Identification with Math Boosts Women's Performance on Mathematical Tasks under Stereotype Threat

Benjamin A. Saunders
Long Island University – Brooklyn

Abstract

Stereotype threat research (Steele & Aronson, 1995) suggests that people often underperform on tasks when they face the prospect of confirming a negative group stereotype. The adverse effect of stereotype threat on performance, moreover, should be stronger for people who identify with the domain in which the stereotype is relevant. The notion that domain identification—the degree to which one's self-worth depends on achievements in a specific domain—further exacerbates performance under stereotype threat warrants further examination. This experiment tested whether identification with math burdened or boosted women's performance on a math test under conditions of stereotype threat. The results revealed no relationship between identification with math and math performance when the test was not diagnostic of participants' ability. However, identification with math positively predicted participants' math performance when the test was diagnostic of ability. These results demonstrate that identification with math boosts (rather than burdens) women's math performance under stereotype threat, and they call for further research on the relationship between domain identification and performance—in stereotype threat situations—across the different social identities, domains, and stereotypes that people share.

Keywords: stereotype threat, domain identification, test performance, stereotypes

The stereotype that women have fewer quantitative skills than men is pervasive in the U.S., and research suggests that it may undermine women's performance in math unless they are explicitly told that math tests do *not* produce gender differences (Spencer, Steele, & Quinn, 1999). Specifically, s*tereotype threat* (Spencer et al., 1999; Steele & Aronson, 1995) indicates that the prospect of confirming negative group stereotypes may depress performance on tasks for which the stereotype may apply. Steele and Aronson (1995) first articulated the parameters of stereotype threat by examining how (a) the effect can be induced and (b) what effect the threat might have on test performance. In developing a social-psychological account for the racial academic achievement gap, they reasoned that Black students may at times become concerned with confirming stereotypes regarding their intellectual inferiority or academic incompetence, and that these concerns may adversely affect performance. Consistent with this idea, Black students underperformed relative to their White counterparts on a difficult verbal test when it was framed as being diagnostic of ability or when asked to indicate their race prior to the test. When the same test was characterized as *not* being diagnostic of ability (or in the absence of race primes), both groups performed similarly (Steele & Aronson, 1995). Additional research revealed comparable effects for women in the domain of mathematics. Whereas women and men performed equally well on easy math tests, women underperformed relative to men on difficult tests unless explicitly told that the test did not produce gender differences (Spencer et al., 1999). Recent research suggests that people underperform under stereotype threat because the experience of stereotype threat disrupts cognitive, physiological, and affective processes which, in turn, undermine one's working memory (Schmader & Beilock, 2012; Schmader, Johns, & Forbes, 2008).

I would like to thank Joshua Haag, David Mazer, Alex Nelson, and Adriana Rodriguez for their assistance with data collection.

Correspondence concerning this article should be addressed to Benjamin A. Saunders, Department of Psychology, LIU – Brooklyn, 1 University Plaza, Brooklyn, NY 11201-5372; email: benjamin.saunders@liu.edu.

One relatively under-examined tenet of stereotype threat theory purports that either the phenomenon only exists among people who identify with the performance domain in which they are stereotyped, or that identifying with a stereotype-relevant domain exacerbates the already-adverse effect (Steele, 1997; Steele, Spencer, & Aronson, 2002). According to the theory, domain identification should therefore worsen performance in stereotype threat situations. Although there is some support for this assumption (Aronson, Lustina, Good, Keough, Steele, & Brown, 1999; Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Keller, 2007; Lawrence, Marks, & Jackson, 2010; Stone, Lynch, Sjomeling, & Darley, 1999), recent theoretical and empirical developments leave open the possibility that domain identification may not always impair performance under conditions of stereotype threat. This paper proposes that—at least for some cases—people with high domain identification will perform better than people with low domain identification (i.e., *the boost hypothesis*). This paper will review the literature on stereotype threat and domain identification before reporting an experiment that tests whether identification with math boosts or burdens female students' math performance when they face the negative stereotype about women and math.

Domain Identification

Domain identification refers to the degree to which people view a given domain (e.g., mathematics) as desirable, consequential, and yielding positive outcomes (i.e., intrinsic enjoyment, praise and achievements; Steele, 1997; Steele, Spencer, & Aronson, 2002). Accordingly, when people identify with a domain like math, they view math as self-relevant and self-defining, and likely consider successes and failures in something like mathematics courses as central to their self-regard. One long-standing tenet of stereotype threat theory posits that domain identification exacerbates performance when people encounter stereotype threat (Steele, 1997; Steele et al., 2002). In other words, domain identification—under conditions of stereotype threat—should negatively predict performance. Although the available literature supports this part of stereotype threat theory, the inconsistent manner in which researchers measured domain identification across studies suggests that its role in stereotype threat research requires further investigation.

The first wave of stereotype threat studies to examine the role of domain identification found support for the idea that identification burdens performance. One of the first studies to test the *burden hypothesis* assigned participants to two groups based on their level of endorsement on a single-item measure of domain identification (Aronson et al., 1999). In examining the stereotype that White men underperform on math tests compared to Asians, the results revealed that highly-identified White male participants performed worse under stereotype threat than their moderately-identified counterparts. Other research found support for the burden hypothesis in demonstrating that athletically engaged White participants performed worse on an athletic task than their disengaged counterparts when contending with the stereotype that White people have poor athletic ability (Stone et al., 1999).

Subsequent research corroborated the burden hypothesis among women in the domain of mathematics (Cadinu et al., 2003); when female participants rank-ordered logical-mathematical ability as their most- or second-most personally important ability (compared to social ability and creative ability), they performed less well than female participants who ranked it third under conditions of stereotype threat. Keller (2007), moreover, found that when female high school students' domain identification scores were split at the median, high-identifiers outperformed low-identifiers under control conditions, but not under stereotype threat, where the pattern reversed such that low-identifiers outperformed high identifiers.

The limited stereotype threat research that has tested the burden hypothesis with Black students has yielded mixed results. One study found no evidence that domain identification was a performance burden for Black students under stereotype threat (McFarland, Lev-Arey, & Ziegert, 2003). However, another study found that Black students' verbal identification (i.e., the importance of verbal abilities to one's sense of self) negatively predicted their performance on a verbal task under stereotype threat but not under control conditions (Lawrence et al., 2010).

In summary, the bulk of research that has tested the burden hypothesis suggests that people highly identified with a stereotype-relevant domain underperform in stereotype threat predicaments compared to their moderately- or low-identified counterparts. However, more-recent research suggests that the role of domain identification in stereotype threat merits further investigation. Specifically, the largest stereotype threat meta-analysis to date noted that—on an aggregate level—"moderately math-identified women were surprisingly affected more severely by stereotype threat than highly math-identified ones" (Nguyen & Ryan, 2008, p. 1327). In other words, meta-analytic work on stereotype threat research leaves open the possibility that domain identification might *boost* students' performance rather than burden it. In accommodating their observation with previous stereotype threat research, Nguyen and Ryan suggested that their theory-discrepant results might shed light on the general inconsistency with which domain identification has been operationalized, and, perhaps equally important, a relatively narrow focus on highly-identified participants at the expense of gaining a better understanding of the

general role that domain identification may play in stereotype threat research. Both of these issues warrant further elaboration, and the next section will discuss why these often-overlooked aspects of stereotype threat research are problematic.

Reconciling the Role of Domain Identification

Earlier stereotype threat research may have supported the burden hypothesis for two related reasons. First, the finding that domain identification negatively predicts performance under conditions of stereotype threat might have emerged on the basis of an overreliance on specialized populations (Smith & White, 2001; Steele, 1997). Support for the burden hypothesis often emerged in experiments where the inclusion criteria limited participation to students considered among the *academic vanguard* (i.e., those with the highest level of preparation for academic success; Steele, 1997). In limiting participation to students who, for example, had high GPAs, had high standardized test scores, or were enrolled in advanced mathematics courses, early stereotype threat research restricted the range of academic preparation to the scholarly elite. The exclusivity of early stereotype threat research, on one hand, demonstrated that the apprehension associated with confirming negative stereotypes about one's group affects even the best students. On the other hand, restricting the range of students' academic preparation limited the generalizability of stereotype threat research.

The fact that early stereotype threat researchers often limited the range of participation to the academic vanguard resulted in a considerable amount of research on how stereotype threat affects elite students, perhaps at the expense of understanding how the joint and interactive effects of stereotype threat and domain identification affect students more broadly. The dearth of stereotype threat research on students in general is problematic given that researchers have known for at least twenty years that the burden of stereotype threat might affect students differentially depending on their preparation (Strenta, Elliott, Adair, Scott, & Mailer, 1993). Supporting this idea, the mismatch between one's academic performance and what one's standardized test scores might predict occurs consistently among talented female students who major in subjects like mathematics. The same effect, however, tends not to occur among women with less math and science preparation, or among women who took math and science courses intended for non-majors. Thus, the assumption that domain identification hinders performance seems to result from a relatively narrow focus on those students who are most talented and most prepared for academic success rather than on research that examined the full spectrum of academic achievement.

Another reason why support for the burden hypothesis might have emerged is because there was a lack of consensus or agreement on the best practices for measuring domain identification. Numerous early stereotype threat studies did not measure domain identification at all; instead, these studies used standardized test scores, GPA, and/or enrollment in advanced math classes as their criteria for participant inclusion or as a proxy for domain identification (Smith & White, 2001). In other words, rather than measuring domain identification, early stereotype threat research often presumed that belonging to the academic vanguard corresponded with high domain identification.

Other studies used a handful of ad hoc items seemingly related to domain identification to select participants for their research. Domain identification, for example, has been assessed with a single-item measure (Aronson et al., 1999), an ordinal ranking of abilities (Cadinu et al., 2003), and with two- or three-item *post-experimental* measures (Leyens, Désert, Croizet, & Darcis, 2000; McFarland et al., 2003). Although there may not have been better alternatives for measuring identification at the time, a psychometrically sound measure of domain identification—designed with stereotype threat research in mind (Smith & White, 2001)—appeared just as accord about the role of domain identification as a burden emerged. The domain identification measure has therefore been employed sparsely in stereotype threat research and deserves more empirical attention as the role of domain identification in stereotype threat continues to materialize.

Support for the burden hypothesis, moreover, often emerged in research that split domain identification at the median (Keller, 2007; Stone et al., 1999), or in studies where only participants scoring in the upper or lower third on measures of domain identification were included (Aronson et al., 1999). Dichotomizing or splitting continuous variables is problematic for a few important reasons. First, the practice of dichotomizing continuous variables often loses information about individual differences (MacCallum, Zhang, Preacher, & Rucker, 2002). For example, when researchers split a continuous variable at the median, they risk categorizing participants with very similar scores (i.e., participants with scores just below or just above the median) into different groups. Dichotomizing continuous variables also reduces effect size, power, and measurement reliability. Thus, rather than dichotomizing continuous variables, applying regression and correlational analysis to original (i.e., un-dichotomized) variables minimizes the above issues.

Consistent with Nguyen and Ryan's (2008) observation that highly math-identified women outperformed moderately math-identified women under stereotype threat, Lawrence and Charbonneau (2009) observed that

identification with math positively predicted performance on math tests, and that the positive relationship between identification with math and test performance occurred in ability-diagnostic (i.e., stereotype threat) conditions as well as non-diagnostic ones. Although these results were observed without manipulating stereotype threat (i.e., the setting in which participants completed the math test varied across two separate experiments) they nonetheless introduce the possibility that domain identification may boost performance under stereotype threat. The fact that these results conflict with later work supporting the burden hypothesis (Lawrence et al., 2010) further suggests that the interplay between stereotype threat and domain identification may vary across social identities and stereotyped domains.

In line with this idea, Shapiro and Neuberg (2007) posit that identifying with a stereotyped domain may be tied to the emergence of stereotype threat for some groups but not for others, and recent research suggests that people belonging to different negatively-stereotyped groups experience different patterns of stereotype threat (Shapiro, 2011). According to the *multi-threat framework* (Shapiro & Neuberg, 2007), people's concerns about whether their performance confirms negative group stereotypes might lead to an array of possible *threats* (i.e., plural) rather than one diffuse threat (Shapiro, 2012). The relative impact of potential stereotype threat situations, then, depends on how one considers two distinct dimensions—the target of the stereotype threat (i.e., whether the threat is to one's self or to one's group), and the source of the threat (i.e., whether one's self, out-group others, or in-group others will draw conclusions about one's performance). Within the multi-threat framework, the above dimensions might produce one of six distinct stereotype threats, including self-concept threat (i.e., concerns that one's performance will confirm—in one's own mind—that a given stereotype is true of one's self), out-group own-reputation threat (i.e., concerns that one's performance will confirm—in the minds of out-group members—that a given stereotype is true of one's self), in-group own-reputation threat (i.e., concerns that one's performance will confirm—in the minds of in-group members—that a given stereotype is true of one's self), group-concept threat (i.e., concerns that one's performance will confirm—in one's own mind—that a given stereotype is true of one's group), out-group reputation threat (i.e., concerns that one's performance will confirm—in the minds of out-group members—that a given stereotype is true of one's group), and, last, in-group reputation threat (i.e., concerns that one's performance will confirm—in the minds of in-group members—that a given stereotype is true of one's group).

Initial research on the multi-threat framework revealed, for example, that participants who agreed that others associate their race or ethnicity with a negative stereotype tended not to endorse the stereotype themselves (Shapiro, 2011). Furthermore, they experienced significantly less concern about confirming that the stereotype was true of themselves or of their negatively-stereotyped group; instead they tended to be more worried about reinforcing the stereotype to others. Conversely, participants who agreed that others associate their weight with negative stereotypes tended not to identify strongly with being overweight, and were more worried about confirming the stereotype to themselves (i.e., in their own eyes) than they were over confirming the stereotype to or about others. What these different patterns of stereotype threat may mean for performance on domain relevant tasks requires further study, as does the role of gender in the broader multi-threat framework. However, Shapiro's (2011) findings suggest that it is at least possible that the threat that women may encounter in the domain of math is of a different nature than the threat that Black students may encounter in the domain of academics or intellect. Importantly, the multi-threat framework posits that domain identification might only burden performance in the case of self-concept threats, but not the other ones (Shapiro & Neuberg, 2007). That domain identification might only apply to self-concept threats, then, offers yet another explanation for the discrepancy between stereotype threat theory (Steele et al., 2002) and recent meta-analytic reviews (Nguyen & Ryan, 2008).

In summary, the above findings suggest that the construct of domain identification deserves more attention as an individual difference variable within stereotype threat studies, and that there is reason to examine the idea that domain identification may boost female students' performance rather than undermine it.

Method

Participants

One hundred eleven female undergraduate students at a multicultural, urban university participated in this experiment for partial fulfillment of course requirements. Their ages ranged from 18 to 49 years (M = 21.92, SD = 6.58). The sample was ethnically diverse: 29% of participants reported their ethnicity as African American or Black, 21% as Asian, 14% as Caucasian or White, 12% as Latino/a or Hispanic, nine percent as Middle Eastern, and eight percent as multi-racial. Four percent reported "Other" and four percent opted to not report their ethnicity.

Procedure

Participants attended a brief pre-experimental session at least 48 hours in advance of the experiment. During the pre-experimental session, they completed basic demographic items as well as the Domain Identification Measure (DIM; Smith & White, 2001). At the conclusion of this session, participants scheduled an appointment with a research assistant for participation in the experiment. Upon arriving at the laboratory, participants were greeted by a White male experimenter who explained that, over the next 30 minutes, they would complete a math test consisting of items similar to those seen on college entrance exams.

Manipulation

Participants were randomly assigned to one of two conditions via Research Randomizer—an online tool that allows researchers to generate random numbers for use in assigning participants to experimental conditions (https://www.randomizer.org). The present research manipulated stereotype threat by informing participants that the test they were about to take was either non-diagnostic or diagnostic of ability (Steele & Aronson, 1995). Participants in the *non-diagnostic* condition listened to instructions explaining that the purpose of the study was to assess the mental processes involved in solving math problems, and that the test was unrelated to math competence and therefore could not assess their math ability. *Ability diagnostic* participants, however, listened to instructions indicating that the purpose of the study was to investigate the mental factors involved in solving math problems, that the test would measure their math ability, and that they would receive feedback on their performance at the end of the study.

Measures

Domain identification. The nine math-related items from the DIM (Smith & White, 2001) assess the extent to which people identify with mathematics, and example items include "Mathematics is one of my best subjects" (1 = *strongly disagree*; 5 = *strongly agree*), "How important is it to you to be good at Math?" (1 = *not at all*; 5 = *very much*), and "Compared to other students, how good are you at math?" (1 = *very poor*; 5 = *excellent*). Items were averaged to create an index of identification with math with excellent internal consistency ($\alpha = .93$), and the Cronbach's alpha reported for the DIM in the present study matches the internal consistency reported in prior research ($\alpha = .93$; Smith & White, 2001).

Test performance. Participants completed the Necessary Arithmetic Operations Test (NAOT; French, Ekstrom, & Price, 1963)—a time-limited (10 minutes) performance measure containing 20 quantitative items. The NAOT instructs test takers to identify the operation(s) needed to solve each arithmetic problem correctly without producing the quantitative solution to the problem. Prior research found support for stereotype threat effects using the NAOT (Danso & Esses, 2001; Wout, Danso, Jackson, & Spencer, 2008). The NAOT, moreover, yields a high loading on *g* (i.e., general intelligence; Marshlek, Lohman, & Snow, 1983), and is strongly correlated with visuospatial tasks like the Raven Progressive Matrices (Raven, 1948) and verbal tasks (i.e., the Sternberg Verbal Analogies task; Ekstrom, French, Harman, & Dermen, 1976) alike (see Snow, Kyllonen, & Marshalek, 1984 for more information). The present research examined test performance using two metrics often employed in stereotype threat research. Number correct refers to the total number of correct responses reported by participants. Since some stereotype threat researchers suggest that test accuracy is a more meaningful measure of performance than number correct (Shih, Pittinsky, & Ambady, 1999), the present study also examined accuracy (i.e., the total number of correct responses divided by the number of problems attempted).

Results

Preliminary Analyses

Means and standard deviations for the primary study variables are presented in Table 1. Participants' overall test performance (i.e., as measured by number correct) overlapped considerably with their test accuracy, $r(109) = 0.74$, $p < .001$, yet they appear to assess slightly different constructs given that they are not correlated more strongly with one another. Together, the preliminary results suggest the following: (1) participants in the diagnostic condition were more accurate than their non-diagnostic counterparts, and (2) identification with math did not vary by condition. Although the preliminary results do not support what stereotype threat theory would predict, they

nonetheless leave open the possibility that test performance might vary as a function of the two-way interaction between identification with math and test diagnosticity.

Table 1

Means and Standard Deviations for the Primary Study Variables

	Non-Diagnostic		Diagnostic				
Measure	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*
Number Correct	9.80	3.11	10.41	3.65	109	-0.95	.34
Accuracy	0.50	0.16	0.59	0.18	109	-2.79	.01
Identification with Math	3.17	1.08	3.43	1.08	106	-1.25	.22

Recall that recent meta-analytic work (Nguyen & Ryan, 2008) suggested that highly-identified participants might show reactance in the face of stereotype threat. If, for example, higher levels of identification with math are associated with increased motivation to persist in the face of stereotype threat, then one might expect the effect of stereotype threat on performance to depend on the degree to which people identify with math. The potential moderating role of identification with math will be discussed next.

Data Analytic Plan

The present research tested the burden and boost hypotheses with a moderated hierarchical regression. In doing so, this research examined the effect of domain identification on performance—at different levels of condition—while keeping domain identification continuous (see Lawrence et al., 2010 for a similar data-analytic test of domain identification), rather than employing the median split technique often used in domain identification research (Aronson et al., 1999; Keller, 2007; Stone et al., 1999). Remember that domain identification as a continuous variable rather than a dichotomous one is advantageous for at least two reasons (MacCallum et al., 2002). First, continuous variables maintain information about individual differences that dichotomous variables may lose. Second, moderated regression with a continuous variable has a relative effect size and power advantage over the median split technique sometimes employed with ANOVA.

Hypothesis Testing

Number correct. Number correct was regressed on condition and the centered identification with math variable. The present analyses employed a two-block hierarchical regression; the main effects of condition and identification with math were entered in the first block, and the two-way condition x identification with math interaction was entered in the second block.

Although condition did not account for any variability in test performance ($\beta = 0.26$, $t(105) = 0.42$, $p = .67$, $sr^2 = .00$), identification with math did ($\beta = 0.84$, $t(105) = 2.90$, $p = .01$, $sr^2 = .07$, $R^2 = .08$, $F[2, 105] = 4.51$, $p = .01$). Specifically, identification with math positively predicted the number of items participants solved correctly. The second block, moreover, revealed that the effect of condition on number correct was qualified by the condition x identification with math interaction ($\beta = 1.29$, $t(104) = 2.26$, $p = .03$, $sr^2 = .05$, $R^2 = .12$, $F[1, 104] = 4.83$, $p = .003$).

Simple slopes analyses revealed that identification with math did not predict women's test performance in the non-diagnostic condition, $\beta = 0.23$, $t(104) = .59$, $p = .56$, $sr^2 = .00$, whereas identification with math positively predicted women's test performance in the ability-diagnostic condition, $\beta = 1.52$, $t(104) = 3.68$, $p < .001$, $sr^2 = .11$ (see Figure 1). These results suggest that identification with math boosted rather than burdened women's math performance when they faced stereotype threat.

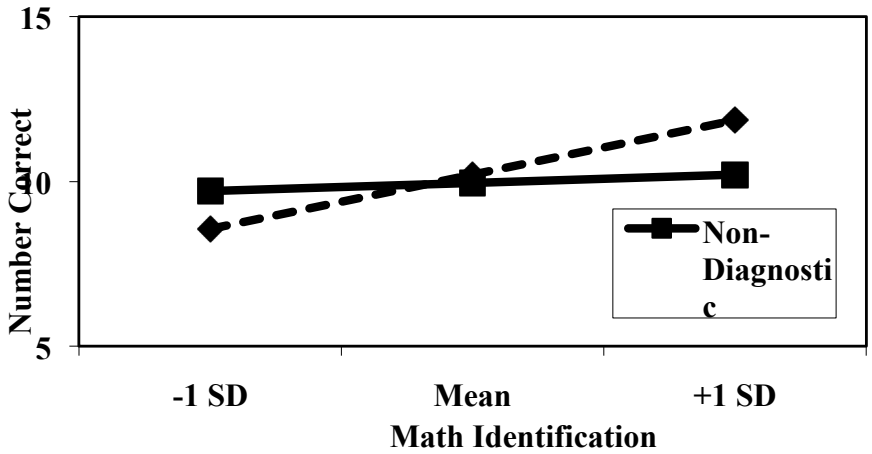

Figure 1. The effect of identification with math on women's test performance by condition.

Accuracy. Recall that because some stereotype threat researchers argue that test accuracy is a more meaningful measure of performance than number correct (Shih et al., 1999), the present research also tested the burden and boost hypotheses with accuracy as the dependent variable in a moderated hierarchical regression. The accuracy results departed from those reported for number correct in two key ways, but they still supported the boost hypothesis. The two ways in which the accuracy results diverged from those reported for number correct were as follows: First, while identification with math again positively predicted performance ($\beta = 0.04$, $t(105) = 2.49$, $p = .02$, $sr^2 = .06$), the effect of condition on accuracy revealed that ability-diagnostic participants were slightly more accurate than non-diagnostic ones ($\beta = 0.07$, $t(105) = 2.27$, $p = .03$, $sr^2 = .05$, $R^2 = .11$, $F[2, 105] = 6.43$, $p = .002$). The finding that participants in the ability-diagnostic condition were more accurate than their non-diagnostic peers likely reflects the fact that ability-diagnostic participants ($M = 17.90$, $SD = 4.32$) attempted fewer items than non-diagnostic participants ($M = 20.15$, $SD = 4.95$, $t(109) = 2.53$, $p = .01$), and the tendency for people to attempt fewer problems when under stereotype threat is supported in prior research (Steele & Aronson, 1995). Second, the condition x identification interactive effect on accuracy did not reach statistical significance, ($\beta = 0.02$, $t(104) = 0.50$, $p = .62$, $sr^2 = .00$). In reconciling the results of this analysis with the same effect reported for total number correct, one plausible-but-speculative explanation is that by attempting fewer problems, ability-diagnostic participants prioritized accuracy at the expense of overall test performance. Attempting fewer problems would be advantageous in situations where important academic outcomes depend exclusively on the items that students attempt, as tends to be the case with standardized tests like the Graduate Record Examination (GRE) and the Scholastic Aptitude Test (SAT). However, this strategy would be disadvantageous in situations where performance is based on students' performance across a full test domain, as tends to be the case in most university settings.

Although the condition x identification interaction on accuracy was not significant, the burden and boost hypotheses require an examination of the condition x identification with math interactive effect and it was examined nonetheless. In analyzing the simple slopes, the pattern of results resembled those reported for number correct. Specifically, there was no relationship between identification with math and accuracy among participants in the non-diagnostic condition, $\beta = 0.03$, $t(104) = 1.46$, $p = .15$, $sr^2 = .02$. But, as with the results reported for number correct and consistent with the boost hypothesis, identification with math positively predicted accuracy in the ability-diagnostic condition, $\beta = 0.05$, $t(104) = 2.06$, $p = .04$, $sr^2 = .04$. These results again support the idea that identification with math boosts women's math performance under conditions of stereotype threat.

Discussion

The present experiment—in contrast to the bulk of stereotype threat research to have examined the effect of domain identification on performance—found that identification with math boosted (rather than burdened) female participants' math performance under stereotype threat. Although these results are inconsistent with stereotype threat theory, they nonetheless resemble a pattern observed in recent meta-analytic work that found that highly-identified women outperformed moderately-identified ones under stereotype threat (Nguyen & Ryan, 2008). These

results are the first to demonstrate that identification with a particular domain may help (rather than hinder) performance on domain-relevant tasks when under stereotype threat.

Perhaps the strongest recent support for the burden hypothesis used a similar methodology and data-analytic strategy, albeit with a different stereotyped group and stereotype-relevant domain (Lawrence et al., 2010). Until recently, stereotype threat research generally made no distinction between how members of different stereotyped groups would perform when faced with the prospect of confirming a negative stereotype (Shapiro, 2011, 2012; Shapiro & Neuberg, 2007). In fact, the effect of stereotype threat on performance, if anything, had been treated as conceptually interchangeable across different social identities (Nguyen & Ryan, 2008). However, recent research (Shapiro, 2011)—as well as the results reported here—call into question whether researchers should assume that everyone experiences stereotype threat the same way. In other words, it may be the case that domain identification boosts performance under stereotype threat in some domains, or with some stereotyped groups, while burdening performance for others. Accordingly, more research should be conducted to parcel out when domain identification will help versus hinder performance under stereotype threat.

Identifying how and why domain identification facilitates performance under stereotype threat (at least in some cases) seems like an area for future research with considerable promise. Prior research demonstrated that framing stereotype threat situations as challenges virtually eliminated the threat; the typical underperformance effect observed under conditions of stereotype threat did not occur when participants viewed the test-taking situation as a challenge (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010). This finding, taken with Nguyen and Ryan's (2008) speculation that highly-identified participants may show reactance in the face of stereotype threat, offer a potential insight into the phenomenology of stereotype threat for the highly identified. It could be the case, for example, that negatively stereotyped people perceive stereotype threat situations as a challenge rather than as a threat and attempt to use the threat as motivation. This explanation seems plausible within the context of the present experiment since identification with math only boosted performance for participants under threat—the relationship between identification with math and test performance was non-significant and weak for participants in the non-diagnostic condition.

In conclusion, the results of this experiment suggested that there are cases where people highly identified with a domain actually thrive (rather than fetter) when faced with the risk of confirming a negative stereotype about their ability. These findings warrant further research, especially given the emerging understanding that the relationship between domain identification and performance may depend on a dynamic and potentially limitless combination of social identities, domains, and stereotypes that people share (Lawrence & Charbonneau 2009; Nguyen & Ryan 2008; Shapiro & Neuberg 2007). Understanding these factors will promote more informed inquiries into the specific role that domain identification may play in stereotype threat situations.

References

Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology, 46*, 166-171. doi:10.1016/j.jesp.2009.09.014

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29-46. doi:10.1006/jesp.1998.1371

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*, 267-285. doi:10.1002/ejsp.145

Danso, H. A., & Esses, V. M. (2001). Black experimenters and the intellectual test performance of White participants: The tables are turned. *Journal of Experimental Social Psychology, 37*, 158-165. doi:10.1006/jesp.2000.1444

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

French, J., Ekstrom, R., & Price, L. (1963). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77*, 323-338. doi:10.1348/000709906X113662

Lawrence, J. S., & Charbonneau, J. (2009). The link between basing self-worth on academics and student performance depends on domain identification and academic setting. *Learning and Individual Differences, 19*, 615-620. doi:10.1016/j.lindif.2009.08.005

Lawrence, J. S., Marks, B. T., & Jackson, J. S. (2010). Domain identification predicts black students' underperformance on moderately-difficult tests. *Motivation and Emotion, 34*, 105-109. doi:10.1007/s11031-010-9159-8

Leyens, J., Désert, M., Croizet, J., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189-1199. doi:10.1177/0146167200262002

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40. doi:10.1037/1082-989X.7.1.19

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radix and hierarchical models of intelligence. *Intelligence, 7*, 107-127. doi:10.1016/0160-2896(83)90023-5

McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16*, 181-205. doi:10.1207/S15327043HUP1603_2

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

Raven, J. C. (1948). The comparative assessment of intellectual ability. *British Journal of Psychology, 39*, 12-19.

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356. doi:10.1037/0033-295X.115.2.336

Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin, 37*, 464-480. doi:10.1177/0146167211398140

Shapiro, J. R. (2012). Types of threats: From stereotype threat to stereotype threats. In T. Schmader & M. Inzlicht (Eds.), *Stereotype threat: Theory, process, and application* (pp. 71-88). New York, NY: Oxford University Press.

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review, 11*, 107-130. doi:10.1177/1088868306294790

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83. doi:10.1111/1467-9280.00111

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*, 1040-1057. doi:10.1177/00131640121971635

Snow, R. E., Kyllonen, C. P., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47 – 103). Hillsdale, NJ: Erlbaum.

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. doi:10.1006/jesp.1998.1373

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

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. doi:10.1037/0022-3514.69.5.797

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 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. doi:10.1037/0022-3514.77.6.1213

Strenta, A. C., Elliott, R., Adair, R., Scott, J., & Mailer, M. (1993). *Choosing and leaving science in highly selective institutions.* Unpublished report to the Alfred P. Sloan Foundation.

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology, 44*, 792-799. doi:10.1016/j.jesp.2007.07.005