COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018) 678–690

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12558

Stereotype Threat Effects on Learning From a Cognitively Demanding Mathematics Lesson

Emily McLaughlin Lyons,[a] Nina Simms,[b] Kreshnik N. Begolli,[c] Lindsey E. Richland[a]

[a]*Department of Comparative Human Development, University of Chicago*
[b]*Psychology Department, Northwestern University*
[c]*Psychology Department, Temple University*

Received 16 February 2017; received in revised form 22 August 2017; accepted 22 August 2017

Abstract

Stereotype threat—a situational context in which individuals are concerned about confirming a negative stereotype—is often shown to impact test performance, with one hypothesized mechanism being that cognitive resources are temporarily co-opted by intrusive thoughts and worries, leading individuals to underperform despite high content knowledge and ability (see Schmader & Beilock, 2012). We test here whether stereotype threat may also impact initial student learning and knowledge formation when experienced prior to instruction. Predominantly African American fifth-grade students provided either their race or the date before a videotaped, conceptually demanding mathematics lesson. Students who gave their race retained less learning over time, enjoyed the lesson less, reported a diminished desire to learn more, and were less likely to choose to engage in an optional math activity. The detrimental impact was greatest among students with high baseline cognitive resources. While stereotype threat has been well documented to harm test performance, the finding that effects extend to initial learning suggests that stereotype threat's contribution to achievement gaps may be greatly underestimated.

Keywords: Stereotype threat; Performance pressure; Learning; Cognitive load; Situational Interest; Achievement gaps

1. Introduction

In the context of an ever-increasing emphasis on school accountability and standardized test performance, many students experience a great deal of pressure at school, which can produce worry about how their performance will be evaluated (Triplett & Barksdale,

Correspondence should be sent to Lindsey Richland, Social Sciences Research Building, Office 103, University of Chicago, 1126 East 59th St., Chicago IL 60637. E-mail: lrichland@uchicago.edu

2005; Watson, Johanson, & Dankiw, 2014). Performance pressure is known to impact test performance, leading some students to underperform relative to their knowledge base, or "choke" (see Beilock & Carr, 2005; Beilock, 2008; Nguyen & Ryan, 2008). While motivation may be increased by performance pressure, performance on cognitively taxing tests that require reasoning or representational manipulation may be impaired (see Maloney, Sattizahn, & Beilock, 2014).

Coupled with this general evaluative pressure, everyday experiences of identity judgments can be frequent and related to overall achievement (Wong, Eccles & Sameroff, 2003). Thus, students of color may be also susceptible to experiencing stereotype threat both during testing events and in everyday learning contexts (Larnell, Boston, & Bragelman, 2014; Nasir, Snyder, Shah, & Ross, 2009, 2012; Steele & Aronson, 1995). Stereotype threat is the phenomenon in which fear of being judged through the lens of a negative stereotype results in intrusive thoughts and worries that tax cognitive resources (Schmader & Beilock, 2012). Extensive research in adults (for review see Nguyen & Ryan, 2008) and growing work with children (e.g., McKown & Strambler, 2009; Neuburger, Ruthsatz, Jansen, & Quaiser-Pohl, 2015; Wasserberg, 2014) documents that when a self-relevant negative stereotype is made salient before a test, performance suffers.

Stereotype threat is most often invoked by giving participants a prompt before a test that includes some combination of increasing identity salience, such as by asking participants to report their race or presenting the study as being concerned with how individual differences influence performance, and raising the stakes of performance, such as presenting the task as diagnostic of ability. For example, in Steele and Aronson's (1995) seminal study, stereotype threat was invoked among African American adult participants by describing the study as "being concerned with 'various personal factors involved in performance'" and as diagnostic of participants' ability. In contrast, in the control condition, the test was described as simply aimed at better understanding factors involved in problem solving. African American participants who received the former instructions underperformed relative to African American participants who received the control prompt and relative to White participants who received either prompt.

Studies have shown that children begin to become aware of negative racial stereotypes as early as 6 years of age, with most children being aware of broadly held racial stereotypes by age 10 (McKown & Weinstein, 2003; McKown, 2004), and that, like adults, stereotype-aware children can be susceptible to prompts that invoke stereotype threat by increasing identity salience and performance consequences (McKown & Strambler, 2009; Wasserberg, 2014). In school-based studies with children, stereotype threat has been invoked by presenting a task as being diagnostic of ability (e.g., "a good way of testing how good you are at different kinds of school problems," McKown & Strambler, 2009) or as being predictive of performance on an upcoming standardized test and asking students to provide their race (Wasserberg, 2014).

1.1. Effects of stereotype threat on learning

Despite this broad literature, much less is known about whether stereotype threat and pressure may also impact *initial learning*. An individual experiencing stereotype threat may not only under-demonstrate her knowledge on a test that requires complex cognition, but worry ideation associated with stereotype threat could also interfere with her engagement in higher cognition during initial learning. In this case, not only would the student's final test performance be impaired, but her initial learning and knowledge formation would be harmed such that content from the lesson was not even available to her at a later point, when the threat of fulfilling a stereotype was no longer salient. This could have potentially systemic effects in a discipline such as mathematics, where the curriculum builds on previously learned material, such that if a student failed to fully understand one content area, s/he may be at a disadvantage for learning the subsequent step, with ever-increasing consequences.

Considering the mechanisms underpinning the relationship between stereotype threat and test performance clarifies why the consequences of pressure on learning, not just on testing, may be an essential part of closing achievement gaps in schooling. Executive function (EF) load is one mechanism that has received much support as an explanation for why stereotype threat may decrease test performance. EF resources may be co-opted for negative ideation about test performance, with looping thoughts such as "I don't want to fail this test, I don't want to fail this test" engaging the same resources that would instead be more productively used to engage with the test material (see Schmader & Beilock, 2012). Test performance for youth and adults under stereotype threat or performance pressure is most affected on challenging, cognitively demanding problems (for review, see Maloney et al., 2014). Stereotype threat effects in particular have been shown to be especially pronounced on cognitively demanding comparison problems that require suppression of a prepotent response, possibly because motivation to perform well made the prepotent response stronger, and EF resources were already taxed, so less available to inhibit that response (Davies, Conner, Sedikides, & Hutter, 2016).

One might assume based on these patterns that having high baseline EF resources would serve as a protective factor against stereotype threat, but it is often the highest EF participants who show the largest effects of stereotype threat on test performance (for reviews see Beilock, 2008; Maloney, Schaeffer, & Beilock, 2013). Much of this work has been conducted in mathematics, and it may be because these are the participants who are most used to engaging high levels of EF resources on challenging tasks (e.g., using strategies that require EF), so they are therefore most likely to have to change their test-taking strategies (e.g., Beilock & Carr, 2005; Beilock & DeCaro, 2007).

A second way that stereotype threat may impact performance is that stereotype threat during testing can negatively impact student interest in and identification with the stereotyped domain (Smith, Sansone, & White, 2007; Woodcock, Hernandez, Estrada, & Schultz, 2012). This could have longer term consequences than solely test performance, since decreased engagement and identification with a domain could impact future school and employment trajectories, including academic coursework selection or career goals.

This study examined how invoking stereotype threat prior to a conceptually demanding mathematics lesson impacted African American fifth-grade students. We examined impacts on learning as well as on situational interest and motivation to engage in non-required math activities. Additionally, we tested whether the impact of stereotype threat experienced while learning was moderated by individual differences in EF, as has been shown to be the case for stereotype threat experienced during testing (Beilock, 2008).

The lesson's objective was to teach students to compare ratios with different denominators. It centered on a comparison between a correct strategy (lowest common multiple) and a misconception (subtraction). The act of comparing and contrasting solutions, including misconceptions, is a recommended educational practice that has been shown to promote deep learning (Durkin & Rittle-Johnson, 2012; Richland & McDonough, 2010; Rittle-Johnson & Star, 2007). At the same time, comparing correct and incorrect solution strategies places high demands on students' cognitive resources, as students must be able to learn from the misconception discussion but ultimately inhibit it as a prepotent response and encode the correct strategy (Zaitchik, Iqbal, & Carey, 2014). Just as performance on challenging test items is most harmed during experiences of stereotype threat or performance pressure, this type of high-quality, high-demand learning could be especially compromised by situational stressors that tax cognitive resources.

2. Method

2.1. Participants

Study participants were 135 fifth-grade students drawn from two K-6 charter schools. All participants were members of minority groups that are negatively stereotyped academically, with the majority of students self-reporting their race/ethnicity to be Black/African American. Participants were nearly evenly split by gender with slightly more number of girls (Control condition: 33 girls, 31 boys; Stereotype Threat (ST) condition: 37 girls, 34 boys). The distribution of participants' race/ethnicity identification was comparable across conditions.

More detailed information on study participants can be found in the supplementary materials.

2.2. Materials and procedure

2.2.1. Design

Procedures were administered during three visits over a 2-week period at each school using a pretest, lesson and immediate posttest, delayed posttest design. Participants were randomly assigned within each classroom to either the Learning under a Stereotype

Threat (ST) condition or Control condition to minimize variability across schools and teachers. In all cases, the primary experimenter was a White female and all but one of the classroom teachers was also a White female. However, random assignment of participants to conditions within classrooms ensured that there were no differences in ethnic backgrounds of experimenters or teachers between experimental groups.

1. At Visit 1, students completed a group-administered pretest to assess their starting understanding of ratio. This assessment, as well as the immediate and delayed posttests, measured conceptual and procedural understanding of ratio as well as misconception usage—the frequency with which students incorrectly made use of subtraction to attempt to solve problems requiring use of ratio concept (for test properties, see Begolli & Richland, 2016).
2. At Visit 2, students interacted individually with a videotaped lesson on ratio, which centered on a comparison between two common solution strategies for solving ratio problems—one a frequent misconception (Subtraction) and one a correct solution (Least Common Multiple, or LCM). Greater details on the lesson and math assessments are provided in the supplementary materials.

Stereotype threat was manipulated prior to the video lesson, before learning. Specifically, at the start of the video lesson, students in the ST condition viewed a video screen in which the following prompt was visible and read aloud:

> Thank you for being part of this study. Please turn to page 2 in your packet and fill in your race. Professors at the University of Chicago are very interested in knowing about you and how you perform on the test after the lesson. Your information is very important to us, because we want to learn how to best help kids like you learn math.

Students in the Control condition viewed a video screen in which the following prompt was visible and read aloud:

> Thank you for being part of this activity. Please turn to page 2 in your packet and write today's date. We're glad that you are doing this activity because it will help us learn about some of the best ways to teach kids math.

In both conditions, the prompt was visible on the screen for 30 seconds. With the exception of this introduction, the video was otherwise identical between the two conditions.

This stereotype threat manipulation is quite similar to other manipulations of stereotype threat—particularly those employed with children in school contexts (e.g., Wasserberg, 2014; McKown & Strambler, 2009). Our manipulation of stereotype threat involved both increasing identity salience (by asking participants to provide their race and describing the study as being focused on better understanding how kids like you learn math), and raising the stakes (by informing students that an important aim of the study was to see how well they performed on a test after the lesson).

All students completed a mathematics posttest immediately following the video lesson. Following the posttest, a subset of the students also completed measures to gather data on their situational interest and mathematics engagement (more detail available in supplementary materials). An abbreviated version of the Situational Interest Survey (Chen, Darst, & Pangrazi, 2001) was administered to a subset of participants in both conditions, with test reliability leaving two measures: instant enjoyment and exploration intention.

A subset of students in both conditions was also given the option to work on non-required math puzzles during free time after finishing the intervention video and posttest.

3. At Visit 3, students completed a delayed posttest and a measure of EF, the d2 test of attention, (Brickenkamp & Zillmer, 1998), followed by a short demographics survey. More detailed descriptions of these measures can be found in the supplementary materials.

The research was conducted with school and the University's IRB approval. Details of the protocol introduction and debriefing are provided in the supplementary materials.

3. Results

3.1. Sample

Results for 135 students for whom we have complete data are included in the analyses reported below. Students at school 2 and at school 1 in year 2 (n = 79; 43 ST) also were administered the situational interest survey, and students at school 2 (n = 40; 21 ST) were administered the optional puzzles.

3.2. Analytic plan

We first report an analysis of pretest performance to ensure that random assignment was successful, and then describe analyses pertaining to our main questions of interest: how student EF and stereotype threat during instruction impacted learning. We examine this using three types of outcome measures. These measures were used to examine main effects on learning, as well as to test the hypothesis that high EF students would be most affected by the threat manipulation (with high EF students defined as those scoring above the median on the d2).

1. First, we examine immediate and sustained uptake of the misconception (subtraction) presented in the lesson, calculating uptake of the misconception immediately (proportion of problems solved with the misconception strategy at immediate posttest minus misconception use at pretest) and sustained over a delay (proportion of problems solved with the misconception strategy at delayed posttest minus at pretest). We anticipate that any gains in use of the misconception after the lesson may indicate that the student memorized the first solution that was presented, without engaging in the higher order comparative processing within the lesson, which would have made clear that this

is an invalid way to solve the problem. Thus, misconception gains could signal that students are motivated to memorize everything presented in the lesson in order to be able to perform well—at the possible expense of the type of deep, critical engagement with comparative instruction that promotes enduring conceptual understanding. Additionally, if EF resources were compromised, these participants may have been unable to engage in that higher level comparative thinking.

2. Second, we examine learning as defined as immediate and sustained gains in accuracy. Accuracy was calculated as correct setup and reasoning on a combination of procedural and conceptual items. Immediate learning was operationalized as accuracy gains from pretest to posttest, and sustained accuracy was operationalized as gains from pretest to the posttest administered a week after instruction.
3. Lastly, we analyzed the impact of experiencing stereotype threat during instruction on the extent to which students reported that they enjoyed the lesson and desired to learn more about the topic covered in the lesson, as well as their likelihood of actually choosing to engage in a non-required math activity.

3.3. Performance at pretest

Neither frequency of misconception use nor overall accuracy differed by condition at pretest (both $p_s > .5$) or between high and low EF students (both $p_s > .4$). Mean misconception use and accuracy at pretest for students in the two learning conditions are reported in Table 1.

3.4. Effects of stereotype threat manipulation overall

In a series of regressions, EF (entered as a continuous variable with values ranging from 136 to 665) and Condition assignment (binary coded: 0 = Control; 1 = ST) were first used to predict immediate and sustained changes in misconception use and content knowledge in the overall sample. Mean immediate and sustained gains in misconception use and accuracy for students in the Control and ST learning conditions are shown in Table 1, and all results and effect sizes from the regression are available in Table 2. First, experiencing stereotype threat while learning predicted significantly greater immediate uptake of the misconception ($p = .009$). This suggests that those students experiencing stereotype threat while learning may have been especially focused on memorizing content from the lesson at the expense of

Table 1
Immediate and sustained changes in use of the misconception and accuracy for students in the control and ST learning conditions

	Pretest Means (SE)		Immediate Gains (SE)		Sustained Gains (SE)	
	Misconception	Accuracy	Misconception	Accuracy	Misconception	Accuracy
Control, $n = 64$	0.46 (.06)	0.15 (.02)	−0.11 (.04)	0.18 (.03)	−0.01 (.05)	0.14 (.03)
ST, $n = 71$	0.42 (.06)	0.14 (.02)	0.05 (.05)	0.16 (.03)	0.08 (.04)	0.08 (.03)

Table 2
Contributions of EF and learning condition to immediate and sustained changes in misconception use and accuracy

Change From Pretest		r^2	p	$EFB_{standardized}$	p	Condition $B_{standardized}$	p
Immediate	Misconception	**.064**	.014	−.132	.126	**.227**	.009
	Accuracy	.036*	.093	**.182**	.038	−.074	.397
One-week delay	Misconception	.023	.223	−.105	.230	.119	.176
	Accuracy	.039*	.076	.15*	.086	−.143	.100

DF for all analyses (2, 129). Boldface value 0.227 has significance of $p < 0.01$. Boldface values have significance of $p < 0.05$.
*$p < .10$.

engaging in the cognitively demanding contrast and comparison process that promotes deep learning and conceptual change. The differences reduced over time, however, most likely because the misconception use of the Control condition increased between immediate and delayed posttest, returning to closer to pretest levels. Effects were less strong in the overall sample for use of the correct solutions, with no immediate differences in accuracy but the condition differences widened over time, revealing a trend to differences in sustained learning ($p = .09$).

Student EF was positively related to immediate learning of correct content ($p = .04$) and somewhat greater ($p = .09$) sustained gains for correct content, indicating that EF may play an important role in learning from this lesson, and that high EF participants learned more overall.

3.5. EF variations in effects of stereotype threat manipulation

While there were some impacts of condition and EF on learning for the full sample of participants, we next investigated the prediction from the testing literature, that the effects of stereotype threat are highest for high EF individuals. Condition assignment was used to predict immediate and sustained changes in misconception use and content knowledge among high and low EF students separately. All results and effect sizes from the regressions are available in Table 3, and mean immediate and sustained changes in misconception use and accuracy for high and low EF students in the Control and ST learning conditions are shown in Fig. 1A and B.

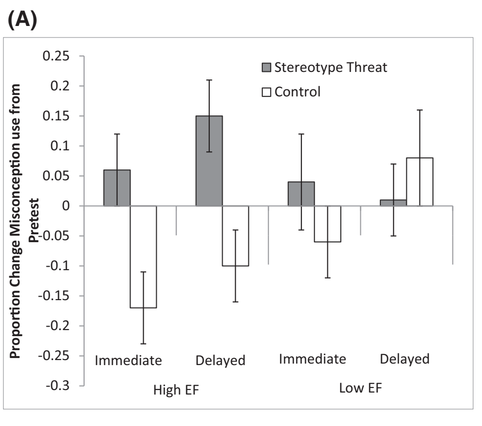
As shown in Table 3, among high EF students, experiencing stereotype threat while learning predicted greater immediate and sustained uptake of the misconception, and smaller sustained gains in correct content, with much larger effect sizes than when considering the sample as a whole. In contrast, for low EF students, learning condition did not impact immediate or sustained gains in either correct content or misconception use.

Post hoc analyses revealed that statistically significant results all had above .75 observed statistical power to detect effects, with the exception of sustained gains in the high EF group (.60). Non-significant results had smaller r^2 and thereby were underpowered to detect relationships, so Type II errors of unidentified effects were possible.

Table 3
Impact of stereotype threat during instruction on immediate and sustained changes in misconception use and accuracy from pretest among high versus low EF students

			r^2	Model p	Condition $B_{standardized}$	Condition p
High EF students	Immediate	Misconception use	**.096**	0.011	**0.310**	0.011
		Accuracy	.015	0.329	−0.122	0.329
	Sustained	Misconception use	**.109**	0.007	**0.331**	0.007
		Accuracy	**.069**	0.033	**−0.263**	0.033
Low EF students	Immediate	Misconception use	.018	0.284	0.134	0.284
		Accuracy	.000	0.977	0.004	0.977
	Sustained	Misconception use	.008	0.488	−0.087	0.488
		Accuracy	.000	0.975	0.004	0.975

Boldface values have significance of $p < 0.05$.
[a]Model degrees of freedom (1,64).



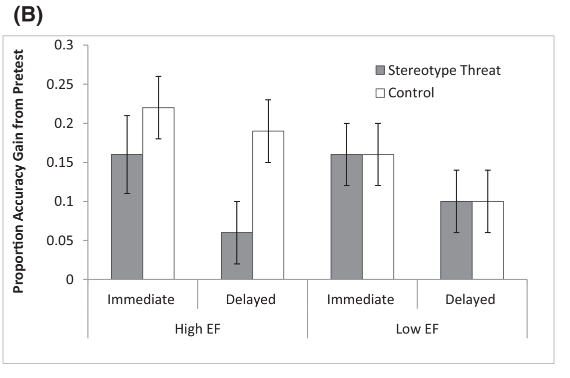


Fig. 1. Immediate and sustained changes in misconception use (A) and accuracy (B) for high vs. low EF students in the control and ST learning conditions.

3.6. Student attitudes

In order to better understand how stereotype threat during learning impacted students' subjective experience of the lesson and likelihood of choosing to engage in a non-required math activity, we compared students' responses to the exploration intention and instant enjoyment subscales of the situational interest survey and number of optional math puzzles completed in the two conditions.

We first examined whether these subscales were related to learning. Exploration intention was related to student learning, with those students who reported greater desire to learn more about the topic showing significantly greater overall gains in accuracy, $r(79) = .22$, $p < .05$. In contrast, instant enjoyment did not predict overall gains in accuracy, $r(79) = .22$, $p = .48$.

We next assessed relations to the manipulation. A one-way ANOVA revealed that students in the ST condition reported both significantly less enjoyment than students in the Control condition (Mean Control: 4.01(0.16); ST: 3.43 (0.19), $F(1, 77) = 5.22$, $p = .03$; *partial* $\eta^2 = 0.063$; observed power = 0.616), and less exploration intention (Mean Control: 3.76 (0.18); ST: 3.25 (0.18), $F(1, 77) = 4.01$, $p = .05$; *partial* $\eta^2 = 0.048$; observed power = 0.496). These differences in students' experience of the lesson are striking because they emerged immediately after the first posttest, when performance did not yet systematically differ across conditions. Moreover, students who had experienced stereotype threat during instruction were less likely to choose to engage in optional math activities after the intervention, attempting significantly fewer math puzzles than students who had not experienced stereotype threat during instruction (Mean Control: 3.47 (0.30); ST: 2.29 (0.43), $F(1, 38) = 5.00$, $p = .03$; *partial* $\eta^2 = 0.116$; observed power = 0.587).

4. Discussion and conclusions

These data provide the first evidence that stereotype threat during conceptually demanding instruction can harm mathematics attitudes and learning. Increasing the salience of race in the context of an evaluative introductory prompt led to increased misconceptions and decreased retention for the conceptual content—particularly among high EF students, who otherwise benefited most from this type of instruction. Overall, EF facilitated learning, predicting greater gains from the lesson. However, stereotype threat had a greater detrimental impact on learning among high EF students. Low EF students learned less regardless of learning condition, while greater between-condition differences were found for high EF students.

Two primary, non-exclusive, mechanisms could explain these results, and both receive some support in this study. First, it is possible that worry ideation based on a fear of confirming negative stereotypes compromised participants' EF capacity during instruction. With reduced EF available for high EF students accustomed to engaging these resources in the sorts of abstract mathematical thinking intended during this lesson, these typically higher performing, high EF students suffered the most. This is consistent with much research on pressure, anxiety, and stereotype threat, which shows the largest effects for high EF or working memory individuals (for reviews, see Beilock, 2008; Maloney et al., 2014). The ST threat may have led these students to switch from concept formation learning strategies that demand high EF resources, to less cognitively effortful memorization strategies.

A second possible contributory explanation for the results was that students who were administered the experimental prompt reported less interest and enjoyment in the mathematics. Additionally, these participants made the choice to spend less time on optional math puzzles. These findings suggest that students who experienced stereotype threat during instruction may have been less likely to think about content from the lesson between sessions or make connections with topics covered during their math classes, leading to worse consolidation of learning and resulting in the lower performance. These data were limited since they were not collected from all participants, but they suggest the potential importance of this pathway impacting both children's learning and engagement in mathematics.

This pathway received some support, as exploration intention predicted student learning, although enjoyment did not. This suggests that differences in exploration intention resulting from experiences of stereotype threat during learning may have contributed to between-condition learning differences, but that differences in enjoyment likely did not. However, even though within the boundaries of this experiment, decreased enjoyment among students who experienced stereotype threat during instruction likely did not drive between-condition learning differences, decreased enjoyment in the face of stereotype threat may have important long-term implications, possibly impacting career goals and leading to less persistence in the content domain. This mechanism is crucial to consider, because it suggests ways that perceptions of stereotype threat could have long-term implications for student engagement with a domain of knowledge, such as mathematics.

This work is a first step in understanding the phenomenon of stereotype threat during higher order thinking and learning, and it raises the possibility that subtle threats could be contributing to real and persistent achievement gaps by race, ethnicity, and gender. Importantly, whereas the effects of stereotype threat during testing are immediate but temporary and do not affect the knowledge base from which children progressively build new understandings, the effects of stereotype threat during instruction persist even when the threat is no longer present and are cumulative, as learning is built on a foundation of prior knowledge. If experiences of stereotype threat prevent students from fully benefitting from instruction at one time point, in a discipline such as mathematics that builds sequentially, subsequent learning opportunities may also be affected, leading to students falling farther and farther behind. Additionally, feelings of performance pressure due to stereotype threats could also lead to disengagement, reduced interest and experience seeking in domains in which these threats are perceived.

As stereotype threat can be induced with very little prompting (McKown & Strambler, 2009), students from academically stereotyped groups may be experiencing it during instruction on a regular basis and the cumulative effect on learning could be even more damaging than underperformance on tests. Future research into what elicits stereotype threat during learning opportunities, for whom, and how often it occurs, will be vital for understanding how to guard against and ameliorate the effects of stereotype threat in the classroom, for both testing and learning opportunities.

Acknowledgments

The research reported here was supported by an NSF Fellowship DGE-1144082 award to Emily Lyons McLaughlin, and grants from the National Science Foundation, SMA-1548292, and the Institute of Education Sciences, U.S. Department of Education, through R305A170488 to the University of Chicago. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

References

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double edged sword. *Educational Psychology*, *107*, 194–213.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*, 339–343.

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

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*, 983–998.

Brickenkamp, R., & Zillmer, E. (1998). *The d2 test of attention*. Seattle, WA: Hogrefe & Huber.

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British Journal of Educational Psychology*, *71*, 383.

Davies, L. C., Conner, M., Sedikides, C., & Hutter, R. C. (2016). Math question type and stereotype threat: Evidence from educational settings. *Social Cognition*, *34*(3), 196. https://doi.org/10.1521/soco.2016.34.3.196.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, *22*, 206–214.

Larnell, G. V., Boston, D., & Bragelman, J. (2014). The stuff of stereotypes: Toward unpacking identity threats amid African American students' learning experiences. *Journal of Education*, *194*(1), 49–57.

Maloney, E., Sattizahn, J., & Beilock, S. L. (2014). Anxiety and cognition. *WIREs Cognitive Science*, *5*, 403–411.

Maloney, E., Schaeffer, M. W., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences and promising interventions. *Research in Mathematics Education*, *15*(1), 115–128.

McKown, C. (2004). Age and ethnic variation in children's thinking about the nature of racism. *Journal of Applied Developmental Psychology*, *25*, 597–617.

McKown, C., & Strambler, M. J. (2009). Developmental antecedents and social and academic consequences of stereotype-consciousness in middle childhood. *Child Development*, *80*, 1643–1659.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype-consciousness in middle childhood. *Child Development*, *74*, 498–515.

Nasir, N. S., Atukpawu, G., O'Connor, K., Davis, M., Wischnia, S., & Tsang, J. (2009). Wrestling with the legacy of stereotypes: Being African American in math class. In D. B. Martin (Ed.), *Mathematics teaching, learning, and liberation in the lives of Black children* (pp. 231–248). London, UK: Routledge.

Nasir, N. S., Snyder, C. R., Shah, N., & Ross, K. M. (2012). Racial storylines and implications for learning. *Human Development*, *55*, 28S–301.

Neuburger, S., Ruthsatz, V., Jansen, P., & Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learning and Individual Differences*, *37*, 169–175.

Nguyen, H. H., & 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*, 1314–1334.

Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*, 28–43.

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*, 561–574.

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

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*, 99–114.

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.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research*, *37*(2), 237–260.

Wasserberg, M. J. (2014). Stereotype threat effects on African American children in an urban elementary school. *Journal Of Experimental Education*, *82*(4), 502–517.

Watson, C. E., Johanson, M., & Dankiw, J. (2014). Effects of high-stakes testing on third through fifth grade students: Student voices and concerns for educational leaders. *Journal of Organizational Learning & Leadership*, *12*(1), 1–11.

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. W. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. *Journal of Personality and Social Psychology*, *103*(4), 635–646.

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*(6), 1197–1232.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*(1), 160–175.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Supplemental Information.