Con	0.962	-2.80	[-37.95, 12.01]
	M-Con vs. F-Con	0.962	-2.30	[-12.02, 37.95]
5	M-StA vs. M-Con	0.446	13.59	[-7.48, 30.19]
	M-StA vs. F-Con	0.446	14.18	[-7.58, 31.46]
	M-StA vs. F-StA	0.446	20.73	[-5.48, 33.72]
	F-StA vs. M-Con	0.738	-9.52	[-19.46, 0.40]
	F-StA vs. F-Con	0.738	-6.55	[-19.55, 6.46]
	M-Con vs. F-Con	0.738	0.59	[-7.28, 8.46]

The table displays the outcome (*p*-values, medians' difference, and its corresponding 95% CI) of between-group comparisons for median accuracy scores conducted by means of a robust ANCOVA that included I-ACD scores as covariates of no interest. The column "covariate value" indicates the points at which the test statistic was evaluated. The reported value of *ps* are corrected for multiple comparisons with the Hochberg's method. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotypes activation condition; CI, confidence interval.

a meta-analysis, see Walton and Cohen, 2003). The enhanced performance observed in the M-StA group could also be due to a stereotype boost directly elicited by their exposure to male-favoring stereotyped information, but this seems less likely because, opposite to the stereotype threat and lift phenomena, stereotype boost is less likely to occur when stereotypes are explicitly activated (Shih et al., 2012).

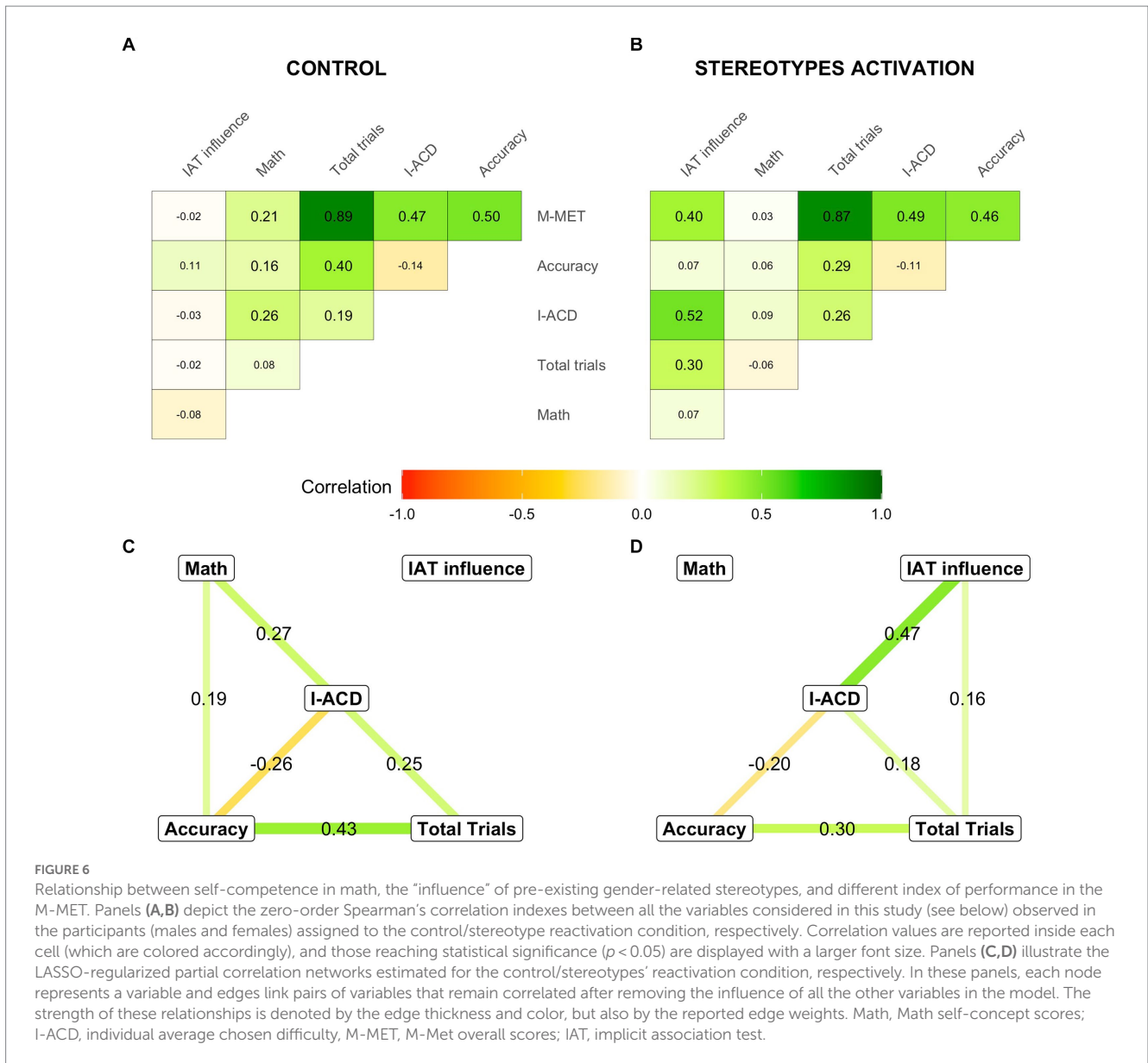
Even though numerous studies have illustrated the effects of stereotype threat (and related phenomena) on cognitive performance, not all studies have found these stereotypes' effects (Stoet and Geary, 2012; Flore and Wicherts, 2015; Flore et al., 2018) and far less is known about how they take place and who may be more prone to experience them (Pennington et al., 2016). Therefore, the present study also attempted to provide more specific information on these two questions.

Regarding the "how" question, the three separate subcomponents of M-MET scores (number of completed trials, chosen difficulty, and arithmetic accuracy) were separately analyzed, hence obtaining

information about the consequences of the experimentally imposed stereotype threat in each of the three main processes underlying M-MET performance (that were interpreted as indicators of persistence/motivation, self-confidence while performing the task, and arithmetic accuracy, respectively). Given that the M-StA group completed more trials (Figure 3) and chose more difficult problems than all other groups (I-ACD scores; Figures 4, 5; Table 4), it can be tentatively concluded that their enhanced M-MET performance stemmed from two complementary effects in motivation/persistence and in-task self-confidence. On the other hand, the F-StA group showed lower I-ACD scores than all the other groups without significantly differing in any other M-MET component, suggesting that the decreased M-MET performance of the F-StA group was primarily due to a reduction of in-task self-confidence during task performance (Figures 4, 5; Table 4). Taken together, these results may be interpreted as indicating that the effects of stereotype threat on M-MET performance were primarily -but not exclusively- conveyed through in-task self-confidence changes. In this regard, it is worth noting that STT assumes self-confidence as a mediator of stereotype effects (e.g., Steele, 1997; Walton and Cohen, 2003; Schmader et al., 2008). Moreover, several studies have reported that stereotype threat reduces task-specific measures of self-confidence in women facing math tasks (Good et al., 2008; Franceschini et al., 2014) while increasing their negative thoughts (Cadinu et al., 2005; Beilock et al., 2007; Schmader et al., 2008), and their anxiety and feelings of dejection (Spencer et al., 1999; Keller and Dauenheimer, 2003; Osborne, 2007; Johns et al., 2008). Stronger and bidirectional, albeit indirect, evidence of a mediatory role of in-task self-confidence also comes from studies assessing the effects of gender-related stereotypes on male and female performance in another highly gender-stereotyped cognitive domain, mental rotation (e.g., Estes and Felker, 2012; Sanchis-Segura et al., 2018).

To further explore the possible mediatory role of in-task self-confidence in M-MET performance and also obtain information about "who" is more affected by the stereotypes' activation, across gender effects were assessed with correlational methods and mediation analyses. To do so, IAT D-scores were first transformed into IAT "influence" scores (see Section 2.3 for details). These scores quantify the strength of the association between "science" and each participant's own gender category and, therefore, the expected positive/negative impact of these implicit associations on the participants' selves (Nosek and Smyth, 2011; Steffens and Jelenec, 2011) and not just the stereotypical/counter-stereotypical direction of their contents.

As could be expected, in the absence of a stereotype activation, neither M-MET scores nor any of its components appeared to be "influenced" by the participants' gender-science implicit associations. Instead, overall M-MET, accuracy, and, even more so, I-ACD scores were directly correlated to the participants' pre-test math self-concept (Figure 6A). Partial correlation networks (Figure 6C) confirmed these exploratory observations and revealed that the relationship between math self-concept and MET scores was directly mediated through I-ACD and, to a lower extent, by accuracy scores. Moreover, through these two mediators, math self-concept also indirectly increased task persistence. That is, participants that perceived themselves as more competent in math before the test were also the ones that felt more self-confident while performing this task and, through several converging processes, ended up obtaining higher M-MET scores. These findings align with the predictions of Bandura's

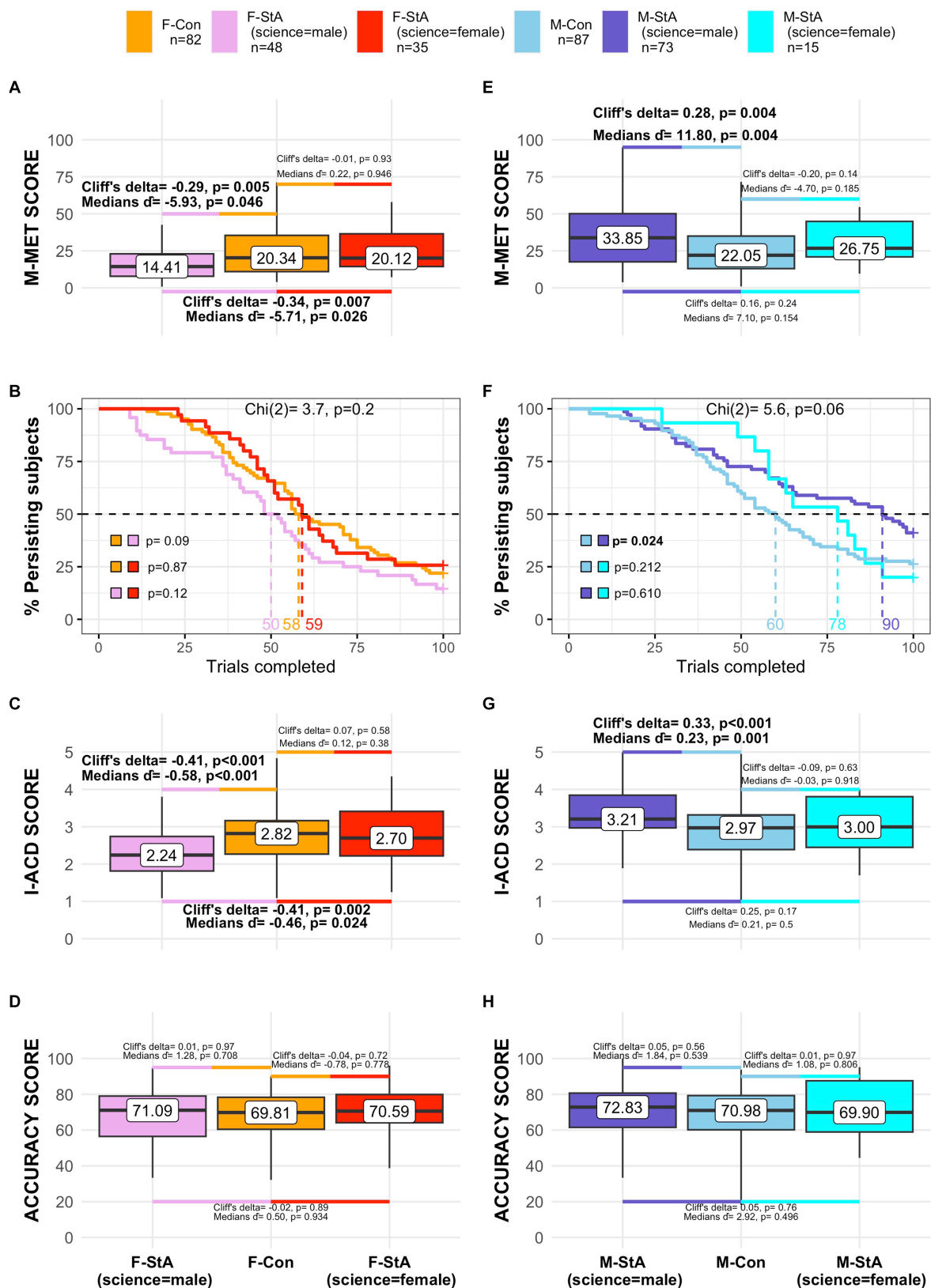


social cognitive theory (Bandura, 1977; Bandura, 1986), the Eccles' expectancy-value model (Eccles, 1983; Wigfield and Eccles, 2000), and SCCT (Lent et al., 1994) and, therefore, with the results of previous studies showing that math self-concept is positively associated to math performance, academic achievement, and persistence in math-related activities (e.g., Multon et al., 1991; Malmivuori, 2006; Brown et al., 2008; Marsh et al., 2009; Parsons et al., 2009; Lee et al., 2015). Moreover, these results also confirm that, as has been previously proposed (Pajares and Miller, 1994; Pajares, 1996), task-specific indexes of self-efficacy have higher and a more direct predictive value on math performance and persistence than math self-concept and other general measures of math self-efficacy.

By contrast, M-MET scores in the StA condition became largely independent on math self-concept and clearly associated to the participants' implicit gender-science associations (Figure 6B). Partial correlation networks (Figure 6D) revealed that the effects of the participants' implicit gender-science associations on M-MET scores were primarily conveyed through I-ACD scores. Thus, the higher the

participants implicitly associated science with their own gender (that is, the higher the positive "influence" of the participants' implicit associations), the higher their self-confidence during task performance, and the higher their final M-MET scores. Positive "influences" of the participants' implicit associations were also directly correlated to the number of completed trials (Figure 6D) and, although this effect appeared to be smaller, it was reinforced by converging indirect effects mediated by I-ACD and accuracy scores, all of which contributed to enhance the overall M-MET scores. These findings are in agreement with the principles of cognitive consistency (Greenwald et al., 2003; Cvencek et al., 2011) and, therefore, with empirical evidence showing that implicit gender-related associations can affect math self-efficacy, achievement, and engagement/persistence (e.g., Nosek et al., 2002a,b; Nosek and Smyth, 2011; Steffens and Jelenec, 2011; Franceschini et al., 2014). Moreover, our results seem to be also in line with studies suggesting that stereotype threat undermines the achievement, interest, and persistence of female STEM students in in math/STEM-related activities through a





**FIGURE 7** Effects of stereotypes' reactivation in participants harboring stereotypical and counter-stereotypical "gender-science" implicit associations. The figure summarizes the M-MET performance after stereotypes' reactivation of females (panels A–D) and males (E–H) holding stereotype-consistent associations ("science = male") or stereotype-inconsistent ("science = female") implicit associations (IAT D-score > 0 and IAT D-score <= 0, respectively). Except for the total number of trials (panels B,F), comparisons between these subgroups of individuals and their respective control groups were conducted by calculating their median differences (*d*) and Cliff's delta and the results are reported within the respective panels. Possible differences in (Continued)

FIGURE 7 (Continued)  
 task persistence (number of completed trials, panels B,F) were assessed by comparing the Kaplan -Meier survival curves with a Long-Rank test and subsequent dyadic comparisons (reported within the panels). Statistically significant effects (uncorrected  $p < 0.05$ ) are highlighted in bold.

TABLE 8 Comparison of the StA subgroups in math self-concept.

	<i>n</i>	Median	Value of <i>p</i>	$\hat{d}$	95% CI	Cliff's delta
F-StA ("science = male") Vs. F-Con	48 82	7 7.67	0.214	-0.67	[-1.5, 0.36]	Delta = -0.13, $p = 0.22$
F-StA ("science = female") Vs. F-Con	35 82	7.56 7.67	0.862	0.11	[-0.80, 0.94]	Delta = 0.01, $p = 0.95$
F-StA ("science = male") Vs. F-StA ("science = female")	48 35	7 7.56	0.320	-0.56	[-1.55, 0.49]	Delta = 0.12, $p = 0.36$
M-StA ("science = male") Vs. M-Con	73 87	7.55 7.17	0.484	0.38	[-0.54, 0.92]	Delta = 0.03, $p = 0.76$
M-StA ("science = female") Vs. M-Con	15 87	6.9 7.17	0.421	0.30	[-0.62, 1.65]	Delta = 0.16, $p = 0.36$
M-StA ("science = male") Vs. M-StA ("science = female")	73 15	7.55 6.9	0.220	0.68	[-0.43, 1.91]	Delta = 0.18, $p = 0.29$

The table displays the outcome (*p*-values, medians' difference, and its corresponding 95% CI) of the median-based comparisons between control and StA subgroups regarding math self-concept. For each gender, the StA subgroups were constructed depending on whether individuals hold stereotypical "science-male / humanities-female" (IAT D scores >0) or counter-stereotypical "science-female / humanities-male" (IAT D-score < 0) associations. M-Con, males assigned to the control experimental condition; M-StA, males assigned to the stereotypes activation condition; F-Con, females assigned to the control experimental condition; F-StA, females assigned to the stereotypes activation condition; CI, confidence interval.

reduction of self-confidence and self-efficacy beliefs (Brown et al., 2008; Wright et al., 2013; Deemer et al., 2014; Lee et al., 2015; Lent et al., 2015; Lin and Deemer, 2021) but probably also through other behavioral/cognitive processes (Woodcock et al., 2012; Thoman et al., 2013; Lewis and Sekaquaptewa, 2016).

To better characterize the modulatory role of gender-science implicit associations on the stereotype threat effects, within-gender comparisons between participants exposed to the StA condition but experiencing distinct (positive vs. negative) "influences" from their implicit associations in M-MET scores were conducted (Figures 7A,E). The results of these comparisons suggest that stereotype threat solely promoted a statistically significant decrease of the M-MET scores in those female students associating "science" with "male" (median IAT D-score = 0.34) and, therefore, suffering a negative "influence" from their implicit associations (median IAT "influence" = -0.34). Complementarily, stereotype activation only promoted a statistically significant increase of the M-MET scores in those male students implicitly associating "science" with "male" (median IAT D-score = 0.65) that is, in those males receiving positive "influences" from their implicit associations (median IAT "influence" = 0.65). By contrast, the M-MET scores of males and females that did not associate "science" with "male" did not significantly differ from those of their respective control groups. More specifically, this subgroup of females exhibited a counter-stereotypical "female-science" association (median IAT D-score = -0.37) that promoted a positive (median IAT "influence" = 0.37) and "protective" influence against the detrimental effects of the stereotype threat without triggering any stereotype reactance effect (Kray et al., 2001), hence showing M-MET scores virtually indistinguishable from those of the F-Con group. By contrast, the male subgroup that did not seem to associate "science" to any gender (median IAT D-score = -0.08) did not seem to receive any

"influence" from their implicit associations (median IAT influence = -0.08) but seemed to "profit" from the male-encouraging stereotype activation, exhibiting M-MET scores that were slightly higher than those of the M-Con group (for similar instructions' effects, see Moè and Pazzaglia, 2006 and Sanchis-Segura et al., 2018).

In agreement with the results of our correlational analyses (Figure 6), the same pattern of within-gender effects was observed on I-ACD scores (Figures 7C,G) and, to a reduced extent, in the number of completed trials (Figures 7B,F), but not in accuracy scores (Figures 7D,H). Therefore, the results of the between, across, and within gender analysis seem to confirm that the effect of stereotype reactivation on M-MET scores arise from changes in self-confidence and task persistence, but also that these effects solely occur or, at least, are more prominent in individuals holding stereotypical implicit gender-science associations. In this way, the results of the present study seem to confirm and extend those of previous reports showing that the effects of stereotype threat and stereotype lift are more prominent in individuals harboring stereotype-consistent implicit associations (Franceschini et al., 2014; Galdi et al., 2014), are probably mediated by in-task self-confidence changes (Franceschini et al., 2014), and may finally affect math-related achievements and the intentions of persisting or quitting math-related activities and math-related studies (Steele and Aronson, 1995; Von Hippel et al., 2011; Woodcock et al., 2012; Thoman et al., 2013).

## 5. Conclusions and implications

The present study provides experimental evidence showing that (at least, some) female engineering students are threatened by

stereotypes about women's math/science abilities. This threat seems to undermine their self-confidence, making them opt for less challenging options, and finally decreasing their achievement in high-demanding math activities. Thus, our findings confirm and extend slowly accumulating evidence showing that stereotype threats may act as a contextual barrier to women's STEM career development (Deemer et al., 2014; Cadaret et al., 2017) and that gender stereotypes do not only make it less likely for women to initially choose engineering and other STEM studies, but can also create adverse environmental conditions for women already enrolled in these studies (Blackburn, 2017; Clark et al., 2021). These hostile conditions may reduce the self-efficacy, achievements, and engagement of female students (Eddy and Brownell, 2016; Blackburn, 2017), making their academic experiences less rewarding and more distressing (Eddy and Brownell, 2016; González-Pérez et al., 2022), and finally increasing their chances of abandoning the field or switching to less math-intensive (i.e., non-STEM) educational/professional options (Isphording and Qendraj, 2019; Clark et al., 2021; González-Pérez et al., 2022).

However, our study also shows that gender stereotypes do not only affect women's performance. In fact, similar or even stronger effects but of the opposite sign (i.e., increased self-confidence and motivation) were observed in (at least, some) male engineering students. Traditionally, these effects have received less attention (Walton and Cohen, 2003; Shih et al., 2012) but they are also relevant as they show that stereotypes affect men and women differently but through similar processes (Cheryan and Plaut, 2010) and that self-affirming conditions can maximize students' potential and performance (Walton and Cohen, 2003; Gawronski et al., 2008), hence providing the basis for designing effective interventions that can equally benefit female and male students (Good et al., 2008; Forbes and Schmader, 2010; Miyake et al., 2010; Walton et al., 2015). Moreover, these effects seem to indicate that interventions should not be solely focused on counteracting the stereotypes held by the disfavored groups (in this case, female STEM students) but should also probably try to change the environmental cues, cultural values, and other people's beliefs that may indirectly promote them (Cheryan and Plaut, 2010; Walton et al., 2015; Rincón and George-Jackson, 2016; Blackburn, 2017).

Finally, our study shows that stereotype effects were not uniform, neither among female nor among male students, but very much moderated by their implicit and pre-existing "gender-science" associations. This observation is important, not only because it contributes to understanding the preconditions under which stereotypes' effects emerge (Pennington et al., 2016; Picho-Kiroga et al., 2021), but also because it highlights that gender categories are far from being homogenous and, therefore, that characterizing these heterogeneous groups through averages, or any other single estimate can be very misleading. In this regard, we advocate to replace analytical strategies based on comparisons between men vs. women averages (which far too often are used to make unwarranted generic statements and conclusions about *all* women and *all* men) by more complex and nuanced ones able to also offer information about within and across gender effects. On the other hand, we also propose that the evaluation of pre-existing implicit gender-related associations may serve to identify individuals at higher risk of suffering the detrimental effects of stereotype threat on performance and career decisions in highly stereotyped academic domains such as math and other "hard" sciences.

## 6. Limitations

The present study is not devoid of limitations. In this regard, the main limitations that we see in our study are:

1. Experimental studies -as the present one- have the strength of allowing us to isolate and manipulate specific factors in order to clarify their role in multicausal and complex phenomena. However, experimental studies are conducted under conditions that are, to some extent, artificial and cannot fully model the complexity of the studied phenomena. Therefore, the ecological validity of experimental studies is necessarily limited and the information provided by this kind of studies must be regarded as complementary and not substitutive of the correlational evidence obtained in real world situations (Aronson and Dee, 2012).
2. The present study was conducted employing a large, gender-balanced, and very homogenous sample (engineering students). While this methodological decision was adequate and convenient in many ways, it could further reduce the generalization of the obtained findings and conclusions. Thus, as hinted in the introduction section, the math self-efficacy and gender-science implicit associations of engineering students probably do not correctly represent those of other (i.e., non-STEM) university students or of the general population. On the other hand, although our sample was large and we used robust statistics that enhanced statistical power, its size could have still been sub-optimal for some analyses conducted with some specific subgroups of participants.
3. To our knowledge, this is the first study using a modified version of the MET task developed by Engle-Friedman et al. (2003). This task allowed us to different measures (overall scores, total number of trials completed, percent of correctly solved problems, and level of difficulty chosen in each trial) that were interpreted as indexes of performance, persistence, arithmetic accuracy, and in-task self-confidence, respectively. However, although we think that our interpretation of these indexes is reasonable, there might be other alternative ones (especially in the case of self-selected difficulty, which could also be measuring more than one single construct). Therefore, future studies should be aimed to better characterize these different aspects of the M-MET task used in the present study and to validate their provisional interpretation.

## Data availability statement

The datasets presented in this article are not readily available because any data sharing should be approved by the Ethics Standards Committees of the Universitat Jaume I. Requests to access the datasets should be directed to CS-S (csanchis@uji.es).

## Ethics statement

The studies involving human participants were reviewed and approved by the study was approved by the Ethics Standards

Committees of the Universitat Jaume I. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

CF and CS-S designed and conceptualized the study. AS-T and SF-E preprocessed the data with which CS-S conducted the statistical analyses. CS-S and CF wrote the manuscript. All authors participated in the experimental sessions, contributed to the manuscript revision and read, edited, and approved the submitted version.

## Funding

This research was supported by a grant (PID2019-106793RB-I00/AEI/10.13039/501100011033) provided by Ministerio de Ciencia e Innovación to CF and CS-S and a grant (UJI B2020-02) awarded to CF and CS-S. SF-E was supported by an FPI grant from UJI

## References

- Ackerman, P. L., Kanfer, R., and Beier, M. E. (2013). Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, stem persistence, and gender differences. *J. Educ. Psychol.* 105, 911–927. doi: 10.1037/A0032338
- Adams, G., García, D. M., Purdie-Vaughns, V., Steele, C. M., Buscher, D., Cabral, M., et al. (2006). The detrimental effects of a suggestion of sexism in an instruction situation. *J. Exp. Soc. Psychol.* 42, 602–615. doi: 10.1016/j.jesp.2005.10.004
- Albert, K. A., and Luzzo, D. A. (1999). The role of perceived barriers in career development: a social cognitive perspective. *J. Couns. Dev.* 77, 431–436. doi: 10.1002/J.1556-6676.1999.TB02470.X
- Aronson, J., and Dee, T. (2012). “Stereotype threat in the real world” in *Stereotype threat: Theory, process, and application*. eds. M. Inzlicht and T. Schmader (Oxford: Oxford University Press), 264–278.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., and Brown, J. (1999). When white men Can't do math: necessary and sufficient factors in stereotype threat. *J. Exp. Soc. Psychol.* 35, 29–46. doi: 10.1006/JESP.1998.1371
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (2000). Social cognitive theory: an agentic perspective. *Annu. Rev. Psychol.* 52, 1–26. doi: 10.1146/annurev.psych.52.1.1
- Bandura, A., Freeman, W. H., and Lightsey, R. (1999). Self-efficacy: the exercise of control. *J. Cogn. Psychother.* 13, 158–166. doi: 10.1891/0889-8391.13.2.158
- Beilock, S. L., Gunderson, E. A., Ramirez, G., and Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1860–1863. doi: 10.1073/PNAS.0910967107/SUPPL\_FILE/STXT.PDF
- Beilock, S. L., Rydell, R. J., and McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *J. Exp. Psychol.* 136, 256–276. doi: 10.1037/0096-3445.136.2.256
- Blackburn, H. (2017). The status of women in STEM in higher education: a review of the literature 2007–2017. *Sci. Technol. Libr.* 36, 235–273. doi: 10.1080/0194262X.2017.1371658
- Blickenstaff, J. C. (2006). Women and science careers: leaky pipeline or gender filter? *Gen. Educ. Staff.* 17, 369–386. doi: 10.1080/09540250500145072
- Brainard, S. G., and Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science. *J. Eng. Educ.* 87, 369–375. doi: 10.1002/j.2168-9830.1998.tb00367.x
- Brown, S. D., and Lent, R. W. (1996). A social cognitive framework for career choice counseling. *Career Dev. Q.* 44, 354–366. doi: 10.1002/j.2161-0045.1996.tb00451.x
- Brown, S. D., Trammayne, S., Hoxha, D., Telander, K., Fan, X., and Lent, R. W. (2008). Social cognitive predictors of college students' academic performance and persistence: a meta-analytic path analysis. *J. Vocat. Behav.* 72, 298–308. doi: 10.1016/J.JVB.2007.09.003

(PREDOC/2020/22). These funding sources did not play any role in designing the study or in the collection, analysis, and interpretation of the data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Buschor, C. B., Berweger, S., Frei, A. K., and Kappler, C. (2014). Majoring in STEM—what accounts for Women's career decision making? A mixed methods study. *J. Educ. Res.* 107, 167–176. doi: 10.1080/00220671.2013.788989

Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., and Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: a multiple-groups analysis. *J. Couns. Psychol.* 57, 205–218. doi: 10.1037/a0018608

Byars-Winston, A. M., and Fouad, N. A. (2008). Math and science social cognitive variables in college students: contributions of contextual factors in predicting goals. *J. Career Assess.* 16, 425–440. doi: 10.1177/1069072708318901

Cadaret, M. C., Hartung, P. J., Subich, L. M., and Weigold, I. K. (2017). Stereotype threat as a barrier to women entering engineering careers. *J. Vocat. Behav.* 99, 40–51. doi: 10.1016/J.JVB.2016.12.002

Cadinu, M., Maass, A., Rosabianca, A., and Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychol. Sci.* 16, 572–578. doi: 10.1111/j.0956-7976.2005.01577.x

Casad, B. J., Petzel, Z. W., and Ingalls, E. A. (2019). A model of threatening academic environments predicts women STEM majors' self-esteem and engagement in STEM. *Sex Roles* 80, 469–488. doi: 10.1007/s11199-018-0942-4

Chen, X. (2013). *STEM attrition: college students' paths into and out of STEM fields. Statistical analysis report. NCES 2014-001*. Washington, DC: National Center for Education Statistics.

Cheryan, S., Master, A., Meltzoff, A. N., Ceci, S. J., Cimpian, A., and Schmader, T. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Front. Psychol.* 6:49. doi: 10.3389/fpsyg.2015.00049

Cheryan, S., and Plaut, V. C. (2010). Explaining underrepresentation: a theory of precluded interest. *Sex Roles* 63, 475–488. doi: 10.1007/s11199-010-9835-x

Clark, S. L., Dyar, C., Inman, E. M., Maung, N., and London, B. (2021). Women's career confidence in a fixed, sexist STEM environment. *Int. J. STEM Educ.* 8, 1–10. doi: 10.1186/s40594-021-00313-z

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

Deemer, E. D., Thoman, D. B., Chase, J. P., and Smith, J. L. (2014). Feeling the threat: stereotype threat as a contextual barrier to Women's science career choice intentions. *J. Career Dev.* 41, 141–158. doi: 10.1177/0894845313483003

Doyle, R. A., and Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: a meta-analysis. *Learn. Individ. Differ.* 47, 103–116. doi: 10.1016/J.LINDIF.2015.12.018

Eccles, J. (2011). Gendered educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Int. J. Behav. Dev.* 35, 195–201. doi: 10.1177/0165025411398185

- Eccles, J. (1983). "Expectancies, values and academic behaviors" in *Achievement and achievement motives*. ed. J. T. Spence (San Francisco, CA: Free man), 75–146.
- Eccles, J. S., Jacobs, J. E., and Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *J. Soc. Issues* 46, 183–201. doi: 10.1111/j.1540-4560.1990.TB01929.X
- Eddy, S. L., and Brownell, S. E. (2016). Beneath the numbers: a review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Phys. Rev. Phys. Educ. Res.* 12:106. doi: 10.1103/PhysRevPhysEducRes.12.020106
- Else-Quest, N. M., Mineo, C. C., and Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychol. Women Q.* 37, 293–309. doi: 10.1177/0361684313480694
- Engle-Friedman, M., Riel, S., Golan, R., Ventuneac, A. M., Davis, C. M., Jefferson, A. D., et al. (2003). The effect of sleep loss on next day effort. *J. Sleep Res.* 12, 113–124. doi: 10.1046/j.1365-2869.2003.00351.x
- Epskamp, S., Borsboom, D., and Fried, E. (2018). Estimating psychological networks and their accuracy: a tutorial paper. *Behav. Res.* 50, 195–212. doi: 10.3758/s13428-017-0862-1
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D. (2012). Qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* 48, 1–18. doi: 10.18637/JSS.V048.I04
- Epskamp, S., and Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychol. Methods* 23, 617–634. doi: 10.1037/MET0000167
- Eriksson, K., and Lindholm, T. (2007). Making gender matter: the role of gender-based expectancies and gender identification on women's and men's math performance in Sweden. *Scand. J. Psychol.* 48, 329–338. doi: 10.1111/j.1467-9450.2007.00588.x
- Estes, Z., and Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Arch. Sex. Behav.* 41, 557–570. doi: 10.1007/s10508-011-9875-5
- European Commission. (2014). Cape Town: Carlos Moedas. Reaching gender equality in science, technology, engineering and mathematics. Available at: [https://ec.europa.eu/commission/2014-2019/moedas/announcements/reaching-gender-equality-science-technology-engineering-and-mathematics\\_en](https://ec.europa.eu/commission/2014-2019/moedas/announcements/reaching-gender-equality-science-technology-engineering-and-mathematics_en) (Accessed March 27, 2023).
- European Commission. (2016). SHE Figures 2015. Available at: [https://ec.europa.eu/research/swafs/pdf/pub\\_gender\\_equality/she\\_figures\\_2015-final.pdf](https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf) (Accessed March 27, 2023).
- Fayer, S., Lacey, A., and Watson, A. (2017). STEM occupations: past, present, and future. *Spotlight Stat.* 1, 1–35.
- Flore, P. C., Mulder, J., and Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: a registered report. *Comprehens. Results Soc. Psychol.* 3, 140–174. doi: 10.1080/23743603.2018.1559647/SUPPL\_FILE/RRSP\_A\_1559647\_SM7401.ZIP
- Flore, P. C., and Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *J. Sch. Psychol.* 53, 25–44. doi: 10.1016/j.jsp.2014.10.002
- Forbes, C. E., and Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *J. Pers. Soc. Psychol.* 99, 740–754. doi: 10.1037/a0020971
- Fouad, N. A., Hackett, G., Smith, P. L., Kantamneni, N., Fitzpatrick, M., Haag, S., et al. (2010). Barriers and supports for continuing in mathematics and science: gender and educational level differences. *J. Vocat. Behav.* 77, 361–373. doi: 10.1016/j.jvb.2010.06.004
- Fouad, N. A., Smith, P. L., and Zao, K. E. (2002). Across academic domains: extensions of the social-cognitive career model. *J. Couns. Psychol.* 49, 164–171. doi: 10.1037/0022-0167.49.2.164
- Franceschini, G., Galli, S., Chiesi, F., and Primi, C. (2014). Implicit gender-math stereotype and women's susceptibility to stereotype threat and stereotype lift. *Learn. Individ. Differ.* 32, 273–277. doi: 10.1016/j.lindif.2014.03.020
- Galdi, S., Cadinu, M., and Tomasetto, C. (2014). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Dev.* 85, 250–263. doi: 10.1111/cdev.12128
- Garriott, P. O., Hultgren, K. M., and Frazier, J. (2017). STEM stereotypes and high school students' math/science career goals. *J. Career Assess.* 25, 585–600. doi: 10.1177/1069072716665825
- Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., and Strack, F. (2008). When "just say no" is not enough: affirmation versus negation training and the reduction of automatic stereotype activation. *J. Exp. Soc. Psychol.* 44, 370–377. doi: 10.1016/j.jesp.2006.12.004
- González-Pérez, S., Martínez-Martínez, M., Rey-Paredes, V., and Cifre, E. (2022). I am done with this! Women dropping out of engineering majors. *Front. Psychol.* 13:4544. doi: 10.3389/fpsyg.2022.918439/BIBTEX
- Good, C., Aronson, J., and Harder, J. A. (2008). Problems in the pipeline: stereotype threat and women's achievement in high-level math courses. *J. Appl. Dev. Psychol.* 29, 17–28. doi: 10.1016/j.appdev.2007.10.004
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197
- Grissom, R. J., and Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications*. New York, NY: Routledge.
- Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2
- Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men. A path analysis. *J. Couns. Psychol.* 32, 47–56. doi: 10.1037/0022-0167.32.1.47
- Hackett, G., and Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *J. Res. Math. Educ.* 20, 261–273. doi: 10.5951/JRESEMATHEM.20.3.0261
- Hand, S., Rice, L., and Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Soc. Psychol. Educ.* 20, 929–945. doi: 10.1007/s11218-017-9408-8
- Hewlett, S. A., Luce, C. B., Servon, L. J., Sherbin, L., Shiller, P., Sosnovich, E., et al. (2008). The Athena factor: Reversing the brain drain in science, engineering, and technology. *Harvard Business Rev.* 86, 22–4, 139
- Hill, C., Corbett, C., and St. Rose, A. (2010). Why So few? Women in science, technology, engineering, and mathematics. *Am. Assoc. Univ. Women* 18:109.
- Inzlicht, M., and Schmader, T. (Eds.) (2011). *Stereotype threat. Theory, process and application*. Oxford: Oxford University Press.
- Isphording, I. E., and Qendrai, P. (2019). Gender Differences in Student Dropout in STEM. *IZA Res. Rep.* 87:9. doi: 10.5157/NEPS:SC5:9.0.0
- Johns, M., Inzlicht, M., and Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *J. Exp. Psychol.* 137, 691–705. doi: 10.1037/a0013834
- Keller, J., and Dauenheimer, D. (2003). Stereotype threat in the classroom: dejection mediates the disrupting threat effect on Women's math performance. *Personal. Soc. Psychol. Bull.* 29, 371–381. doi: 10.1177/0146167202250218
- Kim, A. Y., Sinatra, G. M., and Seyranian, V. (2018). Developing a STEM identity among young women: a social identity perspective. *Rev. Educ. Res.* 88, 589–625. doi: 10.3102/0034654318779957
- Kray, L. J., Thompson, L., and Galinsky, A. (2001). Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *J. Pers. Soc. Psychol.* 80, 942–958. doi: 10.1037/0022-3514.80.6.942
- Lee, H. S., Flores, L. Y., Navarro, R. R., and Kanagui-Muñoz, M. (2015). A longitudinal test of social cognitive career theory's academic persistence model among Latino/a and white men and women engineering students. *J. Vocat. Behav.* 88, 95–103. doi: 10.1016/j.jvb.2015.02.003
- Lemeshow, S., May, S., and Hosmer, D. W. (2011). *Applied survival analysis: Regression modeling of time-to-event data*. Hoboken, NJ: John Wiley & Sons.
- Lent, R. W., and Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: a measurement guide. *J. Career Assess.* 14, 12–35. doi: 10.1177/1069072705281364
- Lent, R. W., Brown, S. D., Brenner, B., Chopra, S. B., Davis, T., Talleyrand, R., et al. (2001). The role of contextual supports and barriers in the choice of math/science educational options: a test of social cognitive hypotheses. *J. Couns. Psychol.* 48, 474–483. doi: 10.1037/0022-0167.48.4.474
- Lent, R. W., Brown, S. D., and Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *J. Vocat. Behav.* 45, 79–122. doi: 10.1006/JVBE.1994.1027
- Lent, R. W., Brown, S. D., and Hackett, G. (2000). Contextual supports and barriers to career choice: a social cognitive analysis. *J. Couns. Psychol.* 47, 36–49. doi: 10.1037/0022-0167.47.1.36
- Lent, R. W., Brown, S. D., and Hackett, G. (1996). "Career development from a social cognitive perspective" in *Career choice and development*. eds. D. Brown and L. Brooks. 3rd ed (San Francisco, CA: Jossey-Bass), 373–421.
- Lent, R. W., Brown, S. D., Schmidt, J., Brenner, B., Lyons, H., and Treistman, D. (2003). Relation of contextual supports and barriers to choice behavior in engineering majors: test of alternative social cognitive models. *J. Couns. Psychol.* 50, 458–465. doi: 10.1037/0022-0167.50.4.458
- Lent, R. W., Brown, S. D., Talleyrand, R., McPartland, E. B., Davis, T., Chopra, S. B., et al. (2002). Career choice barriers, supports, and coping strategies: college students' experiences. *J. Vocat. Behav.* 60, 61–72. doi: 10.1006/JVBE.2001.1814
- Lent, R. W., Miller, M. J., Smith, P. E., Watford, B. A., Hui, K., and Lim, R. H. (2015). Social cognitive model of adjustment to engineering majors: longitudinal test across gender and race/ethnicity. *J. Vocat. Behav.* 86, 77–85. doi: 10.1016/j.jvb.2014.11.004
- Lent, R. W., Miller, M. J., Smith, P. E., Watford, B. A., Lim, R. H., Hui, K., et al. (2013). Social cognitive predictors of adjustment to engineering majors across gender and race/ethnicity. *J. Vocat. Behav.* 83, 22–30. doi: 10.1016/j.jvb.2013.02.006

- Lewis, N. A., and Sekaquaptewa, D. (2016). Beyond test performance: a broader view of stereotype threat. *Curr. Opin. Psychol.* 11, 40–43. doi: 10.1016/j.copsyc.2016.05.002
- Lin, C., and Deemer, E. D. (2021). Stereotype threat and career goals among women in STEM: mediating and moderating roles of perfectionism. *Mental Health Services* 48, 569–583. doi: 10.1177/0894845319884652
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a Meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/A0021276
- Luo, T., So, W. M., Wan, Z. H., and Li, W. C. (2021). STEM stereotypes predict students' STEM career interest via self-efficacy and outcome expectations. *Int. J. STEM Educ.* 8, 1–13. doi: 10.1186/s40594-021-00295-y
- Macphee, D., Farro, S., and Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: gender, ethnic, and social class patterns 2013 the Society for the Psychological Study of Social Issues. *Anal. Soc. Issues Public Policy* 13, 347–369. doi: 10.1111/asap.12033
- Makarova, E., and Herzog, W. (2015). Trapped in the gender stereotype? The image of science among secondary school students and teachers. *Equal. Diver. Inclus. Int. J.* 34, 106–123. doi: 10.1108/EDI-11-2013-0097
- Malmivuori, M. L. (2006). Affect and self-regulation. *Educ. Stud. Math.* 63, 149–164. doi: 10.1007/s10649-006-9022-8
- Manske, D. (2010). European engineering report. Available at: [https://www.engineerseurope.com/sites/default/files/European\\_Engineering\\_Report.pdf](https://www.engineerseurope.com/sites/default/files/European_Engineering_Report.pdf) (Accessed April 23, 2023).
- Marra, R. M., Rodgers, K. A., Shen, D., and Bogue, B. (2009). Women engineering students and self-efficacy: a multi-year, multi-institution study of women engineering student self-efficacy. *J. Eng. Educ.* 98, 27–38. doi: 10.1002/j.2168-9830.2009.tb01003.x
- Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., and Peschar, J. L. (2009). OECD's brief self-report measure of Educational Psychology's Most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *Int. J. Test.* 6, 311–360. doi: 10.1207/S15327574IJT0604\_1
- Master, A., Meltzoff, A. N., and Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proc. Natl. Acad. Sci. U. S. A.* 118. doi: 10.1073/PNAS.2100030118/-/DCSUPPLEMENTAL
- Mejía-Rodríguez, A. M., Luyten, H., and Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. *Int. J. Sci. Math. Educ.* 19, 1229–1250. doi: 10.1007/s10763-020-10100-x
- Miller, D. I., Eagly, A. H., and Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: evidence from 66 nations. *J. Educ. Psychol.* 107, 631–644. doi: 10.1037/EDU0000005
- Miller, M. J., Lent, R. W., Lim, R. H., Hui, K., Martin, H. M., Jezzi, M. M., et al. (2015). Pursuing and adjusting to engineering majors: a qualitative analysis. *J. Career Assess.* 23, 48–63. doi: 10.1177/1069072714523084
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., and Ito, T. A. (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330, 1234–1237. doi: 10.1126/SCIENCE.1195996
- Moè, A., and Pazzaglia, F. (2006). Following the instructions!. Effects of gender beliefs in mental rotation. *Learn. Individ. Differ.* 16, 369–377. doi: 10.1016/j.lindif.2007.01.002
- Morais Maceira, H. (2017). Economic benefits of gender equality in the EU. *Intereconomics* 52, 178–183. doi: 10.1007/s10272-017-0669-4/METRICS
- Multon, K. D., Brown, S. D., and Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: a Meta-analytic investigation. *J. Couns. Psychol.* 38, 30–38. doi: 10.1037/0022-0167.38.1.30
- Murphy, M. C., and Taylor, V. J. (2012). "The role of situational cues in signaling and maintaining stereotype threat" in *Stereotype threat: Theory, process, and application*. eds. M. Inzlicht and T. Schmader (Oxford: Oxford University Press), 17–33.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., and Garrett, J. (2007). Gender and course selection in upper secondary education: effects of academic self-concept and intrinsic value. *Educ. Res. Eval.* 12, 323–345. doi: 10.1080/13803610600765687
- National Association of Colleges and Employers. (2009). Salary survey. Available at: <https://scholarworks.uni.edu/salariesurvey/9> (Accessed April 23, 2023).
- National Science Board. (2010). Science and engineering indicators. Available at: <https://www.heri.ucla.edu/PDFs/NSB.pdf> (Accessed April 23, 2023).
- Nguyen, H. H. D., and Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *J. Appl. Psychol.* 93, 1314–1334. doi: 10.1037/a0012702
- Noland, M., Moran, T., and Kotschwar, B. R. (2016). Is gender diversity profitable? Evidence from a global survey. *SSRN Electron. J.* doi: 10.2139/SSRN.2729348
- Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dyn. Theory Res. Pract.* 6, 101–115. doi: 10.1037/1089-2699.6.1.101
- Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002b). Math = male, me = female, therefore math ≠ me. *J. Pers. Soc. Psychol.* 83, 44–59. doi: 10.1037/0022-3514.83.1.44
- Nosek, B. A., and Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *Am. Educ. Res. J.* 48, 1125–1156. doi: 10.3102/0002831211410683
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2010). Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88. doi: 10.1080/10463280701489053
- OECD (2017). *The pursuit of gender equality: An uphill Battle*. Paris: OECD.
- Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educ. Psychol.* 27, 135–154. doi: 10.1080/01443410601069929
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Rev. Educ. Res.* 66, 543–578. doi: 10.3102/00346543066004543
- Pajares, F., and Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *J. Educ. Psychol.* 86, 193–203. doi: 10.1037/0022-0663.86.2.193
- Parsons, S., Croft, T., and Harrison, M. (2009). Does students' confidence in their ability in mathematics matter? *Teach. Math. Appl. Int. J. IMA* 28, 53–68. doi: 10.1093/TEAMAT/HRP010
- Parsons, S., Croft, T., and Harrison, M. (2011). Engineering students' self-confidence in mathematics mapped onto Bandura's self-efficacy. *Eng. Educ.* 6, 52–61. doi: 10.11120/ENED.2011.06010052
- Pearl, J. (2000). *CAUSALITY models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (2016). Twenty years of stereotype threat research: a review of psychological mediators. *PLoS One* 11:46487. doi: 10.1371/journal.pone.0146487
- Perez, T., Cromley, J. G., and Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *J. Educ. Psychol.* 106, 315–329. doi: 10.1037/a0034027
- Picho-Kiroga, K., Turnbull, A., and Rodriguez-Leahy, A. (2021). Stereotype threat and its problems: theory misspecification in research, consequences, and remedies. *J. Adv. Acad.* 32, 231–264. doi: 10.1177/1932202X20986161
- Picho, K., Rodriguez, A., and Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: a meta-analysis. *J. Soc. Psychol.* 153, 299–333. doi: 10.1080/00224545.2012.737380
- Picho, K., and Schmader, T. (2018). When do gender stereotypes impair math performance? A study of stereotype threat among Ugandan adolescents. *Sex Roles* 78, 295–306. doi: 10.1007/s11199-017-0780-9
- Pronin, E., Steele, C. M., and Ross, L. (2004). Identity bifurcation in response to stereotype threat: women and mathematics. *J. Exp. Soc. Psychol.* 40, 152–168. doi: 10.1016/S0022-1031(03)00088-X
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Rincón, B. E., and George-Jackson, C. E. (2016). Examining department climate for women in engineering: the role of STEM interventions. *J. Coll. Stud. Dev.* 57, 742–747. doi: 10.1353/CSD.2016.0072
- Robbins, S. B., Le, H., Davis, D., Lauver, K., Langley, R., and Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A Meta-Analysis. *Psychol. Bulletin* 130, 261–288. doi: 10.1037/0033-2909.130.2.261
- Rottinghaus, P. J., Falk, N. A., and Park, C. J. (2018). Career assessment and counseling for STEM: a critical review. *Career Dev. Q.* 66, 2–34. doi: 10.1002/cdq.12119
- Rousset, G. A., Pernet, C. R., and Wilcox, R. R. (2017). Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *Eur. J. Neurosci.* 46, 1738–1748. doi: 10.1111/EJN.13610
- Sanchis-Segura, C., Aguirre, N., Cruz-Gómez, Á. J., Solozano, N., and Forn, C. (2018). Do gender-related stereotypes affect spatial performance? Exploring when, how and to whom using a chronometric two-choice mental rotation task. *Front. Psychol.* 9:1261. doi: 10.3389/fpsyg.2018.01261
- Schmader, T., Johns, M., and Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychol. Rev.* 115, 336–356. doi: 10.1037/0033-295X.115.2.336
- Schunk, D. H., and Pajares, F. (2002). "The development of academic self-efficacy," in *Development of achievement motivation*, eds. A. Wigfield and J. S. Eccles. Academic Press. 15–31. doi: 10.1016/B978-012750053-9/50003-6
- Schunk, D. H., and Pajares, F. (2009). "Self-efficacy theory" in *Handbook of motivation at school*. eds. K. R. Wentzel and D. B. Miele. 1st ed (London: Routledge), 49–68.
- Sharpe, D. (2019). Chi-Square test is statistically significant: now what? *Pract. Assess. Res. Eval.* 20:8. doi: 10.7275/tbfa-x148
- Shih, M., Pittinsky, T., and Ho, G. (2012). "Stereotype boost: positive outcomes from the activation of positive stereotypes" in *Stereotype threat: Theory, process, and application*. eds. M. Inzlicht and T. Schmader (Oxford: Oxford University Press), 141–156.

- Skaalvik, E. M., Federici, R. A., and Klassen, R. M. (2015). Mathematics achievement and self-efficacy: relations with motivation for mathematics. *Int. J. Educ. Res.* 72, 129–136. doi: 10.1016/J.IJER.2015.06.008
- 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, 617–629. doi: 10.1007/s11199-012-0209-4
- Smyth, F. L., and Nosek, B. A. (2015). On the gender-science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Front. Psychol.* 6:415. doi: 10.3389/fpsyg.2015.00415
- Spencer, S. J., Logel, C., and Davies, P. G. (2016). Stereotype threat. *Annu. Rev. Psychol.* 67, 415–437. doi: 10.1146/annurev-psych-073115-103235
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and Women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/JESP.1998.1373
- Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. doi: 10.1037/0003-066X.52.6.613
- Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* 69, 797–811. doi: 10.1037/0022-3514.69.5.797
- Steffens, M. C., and 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, 324–335. doi: 10.1007/s11199-010-9924-x
- Stoet, G., and Geary, D. C. (2012). *Can stereotype threat explain the gender gap in mathematics performance and achievement?* vol. 16. Los Angeles, CA: SAGE Publications Sage CA, 93–102.
- Therneau, T. (2022). *A package for survival analysis in R version 3.3–1*. Available at: <https://CRAN.R-project.org/package=survival>
- Thoman, D. B., Smith, J. L., Brown, E. R., Chase, J., and Lee, J. Y. K. (2013). Beyond performance: a motivational experiences model of stereotype threat. *Educ. Psychol. Rev.* 25, 211–243. doi: 10.1007/s10648-013-9219-1
- Treadway, M. T., Bossaller, N. A., Shelton, R. C., and Zald, D. H. (2012). Effort-based decision-making in major depressive disorder: a translational model of motivational anhedonia. *J. Abnorm. Psychol.* 121, 553–558. doi: 10.1037/A0028813
- Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., and Zald, D. H. (2009). Worth the “EEfRT”? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS One* 4:e6598. doi: 10.1371/journal.pone.0006598
- Van Borkulo, C. D., Van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., et al. (2022). Psychological methods comparing network structures on three aspects: a permutation test. *Psychol. Methods*. doi: 10.1037/met0000476
- Von Hippel, C., Issa, M., Ma, R., and Stokes, A. (2011). Stereotype threat: antecedents and consequences for working women. *J. Soc. Psychol.* 41, 151–161. doi: 10.1002/ejsp.749
- Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140, 1174–1204. doi: 10.1037/a0036620
- Walton, G. M., and Cohen, G. L. (2003). Stereotype lift. *J. Exp. Soc. Psychol.* 39, 456–467. doi: 10.1016/S0022-1031(03)00019-2
- Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., and Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women's experience, relationships, and achievement in engineering. *J. Educ. Psychol.* 107, 468–485. doi: 10.1037/A0037461
- Wang, M. T., and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (stem): current knowledge, implications for practice, policy, and future directions. *Educ. Psychol. Rev.* 29, 119–140. doi: 10.1007/s10648-015-9355-x
- Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *Am. Stat.* 70, 129–133. doi: 10.1080/00031305.2016.1154108
- Whitcomb, K. M., Yasemin, Z., Nokes-Malach, T. J., Schunn, C. D., and Singh, C. (2020). Comparison of self-efficacy and performance of engineering undergraduate women and men\*. *Int. J. Eng. Educ.* 36, 1996–2014.
- Wigfield, A., and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.* 25, 68–81. doi: 10.1006/CEPS.1999.1015
- Wilcox, R. R. (2016). *Introduction to robust estimation and hypothesis testing*. Amsterdam: Elsevier.
- Woodcock, A., Hernandez, P. R., Estrada, M., and Schultz, P. W. (2012). The consequences of chronic stereotype threat: domain disidentification and abandonment. *J. Pers. Soc. Psychol.* 103, 635–646. doi: 10.1037/A0029120
- Wright, S. L., Jenkins-Guarnieri, M. A., and Murdock, J. L. (2013). Career development among first-year college students: college self-efficacy, student persistence, and academic success. *J. Career Dev.* 40, 292–310. doi: 10.1177/0894845312455509
- Zander, L., Höhne, E., Harms, S., Pfof, M., and Hornsey, M. J. (2020). When grades are high but self-efficacy is low: unpacking the confidence gap between girls and boys in mathematics. *Front. Psychol.* 11:355. doi: 10.3389/fpsyg.2020.552355
- Zeileis, A., Meyer, D., and Hornik, K. (2007). Residual-based shadings for visualizing (conditional) Independence. *J. Comput. Graph. Stat.* 16, 507–525. doi: 10.1198/106186007X237856