The Effects of Gender Stereotypes for Structure Mapping in Mathematics

Kreshnik Nasi Begolli (keko@temple.edu)
Temple University, Department of Psychological, Organizational, and Leadership Studies, 1301 W. Cecil B. Moore Avenue Philadelphia, PA 19122 USA

Brooke Herd (bherd@uci.edu), Hannah Sayonno (hsayonno@uci.edu), Susanne Jaeggi (smjaeggi@uci.edu)
University of California, Irvine, School of Education, 3200 Education, Irvine, CA 92697 USA

Lindsey Engle Richland (lrichland@uchicago.edu)
University of Chicago, Department of Comparative Human Development, 5730 S. Woodlawn Ave., Chicago, IL 60637

Abstract

Fear of a negative stereotype about one's performance can lead to temporary underperformance on tests; e.g. women may underperform on a math test when prompted to think about gender. The current study extends this literature to examine whether stereotype threat not only leads to underperformance on tests, but also may impact reasoning and learning more broadly. We focus in particular on the effects of stereotype threat on analogical learning, a complex reasoning process that imposes a high working memory load. In this study, we examined the effects of gender stereotypes when females were asked to learn by comparing the mathematical concepts of combinations and permutations. Overall, participants given a threat before learning gained less from the instruction, as reflected by assessments administered immediately after the lesson and after a 1-week delay. This could lead to systematic differences in the quality of abstract representational knowledge for individuals from negatively stereotyped groups.

Keywords: Gender stereotype threat; analogy; comparison; mathematics education; video stimulus; working memory

Stereotype threat refers to the phenomenon in which members of a group fear that others will hold negative assumptions about their skills or capacities, by feeling additional pressure of conforming or disproving such beliefs (Spencer, Steele, & Quinn, 1999). Within education, one existing stereotype is that women naturally have lower abilities in mathematics than men, particularly on higher level mathematics concepts (Maloney & Beilock, 2012; McJunkin, 2009; Spencer, Steele, & Quinn, 1999).

The current literature suggests that working memory (WM) overload may be a primary mechanism explaining detrimental effects of gender stereotype threat on performance (Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003). Participants are compromising their WM ability by anxious ideation about the activated stereotype, thus occupying their WM resources through thoughts such as "I had better perform well to disprove their stereotype about me as a woman." This ideation, while motivational, also engages verbal working memory, thus reducing available resources for task engagement.

Most studies of stereotype threat have focused on impact on test performance, examining the potential for a threat imposed immediately before a test to temporarily reduce performance and produce an unrepresentative measure of participants' performance (see Steele & Aaronson, 1995). Very little work has explored whether these threats might also impact learning, however, which would be a less temporary impact and might lead to more systematic differences in knowledge gains by negatively stereotyped groups. Thus, more research is needed to examine the effects of stereotype threat on learning.

To close this gap in the literature, this study examined the effects of gender stereotype threat on learning mathematics during high working memory load instructional analogies. Instructional analogies are common within mathematics education internationally (Richland, Zur, & Holyoak, 2007), and comparing problems or solutions is a common pedagogical recommendation (e.g., Common Core State Standards Initiative, 2012; National Research Council, 2001). However, for students to gain from these comparisons, they must have adequate available WM resources to represent two or more systems of relations, discover their alignments and map between the systems to draw inferences based on alignments (or misalignments) leading to encoding of a mathematical schema (Begolli, Richland, & Jaeggi, 2015; Gentner, 1983; Gick & Holyoak, 1983; Morrison, Doumas, & Richland, 2011; Waltz, Lau, Grewal, & Holyoak, 2000).

Structure mapping is particularly taxing when learners are not provided with adequate instructional supports to draw attention to key relational correspondences or when the content is not highly familiar (Begolli & Richland, 2016; Richland & McDonough, 2010; Rittle-Johnson & Star, 2007). Under higher WM demands, novices are more likely attend to and retain object features of compared representations rather than abstracting and retaining the common relational structure. In a previous experiment, Richland and McDonough (2010) tested the role of high versus low instructional cues within a video lesson in which a teacher made an analogy between the mathematical concepts of combinations and permutations. Students in the low-cued lesson (hypothesized to require higher WM demands) performed worse on cross-mapping problems on a posttest, suggesting their knowledge was less schematized.

Working memory is theorized to lie at the core of the observed effects of gender stereotype threat and learning from comparisons. Building on this work, the current study examined whether the pressure imposed by a threat of gender stereotypes could reduce WM capacity that students normally rely on to learn a mathematical schema through structure mapping.

To test this hypothesis, this study exposed students to either a threat or neutral video lesson on permutations and combinations concepts. In order to ensure that the instructional analogy imposed high WM demands, we draw our stimuli from the low-cued video lesson utilized by Richland and McDonough (2010).

This work has theoretical implications for informing our understanding of the effects of stereotype threat on learning through structure mapping, possibly due to underlying shared mechanisms. Thus, laying the groundwork for further research to provide evidence linking research on structure mapping and stereotype threat. Further, this work has direct implications for informing teachers about the adverse effects of cultural pressures on learning processes.

Experiment: Impact of Gender Stereotypes for Instructional Analogy

Method

The research study was a randomized experimental study, with half of the participants being administered a gender stereotype threat and the other half receiving no threat. All participants received the same instruction subsequently. Learning was measured on an immediate posttest and one week later. Participants were not cognizant of the study's purpose of examining stereotype threat, as the study was advertised as a GRE math lesson on Combinations and Permutations.

Participants. Participants were 45 women who participated in the study either for course credit or monetary compensation. Only women who had not taken a college level statistics course were invited to participate in the study, to reduce the possibility of prior knowledge of permutations and combinations. The final analyses included 26 females in the threat condition and 19 in the non-threat condition (five participants in the non-threat condition did not complete all phases of the study).

Procedure. Participants were directed to a Google Forms Survey, which collected participants' demographic information and screened them based on their previous knowledge of statistics. If invited to participate, on Day 1, all participants were given 15-minutes to complete a pretest. They next answered a questionnaire, viewed an instructional lesson, and completed a posttest. Participants were randomly assigned to either the threat or non-threat conditions. The videotaped lesson and questionnaire were varied according to condition. Five to seven days later, participants were given 15-minutes to complete a delayed posttest. At the end, participants were debriefed about the true nature of the experiment.

Questionnaire. The questionnaire for the threat (experimental) group, primed female identity in order to reinforce the gender stereotype, taken from (Shih, Pittinsky, & Ambady, 1999). The control group received a similar questionnaire after their pretest, with the same format and number of questions. However, the questions only pertained to neutral topics, such as whether the participant watched cable television (Shih et al., 1999).

Instructional Videos. Each student watched a videotaped lesson on two mathematical concepts commonly included on the GRE: permutations and combinations. The instructional portion of the video lesson was identical to the low-cued video lesson used by Richland & McDonough (2010). From this, two versions of the video lesson were developed: 1) the threat condition containing introductory text providing a stereotype threat, and 2) the non-threat condition with a neutral introductory text. For the threat students, the video lesson began with a screen stating "Please watch the following 15-minute video very carefully," which was followed by "The lesson and the following tests have been shown to display gender differences in the past, exhibiting lower scores for women than men." The wording used to induce threat was modeled according to a study done by McJunkin (2009), which examined females' performance after being exposed to stereotype threat. The words remained on the screen for 30 seconds before the video lesson on combinations and permutations, which continued for approximately 13 minutes.

The video lesson for the control group began with "Please watch the following 15-minute video very carefully," and followed with "This lesson and the following tests have displayed no gender differences in the past; scores between women and men have been relatively equal"

During the instructional portion, an instructor discussed two mathematics problems and their solutions. The first exemplified the concept of combinations and the second the concept of permutations. Permutations involve finding the number of arrangements of a set of objects where the position (or order) is important. Combinations involve calculating the number of arrangements possible in which position (or order) is not important. Importantly, the instructor aligned and mapped between similarities in the permutation and combination concepts, while highlighting that the combination problems require an additional procedural step.

Table 1. Proportion correct at each time point by condition and question type. Standard deviations in parenthesis.

Condition	Pretest		Immediate						Delayed					
			Facilitory		Misleading		Unrelated		Facilitory		Misleading		Unrelated	
Threat	.10	(.16)	.69	(.35)	.33	(.24)	.40	(.32)	.46	(.34)	.40	(.28)	.50	(.28)
No Threat	.14	(.18)	.84	(.29)	.45	(.28)	.55	(.33)	.58	(.34)	.55	(.28)	.61	(.27)

Mathematics Assessment. The mathematics assessments had four types of problems: facilitory similarity, misleading similarity, unrelated, and distractors, discussed next.

Facilitory similarity. These questions refer to problems in which the story context and mathematical structure correspond between the video lesson and the posttest. This meant that the story context of the permutation problem in the instructional lesson (placements in a race), corresponded with the permutation problem on a posttest (placements in a race), and the combination problem story context during instruction was similar to the combination story context on the posttest.

Misleading Similarity. These questions refer to problems in which the story context of problems in the instruction were cross-mapped with problem contexts on the posttest. For instance if students learned the concept of permutations in the context of runners in a race, those students may attempt the same solution strategy on the posttest when faced with a problem that actually involved combinations but was set in the context of runners in a race, or vice versa. This meant that story context during instruction, such as teaching permutations in the context of a race, was followed by a combination problem on the posttest in the context of a race. If participants had coded the abstract relations within the problems they should perform similarly on the facilitory and misleading posttest problems. Those who had retained the surface features should be more likely to preform worse on misleading problems.

Unrelated Context. These problems were dissimilar in surface features to problems that were reviewed in the video, which tested participants' ability to apply combinations and permutations solutions in a completely different context.

Distractor problems. These problems involved factorials and advanced combinatorics problems, and were not analyzed.

Pretest. The pretest had 6 questions, 3 testing combinations and 3 testing permutations concepts. Because students did not watch the video lesson, these questions cannot be discussed in terms of facilitory and misleading similarities and were used as a baseline.

Immediate and Delayed Posttest. Both posttests had 10 questions, which contained the same story contexts, but were presented in an altered order with different numbers. Both tests contained 2 facilitory similarity problems, 2 misleading similarity problems, and 2 unrelated problems. Four problems functioned as distractors and assessed advanced combinatoric problems and factorials.

Answers from participants were coded based on whether the correct solution strategy was used to solve the questions. Correct strategy setup was scored as correct. Calculation errors were not scored, because our interest was to examine whether students knew which solution strategy was appropriate given the context. The assessments and coding was drawn directly from Richland & McDonough (2010).

Results

Raw means of the mathematics measures at each test point are summarized in Table 1. The random assignment of students was successful, with no significant differences in prior knowledge between the two groups (threat vs. no-threat) at pretest $F(1, 43) = .881$, $p = .353$.

We conducted a 2 (time of test: immediate to posttest) x 3 (question type: facilitory, misleading, and unrelated) x 2 (condition: Threat vs. No Threat) repeated measures analysis of covariance with time of test and surface similarity as within subject factors and condition as a between subject factor. Pretest score was utilized as a covariate and was a significant predictor when used in the model $F(1, 42) = 10.388$, $p = .002$, $\eta^2 = .198$.

As expected and shown in Figure 1, students performed worse under threat. There was a main effect of condition $F(1, 42) = 4.312$, $p = .044$, $\eta^2 = .093$ such that students who did not experience the gender threat used the correct strategy more often on all the problems ($M = .57$, SD = .31) than those who were under threat ($M = .44$, SD = .24). Also, problems with the same storyline and mathematical structure on both the lesson and the posttest were easier to solve than problems that were cross-mapped. Thus, there was a main effect of question type $F(2, 84) = 11.794$, $p = .006$, $\eta^2 = .223$, where a larger proportion of facilitory similarity problems were setup correctly than misleading similarity problems or unrelated problems (see Table 1).

Follow up pairwise comparisons confirm that facilitory similarity problems were set up correctly more often than both misleading similarity ($p < .001$) and unrelated problems ($p = .001$). There was also a trend suggesting that misleading similarity problems were more likely to lead to errors when compared to unrelated problems ($p = .051$).
There was no main effect of time $F(1, 42) = 2.041$, $p = .161$, $\eta^2 = .046$, such that students' overall scores were not significantly different between immediate test ($M = .56$, SD

=.21) and delayed test (M =.44, SD =.25). However, there was an interaction between time of test and question type $F(2, 84) = 7.968$, $p = .001$, $\eta^2 = .280$. Students were better at using the correct solution strategies for facilitory problems immediately after the lesson than 1-week later. The inverse was true for misleading and unrelated problems. The decrease for facilitory similarity is likely because the problem story contexts during instruction became less salient over time and thus less helpful at posttest (see Table 1).

There was no interaction between question type and condition F (2, 84) = .005, p = .995, η^2 = .003 and no interaction for time and condition F (2, 42) = .015, p = .682, η^2 = .004. No other interactions were expected to be significant and our results confirmed this expectation.

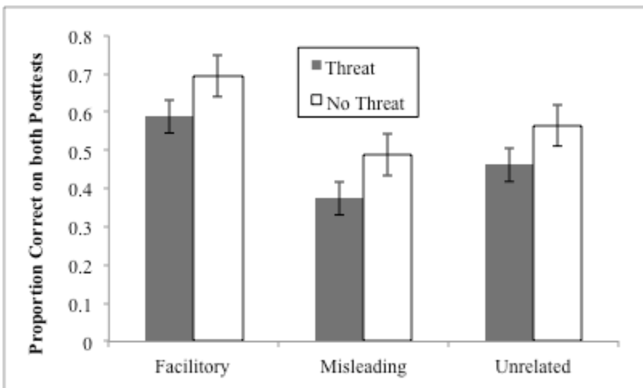

Figure 1. Adjusted means for accuracy by condition for each question type across immediate and delayed posttests using pretest score as covariate. The bars represent standard errors.

Discussion

The study reveals that learning from analogy may be susceptible to pressures from social context, and that effects can be lasting rather than temporary. Many studies have examined the role of stereotype threat on performance, but this study is one of the first to show that threat of negative stereotypes can impact *learning.* Unlike previous accounts documenting the detrimental effects of stereotype threats on women's mathematics test performance (McJunkin, 2009; Steele & Aronson, 1995), inducing the threat before a learning event and examining outcomes after a 1-week delay provides new insight into the impact stereotypes can have for schema formation.

Our data across two time points support that students who were exposed to stereotype threat were significantly worse at learning combinations and permutations concepts in comparison with students who had not been exposed to stereotype threat, and that these differences lasted over a week delay.

The mathematics outcomes align with current views suggesting that concern about negative stereotypes may particularly impair performance on advanced mathematics concepts, more than lower working memory problems that merely require straightforward calculations (see Malone & Beilock, 2012).

A potential concern regarding our paradigm is that participants' immediate posttest performance may have been impaired not due to learning differences, but due to the same immediate performance pressure effects observed in traditional studies of stereotype threat on test performance (Beilock, Rydell, and McConnel, 2007). At the same time, any state differences based on the threat should have dissipated by a 1-week delay. Thus, our results supported our hypothesis, in that stereotype threat impeded the learning of mathematics through structure mapping, rather than artificially decreased test performance. An alternative explanation for decreased performance could be that the effects of threat did not impede structure-mapping processes per se, but broader learning and/or encoding, considering the effects of threat were similar on facilitory and misleading problems.

In line with previous findings in the analogy literature, students in our sample were better able to solve problems where the context between the lesson and the posttest was aligned. This shows that students were successful at encoding the solution strategies involving permutations and combinations concepts. However, students performed worse on problems where the context was cross-mapped between the video lesson and the posttest or the problem context was unrelated to problems in the video lesson.

The question type differences are also consistent with Richland and McDonough's (2010) findings. Misleading similarity problems may have placed a greater demand on WM, particularly inhibitory processes, since students would need to inhibit their inclination to solve based on surface similarities, in order to apply the correct solution strategy. Yet, Richland and McDonough (2010) postulate that higher WM demands, induced by the lack of instructional supports, led students to misapply solution strategies only on misleading similarity problems. It is possible that if students in our study learned from a lesson with high instructional supports, the effects of threat would dissipate for the facilitory questions. Thus, manipulating WM demands through the video lesson in addition to the problem types (facilitory & misleading) and assessing WM or inhibitory processes would strengthen our findings. In general, our results also point towards WM as an important component of analogical thought or broader learning processes, and highlight that these may be mechanisms through which stereotype threat reduces long-term learning and schema formation.

Implications for Future Research

The paradigm used in this study, administering a stereotype threat before an opportunity for analogical learning, provides a platform for ongoing studies that aim to shed light on the impact of pedagogical supports and WM resources on structure-mapping and relational reasoning.

For example, Beilock & Carr (2005) found that stereotype threat affects mathematics performance for women with higher WM capacities. This suggests that women who are most likely to succeed underperform when under threat. While this has tremendous implications for test taking, it may have even greater implications if students face threat throughout their learning experience. Thus, students with greater potential could learn significantly less, which could accumulate over time, resulting in significant achievement gaps. Our data coupled with WM measures and a larger sample size may provide insight into the mechanisms and behavioral effects of stereotype threat on learning.

On the other hand, there is mounting evidence that instructional supports offload WM demands, which may interact with the effects of gender stereotypes. The lack of instructional supports may be particularly beneficial for lower performing students and those with lower WM abilities. Considering that gender stereotypes most significantly affect females with high WM capacities, we may find that the effects of threat differ based on instructional supports and WM capacity. In the current study, we only used the lowly-cued version of Richland & McDonough's (2010) video lesson when inducing stereotype threat. A continuation of this study will compare the effects of stereotype threat when learning from a highly-cued video lesson. The potential implications of this work are that competent teachers, who implicitly induce threat, may help lower performing students at the expense of higher performers.

Overall, these future studies seek to untangle the effects of stereotype threat and its possible cognitive mechanisms, under varying instructional contexts, and clarify their role for learning mathematics through structure-mapping. While we discuss mathematics more broadly, combinations and permutations represent only one mathematical domain, and further studies are needed to understand whether these results generalize to other mathematical concepts. Despite this, it is important to note that the stimulus used (the math video lesson) does approximate a true lesson experience. The instruction was not text-based, and came from a live instructor, thus suggesting greater ecological validity for generalizing to teacher actions in a classroom setting.

This research has important implications for teachers in the classroom. Teachers must be aware of the culture they are creating within their classroom and should support and facilitate learning through teaching methods that utilize high instructional supports and reduce any hindrances of societal pressures that might reduce learners' abilities.

In sum, these research findings not only reveal the potentially profound effects of stigma in a learning context but also suggest various research directions for further studies. This research creates a platform for further investigating the role of working memory as an underlying mechanism of the pressures of social stigma. Furthermore, our results lead to new research questions regarding stereotype threat and teaching methods. Questions are left to be asked regarding the impact of high or low instructional teaching methods in aiding or preventing stereotype threat from impeding learners, as well as whether individual characteristics, such as innate and perceived ability and motivation, can alter the effects stereotype threat can have on a learner.

References

Begolli, K. N., & Richland, L. E. (2016). Teaching Mathematics by Comparison: Analog Visibility as a Double-Edged Sword. *Journal of Educational Psychology*, *108*(2), 194–213.

Begolli, K. N., Richland, L. E., & Jaeggi, S. M. (2015). The Role of Executive Functions for Structure-Mapping in Mathematics. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warfaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (Vol. 1, pp. 1–6). Austin, TX: Cognitive Science Society.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101–105.

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. http://doi.org/10.1037/0096-3445.136.2.256

Common Core State Standards Initiative. Common core state standards for mathematics (2012). Common Core State Standards Initiative.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science*, *7*(2), 155–170. http://doi.org/10.1016/S0364-0213(83)80009-3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. http://doi.org/10.1016/0010-0285(83)90002-6

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406.

McJunkin, L. M. (2009). Effects of stereotype threat on undergraduate women's math performance: Participant pool vs. classroom situations. *Emporia State Research Studies*, *45*(2), 27–31.

Morrison, R. G., Doumas, L. A. A. L. a a, & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, *14*(3), 516–529. http://doi.org/10.1111/j.1467-7687.2010.00999.x

National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics*. The National Academies Press. Retrieved from http://www.nap.edu/catalog/9822/adding-it-up-

helping-children-learn-mathematics
Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*(1), 28–43. http://doi.org/10.1016/j.cedpsych.2009.09.001
Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. http://doi.org/10.1037/0022-0663.99.3.561
Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440.
Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80–83.
Shih, M., Pittinsky, T. L., & Trahan, A. (2006). Domain-specific effects of stereotypes on performance. *Self and Identity*, *5*(1), 1–14.
Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.
Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*. http://doi.org/10.1037/0022-3514.69.5.797
Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*(7), 1205–1212.