Scandinavian Journal of Psychology, 2007, 48, 329–338 DOI: 10.1111/j.1467-9450.2007.00588.x

Personality and Social Sciences

Making gender matter: The role of gender-based expectancies and gender identification on women's and men's math performance in Sweden

KIMMO ERIKSSON and TORUN LINDHOLM

Mälardalen University, Västerås, Sweden

Eriksson, K. & 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. *Scandinavian Journal of Psychology*, *48*, 329–338.

It is well established that an emphasis on gender differences may have a negative effect on women's math performance in USA, Germany and the Netherlands. It has further been found that an individual's identification with the stereotyped group may moderate effects of negative stereotypes. The present study investigated how gender-based expectancies affected the math performance of women and men in Sweden, a nation with a smaller gender gap than in other countries, and a strong cultural emphasis on gender equality. Participants, 112 female and 74 male undergraduate math students from Swedish universities, completed a difficult math test in which their gender was either linked to their test performance or not. Men performed better than women when gender was made relevant among participants who did not see their gender as an important aspect of their identity, while participants high in gender identification were unaffected by gender identity relevance. Moreover, the gender relevance manipulation affected men's performance more than women's. The results deviate from findings on US samples, indicating that the role of group identification as a moderator of stereotype-based expectancy effects is complex, and that factors in the cultural context may interact with individual differences in identification to determine the impact of negative stereotypes.

Key words: Gender-based expectations, math performance, gender identification, intercultural differences.

Torun Lindholm, Department of Psychology, Stochholm University, SE-106 91 Stochholm, Sweden. Tel: +46-8-16 40 72; fax: +46-8-15 93 42; e-mail: tlm@psychology.su.se

INTRODUCTION

Do men perform better than women at math tests? The correct answer is: "It depends." For instance, it is well known that it may depend on where we look in the distribution of test scores. According to the latest report from PISA, the OECD program for international student assessment (OECD, 2004, p. 98), gender differences in math performance (favoring boys) tend to be small, and are much smaller than the gender differences in the area of reading literacy (favoring girls). On the other hand, among the OECD students scoring at the highest of the six levels of math performance differentiated by the study, there are almost twice as many boys as girls (PISA, 2004, table 2.5b).

While this distributional variation is interesting and important, it is not the full story. Over the past years, a growing body of research has demonstrated that gender differences in math performance may also be created, perpetuated, or eliminated by factors in the test situation itself. Specifically, research has shown that the activation of the negative stereotype of women's abilities in the math domain can lead women to underperform on challenging math tests (Inzlicht & Ben Zeev, 2000; Keller & Dauenheimer, 2003; Marx, Stapel & Muller, 2005; Spencer, Steele & Quinn, 1999; Schmader, 2002; Schmader, Johns & Barquissau, 2004). For example, in a study by Spencer and colleagues (1999), male and female US students who were highly skilled at mathematics completed a difficult math test administered in two different test situations; participants were informed in advance either that the test had been shown to produce gender differences, or that it had been shown *not* to produce such differences. The results showed that when told that the test had shown gender differences, women performed more poorly than men. On the other hand, when told that the test was insensitive to gender differences, women in this study performed equal to men. The phenomenon, known as *stereotype threat* (Steele & Aronson, 1995), is assumed to occur when an individual experiences the risk of confirming a negative stereotype about the in-group. Thus, reminding women of the negative stereotype about women's math abilities constitutes a "threat in the air", which signals the possibility of the individual being treated or judged in terms of the stereotype. This threat, and the pressure it causes, may inhibit individuals from performing to their potential.

German and Dutch studies (e.g. Marx *et al.*, 2005; Keller and Dauenheimer, 2003) have shown that stereotype threat effects on women's math performance are not exclusive to the USA. However, to our knowledge no study of this phenomenon has been made in Sweden or any other Scandinavian country. A recent report of World Economic Forum (Lopez-Claros and Zahidi, 2005), measures the size of the "gender gap" in 58 countries. The report measures the

extent to which women have achieved full equality with men in economic participation and opportunities, political empowerment, educational attainment, and health. According to this report, Sweden has the smallest gender gap in the world, followed by the other Scandinavian countries. Germany, the Netherlands and the USA are considerably less gender equal. *A priori*, it seems quite plausible that such cultural differences in gender equality can play a role for gender stereotype threat effects.

For example, one possibility is that the cultural emphasis on gender equality in Scandinavia may have led to a weakening of gender stereotypes. In a recent survey to high school students in Sweden (Brandell, Larsson, Nyström, Palbom, Staberg & Sundqvist, 2005), most students were neutral when asked whether the statement "Think math is difficult" applies more to one sex than the other. However, while 3% responded "women, definitely" and another 25% "women, perhaps", only 7% chose one of the two alternatives that stressed men's difficulties with math. The pattern was the same for male and female respondents. Furthermore, while the proportion of women taking college programs or university courses heavy in mathematics has increased in Sweden during the past decade, the vast majority of math students in higher education are still men. Thus, although the majority does not seem to endorse stereotypes regarding the math abilities of women and men, there are certainly differences in the way Swedes associate mathematics with men and women, respectively.

Even if the math and gender stereotype is sufficiently pervasive in Sweden, the Swedish norm of gender equality may have other consequences for the way both women and men perceive and react to stereotype-based expectations, making generalizations of findings across cultures unwarranted. For instance, compared to women in other countries Swedish women may have a stronger sense that they have the same general rights to opportunities and training as men, and such a belief may function as a protection against negative effects of the math stereotype. In the same vein, compared to men in other countries Swedish men may perceive the male gender role as less stable and secure, which could make them more sensible to stereotype-based performance expectations. Such effects could go both ways. Some studies have found that members of a positively stereotyped group may show a performance boost, or *stereotype lift* (Walton & Cohen, 2003), when the stereotype is made salient. But high public performance expectancies can also lead to suboptimal performance, *choking under pressure* (Baumeister *et al.*, 1985). For instance, Beilock, Kulp, Holt and Carr (2004) found that performance on working memory-demanding math problems declined for participants who were solving the problems in a high-, as compared to low-pressure environment (whereas performance on well-learned arithmetic problems was unaffected by pressure).

The aim of the current study was to examine the effects of stereotype-based gender expectations on women's and men's math performance in Sweden. We report on an experimental study carried out on Swedish students taking college math courses. Our design closely follows that developed by Schmader and colleagues (2002, 2004) in terms of math test and threat manipulation. The essence of the manipulation is that subjects are told either that the researchers are interested in how women score on the test relative to men, using each individual's score as an indicator of women's or men's math ability in general (the "gender identity relevant" condition), or that the test has previously exhibited no sex differences and that the researchers are not looking for any (the "gender identity not relevant" condition).

Gender-based expectancies in math performance: The role of gender identification

Stereotype threat effects have been replicated among several groups and in a variety of domains. While the effects of gender stereotypes on women's math performance probably is one of the most well documented, activating a negative stereotype has been found to decrease the performance of African-Americans and Latinos in intellectual domains, European Americans in sports, and elderly in memory tasks. Although the effect has been replicated many times, the phenomenon is only partially understood. A number of possible moderating and mediating factors have been proposed and studied. For math performance, these include self-reported effort (Aronson, Lustina, Good, Keough, Steele & Browne, 1999), self-handicapping (Keller & Dauenheimer, 2003), anxiety (Aronson *et al.*, 1999), self-esteem (Oswald & Harvey, 2001), stereotype endorsement (Schmader *et al.*, 2004), and group identification (Schmader, 2002).

While not all of the proposed mechanisms have received empirical support, one factor that has been found to moderate effects of stereotype-based expectancies is *group identification* (Schmader, 2002). Group identification refers to the extent to which individuals consider their social groups to be an important aspect of their self-concept. It has been conceptualized as part of an individual's social identity (Tajfel & Turner, 1986), and is often measured by one of the subscales of the Collective Self-esteem Scale developed by Luhtanen and Crocker (1992). Schmader (2002) reworded this scale to assess the perceived importance of gender identity to self-definition, and investigated the role of gender identification on women's vulnerability to stereotype threat effects in a sample of white US participants. In her study, Schmader demonstrated that the negative stereotype threat effect on women's math performance only affected those women who were strongly identified with their gender. According to Schmader, the explanation to this effect of gender identification may be that women who are highly identified with their gender are more motivated to maintain a positive image of that identity than low identifiers. Hence, highly gender identified women may experience greater threat, and resulting impairments at the suggestion that their in-group is inferior to

other groups. However, in the literature on stereotyped minorities, there is also evidence suggesting that high group identification can help offset the effects of stereotype threat by providing social support and motivation to challenge the stereotype (Branscombe, Schmitt & Harvey, 1999; Ellemers, Spears & Doosje, 1997; Spears, Doosje & Ellemers, 1997; Wong, Eccles & Sameroff, 2003). Of particular interest in the current context is a study by Cohen and Garcia (2005) on collective threat, a version of stereotype threat. In a series of experiments, these authors found various protective effects of group identification among black students, but also a tendency towards a protective effect of gender identification among women in a mathematical setting.

From the extant literature then, it appears that high gender identification can serve either as a buffer against, or as an amplifier of women's vulnerability to math stereotype threats. One factor that could affect whether gender identification will harm or protect against threat effects, is the gender equality norm in a given context. Specifically, it seems reasonable that those women who are highly identified with their gender are also those who most strongly believe in women's equal rights. In a society as Sweden, where gender equality has long been emphasized, women who hold such beliefs may experience that their claim of equality has strong societal support. Such perceived support may serve as a protection against negative stereotype-based expectations, and may even make those women view a potential threat as an opportunity to defend the image of their group, and challenge the negative stereotype (Cohen & Garcia, 2005).

In a society with strong gender equality norms, where men may often perceive their gender identity threatened, it also seems conceivable that an emphasis on gender differences in a situation may not have uniformly positive effects on men's performance. In particular, one could speculate that men who strongly identify with being men may be especially likely to choke under pressure caused by expectancies related to their gender in such situation. In the present study, we set out to test the moderating role of gender identification to threat effects on women's and men's math performance in a sample of Swedish math students.

Stereotyped-based expectancies and social monitoring

Stereotype threat is assumed to be caused by the recognition of cues in the social context that activate an individual's beliefs and fears of how others view their group and, by extension, themselves. However, people differ in the extent to which they care about social cues, or are sensitive to other persons' reactions. Some individuals are highly sensitive to other people and alter their responses to others' cues. Other persons do not pick up cues from the situation as easily, but rather use their own values and motives across situations to guide their behavior. It seems reasonable that such difference in *social monitoring* (Snyder, 1974) may influence the effects of stereotype threat. Specifically, one could speculate that individuals high in social monitoring would be those most sensitive to cues about other's expectations on their performance, and hence those most vulnerable to stereotype threat effects. On the other hand, it also seems possible that people high as compared to low in self-monitoring may find it easier to cope with the risk of having others judge them stereotypically, because high self-monitors are more accustomed to deal with self-representational concerns. Thus, it is possible that high self-monitors are those most resilient to stereotype threat effects. In fact, this latter notion received support in a recent study by Inzlicht and colleagues (2006), where high self-monitoring was found to predict better performance in a threat situation, while having no effect on participants' performance in a no-threat condition. In the present study we tried to replicate this result by investigating whether responses to stereotype threat varied as a function of an individuals' level of social monitoring.

Stereotyped-based expectancies and stereotype endorsement

Some theorists have argued that to experience stereotype threat, one need not believe the stereotype nor even be worried that it is true of oneself. Thus, the mere awareness that a negative stereotype might be applied to one's performance, and not an individual's own endorsement of the stereotype, is sufficient for stigmatized individuals to experience stereotype threat (Steele, 1997; Steele, Spencer & Aronson, 2002; Leyens, Desert, Croizet & Darcis, 2000). However, research does not uniformly support this contention. For example, Schmader *et al.* (2004) showed that women who indicated belief that the stereotype about women's math abilities might be accurate performed worse when their gender identity was made salient during a math test than when gender was not mentioned. However, women who rejected the stereotype showed no effect of stereotype threat at all. While this finding indicates that individual differences in stereotype endorsement may moderate the effects of stereotype threat, the empirical evidence for the moderation is still limited. A final aim of the current study was to test the reliability of Schmader *et al.*'s (2004) findings by examining the relation between stereotype endorsement and stereotype threat.

To summarize, the main purpose of the present study was to examine whether previous findings of stereotype threat effects on women's math performance would be obtained in a sample of Swedish students taking college math courses. Based on the assumption that the gender stereotypes about math abilities in Sweden largely resemble those prevalent in other Western societies, one would expect Swedish women to perform more poorly relative to men when their gender identity was linked to their performance on a math test, than when gender is not relevant. However, because the strong gender equality norm in Sweden may have made both Swedish women and men react differently to math stereotype threat than participants in other countries, this prediction was only tentative. In line with previous findings (Cohen &

Garcia, 2005; Schmader, 2002), we further hypothesized that potential effects of gender identity relevance on women's math performance would be moderated by group identification. However, rather than increasing vulnerability to gender relevance effects, we tentatively expected high gender identification among women in Sweden to protect against such effects. Moreover, we tested the hypothesis that high gender identification among Swedish men would be associated with performance decrement when gender identity is made relevant to test performance, as compared to when it is not. We further hypothesized that women high in self-monitoring would experience less stereotype threat than women low in self-monitoring, and that women who endorsed the stereotype about women's poor math abilities would show stronger effects of stereotype threat than women who rejected this stereotype.

METHOD

Participants and design

Participants were 112 female and 74 male Swedish first-year undergraduates between 18 and 42 years (*M* and *SD* = 23.8 (6) years) taking university mathematics courses. Participants were recruited in math classes at four Swedish universities (Linköping University, Mälardalen University, Stockholm University and the Royal Institute of Technology). All participants were studying to become either chemistry engineers (67 men, 70 women) or math teachers (7 men, 42 women), these being the only two major Swedish university programs heavy on mathematics where women are at least as common as men. Participants were randomly assigned to one of two conditions in which their gender identity was either linked to their performance on a math test or not (gender identity relevant/gender identity not relevant). Two different procedures for random assignment were used. For the majority of participants, a class was randomly split in two groups and each group was then assigned at random to one of the two experimental conditions. In classes with few participants, the entire group was assigned to the same condition (determined randomly).[1]

All analyses in the study included participants' gender and experimental condition (gender identity relevant/not relevant) as independent variables. A third predictor was in separate analyses participants' gender identification, self-monitoring, and stereotype endorsement. Gender identification and stereotype endorsement were continuous variables in the design, whereas social monitoring was analyzed as a categorical variable (high/low).

Procedure

Participants were tested in their ordinary classrooms during math classes. Teachers of these classes had been informed in advance that a researcher wanted to use a half-hour to administer a math test as part of a research project. A male experimenter arrived at the classroom and introduced himself as a researcher developing a standardized math exam. After he explained that participation was voluntary and that all individual results would be confidential, he read out loud instructions adopted from Schmader (2002). In both experimental conditions, he explained that he was interested in each individual's performance on the test and that he would be comparing participants' individual scores to those of other students, using test scores as an indicator of personal math ability. Participants were asked to write their first initials and last names on the cover sheets of their tests. In the Gender Identity Not Relevant condition the researcher mentioned that this test had previously exhibited no sex differences, whereas in the Gender Identity Relevant condition the researcher went on to explain that he was also interested in how women score on the test relative to men, using each individual's score as an indicator of women's or men's math ability in general. In line with this instruction, participants in the Identity Relevant condition identified their gender on the cover sheets of their tests. In the Identity Not Relevant condition gender was only asked for in the ensuing questionnaire.

After finishing the math test, participants completed a questionnaire, the contents of which are described below. Participants received a cinema ticket voucher (valued at approximately US$10) in exchange for their participation.

After the last experimental session a letter was sent to teachers and participants in which the true purpose of the study was explained.

Math test

The test consisted of 20 multiple-choice questions, translated and adapted from the set of questions used by Schmader (2002). These questions are word problems with a potentially frustrating quality: although only very elementary mathematics (basic arithmetic) is required, the reasoning needed to formulate which calculations to perform is non-trivial. The adaptation consisted of simplifying the numbers (while maintaining conceptual difficulty), since otherwise Swedish students seem to be at loss if they are not allowed to use a calculator.

Gender identification

A Swedish translation of the four-item identity importance subscale from the Collective Self-Esteem Scale as modified by Schmader (2002) was used to assess the perceived importance of gender identity to self-definition. Participants rated the following four items on a scale ranging from 1 (strongly disagree) to 5 (strongly agree): "Being a woman/man is an important part of my self-image," "Being a woman/man is unimportant to my sense of what kind of person I am" (reverse scored), "Being a woman/man is an important reflection of who I am," and "Being a woman/man has very little to do with how I feel about myself" (reverse scored). Participants' responses to the four items were averaged to form an index of gender identification ($\alpha = 0.66$). A *t*-test showed that women scored marginally higher than men on this gender identification index (*M*s and range for men and women were 3.21 (1.25–5), and 3.45 (1–5), respectively), $t(184) = -1.88$, $p = 0.06$.

Math grade from high school

In Swedish secondary schools the system for grading was changed in the 1990s, from numbers 1–5, with higher numbers indicating better grades, to letters (IG and G indicating low grades; VG indicating medium grade; and MVG indicating highest grade). In line with how comparisons are usually made between these grade systems, we have mapped the *highest* grades (5 and MVG) together, the *medium high* grades (4 and VG) together, and the *low* grades (1, 2, 3, IG, G) together.

Stereotype endorsement

Stereotype endorsement was assessed with a Swedish translation of the three items used by Schmader (2004): "It is possible that

men have more math ability than do women", "In general, men may be better than women at math", and "I don't think that there are any real gender differences in math ability". Responses were made on a scale from 1 (strongly disagree) to 5 (strongly agree). After reversing responses to the last item, all three items were averaged to construct a stereotype endorsement index ($\alpha = 0.65$).

Self-monitoring scale (SMS)

The SMS, developed originally by Snyder (1974), and later modified by Gangestad and Snyder (1985; Snyder and Gangestad, 1986) was translated to Swedish and used to assess self-monitoring. The scale consists of 18 items to which participants respond "True" or "False". Sample items included "I find it hard to imitate the behavior of other people", and "I am not particularly good at making people like me". The scale showed good internal reliability ($\alpha = 0.73$). In accordance with the recommendations of Gangestad and Snyder (1985), a median split ($Mdn = 9.00$) was used to classify participants into groups of low (24 men and 59 women, $M = 5.77$, $SD = 1.88$) and high (50 men and 53 women, $M = 11.45$, $SD = 2.02$) self-monitors.

Manipulation check

A check of the experimental manipulation was measured using two items, $r = 0.70$, $p < 0.001$: "I am concerned that people will judge persons of my gender as a whole based on my performance on this test" and "People will think that persons of my gender as a whole have less ability if I did not well on this test". Responses were made on the same five-point scale as was used for gender identification and stereotype endorsement.

RESULTS

Prior to the main analyses, a set of ANOVAs confirmed that participants in the two experimental conditions (Gender identity relevant/not relevant) did not differ in any of the independent variables (gender identification, $t(184) = -0.52$, $p = 0.60$; math grades from high school, $t(184) = 1.15$, $p = 0.25$, social monitoring, $t(184) = 0.74$, $p = 0.42$; stereotype endorsement, $t(184) = 1.05$, $p = 0.29$). Analyses further revealed that women had slightly lower math grades than men from high school (*M*s and *SD*s for men were 4.14 (0.82), and for women 3.92 (0.81), respectively), $F(1, 184) = 3.15$, $p = 0.078$, and that men scored significantly higher than women in social monitoring (*M*s and *SD*s for men were 9.76 (3.10), and for women 8.37 (3.55), respectively), $F(1, 184) = 6.08$, $p = 0.015$. Although men tended to endorse the negative stereotype of women's math abilities to a higher degree than women (*M*s and *SD*s for men were 2.20 (1.04), and for women 1.96 (0.94), respectively), this difference did not reach significance, $p = 0.11$. There was no correlation between gender identification and math grade from high school, $r = -0.06$, $p = 0.43$, or social monitoring, $r = 0.036$, $p = 0.67$. The correlation between gender identification and endorsement of gender differences in math ability was non-significant among both men, $r = 0.18$, $p = 0.12$, and women, $r = 0.12$, $p = 0.20$.

Manipulation check of gender identity relevance

A 2 (participant gender) × 2 (experimental condition: Gender identity relevant/not relevant) ANOVA showed that women experienced more threat overall ($M = 4.10$, $SD = 2.5$) than did men ($M = 2.86$, $SD = 1.57$), $F(1, 177) = 11.53$, $p = 0.001$. Confirming the effectiveness of the experimental manipulation, women reported higher levels of threat in the Gender identity relevant ($M = 4.51$, $SD = 2.74$) than in the Identity not relevant condition ($M = 3.68$, $SD = 2.17$), $F(1, 177) = 4.08$, $p = 0.045$. Men showed similar threat levels regardless of experimental condition (gender identity relevant: $M = 2.74$, $SD = 1.39$; not relevant: $M = 3.00$, $SD = 1.77$), $p = 0.49$. There were no significant relations among women between perceived threat and math performance, $r = -0.14$, $p = 0.15$, nor between perceived threat and gender identification, $r = 0.14$, $p = 0.14$.

Gender-based expectancies in math performance: The role of gender identification

Math test performance. Participants' performance on the math test was operationalized as the number of items answered correctly minus one-fifth of one point for each incorrect answer to adjust for guessing (Quinn & Spencer, 2001). These scores were analyzed using hierarchical multiple regression. Since math grade was expected to influence performance on the test, participants' math grades from high school, and the interaction between math grade and experimental condition (Yzerbyt, Muller & Judd, 2004) were entered as covariates in the first step of the analysis.[2] The second step included the three main effects (dummy coded variables for gender and experimental manipulation, and the continuous gender identification measure). All possible two-way interactions between the three predictors were entered in step 3, followed by the three-way interaction between those variables in step 4. The interactions were created with the cross-products of the variables. Following the recommendations of Aiken and West (1991), the continuous predictor variable was centered before the cross-products were calculated in order to avoid artificial multicollinearity. Results of the analysis are presented in Table 1.

As predicted, math grades from high school was significantly related to participants' test performance in step 1, $\beta = 0.22$, $p = 0.046$, with higher grades predicting higher scores. There was a significant main effect of participant gender in step 2, indicating lower test performance for women as compared to men, $\beta = -0.27$, $p < 0.001$. In step 2, there was also a significant effect of gender identification, with higher degrees of identification predicting higher test performance, $\beta = 0.18$, $p = 0.009$. However, the two main effects were qualified in step 4 by a three-way interaction between experimental condition, participant gender, and gender identification, $\beta = 0.30$, $p = 0.041$, (see Fig. 1). No other main or interaction effects were significant.

Table 1. *Hierarchical regression analyses predicting math performance and number of items attempted from gender, gender identification, and gender identity relevance*

		Math performance		Number attempted	
		β	ΔR^2	β	ΔR^2
Step 1			0.07**		0.02
	Math grade	−0.22*		0.20	
	Math grade × Gender identity relevance	0.06		−0.10	
Step 2			0.10***		0.05*
	Gender[a]	−0.27***		−0.19*	
	Gender identification	0.18**		0.07	
	Gender identity relevance[b]	0.06		−0.12	
Step 3			0.02		0.02
	Gender × Gender identification	0.15		−0.03	
	Gender × Gender identity relevance	−0.16		−0.19	
	Gender identity relevance × Gender identification	0.001		0.17	
Step 4			0.02*		0.02*
	Gender × Gender identification × Gender identity relevance	0.30*		0.32*	

[a] Coded as 0 = men, 1 = women.
[b] Coded as 0 = Gender identity not relevant, 1 = Gender identity relevant.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

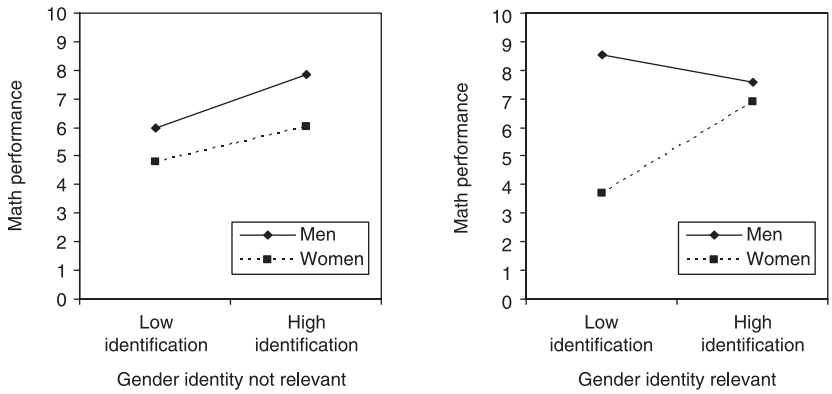

Fig. 1. Gender identification as a moderator of gender identity relevance effects on women's and men's math performance in a sample of Swedish math students.

Using the procedures described by Aiken and West (1991), the three-way interaction was decomposed by using simple slopes analyses. A first set of analyses tested the simple slopes representing gender differences within each experimental condition and evaluated at one standard deviation above and below the mean of gender identification ($M = 3.35$, $SD = 0.86$). In contrast to Schmader's (2002) findings, test performance was not predicted by gender for participants high in gender identification when gender identity was linked to test performance, β = –0.10, $p = 0.48$. However, for participants low in gender identification, analyses revealed lower performance for women as compared to men when gender was made explicitly relevant in the test situation, β = –0.59, $p < 0.001$. When gender was not relevant to test performance, women performed comparably to men regardless of whether they were high, β = –0.22, $p = 0.12$, or low β = –0.14, $p = 0.30$, in gender identification.

Analyses were also conducted to examine the effect of experimental condition separately among men and women and evaluated at high and low levels of gender identification. These analyses showed no performance differences as a consequence of experimental manipulation among women high, β = 0.09, $p = 0.42$, or low, β = –0.13, $p = 0.31$, in gender identification. Men high in gender identification were also unaffected by the manipulation, β = –0.03, $p = 0.84$. However, men low in identification performed better on the math test when gender identity was made relevant rather than not relevant, β = 0.32, $p = 0.03$.

A final set of analyses examined the effects of gender identification separately within each gender and experimental

condition, respectively. When gender was not relevant to the test, gender identification was not significantly related to test performance either for women, β =0.15, $p = 0.20$, or for men, β = 0.24, $p = 0.12$. When gender was made relevant, lower levels of gender identification predicted poorer performance among women, β = 0.38, $p = 0.004$, while differences in identification among men failed to predict test performance, β = –0.12, $p = 0.46$.

Number of items attempted. Following Schmader (2002), we also analyzed the number of items that participants attempted to solve in the math test (see Table 1). A hierarchical regression with the same independent variables and covariates as was used in the analysis of test performance revealed an effect of math grade bordering on significance in step 1, β = 0.20, $p = 0.075$. In step 2, there was also a significant effect of participant gender, β = –0.19, $p = 0.01$. Mirroring the pattern observed for test performance, this effect was moderated in step 4 by a significant three-way interaction between gender, gender identification, and experimental condition, β = 0.32, $p = 0.041$. There were no other significant main or interaction effects.

For participants high in gender identification, follow-up, simple slopes analyses showed a tendency towards fewer attempts for women than for men when gender was not relevant to test performance, β = –0.25, $p = 0.09$, while gender had no effect among participants in this group when gender was made relevant, β = –0.14, $p = 0.34$. However, in parallel to observation with regard to performance accuracy, women low in identification with their gender attempted fewer items than men low in identification when gender was made relevant to their performance β = –0.43, $p = 0.003$, while men and women in this group attempted the same number of items when gender was not relevant, β =0.06, $p = 0.67$.

Simple slopes analyses were further used to examine the effects of experimental condition separately for men and women high and low in gender identification. Results revealed that women low in identification attempted significantly fewer items overall when gender was made relevant as compared to when it was not relevant, β = –0.45, $p = 0.001$, while experimental manipulation had no effect on the number of attempts among women high in gender identification, β = 0.03, $p = 0.79$. There were no effects of the manipulation on men in any of the gender identification groups (low identification: β = 0.05, $p = 0.76$; high identification: β = –0.08 , $p = 0.64$).

A final set of analyses examined the effects of gender identification separately for men and women in the two experimental conditions, respectively. For women, high gender identification predicted more attempts when gender identity was made relevant, β = 0.34, $p = 0.019$, while identification had no effect on women's number of attempts when gender was not relevant, β = –0.16, $p = 0.20$. Gender identification was unrelated to the amount of attempts among men in both when gender was made relevant to test performance, β = 0.03, $p = 0.84$, and when it was not, β = 0.16, $p = 0.31$.

Stereotype-based expectancies and social monitoring

Test performance and number of attempts. To examine whether gender salience effects varied as a function of participants' degree of social monitoring, we conducted two ANOVAs with experimental condition, participants' gender, and social monitoring (high/low) as independent variables, using test performance and number of items attempted as dependent variables in separate analyses. As before, math grade was entered as covariate in both analyses. The results showed the same main effect of math grade and gender on both test performance and number of attempts that appeared in the previous analyses. No other main or interaction effects were significant on any of the two dependent measures.

Stereotype-based expectancies and stereotype endorsement

Test performance and number of attempts. To assess the role of stereotype endorsement on gender salience effects, two hierarchical multiple regressions were conducted with test performance and number items attempted as dependent variables in separate analyses. In both analyses, math grades and the interaction between math grade and experimental manipulation were entered in the first step. The second step included three main effects (gender and experimental manipulation dummy coded as in previous analyses, and the continuous measure of stereotype endorsement), step 3 and step 4 included two-way and three-way interactions between those variables, respectively. As in previous analyses, the continuous predictor was centered before cross-products between variables were created. Apart from the effects of math grade in step 1 (identical to those found in analyses of gender identification effects), step 2 showed a main effect of participant gender on both performance accuracy, β = –0.25, $p = 0.001$, and number of items attempted, β = –0.18, $p = 0.017$. There were no other main or interaction effects on any of the two dependent measures.

DISCUSSION

The main purpose of the present study was to examine whether previous findings of effects of gender-based expectations on women's math performance compared to men hold also in Sweden, with its particular cultural emphasis on gender equality. We further investigated whether potential effects of stereotype-based expectancies would be moderated by gender identification, social monitoring, and stereotype endorsement.

Consistent with earlier findings of a male advantage in the domain of arithmetic word problems (Johnson, 1984; Hyde, Fennema & Lamon, 1990; Geary, 1996), the current study showed that women were less accurate than men on a math test at an overall level. However, also in line with previous results (Schmader, 2002), we found that the effect of gender

on math performance was moderated by a combination of personal and situational variables. Thus, the gender difference in test performance appeared only among certain groups of participants in certain situations. Specifically, we found that a test situation in which gender identity was made relevant produced gender difference between men and women who did *not* see their gender as an important part of their self-definition. Moreover, the effect of the experimental manipulation on participants low in gender identification was more pronounced among men than among women. In the introduction we argued that the gender gap favoring men is considerably smaller in Sweden than in most other countries (including the USA). We further speculated that this may have consequences for both men and women in situations where performance expectations are explicitly linked to gender. We will discuss the effects of gender identity relevance and gender identification on the performance of men and women separately.

The role of gender identity relevance and gender identification on women's math performance

Our analysis showed that women who conducted a math test in a situation with negative stereotype-based expectancies performed better the more they identified with their gender. While contrary to the findings of Schmader (2002), this result is in line with recent research on minority groups, showing that group identification may serve as a buffer against negative stereotypes (Cohen & Garcia, 2005). Because the equality norm has been emphasized more strongly in Sweden than in other countries, Swedish women may have developed a stronger belief in their right to equal treatment, and the perceived societal support for these rights may make them able to ignore threats of confirming a negative stereotype in a specific domain. However, awareness of women's rights, and hence the extent to which the individual can overcome negative stereotypic expectations, is likely associated with the importance of gender to the individual's identity. Specifically, women who are highly identified with their gender may be those who are most aware of women's equal rights and the societal support for these rights, while women lower in gender identification may more often doubt this principle. While both groups of women thus to an equal extent may endorse the negative stereotype of women's math abilities, and recognize a threat of confirming a negative stereotype about women's math abilities, those high in gender identification can use their strong, general belief and their experience of support for this belief as a protection against these expectancies. However, if women who identify less with their gender are more uncertain of their rights in general, they lack this means of overcoming the pressure of threat in the negative expectancy condition, and their performance will suffer accordingly.

Analyses of the number of items attempted in the test showed that women low in gender identification attempted to solve fewer items when gender was made relevant to test performance than when gender was not relevant. The same pattern has been found among individuals from stigmatized groups in other stereotype threat studies (Schmader, 2002; Steele & Aronson, 1995). The result suggests that the gender identity relevance manipulation made these women more careful and thorough on any given problem, or alternatively, that they experienced more difficulties solving each problem because their cognitive resources were distracted.

The role of gender identity relevance and gender identification on men's math performance

Our results showed further that the salience of gender differences had little effect on the performance of highly gender identified men. However, in line with expectations, the tendency for this group was to perform worse in a situation with explicit positive expectations, which would be consistent with the notion of choking under pressure (cf. Beilock *et al*., 2004). Interestingly, the group that was most strongly affected by gender identity relevance in the test situation was men low in gender identification. Specifically, men in this group seemed to profit from being reminded of gender differences in math performance, and scored significantly higher when gender was made salient rather than when it was not. This finding is clearly consistent with the notion of stereotype lift (Walton & Cohen, 2003). Thus, in terms of expectancies, the general pattern of identity relevance and gender identification effects among men parallel those found among women; Under more positive expectancies, which for men is the condition where gender is made relevant to test performance, and for women is the condition where gender is not relevant, low identifiers of both genders tend to perform better while high identifiers, if anything, tend to perform worse. The larger experimental effect on men as compared to women deviates from the typical finding of gender-based expectancies on math performance (Spencer *et al.*, 1999; Schmader, 2002). It seems possible that Swedish men may experience more threats to their male identity, and perceive their male gender role as less stable and secure as compared to men in other countries. This may make them more attentive to any gender stereotype-based expectancies than men in less gender equal countries. In the introduction, we hypothesized that men who identify strongly with their gender would be those most affected by stereotype-based expectancies. However, one could speculate that those most responsive to threats and criticism towards the male gender are those who have dissociated from their male identity, i.e., men low in identification. When these individuals perceive that the environment signals positive associations to men, they may also be those who respond most strongly to the signal.

The role of social monitoring and stereotype endorsement

In contrast to the findings of Inzlicht *et al.* (2006) we found no support for the notion that social monitoring would

moderate stereotype threat effects in our study. Clearly, more research is needed to explore factors that may moderate the influence of this factor. Finally, our study found no evidence that endorsement of the stereotype of women's poor math ability would be related to stereotype threat effects. Although inconsistent with the findings of Schmader (2002), this fits well with the theoretical assumption that an individual's own endorsement of a negative stereotype is unrelated to the impact of the effect of the stereotype.

Conclusions

The present study adds to a body of research showing that gender differences in math performance can vary as a function of subtle cues in the test situation. In Sweden, the country with allegedly the smallest gender gap in the world and with a particularly strong emphasis on gender equality, stereotype threat still has the power to affect math test performance. However, the specific pattern of influence of this stereotype differs somewhat from those found in cultures where gender inequality is less contested. For example, our results indicated that individuals who report that their gender is relatively unimportant to their identity were those most affected by manipulation of gender identity relevance in a math test situation. Moreover, in our study the effect of gender stereotype threat was larger for the positively stereotyped gender. Further research is needed to delineate the mechanism behind the differential effects of group identification in this and other studies. An understanding of how identification with stereotyped groups can moderate the effects of stereotype threat may prove useful for the development of strategies to counter the impact of these stereotypes.

This research was partially supported by grants to the first author from the Wallenberg Foundation and the Swedish Research Council. The authors are grateful to Sylvia Bergvall for assistance in data coding.

NOTES

[1] Of the 186 participants, 49 belonged to the two groups that were assigned entirely to one condition (24 to Gender identity relevant, 25 to Identity not relevant). To ensure that our results are not an artifact of group-level effects, we analyzed the results of those 49 participants and the 137 properly randomized participants separately. Although the statistical power is no longer sufficient for effects to reach significance in these separate analyses, we obtained the same overall pattern with respect to main and interaction effects in the two subsamples.

[2] As pointed out by an anonymous reviewer, math grades of women from school may tend to misrepresent these students' actual abilities in systematic ways. Hence, using math grades as a covariate to adjust for prior ability may conflate any true differences between men and women with the effects of stereotype threat. In response to this notion, we also tested our predictions by first regressing participants' performance on their math grades (separately for each gender), and then using the residuals from these regressions as the criterion in the hierarchical regression analysis (Brown & Joseph, 1999). The results of this analysis showed essentially the same results as was found in the analysis using math grade as a simple covariate. Most importantly, the three-way interaction between gender, threat, and gender identification was still present, $\beta = 0.34$, $p = 0.032$.

REFERENCES

Aihen, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M. & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.

Baumeister, R. F., Hamilton, J. C. & Tice, D. M. (1985). Public versus private expectancy of success: Confidence booster or performance pressure? *Journal of Personality and Social Psychology*, *48*, 1447–1457.

Beilock, S. L., Kulp, C. A., Holt, L. E. & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*, 584–600.

Brandell, G., Larsson, S., Nyström, P., Palbom, A., Staberg, E.-M. & Sundqvist, C. (2005). *Kön och matematik*. (Preprints in Mathematical Sciences, 2005:20). Lund, Sweden: Centre for Mathematical Sciences, Lund University.

Branscombe, N. R., Schmitt, M. T. & Harvey, R. D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology*, *77*, 135–149.

Brown, R. P. & Joseph, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, *76*, 246–257.

Cohen, G. L. & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology*, *89*, 566–582.

Ellemers, N., Spears, R. & Doosje, B. (1997). Sticking together or falling apart: In-group identification as a psychological determinant of group commitment versus individual mobility. *Journal of Personality and Social Psychology*, *72*, 617–626.

Gangestad, S. & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, *92*, 317–349.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, *19*, 229–284.

Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.

Inzlicht, M., Aronson, J., Good, C. & McKay, L. (2006). A particular resiliency to threatening environments. *Journal of Experimental Social Psychology*, *42*, 323–336.

Inzlicht, M. & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.

Johnson, E. S. (1984). Sex differences in problem solving. *Journal of Educational Psychology*, *76*, 1359–1371.

Keller, J. & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect on women's math performance. *Personality and Social Psychology Bulletin*, *29*, 371–381.

Leyens, J. P., Desert, M., Croizet, J. C. & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, *26*, 1189–1199.

Lopez-Claros, A. & Zahidi, S. (2005). *Women's Empowerment: Measuring the Global Gender Gap*. Geneva: World Economic Forum.

Luhtanen, R. & Crocker, J. (1992). A collective self-esteem scale:

Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*, 302–318.

Marx, D. M., Stapel, D. A. & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology*, *88*, 432–446.

OECD (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publications.

Oswald, D. L. & Harvey, R. D. (2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology*, *19*, 338–356.

Quinn, D. M. & Spencer, S. S. (2001). The interference of stereotype threat with women's generaton of mathematical problem-solving strategies. *Journal of Social Issues. Special Issue: Stigma: An Insider's Perspective*, *57*, 55–71.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201.

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

Smith, J. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review*, *16*, 177–206.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526–537.

Snyder, M. & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, *51*, 125–139.

Spears, R., Doosje, B. & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin*, *23*, 538–553.

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

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, 34 (pp. 379–440). San Diego, CA: Academic Press.

Stone, J., Lynch, C. I., Sjomeling, M. & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, *77*, 1213–1227.

Tajfel, H. & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel and L. W. Austin (Eds.), *Psychology of intergroup relations*. Chicago: Nelson-Hall.

Walton, G. M. & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Psychology*, *39*, 456–467.

Wong, C. A., Eccles, J. S. & Sameroff, A. (2003). The influence of ethnic discrimination and ethnic identification on African American adolescents' school and socioemotional adjustment. *Journal of Personality*, *71*, 1197–1232.

Yzerbyt, V. Y., Muller, D., Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*, 424–431.

Received 2 December 2005, accepted 22 November 2006