Authors:
Una Tellhed, Ph.D., Senior Lecturer, Department of Psychology, Lund University.
Caroline Adolfsson, M.S., Lund University.

Title

Competence and confusion: How stereotype threat can make you a bad judge of your competence.

Contact information
Postal address: Una Tellhed, Department of Psychology, Lund University, Box 213, SE-221 00 LUND, Sweden
E-mail: una.tellhed@psy.lu.se
Telephone: +46 46 222 86 87
Telefax: +46 46 222 42 09

Abstract

Women tend to have competence doubts for masculine-stereotyped domains (e.g. math), while men tend to think they can handle both feminine-stereotyped and masculine-stereotyped domains equally well (e.g. Watt, 2010). We suggest that perhaps women's more frequent experience with stereotype threat (Pillaud et al., 2015) can partly explain why. Our results showed that when stereotype threat was primed in high school students (n = 244), there was no relationship between their performance on an academic test (the SweSAT) and their assessment of their performance (how well they did), while in a non-stereotype threat condition, there was a medium-sized relationship. The effect was similar for both men and women primed with stereotype threat. The results imply that stereotype threat undermines performance assessments.

Keywords

Stereotype threat, competence beliefs, SAT, performance assessment, gender differences

Author note: The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. For access to the study's material, please contact the first author.

Introduction

Although meta-analyses find that men and women are highly psychologically similar and tend to perform roughly equally on most ability tests (Halpern, 2012; Hyde, 2005; 2014), men and women have very different competence perceptions. Women tend to have gender stereotypical competence beliefs, where they doubt their capability for male-stereotypical domains (e.g. math) but think highly of their feminine-stereotyped ability (e.g. empathy; Betz & Hackett, 1981; Bridges, 1988; Matsui, Ikeda, & Ohnishi, 1989; Tellhed, Bäckström & Björklund, 2016). Women's perceptions contrast with men's, who tend to believe they can handle both masculine-stereotyped and feminine-stereotyped domains equally well (Betz & Hackett, 1981; Bridges, 1988; Matsui, et al., 1989; Tellhed, et al., 2016). For masculine-stereotyped abilities like math, there is evidence that women underestimate and men overestimate their performance (Bench, Lench, Miner, Flores, & Liew, 2015; Watt, 2010). This creates gender stereotypical differences in men's and women's math ability perceptions, despite the gender similarity in performance (Ceci, Williams & Barnett, 2009; Else-Quest, Hyde, & Linn, 2010; Hackett, 1995; the Organisation for Economic Co-operation and Development (OECD) 2015; Weber, Skirbekk, Freund & Herlitz, 2014).

It's important to learn why women doubt their competence for masculine-stereotyped domains despite their good performance. Not only because competence doubts can be psychologically painful, but also because competence beliefs are an important predictor of career interest and career choice (e.g., Betz & Hackett, 1983; Hackett, 1995; Lent, Brown, & Hackett, 1994; Tellhed et al., 2016). Women's lower competence beliefs for male-dominated domains contribute to the gender segregation in the labor market, which has been identified as problematic from both a societal, employer and an individual perspective (European Commission, 2014).

We theorize that perhaps women's more frequent experience with *stereotype threat* (Steele & Aronson, 1995; Steele, 1997) as compared to men's (Pillaud, Rigaud & Clémence, 2015) is part of the reason why women are more prone to doubt their competence than men are. In the current study we test the novel prediction that perhaps stereotype threat confuses our competence assessments.

Stereotype threat and gender

After two decades of research on stereotype threat, it is now well-established that stereotypes which portray one group as inferior to another can threaten the negatively stereotyped group members in contexts where they become salient (Steele & Aronson, 1995; Steele, 1997). This situational predicament, termed stereotype threat, can give rise to a number of cognitive and affective responses (see Inzlicht & Schmader, 2012; Spencer, Logel & Davies, 2016 for reviews). The most researched effect is the underperformance of threatened group members on ability tests, thereby ironically confirming the stereotype prophecy (Schmader & Beilock, 2012).

Stereotype threat is a general phenomenon; applicable to any group that is somehow negatively stereotyped (Inzlicht & Schmader, 2012). However, studies have shown that women are more susceptible to gender-related stereotype threat cues than men are (Pillaud et al., 2015). For women, stereotype threat seems to be "in the air", as Claude Steele (1997) put it: Women may sense stereotype threat from very subtle, or implicit cues, such as being subtly reminded of their gender identity prior to a test (e.g. Ambady, Shih, Kim, & Pittinsky, 2001; McGlone & Aronson, 2006), or simply being told that a test is diagnostic of a male-stereotyped ability, such as negotiation (e.g., Kray, Thompson & Galinsky, 2001; Tellhed & Björklund, 2010). Women's high susceptibility to stereotype threat contrasts with men's apparent resilience to gender-related stereotype threat cues (Pillaud et al., 2015). The much fewer studies that demonstrate gender-related stereotype threat in men, have used quite strong explicit primes of explicitly telling men that they are inferior to women (Hartley & Sutton, 2013; Koenig & Eagly, 2005; Nguyen & Ryan, 2008; Pillaud et al., 2015).

What explains this gender difference in stereotype threat susceptibility? It has been suggested that the difference relates to structural gender differences in status and power, and that men's higher status shields them from stereotype threat (Pillaud et al., 2015). It is unclear how this happens, but according to stereotype content model (Fiske, Cuddy, Glick & Xu, 2002) higher-status groups are stereotypically portrayed as more competent than others. This is apparent in gender stereotypes, where men are perceived as more competent than women (Fiske et al., 2002; Prentice & Carranza, 2002; Williams & Best, 1990). Stereotype content model and also social dominance theory (Sidanius & Pratto, 1999) state that portraying a group as less competent can function as a "legitimizing myth" that rationalizes structural inequality in status and power, such as men's higher power and status as compared to women's. As women's competence is more frequently questioned than men's (Fiske, et al.,

2002), competence-related threats perhaps become chronically accessible for women and are therefore more easily activated.

Gender differences in the SAT and stereotype threat. One example where women are more susceptible to stereotype threat is in relation to scholastic aptitude tests (SAT). Although women tend do better than men in school in many countries (OECD, 2015), a meta-analysis has shown that women tend to underperform on SATs, which has been explained as a stereotype threat-effect (Walton & Spencer, 2009; Steele, 2012). The SAT may cause stereotype threat in women, since it is perceived as indicative of a broad intellectual ability that is stereotypically associated more strongly with men than with women (Walton & Spencer, 2009; Steele, 2012). Since women are stereotyped as less competent than men, female test takers may *implicitly* activate negative female competence stereotypes when taking the SAT, without any explicit mentioning of gender stereotypes. That is, the SAT may implicitly trigger stereotype threat in women. Interestingly, in a study of British children, boy's typical advantage on the SAT, was reduced when children were *explicitly* told that "girls do better in school" (Hartley & Sutton, 2013). This explicit prime triggered stereotype threat in the boys, which significantly reduced their test scores, while leaving the girls' scores unaffected.

In the current study we will replicate and expand the design of Hartley and Sutton (2013) in a sample of Swedish high-school students. We predict that men will outperform women on an SAT when gender differences are not explicitly mentioned, since the SAT may implicitly induce stereotype threat in women (Walton & Spencer, 2009; Steele, 2012). We expect this common gender difference to be smaller for a group that is explicitly told that women tend to do better in school, since it may induce stereotype threat in men (Hartley & Sutton, 2013). The main aim of the present study is however to test if stereotype threat affects how people *perceive* their test performance. Could stereotype threat perhaps confuse our performance judgments? We test the hypothesis that assessment of test performance is more weakly related to actual test performance in a stereotype threat condition, as compared to a non- stereotype threat condition. We will next expand on our reasoning for this novel hypothesis.

Stereotype threat and performance assessments

Previous research has associated stereotype threat with a number of psychological processes which disrupts test performance. For instance, stereotype threat causes feelings of uncertainty

and self-doubts, increases arousal and triggers anxiety (see Schmader & Beilock, 2012 for a review). The affected becomes vigilant to negative cues in the environment and attention shifts automatically towards negative and anxiety-related cues (e.g Johns, Inzlicht, & Schmader, 2008). There is also evidence that the threatened individual tries to self-regulate and control the negative thoughts and feelings. There is neurological evidence that brain regions typically involved in emotion regulation (the ventral anterior cingulate cortex) are activated during stereotype threat (Krendl, Richeson, Kelley & Heatherton, 2008; Wraga, Helt, Jacobs, & Sullivan 2007). The rumination, increased arousal and attempts to self-regulate have a depleting effect on working memory capacity (e.g. Beilock, 2008; Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003). As working memory is crucial for controlled thought processes and deliberate decision making, the cognitive overload can impair performance on difficult tasks, such as challenging academic tests (Schmader & Beilock, 2012).

Although stereotype threat is a well-researched phenomenon, we have not seen any discussion in the literature as to how it may relate to *performance assessments* in the literature, which is the aim of the current study. To make a correct assessment of our performance, we need to compare it to target performances or estimations of the normal distribution. This requires controlled thought processes, which are highly reliant on working memory capacity (see Hassin, 2005; Miyake & Shah, 1999 for reviews). Previous research has shown that when cognitive resources are depleted (like under stereotype threat) people use less controlled processing and instead rely on automatic processing and salient heuristics (see Bodenhausen & Todd, 2010 for a review). These are less demanding of working memory capacity but may be biased by, for instance, salient stereotypes (e.g. Bodenhausen, Macrae & Sherman, 1990). In the current study we will test the hypothesis that performance assessments are more weakly related to test performance for people under stereotype threat, as compared to non-threatened people. We propose that the cognitive depletion brought on by stereotype threat perhaps confuses our performance assessments and makes them more susceptible to bias. If this is the case, it may perhaps contribute to the explanation of why women, who are often exposed to stereotype threat, tend to suffer from competence doubts in male-dominated domains. If stereotype threat makes you confused as to how well you perform in a domain, your overall competence beliefs may be more susceptible to bias by salient stereotypes, which in the case of gender, portray women as less competent than men.

To summarize, the hypotheses are:

1. Men outperform women on the SweSAT when it is presented as a standardized academic test with no mentioning of gender differences (women's stereotype threat), but gender differences are reduced in size when participants are explicitly told that women outperform men in school (men's stereotype threat).
2. Performance assessments are less strongly related to test performance on the SweSAT for men and women in the respective stereotype threat condition for each gender.

Method

Participants, design and procedure

96 male and 149 female Swedish senior high school students were recruited for the study, mean age was 17.8, *SD* = .09. The study had a 2 (men vs. women) x 2 (men's stereotype threat vs. women's stereotype threat) between subjects design. The participants were randomized to conditions. As more women than men volunteered to participate, the gender distribution was somewhat uneven over conditions (see Table 3).

All participants were initially told that they were about to take a standardized academic test (the SweSAT) and then answer some questions. Half of the participants were next explicitly told by the experimenter that "The purpose of this study is to investigate why women typically perform better than men in school." In addition to being presented orally, this information was also written on the top of the informed consent sheet for this participant group. This manipulation was meant to induce stereotype threat in men and is hereafter called "men's stereotype threat condition".

For the other half of the participants, gender was *not* mentioned prior to the test taking. They could therefore be considered a standard "control group". However, as described in the introduction, it has been repeatedly shown that women experience stereotype threat when taking the SAT with no mention of gender differences (Walton & Spencer, 2009; Steele, 2012). Therefore we have deemed this experimental condition as a "women's stereotype threat condition". The test packets were the same in both conditions; it began with the SweSAT followed by the questions described below.

Materials

The SweSAT. Sixteen questions were selected from the 2011 fall version of the Swedish Scholastic Aptitude Test (SweSAT, 2013). Half were selected from the verbal section and half from the quantitative section of the SweSAT. All questions were multiple choice and a total score was calculated. The participants were also asked to indicate their previous experience with the SweSAT; if they had taken the test before and if so, how many times.

Performance assessment. The participants were asked to estimate how well they believe they performed on the test on a Likert-scale from (1) "Not well at all" to "Extremely well" (7).

Manipulation check. Last in the survey, as a manipulation check, the participants were asked to indicate what the experimenter had orally stated prior to them taking the SweSAT of the following alternatives: "Women do better in school", "Men do better in school" or "None of the above". They were also asked to indicate their personal beliefs on gender differences in academic performance.

Ethical considerations

The material represents a subset of a dataset collected for one of the authors' master's thesis (Adolfsson, 2015). The results of the current article have not been previously published. The study was approved as following the APA ethical standards by the student thesis examiner at the Department of Psychology at Lund University in Sweden, prior to data collection. The participants gave informed consent and the instructions followed standard protocol regarding voluntary participation, anonymity, confidentiality and the right to discontinue the study without repercussion. No personal data was collected that could disclose the participants identity. Participants were debriefed and received a lecture on stereotype threat after the data collection. The priming statement, "women do better in school than men", has been found to be true (OECD, 2015) and therefore was not found to be deceptive. Further, due to the thorough debriefing offered by researchers, no long term, serious psychological effects were suspected, and no participants disclosed of being negatively affected by the participation.

Results

Preliminary analyses

One outlier was detected on the SweSAT and was excluded from further analysis as he had taken the SweSAT four times previously and was therefore not considered representative of the population. The mean previous experience with the SweSAT (number of times taking the test) was .34, SD = .77. All other assumptions for statistical tests were met.

The majority registered the oral experimental prime correctly. 84 % of the men and 96 % of the women that were not explicitly primed prior to the test (women's stereotype threat), correctly indicated that the experimenter had not mentioned gender differences. 70 % of the men and 71 % of the women who had been told by the experimenter that women do better in school (men's stereotype threat condition) correctly remembered hearing this after the test was over. The remainder stated not remembering hearing an oral prime.

75 % of the participants (similarly across conditions) reported believing that women generally do better in school, 3 % believed that boys do better and the rest believed in no gender difference.

The SweSAT

A two-way between-groups ANOVA tested the interaction between the experimental condition and participant gender on the SweSAT. Men were coded as 0 and women as 1. Women's stereotype threat condition (no explicit prime) was coded as 0 and men's stereotype threat condition (explicit prime) was coded as 1. There were no significant main effects, *F*:s < 1, but a small interaction effect, $F(1, 240) = 5.26$, $p = .023$, $\eta^2 = .02$, that supported hypothesis 1: The men performed slightly better than the women in the women's stereotype threat condition, $t(118) = 2.11$, $p = .037$, $\eta^2 = .05$, but there was no gender difference in the men's stereotype threat condition, $t(122) = 1.13$, $p = .260$.

Men's performance was marginally statistically lower in the men's stereotype threat condition as compared to the women's stereotype threat condition, $t(93) = 1.82$, $p = .072$, $\eta^2 = .03$. However, women's scores were not significantly different across conditions, $t(147) = -1.30$, $p = .196$, $\eta^2 = .01$. See Table 1 for descriptive statistics and Figure 1 for a display of the interaction effect.

Relationship between performance and performance assessment

A two-way between groups ANOVA showed no significant effect of condition, $F(1,237) = 2.47$, $p = .117$, $\eta^2 = .01$, or interaction effect, $F(1,237) = 2.08$, $p = .150$, $\eta^2 = .01$, on the performance assessment. However, the men generally thought they performed better on the SweSAT as compared to the women, $F(1, 237) = 23.49$, $p = < .001$, $\eta^2 = .09$. See table 1 for descriptive statistics.

To test the hypothesis (2) that performance assessments are less strongly related to test performance for men and women while under stereotype threat, we performed a hierarchical multiple regression analysis. The aim was to test if the relationship between the SweSAT scores and the performance assessment was moderated by the interaction between the experimental condition and participant gender (i.e. a three-way interaction of the SweSAT score x gender x condition on performance assessment).

One multivariate outlier was identified and removed from analysis, as this violates the assumptions. In step one, the standardized SweSAT test scores, participant gender and experimental condition were regressed on the participants' performance assessments. All interaction possibilities were next introduced step by step in the subsequent four blocks (see Table 2). The overall model was significant, $R^2 = .17$, $F(7, 232) = 6.95$, $p = < .001$.

As seen in table 2, there was a significant three-way interaction of participant gender, the experimental condition and the SweSAT test results on performance assessments, $B = .72$, $\beta = .33$, $p = .020$, in the final model. This moderation effect supported hypothesis 2: There were significant medium sized relationships between the SweSAT test scores and the performance assessments in the non-stereotype threat conditions, but no relationship in the stereotype threat conditions (see Table 3 for correlation statistics and Figure 2 that displays the moderation). This means that in the women's stereotype threat condition, only the men had a relationship between their performance on the SweSAT and their assessment of it. And in the men's stereotype threat condition, only the women had a relationship between their performance on the SweSAT and their assessment of it. This perhaps indicates that stereotype threat confuses our ability to assess our test performance.

Discussion

The results of the current study imply that performance assessments may become confused under stereotype threat. In a stereotype threat condition, there was no relationship between high school students' performance on the SweSAT and their assessments of how well they performed. This contrasted with the non-stereotype threat conditions where there were moderate sized relationships between the students' performance and their assessment of it. The results were similar for both men and women, which strengthens the result.

To our knowledge this is the first study that shows that stereotype threat may disrupt performance judgments. We will discuss how this result perhaps can contribute to explanations of why women suffer from competence doubts more often than men (Betz & Hackett, 1981; Bridges, 1988; Matsui, et al., 1989; Tellhed, et al., 2016).

Women's stereotype threat

Women are more susceptible to stereotype threat as compared to men, as it is often implicitly activated in women, while men need explicit statements of inferiority to be threatened (Nguyen, & Ryan, 2008; Pillaud et al., 2015). In the current study, we replicated previous research that finds gender differences in favor of men on the SAT (Walton & Spencer, 2009; Steele 2012). Also in the current study, the women performed slightly worse than the men on the SweSAT, when no gender stereotypes were mentioned. The gender differences on the SAT contrast with women's otherwise superior academic performance (e.g. OECD, 2015) and has previously been explained as a stereotype threat effect (Walton & Spencer, 2009; Steele 2012).

The results of the present study expand the literature by demonstrating that stereotype threat not only affects performance, but might also confuse performance assessments. Interestingly, in the condition where gender was not mentioned (women's implicit stereotype threat condition), only the men had a relationship between test performance and performance assessment.

This finding might help us better understand women's competence doubts in male-stereotyped domains (Bench et al., 2015; Watt, 2010). If stereotype threat confuses individuals' performance assessments, and women frequently experience stereotype threat, then women's competence beliefs may become disconnected from their ability in male-stereotyped domains. If women are unsure of their individual performance in male-stereotyped domains, their competence beliefs may become susceptible to gender bias from

the commonly known stereotypes that portray women as less competent than men (Fiske, et al., 2002; Halpern, 2012). Thus, a long-term outcome of the confusing effects of repeated stereotype threat exposure might be that women come to underestimate their ability in male-stereotyped domains. Paraphrasing Bandura (1997): Interpretation of our performance is the most important influence on competence beliefs (self-efficacy). Future studies may want to examine if women's overall competence judgments in male-stereotyped domains relate to the confused performance assessments under stereotype threat that we have demonstrated here.

In the current study we further attempted to remove women's implicit stereotype threat (and instead threaten men) with the explicit prime that women do better than men in school. This could be categorized as a *stereotype lift/boost* prime for women, in that it portrays them as superior to the outgroup men (Shih, Pittinsky, & Ho, 2012; Walton & Cohen, 2003). In the literature, stereotype lift/boost sometimes improves performance, but tends to have smaller effect sizes than stereotype threat (Shih et al., 2012; Walton & Cohen, 2003). However, like in the study by Hartley and Sutton (2013), the prime did not affect the women's test performance nor did it affect the participants mean performance assessments. One speculation is that a single exposure to an explicit prime of female superiority might be a too weak manipulation to completely lift women's implicit stereotype threat experience.

Importantly though, the explicit prime seemed to make the women's individual performance assessments match their performance level better. Thus, hearing that women do better than men in school seemed to make the women less confused as to how well they individually performed on the SweSAT. Presence of the prime made the women's relationship between performance and performance assessment similar to the men's relationship in their non-stereotype threat condition.

Men's stereotype threat

Men are stereotypically associated with high competence (Fiske et al., 2002) and subsequently suffer less frequently from gender-related stereotype threat, as compared to women (Nguyen, & Ryan, 2008; Pillaud et al., 2015). However, if explicitly primed with a negative in-group stereotype, men can also experience stereotype threat (Hartley & Sutton, 2013; Koenig & Eagly, 2005; Nguyen, & Ryan, 2008; Pillaud et al., 2015). In the current study, explicitly stating that women do better in school eliminated the commonly found gender difference on the SweSAT. Like in the study by Hartley and Sutton (2013), the men's scores decreased by this prime.

Even though men seldom suffer from gender-related stereotype threat, there is a risk that an increased awareness of women's' superior performance in school (OECD, 2015), may cause more frequent stereotype threat experiences for men in school. Young men's underperformance in school has recently come into public attention (SOU 2014:6), and an increased stereotype threat risk for men might further increase the gender gap.

Like the women under implicit stereotype threat, the men who were primed with stereotype threat also seemed confused in their performance assessments. When women's superior performance was mentioned, the men had a zero correlation between their performance on the SweSAT and their performance assessment.

If stereotype threat becomes more frequent for young men in the future, will men's competence doubts increase to similar levels as women's? Not necessarily. Our results demonstrate that the male participants judged their performance as consistently better than the female participants', also in the condition that lacked a gender difference in performance (men's stereotype threat). This does not tell us how repeated exposure to stereotype threat may affect men's long-term competence beliefs. However, we speculate that the common association of men with competence (Fiske, et al., 2002), may have contrasting long-term influences on men's versus women's competence beliefs, when confused as to how well one performs. If performance assessment relies on gut-feelings during stereotype threat, they may become biased by gender stereotypes, ultimately causing men's guesses to exceed women's.

Limitations and future directions

Since this is the first study suggesting that performance assessment is affected by stereotype threat, the result needs to be interpreted with caution until thoroughly replicated. More research is also needed to investigate what the processes are that lead up to the effect. We have speculated that the cognitive overload, commonly associated with stereotype threat, perhaps makes threatened individuals rely more on automatic processing for their performance assessments, in contrast to careful controlled consideration. If people rely on their "gut feeling" of how well they performed on a test, this can mismatch their actual performance, and be susceptible to bias by for instance salient gender stereotypes. This hypothesis must of course be tested empirically and with care as to what the key components are in the process. Alternative explanations are conceivable: Does the confusion stem from a lack of monitoring of one's performance during the test, or from inadequate memory encoding of the assessment, or perhaps from biased memory retrieval of it, after the test? And

if performance assessment is biased, what are the main bias factors? We have discussed salient stereotypes as a source of bias, but it could also be, for instance, anxiety that distorts assessment.

Further, it is important to point out that it is difficult to assess one's performance even in the absence of stereotype threat (Kruger & Dunning, 1999). Relatedly, the relationship between performance and performance assessments were not strong in the non-threat conditions in the current study, but only moderate in size (see Table 3).We recommend that replications use complementary measures of performance assessments, like having the participants' estimate their number of correct answers on the test, and perhaps their performance rated in comparison to different standards; other class pupils, or the general public. A more precise performance assessment measure could shed light upon the question as to whether the participants perhaps over- or underestimate their performance during stereotype threat, and if this shows a gender difference, as previous studies have found (Bench et al., 2015; Watt, 2010).

Future studies may also want to be creative in the construction of control/comparison groups to the stereotype threat conditions. In the current design we compared men's and women's stereotype threat to presumed non-stereotype threat conditions (i.e. control) for the respective genders. However, as we've discussed, one could argue that our non-threat conditions are rather stereotype boost/lift conditions (explicit for women and implicit for men, Shih et al., 2012; Walton & Cohen, 2003). This opens for the possibility to interpret our results as caused by either stereotype threat, by stereotype boost/lift, or by a combination of both effects. To determine this, future studies may want to add a non-threat control condition that attempts to nullify all gender stereotypes. This is difficult as gender often is our most accessible social identity (Rudman & Glick, 2008; Stangor, Lynch, Duan, & Glass, 1992), but one may for instance attempt to state, in a non-conspicuous manner, that the test typically shows no gender differences (see Murphy & Jones Taylor, 2012 for a review).

Lastly, as described, not all participants declared that they had registered the oral prime correctly. Since we failed to ask if they remembered the written prime, we cannot determine if these participants registered it correctly, or missed both primes. We therefore advise future replications to be cautious in the manipulation checking.

Conclusions

Stereotype threat has many adverse effects for negatively stereotyped groups. In addition to disrupting test performance, the present study shows that stereotype threat may also confuse people's performance assessments. Lowering the relationship between performance and performance assessment may have different implications for men versus women in the long-term: While it perhaps contributes to the development of women's gender-stereotypical competence beliefs, men's competence beliefs might be protected by the strong association of men with competence.

References

Adolfsson, C. (2015). *Saying "neigh" to being a study horse: Creation of the Anti-studying Norm Scale (ANS).* Unpublished master's thesis, Lund University, Lund, Sweden

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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, (5). 339. doi: 10.1111/j.1467-8721.2008.00602.x

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

Bench, S. W., Lench, H. C., Miner, K., Flores, S. A., & Liew, J. (2015). Gender gaps in overestimation of math performance. *Sex Roles*, 536-546. doi: 10.1007/s11199-015-0486-9

Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology, 28*(5), 399-410. doi: 10.1037//0022-0167.28.5.399.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*(3), 329-345. doi: 10.1016/0001-8791(83)90046-5

Bodenhausen, G. V., & Todd, A. R. (2010). Automatic aspects of judgment and decision making. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. (pp. 278-294). New York: The Guilford Press.

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1990). On the dialects of discrimination: Dual processing in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 271-290). New York: The Guilford Press

Bridges, J. S. (1988). Sex differences in occupational performance expectations. *Psychology of Women Quarterly, 12*(1), 75-90. doi: 10.1111/j.1471-6402.1988.tb00928.x

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*(2), 218-261. doi: 10.1037/a0014412

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. doi: 10.1037/a0018053

European Commission. (2014). *A new method to understand occupational gender segregation in European labour markets*. Luxembourg: Publications Office of the Euoropean Union Retrieved from http://ec.europa.eu/justice/gender-equality/files/documents/150119_segregation_report_web_en.pdf.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902. doi: 10.1037/0022-3514.82.6.878

Hackett, G. (1995). Self-efficacy in career choice and development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 232-258). New York: Cambridge University Press.

Halpern, D. F. (2012). *Sex differences in cognitive abilities*. New York, Hove: Psychology Press Taylor & Francis Group.

Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development, 84*(5), 1716-1733. doi:10.1111/cdev.12079

Hassin, R. R. (2005). Nonconscious control and implicit working memory. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconsious* (pp. 196-222). Oxford: Oxford University Press.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592. doi: 10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*(1), 373-398. doi: 10.1146/annurev-psych-010213-115057

Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford University Press.

Johns, M. J., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137*(4), 691-705. doi:10.1037/a0013834

Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles, 52*(7/8), 489-496. doi: 10.1007/s11199-005-3714-x

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80*(6), 942-958. doi:10.1037/0022-3514.80.6.942

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A Functional Magnetic Resonance Imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science,* (2). 168-175. doi: 10.1111/j.1467-9280.2008.02063.x.

Kruger, J., & Dunnig, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi: 10.1037/0022-3514.77.6.1121

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79-122. doi: 10.1006/jvbe.1994.1027

Matsui, T., Ikeda, H., & Ohnishi, R. (1989). Relations of sex-typed socializations to career self-efficacy expectations of college students. *Journal of Vocational Behavior, 35*(1), 1-16. doi: 10.1016/0001-8791(89)90044-4

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, *27*486-493. doi:10.1016/j.appdev.2006.06.003

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambrige University Press.

Murphy, M. C., & Jones Taylor, V. (2012). The role of situational cues in signaling and maintaining stereotype threat. In M. Inzlicht, & Schmader, T. (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34-50). New York: Oxford University Press.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314-1334. doi:10.1037/a0012702

OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. Pisa, OECD Publishing Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-gender-eng.pdf.

Pillaud, V., Rigaud, D., & Clémence, A. (2015). The influence of chronic and situational social status on stereotype susceptibility. *PLOS ONE, 10*(12), 1-15. doi:10.1371/journal.pone.0144582

Prentice, D., & Carranza, E. (2001). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, *26*(4), 269-281. doi: 10.1111/1471-6402.t01-1-00066

Rudman, L. A., & Glick, P. (2008). Development of gender relations. In L. A. Rudman & P. Glick (Eds.), The social psychology of gender: How power and intimacy shape gender relations (pp. 54–80). New York: The Guilford Press.

Schmader, T., & Beilock, S. (2012). An integration of processes that underlie stereotype threat theory. In M. Inzlicht, & Schmader, T. (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34-50). New York: Oxford University Press.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440-452. doi:10.1037/0022-3514.85.3.440

Shih, M. J., Pittinsky, T. L., & Ho, G. C. (2012). Stereotype boost. In M. Inzlicht, & Schmader, T. (Eds.), *Stereotype threat: Theory, process, and application* (pp. 141-156). New York: Oxford University Press.

Sidanius, J., Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

SOU (2014:6). *Män och jämställdhet. Betänkande av Utredningen om män och jämställdhet*. [Men and Gender Equality. Report from the Government Inquiry of Men and Gender Equality]. Stockholm: Fritzes offentliga publikationer. Retrieved from http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2014/02/sou-20146/

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415-437. doi: 10.1146/annurev-psych-073115-103235

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62*(2), 207-218. doi: 10.1037/0022-3514.62.2.207

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629. doi:10.1037/0003-066X.52.6.613

Steele, C. M. (2012). Conclusions. In M. Inzlicht, & Schmader, T. (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34-50). New York: Oxford University Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811. doi:10.1037/0022-3514.69.5.797

SweSAT. (2013). *The Swedish Scholastic Aptitude Test – SweSAT*. Retrieved from http://www.edusci.umu.se/english/swesat/?languageId=1

Tellhed, U., & Bjöklund, F. (2011). Stereotype threat in salary negotiations is mediated by reservation salary. *Scandinavian Journal of Psychology*, *52*(2), 185-195. doi:10.1111/j.1467-9450.2010.00855.x

Tellhed, U., Bäckström, M., & Björklund, F. (2016). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex Roles*. doi: 10.1007/s11199-016-0694-y

Walton, G. M., & Cohen, G. L., (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-467. doi: 0.1016/S0022-1031(03)00019-2

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, (9). 1132-1139. doi: 10.1111/j.1467-9280.2009.02417.x.

Watt, H. G. (2010). Gender and occupational choice. In J. C. Chrisler & D. R. Watt (Eds.), *Handbook of gender research in psychology* (pp. 379-400). New York: Springer.

Weber, D., Skirbekk, V., Freund, I., & Herlitz, A. (2014). The changing face of cognitive gender differences in Europe. *Proceedings of the National Academy of Sciences, 111*(32), 11673-11678. doi:10.1073/pnas.1319538111

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study*. Newbury Park, CA: Sage.

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive & Affective Neuroscience*, *2*(1), 12-19. doi:10.1093/scan/nsl041.

Table 1
Descriptive statistics on the SweSAT score and performance assessments for men's stereotype threat (ST) and women's stereotype threat.

		Men		Women	
Variable	Prime	*M*	*SD*	*M*	*SD*
SweSAT score					
	Men's ST	5.52	2.15	5.96	2.10
	Women's ST	6.36	2.32	5.53	1.90
Perform. asst.					
	Men's ST	3.02	1.45	2.50	1.09
	Women's ST	3.00	1.24	2.01	1.04

Table 2

Results from the final model (5) of the hierarchical regression of the standardized SweSAT-score, Gender, Condition, and their interactions on participants' performance assessments. R^2 change by the 5 models.

	B (SE)	β	*p*	*CI 95 % for B*	*ΔR^2*
SweSAT	.560 (.179)	.448	.002	.208, .911	
Gender	-.857 (.224)	-.334	<.001	-1.298, -.415	
Condition	.135 (.243)	.054	.580	-.345, .615	.149**
Gender×Condition	.290 (.310)	.107	.350	-.320, .900	.004
Gender×SweSAT	-.417 (.230)	-.256	.071	-.870, .037	<.001
Condition×SweSAT	-.520 (.238)	-.308	.030	-.989, -.051	.001
Gender×Cond×SweSAT	.724 (.309)	.328	.020	.116, 1.332	.020*

Note: */** = Significant at the .05 /.001 level.

Table 3
Correlations between SweSAT-scores and performance assessments by condition (Men's stereotype threat (ST) vs. Women's stereotype threat) and participant gender. The relationship is only significant when a gender group is not under stereotype threat.

Prime	Gender	*r*	*p*	*n*
Men's ST	Men	.023	.876	49
Men's ST	Women	.326	.005	74
Women's ST	Men	.445	.003	43
Women's ST	Women	.136	.249	74

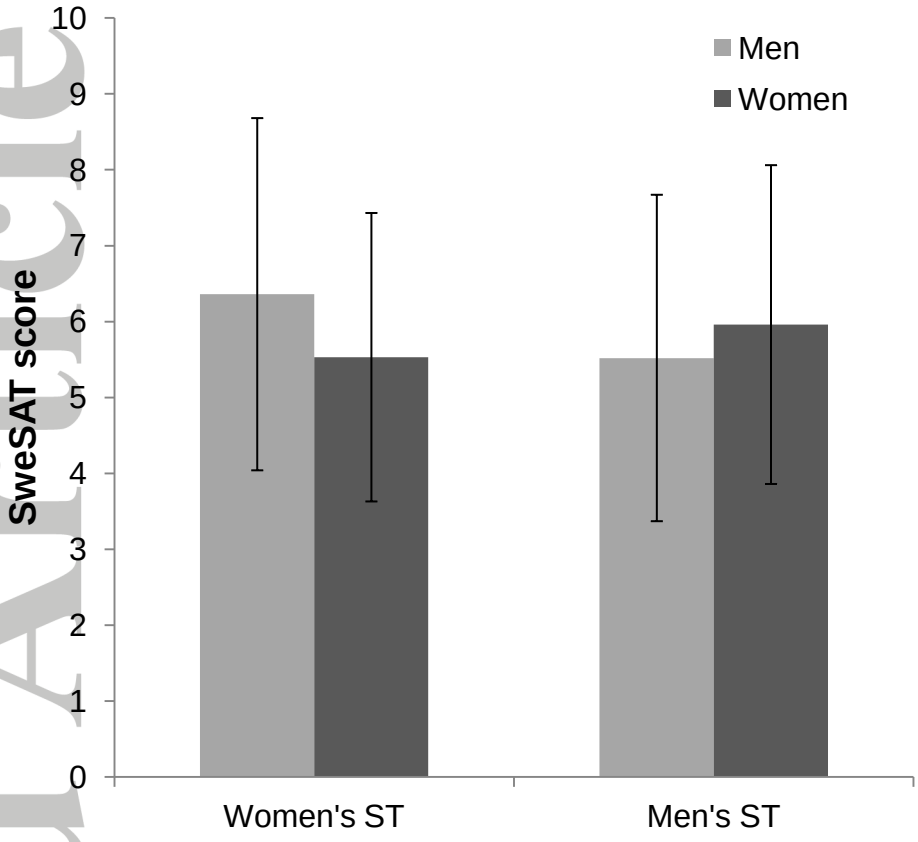

Figure 1. *SweSAT score for men and women in the women's stereotype threat (ST) condition and men's stereotype threat condition.*

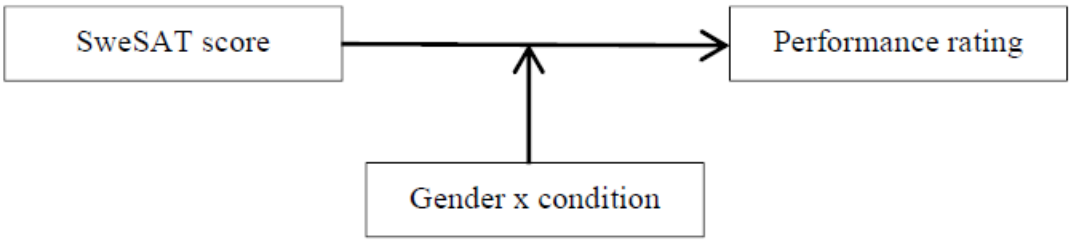

Figure 2. *The relationship between the participants' score on the SweSAT and their ratings of their performance as moderated by the interaction of participant gender and the experimental condition. While under stereotype threat, neither men's nor women's test scores related to their performance assessment, which was otherwise related.*