2011

Confirming the Stereotype: How Stereotype Threat, Performance Feedback, and Academic Identification affect Identity and Future Performance

Tessa L. Dover
Claremont McKenna College

CLAREMONT McKENNA COLLEGE

CONFIRMING THE STEREOTYPE: HOW STEREOTYPE THREAT, PERFORMANCE FEEDBACK, AND ACADEMIC IDENTIFICATION AFFECT IDENTITY AND FUTURE PERFORMANCE

SUBMITTED TO

PROFESSOR SHANA LEVIN

AND

DEAN GREGORY HESS

BY

TESSA DOVER

FOR

SENIOR THESIS

FALL 2010/SPRING 2011

APRIL 25, 2011

Table of Contents

Confirming the Stereotype: How Stereotype Threat, Performance Feedback, and Academic Identification affect Identity and Future Performance

Tessa Dover, Claremont McKenna College

Author Note

The author would like to thank her reader, Dr. Shana Levin, for her insightful and conscientious input and comments throughout the planning, conducting, analyzing, and reporting phases of the project. She would also like to thank the CMC Psychology Department for a well-rounded education in psychological theory and methods. Finally, she would like to thank her friends and family for their support throughout the project and her four years at CMC.

Correspondence concerning the project can be addressed to Tessa Dover, 1136 Rosemary Ct., San Luis Obispo, CA, 93401. (805)748-3493. Email: TDover11@gmail.com

Abstract

This study investigates the post-performance effects of stereotype threat. Undergraduate students (N = 130) classified as either strongly- or weakly- identified with academics were told a diagnostic anagram task either typically shows poorer performance for their gender (stereotype threat) or no gender differences (no stereotype threat), and received arbitrary positive or negative feedback on an initial task. They later performed a second anagram task. Results indicate a 2-way interaction between stereotype threat and academic identification among those who received negative feedback. Negative feedback under stereotype threat did not harm performance for participants strongly-identified with academics, but did harm performance for participants weakly-identified with academics. This same 2-way interaction within the negative feedback condition also predicted post-feedback levels of identification as a college student, though it did not seem to affect post-feedback levels of academic identification. Strongly-identified participants receiving negative feedback identified less as a college student if they were under stereotype threat while weakly-academically identified participants identified more. Levels of post-feedback identification as a college student negatively predicted performance.

Confirming the Stereotype: How Stereotype Threat, Performance Feedback, and Academic Identification affect Identity and Future Performance

Threats to social identity remain one of the most powerful but publically underestimated social processes in contemporary society. The worry that we will be evaluated based on our group memberships rather than out own personal characteristics can feel just like threats to our own selves—both psychologically and physiologically. The implications of social identity threat have been researched both in the long term and the short term, and it has been found to affect many domains including attributions to discrimination (Eccleston & Major, 2006; Kaiser, Dyrenforth, & Hagiwara, 2006; Verkuyten, 1998), academic performance (Osborne & Walker, 2006; Steele & Aronson, 1995) interpersonal interactions (Castelli, Pavan, Ferrari, & Kashima, 2009; Leary & Schreindorfer, 1998; Romero-Canyas, Anderson, Reddy, & Downey, 2009), and even sports performance (Beilock, & McConnell, 2004; Stone, Lynch, Sjomeling, & Darley, 1999; Stone & McWhinnie, 2008). While long-term exposure to social identity threat (e.g., having a stigmatized identity) has been hypothesized to cause several detrimental outcomes (including poor health, reduced involvement in academics, etc.; Allison, 1998; van Laar, 2000), short-term-effects of social identity threat have more immediate and measurable deleterious effects.

One of the most powerful short-term effects of social identity threat involves the extent to which negative stereotypes affect performance on important academic tasks and tests. Known as stereotype threat (Steele, 1997; Steele & Aronson, 1995), this form of social identity threat specifies that cuing negative stereotypes about one's group in an evaluative setting will cause depressed performance because negatively-stereotyped group members are

threatened by their devalued social identity. While the phenomenon of stereotype threat has been replicated among many different social identities and skills, little research has attempted to link the short-term effects of experiencing stereotype threat (i.e., depressed performance) to the more long-term effects of social identity threat (i.e., less academic striving and lower engagement with academics). This study will investigate what happens when we either confirm or disconfirm the negative stereotypes that exist about our groups in academic settings. By looking at how receiving positive or negative performance feedback on diagnostic tasks presented as either gender-fair or gender-biased affects subsequent performance on similar tasks, we can better understand how short-term experience with social identity threat may affect whether we are able to strive on similar tasks in the future, or whether we disengage.

Stereotype Threat

Both heartbreaking and hopeful, stereotype threat theory (Steele, 1997; Aronson, Quinn, & Spencer, 1998) presents a purely situational account for the gaps in academic achievement between majority and minority groups. Stereotype threat is heartbreaking because it can affect seemingly anyone, its effects are still found when students do not endorse the stereotype (Aronson et al., 1999), and its effects are especially strong when the individual values and identifies strongly with the relevant domain (Keller, 2007a). However, the situational nature of ST also provides hope because it accounts for the gap in achievement without implying that minority groups are naturally inferior. In fact, most identities have been found to be susceptible to stereotype threat: African Americans (Steele & Aronson, 1995), Latinos (Gonzales, Blanton, & Williams, 2002), women (Spencer, Steele, & Quinn, 1999), people with low socio-economic status (Croizet & Claire, 1998; Harrison,

Stevens, Monty, & Coakley, 2006), homosexual men (Bosson, Haymovitz, & Pinel, 2004), overweight women (Seacat & Mickelson, 2009), White men (Aronson et al., 1999; Frantz et al., 2004; Stone et al., 1999), and even psychology majors (Croizet et al., 2003). This implies that perhaps *all* groups are susceptible to stereotype threat effects in important evaluative settings so long as a negative stereotype exists and the situational cues prime that stereotype (for an overview of the necessary factors hypothesized to produce stereotype threat, see Aronson et al., 1999). Thus, minority groups and women in certain domains are more susceptible to performance decrements not because they think differently, but rather because they contend with more stereotypes in academic settings.

Many mechanisms explaining the effects of stereotype threat have been postulated and experimentally tested to explain exactly why these deficits in performance occur. Recently, Schmader, Johns, and Forbes (2008) presented a process model of the various mechanisms affecting performance, taking into account both how stereotype threat is cued and the cognitive processes that are affected by ST and subsequently impair performance. According to their model, for tasks depending heavily on working memory (such as difficult math exams, Stroop-like inhibition tasks, anagram tasks, or challenging problem-solving tasks), a combination of physiological stress, attempts to monitor performance, negative affect, and attempts to suppress that negative affect all co-conspire to deflate working memory and subsequent performance—much like cognitive load. For tasks that do not depend on working memory but rather more automatic processing (e.g., golf putting; Beilock, Jellison, Rydel, McConnell, & Carr, 2006), stereotype threat affects performance through a different process (hyper-vigilance and explicit monitoring, essentially). However, for the academic domains of interest in the current study, only the processes behind

performance decrements found on academic tasks that depend heavily on working memory need be considered.

Interventions for stereotype threat have been developed and met with some success, at least in laboratory settings. For example, Marx and Roman (2002) found that providing women examples of women succeeding in math reduced the effect of stereotype threat for women taking a math test, and Johns, Schmader, and Martens (2005) found that simply educating people about stereotype threat may decrease its effect. Even more subtle, Steele and Aronson (1995) found that indicating race following a test (as opposed to before the test) limited the performance decrements usually seen with stereotype threat. Making self-affirmations about more global abilities and qualities has also been found to decrease stereotype threat's effects (Martens, Johns, Greenberg, & Schimel, 2006; Miyake et al., 2010). These interventions are likely successful because of the situational nature of stereotype threat—slight differences in testing protocol or information given prior to tests can reduce the anxiety of confirming stereotypes. As such, the creators of major tests such as the SAT and AP have responded to stereotype threat research with reforms in their testing protocol (usually recording ethnicity and gender after the test). Doubts have been raised as to whether these changes have successfully reduced stereotype threat: Stricker and Ward (2004) examined the effects of indicating gender and ethnicity before or after taking the Advanced Placement Calculus AB exam and a computerized placement test and found no evidence that the slight manipulation affected performance. However, two other researchers re-analyzed the data and found that the very simple change could potentially allow almost 5,000 women to receive calculus credit each year (Danaher & Crandall, 2008). Others have pointed out that the Stricker and Ward (2002) study did not find ethnicity effects because the negative

stereotypes are already “in the air” and triggered automatically in high-stakes testing situations. Nevertheless, attempts to influence testing policy and raise awareness about the short-term effects of stereotype threat on the academic performance of those plagued with academic stereotypes are ongoing.

Beyond providing hope that stereotype threat can be “cured” or reduced with rather simple and quick interventions or changes to test protocol, the perspective that situational cues explain a large percent of performance differences between groups also de-emphasizes an ingrained cultural explanation of achievement disparities. However, while it may be primarily the situation that accounts for stereotype threat’s effect on performance on an important test, the way in which people perceive feedback on those tests is likely to be affected past experiences. And a person’s reaction to that feedback is likely to be projected into how they treat similar tasks in the future. The downstream effects of stereotype threat—or more specifically, what happens when we receive feedback on stereotypic tasks and then face the same situation again—have been studied less, and may be affected more by individual differences in past academic experiences. Some early research and reviews on stereotype threat suggest that over time, stereotype threat may cause disengagement with academics and lower academic striving as those facing stereotypes find it difficult to maintain motivation on academic tasks (the “disidentification hypothesis;” Steele, 1997). However, surprisingly little direct evidence of this phenomenon has been conducted within stereotype threat research. Stereotype threat theory may provide a good explanation of differences seen in a single testing situation, but it does not present a full theory of why some people persist despite the stereotypes facing them and why some disengage with academic

domains. Luckily, other areas of social identity threat research can help inform on how repeated exposure to stereotype threat may affect performance on the long term.

Stigma and Social Identity Threat in the Long Term

The primary tenet of social identity theory asserts that we associate some our own self-worth with the evaluations of our groups, and that we want our groups to be evaluated positively (Tajfel & Turner, 1986). When we are randomly assigned to a group in a laboratory, we tend to reward our in-group more and the out-group less, even when it is not benefit us personally (e.g., Tajfel, Billig, Bundy, & Flament, 1971). This suggests that our social identities—even those that are determined randomly for a short amount of time—hold incredible importance to how we perceive ourselves, other people, and the feedback we receive. When our ad-hoc group is negatively evaluated, it can also produce threat—termed social identity threat—with the same physiological profile of other types of threat (Scheepers, 2009). And when our social identities are well-ingrained and chronically devalued in society, this chronic exposure to social identity threat has the potential to alter both our physiological and psychological profiles. For example, African Americans in the United States have been found to have more physical and mental health problems than European Americans (Allison, 1998; Clark, Anderson, Clark, & Williams, 1999), with evidence that being under social identity threat affects cardiovascular functioning (Blascovich, Spencer, Quinn, & Steele, 2001) and motivates us to perceive many interpersonal and academic experiences in terms of our group membership (see Major & O'Brien, 2005). Research investigating the effects of long-term social identity threat has been conducted primarily by comparing reactions to negative events between de-valued minority group members (i.e., African-Americans or, in some cases, Latinos) and non-stigmatized

majority group members (i.e., European-Americans), with the ethnicity of the participant serving as a proxy for experiences with chronic identity threat. However, women facing stereotypes, stigma, and discrimination in workplace and academic settings also have been studied to help inform reactions to long-term exposure to social identity threat. Under these research paradigms, several interesting—and oftentimes counterintuitive—methods of coping with the experience of stigmatization and chronic social identity threat have been observed.

Attributions to group membership. A non-academic domain that may prove useful for understanding how long-term social identity threat affects reactions to academic feedback and evaluation involves attributions to discrimination. When minority group members receive negative treatment in a blatantly prejudicial environment, the ability to attribute the unfair treatment to group membership (as opposed to personal characteristics) allows them to spare their self-esteem. For example, when minority group members read an article describing discrimination as prevalent on a college campus, their self-esteem was harmed less following rejection compared to when discrimination was not described as prevalent (Major, Quinton, & Schmader, 2003).[1] While this may not make intuitive sense—we expect that those who face discrimination to feel bad about themselves—the ability for those who chronically suffer social identity threat to attribute their feedback to their group memberships is likely adaptive and may have important implications for how stigmatized individuals respond to negative feedback in academic settings.

[1] While stigmatized minorities are not *likely* to report that they think their negative treatment is due to discrimination (likely because claiming discrimination has severe negative social consequences in work or other settings; Kaiser & Miller, 2001), differential physiological responses have been recorded when unfair treatment comes from out-group versus in-group members (Mendes, Major, McCoy, & Blascovich, 2008) and they *do* report that they perceived feedback as discrimination in comfortable, in-group settings (Stangor, Swim, Van Allen, & Sechrist, 2003).

For the most part, the evidence that attributing negative events to discrimination suggests a self-protective quality of having a negatively-evaluated social identity. Indeed, those unable to blame their treatment on their group membership suffer from lowered self-esteem when faced with negative information about their group (Major & Sawyer, 2009). This attributional account has been applied to academic domains with both cross-sectional and longitudinal research. For example, van Laar (2000) observed that compared to White students, African-American students experienced more decreases in expectations for the future as they moved from high-school through their first year in college; yet their self-esteem remained at similar levels. She also found that Black students who naturally attribute failure to external sources could remain motivated if they were able to attribute *internally* for successes. Together, these studies suggest that despite having academic experiences that lower expectations, Black students were able to maintain their self-esteem and spare their academic motivation if they could attribute externally for their failures. These studies, however, do not directly test whether negative academic experiences were the cause of the lowered expectations of Black students—the author only claims that Black college students face disappointment during their first year of college because their post-high school optimism was undercut by failure experiences. Further, the study does not directly assess whether the Black students were attributing their failures to discrimination or their group membership, but rather only that they attributed it to some external source. Nevertheless, the findings provide support for the possibility that the ability to attribute negative academic experiences to group membership may preserve self-esteem and motivation in academics.

Disengagement from the domain. The ability to attribute negative group-relevant feedback to group membership is one of a few "tools" that those exposed to chronic social

identity threat have adapted in order to preserve a positive personal identity. However, while the ability to externally attribute may prevent negative self-evaluations and maintain motivation, continued exposure to social identity threat in academic settings has been hypothesized to promote a second adaptive function: disengagement. This has been demonstrated among women exposed to negative gender stereotypes by Davies, Spencer, Quinn, and Gerhardstein (2002): female students chose to answer fewer math problems and more verbal problems on a difficult test. They also reported less interest in working in professions depending on quantitative skills compared to their female counterparts not exposed to negative stereotypes. Somewhat similarly, correlational studies indicate that Black students value academics outcomes less if they believe that the status differences in society were unjust, and that Latino students discount the validity of academic outcomes more when they perceive ethnic injustice in society (Schmader, Major, & Gramzow, 2001). This has a logical adaptive function: if we have negative experiences with certain tasks, and we do not perceive that there is a likelihood of future success (because other members of our group are negatively evaluated in that specific domain), we pull our efforts from those areas so we no longer feel bad for our failures.

From the perspective of society at large, however, this disengagement has very problematic consequences: it causes those with stigmatized identities to withdraw from important academic pursuits and achieve less, thus perpetuating stereotypes and enforcing the cultural perception that minority groups are unmotivated and less skilled. So while it may be adaptive to disengage from the domains that carry the likelihood of negative evaluation, investigating how to reduce such disengagement and maintain motivation remains an important goal for social psychological research.

Long-Term Social Identity Threat and Stigma Applied to Stereotype Threat

The two major strategies adapted to cope with long term exposure to social identity threat—attributions to group membership and disengagement from the domain—can help inform us about how people who confirm negative stereotypes about their group's academic performance will respond to future tasks. Regarding attributions, research has shown that encouraging individuals under stereotype threat to attribute their anxiety/arousal to an external source makes them less susceptible to performance decrements (Ben-Zeev, Fein, & Inzlicht, 2005). This suggests that attributions may play a part in dealing with threatening academic environments. However, the stereotype threat studies investigating attributions look primarily at how external attributions spare the working memory decrements found to affect performance, *not* how attributions may be used to interpret the feedback following that performance. Another study also found that when facing stereotype-confirming feedback on a computer task, women were not as able to attribute their failures externally (compared to men; Koch, Müller, & Sieverding, 2008). This suggests that the ability to attribute failures externally when facing stereotypes may be a tool some people are able to employ, while others who believe that their natural capabilities caused them to fail continue to fall pretty to stereotype threat. However, with both of these studies, the attributions were toward external events, *not* group membership. To the author's knowledge, no studies exist that investigate whether people who can attribute academic successes or failures to their group membership are differentially affected by stereotype threat.

Regarding the strategy of disengagement, stereotype threat theory hypothesizes that over time, disidentification with academics (i.e., disengagement from the academic domain) will result from repeated exposure to stereotype threat (see Steele, 1997). However, very

little empirical work has explored this link directly. Indeed, the evidence demonstrating minority students' reduced academic identification throughout college years (e.g., van Laar, 2000) and ethnic differences in the link between self-concept and academic performance (e.g., Crocker & Major, 1989) still provide the main support for the theory. Within studies of stereotype threat, it has been found that being less engaged with academics and less identified with academic domains seems to make people less susceptible to stereotype threat. For example, women who were highly-identified with math faced the largest performance decrements on a difficult math test while under stereotype threat compared to their non-math identified counterparts (Keller, 2007a). Steele's theory (1997) explains this effect well: those who do not care about academic outcomes are not trying to perform well and thus do not worry about confirming negative stereotypes. This suggests that learning to disengage from academic domains after chronic experience with stereotype threat serves a protective function. However, because identification was measured as an individual difference variable and not as an outcome variable, it cannot be concluded that the experiences of stereotype threat actually cause such disidentification. Thus, we know that reduced engagement with academics spares people from the negative effects of stereotype threat, but we do not know if academic engagement is just a naturally varying individual difference or whether it can be linked to previous stereotype threat experiences. In fact, in the original stereotype studies (Steele & Aronson, 1995), levels of academic identification reported by African Americans after taking a test did not change depending on stereotype threat condition. While this does not necessarily mean that disengagement does not result from stereotype threat (because actual academic identification might differ from self-reported academic identification when participants are worried about their self-presentation), it does suggest that more research

needs to address whether stereotype threat directly reduces academic identification. Moreover, the effect of feedback that either confirms or disconfirms those stereotypes (i.e., negative or positive feedback) on academic identification needs to be investigated in order to establish the link between stereotype threat and engagement or disengagement on future academic tasks.

Based on a review of the available research, it is likely that individuals facing stereotype threat may be responding to negative performance feedback by employing coping strategies similar to the strategies used with other forms of chronic social identity threat. However, it is not clear exactly when these strategies might be utilized, or whether specific experiences of stereotype threat can trigger these responses. Further, it is possible that individual differences affect how people respond to feedback that confirms (or disconfirms) the negative stereotypes that exist about their groups. In fact, while research on stigma certainly indicates that it is difficult to remain engaged and motivated in academic domains when facing stereotypes and negative evaluations chronically, some students continue to strive and succeed. It is possible that it is the high level of natural skill that buffers these individuals from receiving negative evaluations that they believe might be due to their group membership. Or, it is possible that some people are able to successfully attribute their failures to other sources and remain motivated. Another alternative to disengaging from academic domains when facing stereotype-confirming feedback that has been suggested involves a different, more temporary form of disengagement, and may help explain how some individuals continue striving in the face of negative stereotypes and negative feedback.

Situational Disengagement versus Chronic Disengagement

As opposed to chronic disengagement that causes members of stigmatized groups to

become less identified with academics and care less about their performance on academic tasks (described above), situational disengagement is described by Nussbaum and Steele (2007) as a coping strategy specific to those facing stereotype threat that also have high engagement and identification with academics. The researchers found that among high-level minority students, negative feedback given in a diagnostic (versus non-diagnostic) task led to more perseverance in a subsequent task. Importantly, they found that this relationship was mediated by their level of disengagement with the feedback (i.e., saying that the test did not say much about ability). The more these highly-identified students reported that the task was not a meaningful indication of ability, the more they chose to persist when given the opportunity to complete a similar task. In this way, highly-identified students were able to "situationally disengage" with the feedback and continue to strive. This account contrasts with the findings that stigmatized individuals disengage from the task or the domain and choose to withdraw their efforts in order to save self-esteem (Crocker & Major, 1989). They key, according to Nussbaum and Steele, is the high-level of academic identification.

The Nussbaum and Steele study (2007) provides a first look at how stereotype-confirming feedback affects subsequent approaches to similar tasks, and it suggests that situational disengagement may be another "tool" that students who are highly-identified with academics can use to spare future academic motivation. However, the study did not measure baseline levels of academic identification and simply assumed that all participants were highly engaged in academics because they were students at a top university. Thus, it is unclear whether a weakly-identified participant would react similarly. If it is true that their sample consisted only of strongly-identified participants, there is reason to believe that those who are weakly-identified with academics would be less able to remain persistent in the face

of feedback that confirmed negative academic stereotypes. For example, Forbes, Schmader, and Allen (2008) found that for minority students who placed little value on academics, their brains reacted less to errors, and they attended to stimuli at a slower rate (compared to students with strong academic identification). This suggests that at an implicit level, those who are weakly-identified with academics discount their errors (or are less attentive to them), but still suffer from performance decrements. This contrasts from the findings by Nussbaum and Steele among strongly-identified students who were able to *spare* their performance by discounting feedback.

The Nussbaum and Steele (2007) study, as well as the Forbes et al. (2008) study, suffers from a further limitation that involves their use of the common stereotype threat manipulation of informing participants that the test is diagnostic of ability. While this manipulation is shown to elicit negative stereotypes automatically in a testing situation and may be more ecologically valid (Steele & Aronson, 1995), the manipulation has not been established to continue eliciting stereotypes as feedback on performance is given, and it may have affected persistence by *more* than just eliciting stereotypes. For the minority participants in the Nussbaum and Steele study, perhaps they were more eager to persist in the diagnostic condition because their feedback was more meaningful and reflecting of their own personal ability. The researchers did not find a difference in persistence between diagnostic condition and control condition for White participants, indicating that minority status (and the accompanying stereotypes) did make a difference in how participants reacted to the feedback. However, we still cannot be sure that stereotypes were activated, so we cannot conclude that the difference in persistence between minorities in the diagnostic and non-diagnostic condition was due to the group-relevance of the feedback. For example, it is

possible that Black students persisted longer in the diagnostic condition *not* because they considered the feedback group-relevant, but because their past academic experiences encourage them to persist longer when their feedback is diagnostic of ability.

Finally, the Nussbaum and Steele study does not provide a complete picture of when students strive despite negative stereotypes because they did not address whether those who persisted would actually perform better on a second trial. The dependent variable of persistence is important for understanding how confirming stereotypes affects how we approach future tasks. However, especially in light of the research demonstrating that stereotype threat can produce inflexible perseverance (persevering on tasks with less efficient and stagnant problem solving strategies as opposed to adaptive ones; Carr & Steele, 2009), the motivation to complete more of the similar problems does not necessarily mean that such persistence would be productive. The account of situational disengagement is an important one, though further research needs to better specify the outcomes of such disengagement and whether weakly-identified students would be able to employ the same protective strategy.

The Present Study

Much of the research on stereotypes, stigmatization, and long-term exposure to social identity threat attempts to understand and explain the academic differences we see between members of minority and majority ethnic groups. These differences are clearly important as gaps between ethnic groups remain some of the most pervasive and stubborn in the United States today. However, in this study we observe how gender stereotypes may affect men and women in stereotypic domains. It has been clearly demonstrated that stereotype threat can affect both men and women based on gender stereotypes (e.g., Keller, 2007b; Spencer et al., 1999), and gender gaps in representation in gender-stereotypical academic domains (e.g.,

women in engineering or men in literature) may be partly due to how men and women perceive their feedback in these stereotypic subjects. Of course, debate may be raised as to whether the lack of females in science, technology, engineering, and mathematics (STEM) fields and the lack of men in humanities like literature are problematic enough to warrant intervention. Furthermore, innate skills and motivational differences *beyond* coping with stereotypes also help define who goes into what field of study or profession. However, it is important to explore how coping with social identity threat on tasks that depend on stereotypic traits influence a man's or woman's motivation to continue persisting on tasks and believing that they are capable of performing well in the future.

This study aims to provide a link between experiences of stereotype threat and subsequent disengagement or striving in academic domains that are stereotypically female-oriented or stereotypically male-oriented. By comparing how those facing gender stereotypes and those not facing gender stereotypes respond to positive and negative feedback when given a chance to perform a similar task, we can better understand the role of disengagement and attributions among men and women in academic settings. Unlike most studies of disengagement and attributions in academic domains (e.g., van Laar, 2000; Osborne & Walker, 2006), this study investigates the *direct* link between academic experiences and future academic performance. And unlike most stereotype threat studies that end after the participants take a single test, this study continues to investigate the effects of stereotype threat after students receive feedback on the stereotypic tasks. By adding the element of differential performance feedback and giving participants the chance to complete a second, similar task, we can observe whether strategies found to preserve self-concepts in other areas of social identity threat research can be utilized to preserve performance in stereotype-laden

academic settings. Probing post-feedback levels of academic identification and identification as a college student will also allow us to assess possible mechanisms responsible for spared or decreased performance. In addition, by comparing those with strong and weak baseline identification with academics, we can assess whether our engagement with academics might influence the strategies we are able to employ to spare performance. Based on a review of the literature, several hypotheses are offered regarding how students with strong and weak identification with academics will respond to stereotype-confirming and stereotype-disconfirming feedback:

Hypothesis 1. There will be a 3-way interaction between stereotype threat condition (stereotype threat vs. no stereotype threat), performance feedback (positive vs. negative feedback), and academic identification (strong vs. weak identification) for participants' performance on a subsequent task:

Hypothesis 1a. For those who receive *positive feedback*, there will be a 2-way interaction between stereotype threat condition and academic identification:

For *strongly-identified students* receiving positive feedback, subsequent performance will differ depending on whether they are under stereotype threat or not. Those under stereotype threat will remained worried about confirming group stereotypes even after receiving positive feedback, thereby depressing their performance. This is hypothesized because strong identification with stereotypical academic domains has been found to increase the effects stereotype threat (Keller, 2007a). Worry about confirming stereotypes does not seem to decrease as students continue to find success in an academic domain, so even though these strongly-identified students have disconfirmed a negative stereotype, they will remain worried about fulfilling it when they perform again. However, those not under stereotype

threat will not perform poorly on the second task because their positive feedback should increase their confidence and reduce any worry that they will not perform well.

For *weakly-identified students* receiving positive feedback, future performance will not change depending on whether they were under stereotype threat or not. In other words, weakly-identified participants should perform similarly after receiving positive feedback no matter if their feedback was group relevant (i.e., stereotype threat condition) or not group-relevant (i.e., no stereotype threat condition). This is hypothesized because weakly-identified students have been found to be less worried about confirming stereotypes about their groups (Keller, 2007a). If they are not that anxious about their performance to begin with, positive feedback should feel good and increase confidence going into the second task regardless of stereotype threat condition.

Hypothesis 1b. For those receiving *negative feedback*, there will be a 2-way interaction between stereotype threat condition and academic identification:

For *strongly-identified students* receiving negative feedback, making that feedback group-relevant (i.e., being in the stereotype threat condition) should spare performance on the future task. Strongly-identified students who are under stereotype threat should perform better compared to when not under stereotype threat. This is expected because strongly-identified students are expected to be better able to attribute their negative feedback to their group membership (see Major et al., 2003) and situationally disengage (see Nussbaum & Steele, 2007) in order to spare performance. Receiving negative feedback that is not tied to group membership (i.e., not under stereotype threat), however, should result in lower relative performance because strongly-identified students will have less confidence in their ability to perform well and feel threatened about their future performance.

For *weakly-identified students* receiving negative feedback, the opposite pattern is expected. When feedback is not group-relevant (i.e., not under stereotype threat), the negative feedback should not depress future performance because those who already devalue academics and tests will not be impacted by what the test says about their cognitive abilities. When the feedback confirms the negative stereotypes about their gender, however (i.e., under stereotype threat), they are likely to disengage from their task (much like those who suffer chronic social identity threat; Crocker & Major, 1989) and not continue to try hard on future tasks. In this way, weakly-identified students in the stereotype threat condition will see the negative feedback as "proof" that they will not be successful in the domain, thus legitimizing the feedback, and causing them to give up (disengage).

Hypothesis 2. Students will report higher levels of academic identification after receiving positive feedback regardless of baseline levels of academic identification or stereotype threat condition. However, this main effect will be qualified by a 3-way interaction between stereotype threat condition (stereotype threat vs. no stereotype threat), performance feedback condition (positive vs. negative feedback), and baseline academic identification (strong vs. weak identification):

Hypothesis 2a. Within the *positive feedback* condition, there will be a 2-way interaction between stereotype threat condition and baseline academic identification:

For *weakly-identified students* receiving positive feedback, post-feedback academic identification will be similar regardless of whether they were under stereotype threat or not. This is expected because weakly-identified students will not feel much more positivity about their performance after disconfirming their stereotypes compared to when just getting positive feedback because they were not worried about confirming negative stereotypes to

begin with. Thus, they should be more identified with academics following stereotype-disconfirming feedback (i.e., stereotype threat, positive feedback) compared to simple positive feedback (i.e., no stereotype threat, positive feedback)

For *strongly-identified students* receiving positive feedback, post-feedback academic identification will be higher if the feedback disconfirms negative stereotypes about their gender (i.e., under stereotype threat) compared to when it is not group-relevant (i.e., not under stereotype threat). This higher post-feedback academic identification is expected for strongly-identified students because performing well despite being stereotyped on the task will convince them that academics are an important part of their identity. However, this increased identification with academics will contribute to their future depressed performance because it will increase their susceptibility to stereotype threat (see Hypothesis 1a).

Hypothesis2b. Within the *negative feedback* condition, there will be a main effect of stereotype threat condition on post-feedback college student identification.

For *weakly-identified students* receiving negative feedback, post-feedback academic identification will be lower if they are under stereotype threat compared to when they are not under stereotype threat. This is hypothesized because when they confirm the negative stereotypes about their group (i.e., negative feedback in stereotype threat condition), weakly-identified students will be convinced that they are not good at the task and disengage more from academics. This lowered identification with academics is expected to contribute to their depressed performance on the subsequent anagram task (see Hypothesis 1b) because their disengagement leads them to stop trying.

For *strongly-identified students* receiving negative feedback, post-feedback academic identification is also expected to decrease when under stereotype threat compared to when

not under stereotype threat. This is hypothesized because strongly-identified students are able to situationally disengage from their feedback, which is manifested through lower scores on post-feedback academic identification. However, unlike for weakly-identified students, this disengagement will contribute to strongly-identified students' spared performance on the subsequent task (see hypothesis 1b).

Hypothesis 3. Post-feedback identification as a college student will differ depending on stereotype threat condition (stereotype threat vs. no stereotype threat), the valence of the feedback (positive vs. negative), and baseline academic identification (strong- vs. weak-identification). There is not enough literature to make clear predictions for how participants receiving positive feedback will respond, or how weakly-identified students receiving negative feedback will respond. Thus, the effect of the independent variables on post-feedback college student identification will be largely exploratory. For strongly-identified students in the negative feedback condition, however, there is hypothesized to be a difference between those under stereotype threat and those not under stereotype threat:

For *strongly-identified students* receiving negative feedback, being under stereotype threat will result in lower identification with college students compared to when not under stereotype threat. This is hypothesized because, much like how stereotype-confirming feedback (i.e., negative feedback under stereotype threat) is hypothesized to decrease academic identification, strongly-identified students may be able to buffer themselves from the feedback by de-emphasizing their identity as a college student. This disidentification may be more likely when under stereotype threat because the feedback is less "personal," and strongly-identified students may be able to convince themselves that they are not like other

colleges students. When the feedback is not tied to their group-membership, however (i.e., no stereotype threat), discouragement may lead them to identify more with a positive identity.

Method

Participants and Design

Undergraduate women (*n* = 92, 63.4%) and men (*n* = 53, 36.6%) at a Southern California college participated partial credit toward a lower-division psychology course. The mean age was 19.26 (SD = 1.17), and ranged from 18 to 22. Ninety-two participants identified as White (63.4%), 23 as Asian (15.9%), 6 as African American (4.1%), 6 as Latino (4.1%), 9 as Multiracial (6.2%), and 9 as other/missing (6.2%). Participants were randomly assigned to one of four experimental conditions in a 2 (stereotype threat v. no stereotype threat) x 2 (positive feedback v. negative feedback) between-subjects factorial design.

Procedure

Prior to coming into the lab, participants completed an online pre-test survey to assess their level of identification with academics. Later, they came into the lab in small groups (up to 5) of mixed gender. There, they were told that they would be completing two anagram tasks and answering questions in order to help researchers understand factors that influence how people learn from past performance. They were also told either that their gender generally performs worse on anagram tasks (stereotype threat), or that the task was gender fair (no stereotype threat). They then completed an initial anagram task that all participants were told measured cognitive abilities. While the experimenter ostensibly graded the test, participants completed a filler task that consisted of writing a paragraph about skills and qualities important for success in anagram tasks. Participants then received arbitrarily positive or negative feedback about their performance (performance feedback manipulation)

and completed a post-feedback survey. They were then introduced to a second, slightly different anagram task, which they completed after recording their expectations for their performance. The participants were timed on how long it took them to solve each anagram, and the average time was used as a measure of performance. They were then fully debriefed and probed for suspicion.

Stereotype threat manipulation. The stereotype threat manipulation appeared on a task information sheet explaining first anagram task. Participants were randomly assigned to learn that their gender typically performs worse on anagram tasks, or that anagram tasks were considered fair regarding gender and other group differences. In order to make the stereotype threat manipulation believable, men under stereotype threat were told that anagram tasks measured verbal abilities (a domain that women are typically considered superior in), and women under stereotype threat were told it measured spatial abilities (a domain that men are typically considered superior in). Men under stereotype threat read:

> This task requires that you have a strong ability to understand words and their properties. For this reason, researchers commonly use this task to determine and compare individuals' verbal abilities. Also for this reason, women tend to do slightly better in this task because they are generally more verbally-oriented than men.

Women under stereotype threat read:

> This task requires that you have a strong ability to rearrange objects in your head. For this reason, researchers commonly use this task to determine and compare individuals' spatial abilities. Also for this reason, men tend to do slightly better in this task because they are generally more spatially-oriented than women.

Women and men in the no-threat condition read:

> This task requires that you engage in multiple cognitive strategies as you unscramble and form the words. For this reason, researchers commonly use this task to determine and compare individuals' general cognitive ability. This task is also commonly used because it has been shown to be impervious to gender and other group differences.

Participants were also given ostensible national averages of performance that reflected the stereotype threat manipulation: women in the stereotype threat condition were told that men average 88 on this test while women average 82; men in the stereotype threat condition learned that women average 88 and men 82; those in the no-threat condition were told just that the national average is 85. In order to ensure that gender was salient for those in the stereotype threat condition, participants in the stereotype threat condition were asked to indicate their sex before completing each anagram task while participants in the no-threat condition wrote down their participant number instead. The national averages also appeared on the answer sheets so participants could reference them throughout the task and after receiving feedback.

Initial anagram task. Eight 8-letter anagrams were presented on a large screen for 40 seconds each. Participants were instructed to write down as many English words possible that could be made from combinations of the presented letters (e.g., for the anagram AEIBORVH, participants could write down ARE, EAR, BORE, HIVE, BEHAVIOR, etc.). They were instructed that more points would be awarded for longer words and for words that were more difficult to find. This would allow for enough ambiguity in perceived performance so that they would not suspect their feedback was arbitrary. The full list of anagrams presented to participants in the initial anagram task is presented in Appendix A.

Feedback manipulation. Those in the positive feedback condition were told they scored a 94.1 on the first trial of the anagram task, about 9 points higher than the national

average and 12 points higher than their gender's average (if they were under stereotype threat). Those in the negative feedback condition were told they scored a 78.6 on the first trial of the anagram task, about 6.5 points lower than the national average and 3.5 points lower than their gender's average (if they were under stereotype threat).

Second anagram task. Eight anagrams of 5 or 6 letters were presented on a large screen from between 30 and 60 seconds, depending on the difficulty of the anagram. Each anagram had between one and three solutions, though participants needed only to come up with one single solution that used all the letters in the anagram (e.g., for the anagram KAREB, solutions include BAKER, BRAKE, and BREAK. For the anagram TERNE, the word ENTER would be the only solution). Participants were timed for how long it took them to solve each anagram, and if they were unable to find a solution in the time given, their time was recorded as the maximum time allowed. The full list of anagrams presented to participants for the second anagram trial is presented in Appendix B.

Measures

Online pre-study survey. Prior to coming into the lab, participants completed a short online survey. Among other scales, they completed the Academic Identification Scale (Osborne, 1997) that measures how engaged students are with academics and their performance on academic tasks. The 13 items included:

1. Being a good student is an important part of who I am.
2. I feel that the grades I get are an accurate reflection of my abilities.
3. My grades do not tell me anything about my academic potential. [*reverse coded*]
4. I don't really care about what tests say about my intelligence. [*reverse coded*]

5. School is satisfying for me because it gives me a sense of accomplishment.

6. If the tests we take were fair, I would be doing much better in school. [*reverse coded*]

7. I am often relieved if I just pass a course. [*reverse coded*]

8. I often do my best work in school.

9. School is very boring for me, and I'm not learning what I feel is important. [*reverse coded*]

10. I put a great deal of myself into some things at school because they have special meaning or interest for me.

11. I enjoy school because it gives me a chance to learn many interesting things.

12. I feel like the things I do at school waste my time more than the things I do outside of school. [*reverse coded*]

13. No test will ever change my opinion of how smart I am. [*reverse coded*]

Responses ranged from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). After completing the pre-study questionnaire, participants generated a unique code used to connect online responses to the in-lab portion of the study. The full pre-study questionnaire is reprinted in Appendix C.

Post-feedback survey. Immediately after receiving performance feedback, participants completed a post-feedback survey. Among other items, participants responded to 5 items measuring their perceptions of their performance and the test:

1. I did **better** / **much better** / **about the same** / **worse** / **much worse** as I expected on the test. (Circle One)

2. I performed well on this test.

3. I am happy with my score on the test.

4. My score is an accurate representation of how I felt I performed on the test.

5. I enjoyed the test.

Responses on items 2-5 ranged from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). The participants then completed the same Academic Identification scale they had previously completed on the online pre-study questionnaire, as well as a measure of their identification as a college student (adapted from Schmader, 2002, from Luhtanen & Crocker, 1992). The four items included:

1. Being a college student is an important part of my self-image.
2. Being a college student is unimportant to my sense of what kind of person I am. [*reverse coded*]
3. Being a college student is an important reflection of who I am.
4. Being a college student has very little to do with how I feel about myself. [*reverse coded*]

Responses ranged from 1 (*strongly disagree*) to 7 (*strongly agree*). The full post-feedback questionnaire can be found in Appendix D.

Results

Data Preparation

Eleven participants did not complete all the necessary dependent measures or did not complete the online pre-study survey and were excluded from the analyses. Additionally, four participants expressed suspicion during debriefing that their performance feedback was manipulated or that the stereotype threat manipulation (i.e., that men or women tended to do better on the task) was untrue. These participants were also excluded from the analyses, leaving a final sample of 130 participants. The age range remained the same (18-22), M = 19.33, SD = 1.16. Demographic variables and experimental condition for the final sample are presented in Table 1.

The academic identification scale (pre-study and in-lab), college student identification scale, and post-feedback performance perception scale were created after reverse-coding the appropriate items. Because items on the post-feedback performance perception scale were on different scales, standardized scores were created. Overall, the scales had adequate reliability: online academic identification, $\alpha = .78$; in-lab academic identification, $\alpha = .82$; college student identification, $\alpha = .87$; gender identification, $\alpha = .87$; post-feedback performance perceptions, $\alpha = .81$. A median split was created for the pre-study academic identification scale that assessed baseline levels of academic identification (median = 5.38 on 7-point scale), and participants were classified as either strongly-identified or weakly-identified with academics. Academic identification classification by condition is also presented in Table 1.

Table 1

Number of Participants by Gender and Ethnicity for Each Condition in Final Sample

	Total	Stereotype Threat		Performance Feedback		Academic ID	
		ST	No ST	Positive	Negative	Strong-ID	Weak-ID
Gender (*n*)							
Female	**84**	42	42	43	41	49	35
Male	**46**	22	24	23	23	20	26
Ethnicity (*n*)							
White/Caucasian	**84**	45	39	40	44	48	36
Asian	**24**	10	14	13	11	11	13
Black/African Am.	**4**	1	3	2	2	2	2
Latino/a	**4**	2	2	3	1	2	2
Multiracial	**10**	5	5	6	4	6	4
Other/Missing	**4**	2	2	2	2	1	3

Manipulation Checks

In order to assess whether the performance feedback manipulation was successful, an independent samples t-test was performed comparing those in the positive feedback condition to those in the negative feedback condition on their perceptions of their performance on the first anagram task (post-feedback survey). As expected, those in the negative feedback condition (M = -.61, SD = .53) perceived that their performance was worse than those who received positive feedback (M = .57, SD =.53), $t(127) = 12.71$, $p < .01$, d = 2.23). Importantly, however, it appears that most participants, regardless of feedback condition, believed that their scores accurately represented their performance ($M_{positive} = 4.68$, SD = 1.48; $M_{negative} = 4.43$, SD = 1.44 on 7-point scale; $t(127) = .99$, $p = .33$, $d = .17$), meaning that participants believed that their feedback was similarly genuine regardless of performance feedback condition.

In order to assess whether participants under stereotype threat believed that the anagram tasks depended on the skills described before the initial task (i.e., verbal abilities for men and spatial abilities for women), the filler task completed between the initial anagram task and the feedback manipulation was coded for mentioning verbal or spatial abilities, and a chi-square tests of independence were performed for men and women separately. We would expect that men under stereotype threat would more often mention verbal abilities (e.g., knowing qualities of words, or having an expansive vocabulary) in their paragraph explaining skills necessary for good performance on anagram tasks because they were told that the task measures verbal abilities. For the same reason, we would expect women under stereotype threat to mention spatial abilities (e.g., ability to move things around in your head) more often than those not under stereotype threat. Results indicated that men under

stereotype threat did not mention verbal abilities more than men in the non-stereotype threat condition, $\chi^2(1) = 1.44$, $p = .23$, or spatial abilities less often than men not under stereotype threat, $\chi^2(1) = 1.17$, $p = .56$. Similarly, women under stereotype threat did not mention spatial abilities more often than women not under stereotype threat, $\chi^2(1) = .81$, $p = .37$, or verbal abilities less often than when not under stereotype threat, $\chi^2(1) = .47$, $p = .49$. These results are contrary to expectations, and indicate that the stereotype threat manipulation may not have been successful in convincing participants that the anagram task depended on sex-stereotypical abilities. However, because all participants were more likely to mention both spatial and verbal abilities than to not mention them ($\chi^2_{verbal}(1) = 19.23$, $p < .01$; $\chi^2_{spatial}(1) = 6.92$, $p = .01$), it is likely that most participants thought that the task involved both of these skills regardless of whether they were told so or not.

Additionally, a 2-way interaction between stereotype threat and performance feedback on the post-feedback performance perceptions scale on the first anagram task also indicates that the stereotype threat manipulation was successful in influencing how participants viewed their feedback. The main effect of performance feedback (good v. bad) on post-feedback performance perceptions was moderated by the stereotype threat manipulation, $F(1, 125) = 1.41$, $p = .03$, $\eta^2 = .04$. This 2-way interaction is presented graphically in Figure 1. Simple effects testing revealed that such those who received positive feedback while under stereotype threat felt especially positive about their feedback ($M = .76$, $SE = .09$) compared to those who were not under stereotype threat ($M = .40$, $SE = .09$), $F(1, 125) = 7.91$, $p < .01$, $\eta^2 = .06$. The difference between those in the stereotype threat condition ($M = -.65$, $SE = .09$) and those not in the stereotype threat condition ($M = -.58$, $SE = .09$) was not significant when receiving negative feedback, $F(1, 125) = .28$, $p = .60$, $\eta^2 < .01$.

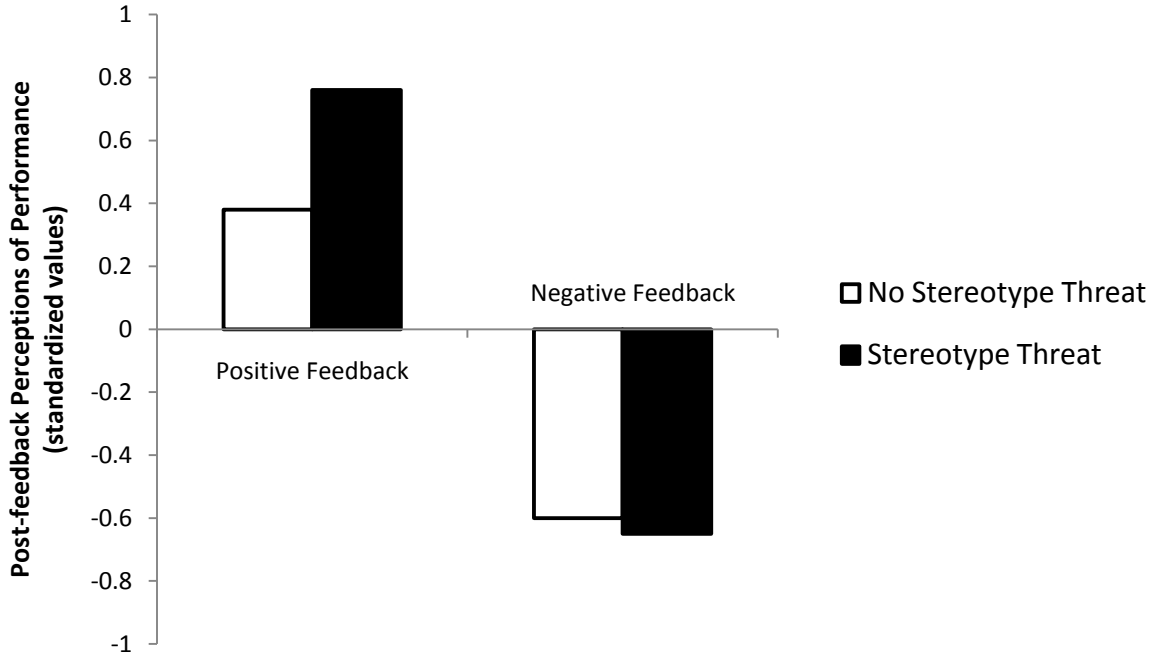

Figure 1. Effect of feedback condition and stereotype threat condition on post-feedback performance perceptions. Values are standardized scores.

Also to test whether those in the stereotype threat condition actually experienced stereotype threat, an independent samples t-test was performed assessing whether stereotype threat condition influenced performance on the initial anagram task. Contrary to the effects usually found in stereotype threat research, participants who were not under stereotype threat (M = 186.21, SD = 59.42) did not perform better on the initial task compared to those who were under stereotype threat (M = 187.40, SD = 66.29; $t(127) = -.11$, $p = .92$, $d = .02$). This is potentially problematic because it indicates that the stereotype threat manipulation was unsuccessful in causing the participants to worry about their performance enough to hinder it. However, this initial anagram task was designed to be ambiguous enough to provide believable manipulated feedback—not to measure cognitive abilities or tap working memory.

The second anagram task (i.e., the task used as the primary performance dependent variable) was designed to rely more on thinking quickly and adjusting problem-solving strategies was designed to measure the effects of stereotype threat and performance feedback; however it was unsuitable for use on the initial task because participants would have a better idea of how they performed and might be less likely to believe their manipulated performance feedback. Thus, while the stereotype threat manipulation does not seem to affect performance on the initial task, it does not necessarily mean that the participants were not "threatened" during this and the subsequent anagram task. Further, the similar performance for those under stereotype threat and those not under stereotype threat will allow us to draw more internally valid conclusions about the true causes of performance differences found for the subsequent task. This is because any effects found for subsequent performance will not be due to actual performance on the initial task, but rather the manipulations.

Ethnicity and Gender Effects

In order to assess whether there were important differences in the dependent variables based on gender or ethnicity, t-tests for gender and one-way ANOVAs for ethnicity were performed on all variables used in the study. The one-way ANOVAs for ethnicity revealed no significant differences based on reported ethnic background (all *p*s > .05). T-tests assessing gender effects indicate that men and women differed significantly on measures of baseline (pre-study) and in-lab academic identification: baseline: M_{women}= 5.40, *SD* = .69, M_{men}= 5.15, *SD* = .73; $t(128) = 2.00$, $p < .05$, $d = .35$. In-lab: M_{women}= 5.44, *SD* = .65, M_{men}= 5.06, *SD* = .73; $t(128) = 3.04$, $p < .01$, $d = .55$. These differences indicate that women were generally more identified with academics than men. Importantly, however, no differences were found between men and women on measures of actual performance. Nevertheless,

because there were differences in levels of academic identification and because study depends directly on gender stereotypes that might affect men and women differently, future analyses will control for gender in order to ensure that the effects seen are not due to differences in how the two sexes respond to the manipulations or identify with academics.

Effects of Stereotype Threat, Performance Feedback, and Academic Identification

Performance. An ANCOVA with gender entered as a covariate was performed with stereotype threat condition (stereotype threat v. no stereotype threat), feedback condition (positive feedback v. negative feedback), and academic identification (strongly-identified v. weakly-identified) entered as independent variables and average time to complete each anagram entered as the dependent variable. No main effects or two-way interactions were significant for the average time spent on solving each anagram in the second anagram task. However, a three-way interaction between stereotype threat condition, feedback condition, and academic identification was found, $F(1, 124) = 5.55$, $p = .02$, $\eta^2 = .04$. This interaction is presented graphically in Figure 2.

Further analyses revealed that this three-way interaction was driven by a significant two-way interaction between stereotype threat condition and academic identification among those who received negative feedback, $F(1, 58) = 5.04$, $p = .03$, partial $\eta^2 = .08$. The two-way interaction for those in the positive feedback condition was not significant, $F(1, 61) = .96$, $p = .33$, partial $\eta^2 = .02$. Among participants who received negative feedback, those who were highly-identified with academics performed better when under stereotype threat ($M = 31.01$, $SE = 2.26$) compared to when they were not under stereotype threat ($M = 37.19$, $SE = 2.66$). The opposite pattern emerged for those low in academic identification: stereotype threat led

to lower performance (*M* = 37.24, *SE* = 2.57) compared to when not under stereotype threat (*M* = 32.54, *SE* = 2.26).

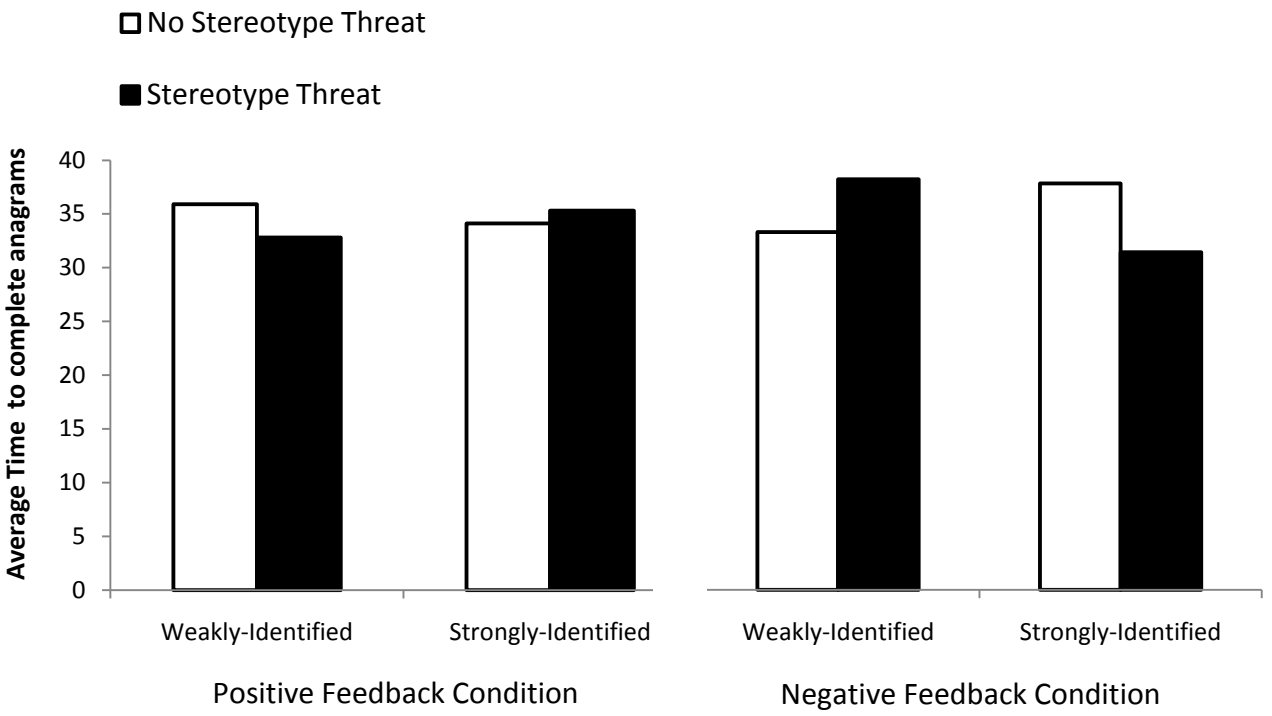

Figure 2. Significant 3-way interaction between stereotype threat condition, baseline academic identification, and performance feedback condition for average time to solve an anagram (performance). The graph on the left contains the non-significant 2-way interaction for those in the positive feedback condition. The graph on the right contains the significant 2-way interaction between stereotype threat condition and academic identification for those in the negative feedback condition. Gender is entered as a covariate on all analyses. Note that higher levels on the graph represent lower levels of performance.

Simple effects of stereotype threat were not significant within the strongly-identified and weakly-identified participants, although the main effect of stereotype threat was

marginally significant for those in the weakly-identified category (strongly-identified: $F(1, 59) = 3.13$, $p = .07$; weakly-identified: $F(1, 59) = 1.89$, $p = .17$). Power analyses revealed that there may not be enough power to detect significant differences. The power to detect significant differences in among those with strong academic identification was .41 (41%), while the power to detect differences among those with weak academic identification was only .27 (27%). This low power is understandable given the fact that the simple effects are comparing one eighth of the sample to another eighth. However, it indicates that there was insufficient power in to detect the differences, especially among weakly-identified participants. While it would be ideal to have more participants or a design that only assessed effects among those who received negative feedback, this study did not have enough participants in each cell. This may mean that the marginally-significant effects might be significant if there were a larger sample.

Post-feedback academic identification. In order to assess whether stereotype threat, feedback, and academic identification affected post-feedback levels of academic identification, an ANCOVA was performed with gender entered as a covariate. Contrary to predictions, no 3-way interaction between the independent variables was found, $F(1, 120) = .01$, $p = .91$, $\eta^2 < .01$. Looking next at whether sets two independent variables interacted to influence post-feedback academic identification, no 2-way interactions were significant: stereotype threat by feedback: $F(1, 120) = 1.23$, $p = .27$, $\eta^2 = .01$; stereotype threat by baseline academic identification: $F(1, 120) = .68$, $p = .41$, $\eta^2 = .01$; feedback by baseline academic identification: $F(1, 120) = .17$, $p = .68$, $\eta^2 < .01$. Finally, assessing whether the independent variables alone predicted post-feedback academic identification, an unsurprising main effect of baseline academic identification was found, $F(1, 124) = 117.49$, $p < .01$, $\eta^2 =$

.49. Those with high baseline academic identification reported higher levels of post-feedback academic identification (M = 5.76, SE = .06), while those with low baseline academic identification reported lower levels of post-feedback academic identification (M = 4.80, SE = .06). Additionally, there was an unsurprising main effect of feedback condition on post-feedback academic identification, such that those who received negative feedback reported less identification with academics (M = 5.16, SE = .06) compared to those who received positive feedback (M = 5.39, SE = .06), regardless of previous level of identification or whether they were under stereotype threat, $F(1, 124) = 6.44$, $p = .01$, $\eta^2 = .06$.

Identification as a college student. An ANCOVA with stereotype threat condition, performance feedback condition, and academic identification entered as independent variables and gender entered as a covariate was performed with post-feedback identification as a college student as the dependent variable A three-way interaction between stereotype threat condition, feedback condition, and academic identification was also found for the extent to which participants identified as a college student after performing the initial anagram task and receiving feedback, $F(1, 124) = 10.35$, $p < .01$, $\eta^2 = .05$. This interaction is presented graphically in Figure 3. Like in the ANCOVA assessing performance, this effect appeared to be driven by a two-way interaction within the negative feedback condition: the interaction between stereotype threat condition and academic identification was found for those who received negative feedback, $F(1, 58) = 5.85$, $p = .02$, $\eta^2 = .09$, but not for those who received positive feedback, $F(1, 61) = 1.76$, $p = .26$, $\eta^2 = .02$. Simple effects tests indicate that within the negative feedback condition, those weakly-identified with academics had higher identification as a college student when under stereotype threat (M = 5.43, SE = .27) compared to when not under stereotype threat (M = 4.63, SE = .24), $F(1, 59) = 4.89$, p =

.03, η^2 = .08. The pattern for those with strong identification with academics was the opposite, with higher levels of college student identification reported for those not under stereotype threat (*M* = 5.58, *SE* = .28) compared to those under stereotype threat (*M* = 5.15, *SE* = .24). The simple effects tests, however, were not significant for those highly-identified with academics, *F* (1, 59) = 1.31, *p* = .26, η^2 = .02.

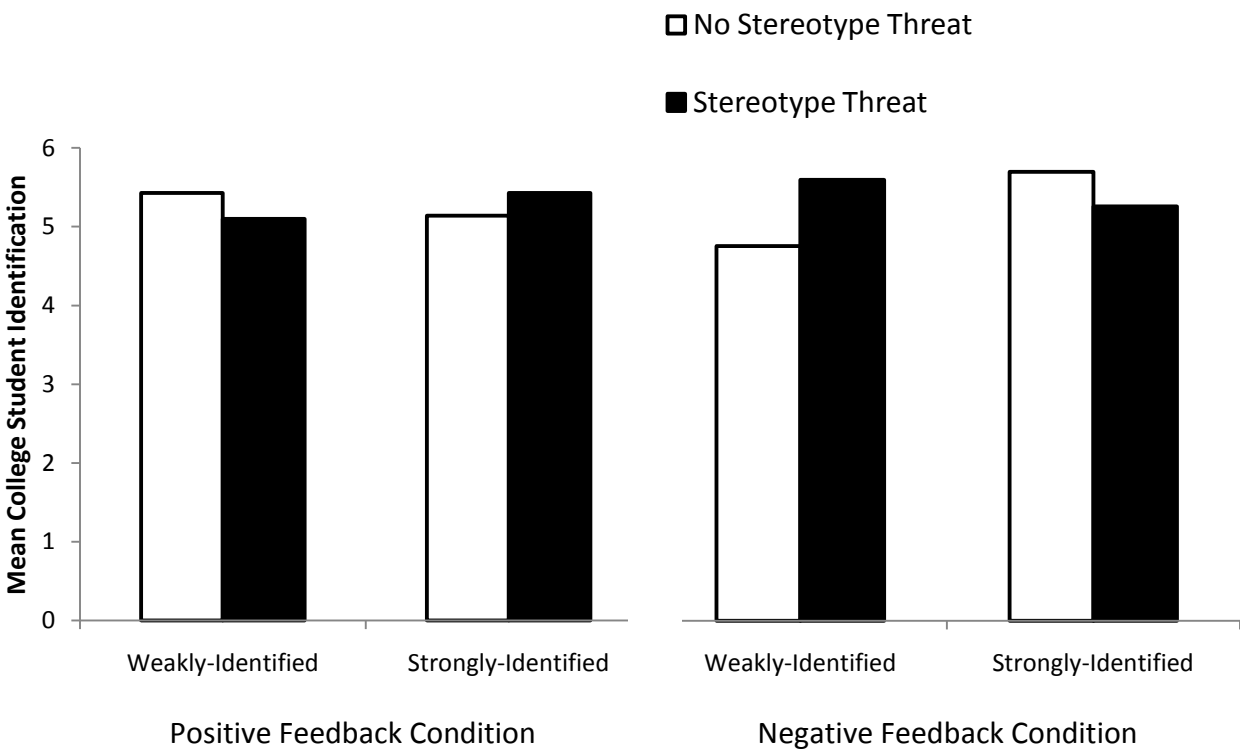

Figure 2. Significant 3-way interaction between stereotype threat condition, baseline academic identification, and performance feedback condition for college student identification. The graph on the left contains the non-significant 2-way interaction for those in the positive feedback condition. The graph on the right contains the significant 2-way interaction between stereotype threat condition and academic identification for those in the negative feedback condition. Gender is entered as a covariate on all analyses.

Regarding the insignificant simple effect, however, it is important to note that like with the performance measure, there was very little power to detect significant differences when comparing these small portions of the entire sample. Power was estimated at .59 (59%) for weakly-identified participants and only .20 (20%) for strongly-identified participants, indicating that there was insufficient power to detect significant differences.

Interestingly, looking at the effects on performance compared to the effects on college student identification, the variables interact to predict college identification in the opposite direction that they predict performance. For example, while weakly-identified students reported more identification as a college student when facing stereotype-confirming feedback, their subsequent performance was lower. And while not all of the simple effects were significant, the direction of the effects were opposite for each condition, at least among those who experienced negative feedback. See Figure 2 and Figure 3 to compare the interactive effects.

Correlational Analyses

In order to assess relationships between the dependent variables, bivariate correlations were performed between post-feedback levels of academic identification, post-feedback levels of college student identification, and performance. Not surprisingly, post-feedback levels of academic identification were positively correlated with levels of college identification, $r = .27$, $p < .01$. However, while post-feedback levels academic identification were not related to performance, $r = .13$, $p = .14$, post-feedback levels college identification were positively correlated with average time to solve an anagram, $r = .18$, $p = .04$. This indicates that the more a participant identified as a college student after receiving their feedback, the poorer they performed on the subsequent task.

Discussion

This study investigated the post-feedback effects of stereotype threat on performance, providing an essential link between experiences with stereotype threat and their approach to future tasks in the same domain. By assessing how levels of academic identification moderated the relationship between the feedback and future performance, we attained a more nuanced view of academic striving with important implications for understanding why some people disengage from academics after facing negative group-relevant feedback while others are able to employ coping mechanisms that spare their future performance.

Partially consistent with Hypothesis 1, the 3-way interaction between stereotype threat condition, performance feedback, and academic identification was significant for performance on the second anagram task. However, within the positive feedback condition, the hypothesized 2-way interaction between stereotype threat and academic identification was not significant (Hypothesis 1a). This lack of significant effects for the positive feedback condition suggests 1) that positive feedback over-rides any effects of the feedback's group-relevance (i.e., stereotype threat or not), and 2) that the effects of positive feedback are less sensitive to differences in baseline academic identification. Perhaps, similar to work done on how positive affirmations can help reduce the effect of stereotype threat (e.g., Martens et al., 2006; Miyake et al., 2010) receiving confirmation that they are capable in a domain allows students to focus on their strengths and reduces the anxiety they may feel due to their group membership on future tasks. And while weakly-identified students may care less about what tests say about their ability, it feels equally good to receive positive feedback regardless of whether they are strongly- or weakly-identified.

Within the negative feedback condition, the 2-way interaction between stereotype threat and academic identification was significant as hypothesized (Hypothesis 1b). Performance after negative feedback was spared for strongly-identified students in the stereotype threat condition, but it was not spared for weakly-identified students who were under stereotype threat. This finding suggests that strongly-identified students may be better able to cope with stereotype-confirming feedback. When strongly-identified individuals learn that the test they performed poorly on often shows gender differences, they may be able to put some of the fault on their group membership, convince themselves that the test is biased, and bounce back to perform well. The ability to attribute failures to something beyond simply natural ability may allow those who are highly-identified with academics to decrease their anxiety and interfering cognitive processes on future tests, sparing their working memory and performance.

Conversely, it appears that students with weak academic identification are more susceptible to performance decrements when receiving stereotype-confirming feedback (i.e., negative feedback while under stereotype threat). This finding supports the disengagement hypothesis (see Steele, 1997)—that those with repeated exposure to stereotypes and negative feedback will disengage their efforts from future tasks in order to avoid feeling poorly about their poor performance (see Major & O'Brien, 2005). While the difference in performance between those under stereotype threat and those not under stereotype threat was not significant for weakly-identified students, the findings indicate that weakly-identified students have fewer adaptive tools to "bounce back" from stereotype-confirming feedback than strongly-identified students. Weakly-identified students have been found to be less susceptible to stereotype threat (Keller, 2007a). However, if these students respond to

stereotype-confirming feedback with lower performance compared to those who are not negatively stereotyped, they may be less likely to continue in stereotypic domains.

While the hypothesized effects were found for how well students performed on a subsequent task, the hypothesized 3-way interaction was not found for post-feedback levels of academic identification (Hypothesis 2). Hypotheses were correct in predicting that positive feedback would lead to higher levels of post-feedback academic identification, but this was not qualified by an interaction between stereotype threat, performance feedback, and baseline academic identification. Within the positive feedback condition, post-feedback levels of academic identification were similar regardless of stereotype threat condition or baseline academic identification (Hypothesis 2a). Also contrary to hypotheses, within the negative feedback condition, strongly-identified participants did not appear to disengage with academics after receiving stereotype-confirming feedback (i.e., while under stereotype threat) compared to when receiving non-group-relevant negative feedback (i.e., not under stereotype threat). And weakly-identified students did not appear to disengage from academics more when they were under stereotype threat compared to when they were not (Hypothesis 2b). This limits our ability to interpret the mechanisms behind the differences found for performance on the second anagram task, though it does imply that changes in academic identification may not be the primary mechanism through which strongly-identified students strive or weakly-identified students fail to strive on future stereotypic tasks.

The significant 3-way interaction for post-feedback college student identification, however, does help inform us about how some students strive in the face of stereotypes while others do not. While exploratory in nature, the analyses investigating how stereotype threat, performance feedback, and academic identification interact to influence post-feedback levels

of college student identification provide an interesting pattern of results that may help us understand how different students respond to feedback. Like with the performance variable, there were no significant effects within the positive feedback condition for college student identification. However, within the negative feedback condition, there was a significant 2-way interaction between stereotype threat condition and academic identification. Partially consistent with Hypothesis 3, for students strongly-identified with academics who received negative feedback, college student identification was lower for those under stereotype threat compared to those not under stereotype threat; however, this difference was not significant. This may indicate that for strongly-identified students, disidentifying with college students may be a coping mechanism when facing stereotypes in academic settings. The evidence that lower levels of post-feedback college student identification are correlated with better performance also suggests that this ability to disidentify with an aspect of one's social identity—even a positive aspect—may also help to spare performance.

For those in the negative feedback condition with weak-academic identification, stereotype threat led to higher levels of college student identification compared to those not under stereotype threat. This indicates that while weakly-identified students may reduce their college student identification when receiving negative feedback that is not linked to their group membership (i.e., not under stereotype threat), their identification as a college student is higher when their feedback is group-relevant (i.e., under stereotype threat). While there is little theory that might explain why this may be the case, perhaps when the negative feedback is group-relevant, weakly-identified students are attempting to maintain a positive identity by identifying more with college students. Because weakly-identified students feel more threatened by feedback that confirms group stereotypes (i.e., negative feedback in the

stereotype threat condition), they may be more likely to latch on to a positive social identity. This ability to focus on a positive social identity, however, does not appear to spare performance, as evidenced by the negative correlation between post-feedback college student identity and performance on the subsequent task.

The findings indicate that weakly-identified students have fewer adaptive tools to "bounce back" from stereotype-confirming feedback than strongly-identified students. Weakly-identified students have been found to be less susceptible to stereotype threat (Keller, 2007a). However, if these students respond to stereotype-confirming feedback with lower performance compared to those who are not negatively stereotyped, they may be less likely to continue in stereotypic domains.

Limitations

Several limitations to the study should be noted. The first involves the insufficient number of participants in the sample that resulted in low power for detecting significant relationships within conditions. Unfortunately, there was not enough time to collect enough data so that each condition could have a sufficient number of participants. While there were expectations for differences within the positive feedback condition, it appears that the positive feedback outweighs the effect of the other variables; as such, the most interesting effects are only found within the negative feedback condition. Future studies could look solely at how people deal with negative feedback in order to simplify the design and reduce the total number of participants needed to attain enough power to detect differences.

A second limitation involves the comparison between those with strong and weak baseline academic identification. While there were clear differences found between the two groups, all participants were drawn from a small, selective group of colleges. Thus, most

participants—regardless of classification as strongly- or weakly-identified with academics—had relatively high levels of academic identification (M = 5.31, SD = .71 on a 7-point scale). It is logical to hypothesize that the differences between strongly- and weakly-identified students would be larger if the study were run with students with a greater range of identification. Future studies could select students with particularly high or low levels of academic identification, or recruit students from a less homogenous participant pool.

Another potential limitation to the study regards the lack of effect found for the stereotype threat manipulation on performance on the initial task. The fact that participants performed similarly on the initial task regardless of stereotype threat condition suggests that the stereotype threat manipulation was unsuccessful in producing stereotype threat-related performance decrements. However, as pointed out above, the main purpose of the study was not to investigate the phenomenon of stereotype threat, but rather to assess how the experiences being under stereotype threat and receiving group-relevant performance feedback influences how we approach and perform on future tasks. It would have been desirable to establish the link between experiences of depressed performance on the initial task (on account of stereotype threat) and performance on the subsequent task. However, the lack of stereotype threat effects for the initial task also allows us to argue that the differences found on performance on the subsequent tasks were due entirely to the manipulations and baseline academic identification, and not *actual* previous performance. While stereotype threat effects were not seen in initial performance, the task was not designed to tap working memory, and enough research has been conducted on stereotype threat to suggest that even if performance did not differ between conditions, the participants exposed to the stereotype threat manipulation likely felt the same anxiety and worry over confirming stereotypes

typical of participants in other stereotype threat studies. Future studies should look into designing an initial task that allows for ambiguous enough performance to make the feedback manipulation believable, yet is still sensitive enough to detect stereotype threat effects.

Conclusion

This study provides an important first step in understanding how students with strong and weak identification with academics react to stereotype-confirming feedback when faced with a similar task in the future. It provides a significant improvement over the previous research in the area (Nussbaum & Steele, 2007) by documenting actual performance differences (as opposed to just differences in persistence) between those facing stereotype-confirming negative feedback and those facing negative feedback that is not relevant to stereotypes. We can see that the type of feedback received has a significant and measurable impact on how participants are able to perform on similar tasks—a finding with tremendous implications for students facing stereotypes in academic settings.

The study also improves on previous research by comparing those who with strong identification with academics and those with weak identification. The findings indicate that strongly-identified students are able to spare their performance when facing stereotype-confirming feedback while weakly-identified cannot: they seem more hindered when their negative feedback is relevant to their group (i.e., under stereotype threat). This gives us insight into why those who are weakly-identified may be at the most risk for chronic disengagement when facing stereotypes, and why those with strong academic identification are able to continue striving. The study also improves upon the previous research by clarifying that the differences found between those in under stereotype threat and those not under stereotype threat were due to the activation of group membership and not simply the

diagnosticity of the task. By explicitly mentioning gender differences when introducing the task to the participants, this study's design is less implicit but more suited to conclude that the participants viewed their feedback as relevant to their group membership.

While this study gives us a nuanced view of how different types of students respond to feedback, it is still limited. As with any first step in research exploring a relatively new but theoretically-rich research paradigm, many questions remain unanswered. Most importantly, the mechanisms behind the performance differences are still inconclusive: evidence from this study suggests the role of disidentification with the college student identity, while previous theory and research suggest that attributions to group membership and decreased engagement might have played a role in allowing some students to strive while others disengaged. Future studies need to assess mediators of relationship between type of feedback and future performance, including how people attribute their performance, and whether some individuals express more physiological markers of threat when receiving stereotype-confirming feedback. Additionally, future studies could explore how differences in identification with one's group (i.e., gender identification) might influence perceptions of feedback. Finally, stereotype based on ethnic group membership or other stigmatizing characteristics should be explored to see whether other identities show the same pattern of effects that we saw when using gender stereotypes.

The findings that strongly-identified students were able to spare their future performance when negative feedback was linked to group membership suggests that strongly-identified students are able to employ coping mechanisms when underperforming in stereotypic domains where weakly-identified students are not. The mechanisms behind such striving need to be more fully investigated, but the present study makes a large contribution

to both the literature on stereotype threat and the research investigating the effect of more long-term social identity threat in academic settings.

References

Allison, K. (1998). Stress and oppressed category membership. In J. Swim and C. Stangor, *Prejudice: The target's perspective* (pp. 145-170). San Diego, CA: Academic Press.

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

Aronson, J., Quinn, D., & Spencer, S. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. Swim and C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). San Diego, CA: Academic Press.

Beilock, S., Jellison, W., Rydel, R., McConnell, A., & Carr, T. (2006). Causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32*, 1059-1071.

Beilock, S., & McConnell, A. (2004). Stereotype threat and sport: Can athletic performance be threatened? *Journal of Sport & Exercise Psychology, 26*, 597-609.

Bem, S. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 196-205.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174-181.

Blascovich, J., Spencer, S., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*, 225-229.

Bosson, J., Haymovitz, E., & Pinel, E. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40*, 247-255.

Campbell, T., Gillaspy, J., & Thompson, B. (1997). The factor structure of the Bem Sex-Role Inventory (BSRI): Confirmatory analysis of long and short forms. *Educational and Psychological Measurement, 57*, 118-124.

Carr, P., & Steele, C. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853-859.

Castelli, L., Pavan, G., Ferrari, E., & Kashima, Y. (2009). The stereotype and the chameleon: The effects of stereotype use on perceivers' mimicry. *Journal of Social Psychology, 45*, 835-839.

Clark, R., Anderson, N., Clark, V., & Williams, D. (1999). Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist, 54*, 805-816.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*, 608-630.

Croizet, J., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.

Croizet, J., Despres, G., Gauzins, M., Huguet, P., Leyens, J., & Meot, A. (2003). Stereotype threat undermines intellectual ability by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721-731.

Danaher, K., & Crandall, C. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655.

Davies, P., Spencer, S., Quinn, D., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615-1628.

Eccleston, C., & Major, B. (2006). Attributions to discrimination and self-esteem: The role of group identification and appraisals. *Group Processes & Intergroup Relations, 9*, 147-162.

Forbes, C., Schmader, T., & Allen, J. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3*, 253-261.

Gonzales, P., Blanton, H., & Williams, K. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659-670.

Harrison, L., Stevens, C., Monty, A., & Coakley, C. (2006). The consequences of stereotype threat on academic performance of White and non-White lower income college students. *Social Psychology of Education, 9*, 341-357.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*, 175-179.

Kaiser, C., Dyrenforth, P., & Hagiwara, N. (2006). Why are attributions to discrimination interpersonally costly? A test of system- and group-justifying motivations. *Personality and Social Psychology Bulletin, 32*, 1523-1536.

Kaiser, C., & Miller, C. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin, 27*, 254-263.

Keller, J. (2007a). 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.

Keller, J. (2007b). When negative stereotypic expectancies turn into challenge or threat: The moderating role of regulatory focus. *Swiss Journal of Psychology, 66,* 163-168.

Koch, S., Müller, S., & Sieverding, M. (2008). Women and computers: Effects of stereotype threat on attribution of failure. *Computers and Education, 51,* 1795-1803.

Leary, M., & Schreindorfer, L. (1998). The stigmatization of HIV and AIDS: Rubbing salt in the wound. In V. Darlega and A. Barbee (Eds.), *HIV and social interaction* (pp. 12-29). Thousand Oaks, CA: Sage Publications, Inc.

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18,* 302-318.

Major, B., & O'Brien, L. (2005). The social psychology of stigma. *Annual Review of Psychology, 56,* 393-421.

Major, B., Quinton, W., & Schmader, T. (2003). Attributions to discrimination and self-esteem: Impact of group identification and situational ambiguity. *Journal of Experimental Social Psychology, 39,* 220-231.

Major, B., & Sawyer, P. (2009). Attributions to discrimination: Antecedents and consequences. In T. Nelson (Ed.) *Handbook of prejudice, stereotyping, and discrimination* (pp. 89-110). New York, NY: Psychology Press.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236-243.

Marx, D., & Roman, J. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28,* 1183-1193.

Mendes, W., McCoy, S., Major, B., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology, 94*, 278-291.

Miyake, A., Kost-Smith, L., Finkelstein, N., Pollok, S., Cohen, G., & Ito, T. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*, 1234-1237.

Nussbaum, A., & Steele, C. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology, 43*, 127-134.

Osborne, J. (1997). Identification with academics and academic success among community college students. *Community College Review, 25*, 59-67.

Osborne, J., & Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be most likely to withdraw. *Educational Psychology, 26*, 563-577.

Romero-Canyas, R., Anderson, V., Reddy, K., & Downey, G. (2009). Rejection sensitivity. In M. Leary and R. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 466-479). New York, NY: Guilford Press.

Rosenberg, M. (1969). *Society and the adolescent self-image*. Middleton, CT: Weslyan University Press.

Scheepers, D. (2009). Turning social identity threat into challenge: Status stability and cardiovascular reactivity during inter-group competition. *Journal of Experimental Social Psychology, 45*, 228-233.

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356.

Schmader, T., Major, B., & Gramzow, R. (2001). Coping with ethnic stereotypes in the academic domain: Perceived injustice and psychological disengagement. *Journal of Social Issues, 57*, 93-111.

Seacat, J., & Mickelson, K. (2009). Stereotype threat and the exercise/dietary health intentions of overweight women. *Journal of Health Psychology, 14*, 556-567.

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

Stangor, C., Swim, J., Van Allen, K., & Sechrist, G. (2000). Reporting discrimination in public and private contexts. *Journal of Personality and Social Psychology, 82*, 69-74.

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

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

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

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology, 44*, 445-452.

Stricker, L., & Ward, W. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34*, 665-693.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1*, 149-178.

Tajfel, H., & Turner, J. (1986). The social identity theory on intergroup behavior. In S. Worchel and W. Austin (Eds.), *The psychology of intergroup relations* (pp. 7-24). Chicago, IL: Nelson-Hall.

van Laar, C. (2000). The paradox of low academic achievement but high self-esteem in African American students: An attributional account. *Educational Psychology Review, 12*, 33-61.

Verkuyten, M. (1998). Perceived discrimination and self-esteem among ethnic minority adolescents. *Journal of Social Psychology, 138*, 479-493.

Appendix A

Anagrams Presented in First Initial Anagram Task

1. AEIBORVH (behavior)
2. ERTESNPE (pretense)
3. ESUBSISN (business)
4. IHMTIDGN (midnight)
5. NOIFAUNTN (fountain)
6. OPDEARPV (approved)
7. SEURPTAS (pastures)
8. VOETRELA (elevator)

Note. All anagrams were presented for 40 seconds. Word in parentheses is the word that uses all eight letters, though participants could write down any words that used any combination of letters in the anagram.

Appendix B

Anagrams Presented in the Second Anagram Trial

1. KAREB (break, brake, baker; 45s)
2. REGEYN (energy; 75s)
3. ECSNER (screen; 60s)
4. TERNE (enter; 60s)
5. VEESIR (revise; 75s)
6. LGEOV (glove; 45s)
7. RENIND (dinner; 60s)
8. TALNS (slant; 45s)

Note. Words in parentheses are acceptable answers for each anagram. Numbers in parentheses represent the number of seconds the particular anagram was shown before automatically switching to the next one.

Appendix C

Full Pre-Study Survey (Online)

Self-Esteem Scale (Rosenberg, 1969)

Responses from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*).

1. I feel that I'm a person of worth, at least on an equal plane with others.
2. I feel that I have a number of good qualities.
3. All in all, I am inclined to feel that I am a failure. [*reverse coded*]
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of. [*reverse coded*]
6. I have a positive attitude toward myself.
7. On the whole, I am satisfied with myself.
8. I wish I could have more respect for myself. [*reverse coded*]
9. I certainly feel useless at times. [*reverse coded*]
10. At times I think I am no good at all. [*reverse coded*]

Academic Identification Scale (Osborne, 1997)

Responses from 1 (s*trongly disagree*) to 7 (s*trongly agree*).

1. Being a good student is an important part of who I am.
2. I feel that the grades I get are an accurate reflection of my abilities.
3. My grades do not tell me anything about my academic potential. [*reverse coded*]
4. I don't really care about what tests say about my intelligence. [*reverse coded*]
5. School is satisfying for me because it gives me a sense of accomplishment.

6. If the tests we take were fair, I would be doing much better in school. [*reverse coded*]
7. I am often relieved if I just pass a course. [*reverse coded*]
8. I often do my best work in school.
9. School is very boring for me, and I'm not learning what I feel is important. [*reverse coded*]
10. I put a great deal of myself into some things at school because they have special meaning or interest for me.
11. I enjoy school because it gives me a chance to learn many interesting things.
12. I feel like the things I do at school waste my time more than the things I do outside of school. [*reverse coded*]
13. No test will ever change my opinion of how smart I am. [*reverse coded*]

Gender Identification Scale (Schmader, 2002; adapted from Luhtanen & Crocker, 1992)

Responses from 1 (*strongly disagree*) to 7 (*strongly agree*).

1. Being a man/woman is an important part of my self-image.
2. Being a man/woman is unimportant to my sense of what kind of person I am. [*reverse coded*]
3. Being a man/woman is an important reflection of who I am.
4. Being a man/woman has very little to do with how I feel about myself. [*reverse coded*]

College Student Identification Scale (Adapted from Schmader, 2002; adapted from Luhtanen & Crocker, 1992)

Responses from 1 (*strongly disagree*) to 7 (*strongly agree*).

1. Being a college student is an important part of my self-image.
2. Being a college student is unimportant to my sense of what kind of person I am. [*reverse coded*]
3. Being a college student is an important reflection of who I am.
4. Being a college student has very little to do with how I feel about myself. [*reverse coded*]

Note. The Self-Esteem scale and Gender Identification scale were not used in the final analyses.

Appendix D

Full Post-Feedback Survey (In-Lab)

Performance Perceptions

Responses for items 2-5 ranged from 1 (s*trongly disagree*) to 7 (s*trongly agree*).

1. I did **<u>better</u>** / **<u>much better</u>** / **<u>about the same</u>** / **<u>worse</u>** / **<u>much worse</u>** as I expected on the test. (Circle One)
2. I performed well on this test.
3. I am happy with my score on the test.
4. My score is an accurate representation of how I felt I performed on the test.
5. I enjoyed the test.

State Self-Esteem Scale (Modified from Rosenberg, 1969)

Responses ranged from 1 (s*trongly disagree*) to 7 (s*trongly agree*).

1. Right now, I feel that I'm a person of worth, at least on an equal plane with others.
2. At the moment, I feel that I have a number of good qualities.
3. At the moment, I am inclined to feel that I am a failure. [*reverse coded*]
4. Right now, I feel that I am able to do things as well as most other people.
5. Right now, I do not have much to be proud of. [*reverse coded*]
6. At the moment, I have a positive attitude toward myself.
7. Right now, on the whole, I am satisfied with myself.
8. Right now, I wish I could have more respect for myself. [*reverse coded*]
9. At this time, I certainly feel useless. [*reverse coded*]
10. Right now, I think I am no good at all. [*reverse coded*]

Bem Sex-Role Inventory, Short Form (Campbell, Gillaspy, & Thompson, 1997; adapted from Bem, 1974)

Responses Ranged from 1 (*not at all like me*) to 7 (*very much like me*). [M] denotes a masculine trait; [F] denotes a feminine trait.

1. Defends own beliefs [M]
2. Independent [M]
3. Affectionate [F]
4. Assertive [M]
5. Strong personality [M]
6. Forceful [M]
7. Sympathetic [F]
8. Has leadership ability [M]
9. Sensitive to others' needs [F]
10. Willing to take risks [M]
11. Understanding [F]
12. Compassionate [F]
13. Eager to soothe hurt feelings [F]
14. Dominant [M]
15. Warm [F]
16. Willing to take a stand [M]
17. Tender [F]
18. Aggressive [M]
19. Loves children [F]

20. Gentle [F]

Academic Identification Scale (Osborne, 1997)

Responses ranged from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. Being a good student is an important part of who I am.
2. I feel that the grades I get are an accurate reflection of my abilities.
3. My grades do not tell me anything about my academic potential. [*reverse coded*]
4. I don't really care about what tests say about my intelligence. [*reverse coded*]
5. School is satisfying for me because it gives me a sense of accomplishment.
6. If the tests we take were fair, I would be doing much better in school. [*reverse coded*]
7. I am often relieved if I just pass a course. [*reverse coded*]
8. I often do my best work in school.
9. School is very boring for me, and I'm not learning what I feel is important. [*reverse coded*]
10. I put a great deal of myself into some things at school because they have special meaning or interest for me.
11. I enjoy school because it gives me a chance to learn many interesting things.
12. I feel like the things I do at school waste my time more than the things I do outside of school. [*reverse coded*]
13. No test will ever change my opinion of how smart I am. [*reverse coded*]

College Student Identification Scale (Adapted from Schmader, 2002; adapted from Luhtanen & Crocker, 1992)

Responses ranged from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. Being a college student is an important part of my self-image.
2. Being a college student is unimportant to my sense of what kind of person I am. [*reverse coded*]
3. Being a college student is an important reflection of who I am.
4. Being a college student has very little to do with how I feel about myself. [*reverse coded*]

Gender Identification Scale (Schmader, 2002; adapted from Luhtanen & Crocker, 1992)

Responses ranged from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. Being a man/woman is an important part of my self-image.
2. Being a man/woman is unimportant to my sense of what kind of person I am. [*reverse coded*]
3. Being a man/woman is an important reflection of who I am.
4. Being a man/woman has very little to do with how I feel about myself. [*reverse coded*]

Trial 2 Performance Expectancies Scale

Responses for items 2-5 ranged from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. I think I will do **much better** / **better** / **about the same** / **worse** / **much worse** on the second trial of this test compared to the average college student. (Circle One)
2. I think I will do **much better** / **better** / **about the same** / **worse** / **much worse** on the second trial compared to how I did on the first trial. (Circle One)

3. I am worried about my performance on the second trial of this test. [*reverse coded*]
4. I am optimistic about my performance on the second trial of this test.
5. I think the feedback from the first test will help me on the second trial of the test.
6. I think my score will be _____/100 on the second trial (remember, average college students is 85).

Note. The State Self-Esteem scale, Bem Sex-Role Inventory, Gender Identification scale, and Trial 2 Performance Expectancies scale were not used in the final analyses.

Appendix E

Final Survey

Performance Perceptions

Responses for items 3-6 ranged from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. I think I did **much better** / **better** / **about the same** / **worse** / **much worse** on the second trial of this test compared to the average college student. (Circle One)
2. I think I did **much better** / **better** / **about the same** / **worse** / **much worse** on the second trial compared to how I did on the first trial. (Circle One)
3. I think I performed well on the second trial of the test.
4. I felt confident when I was completing the second trial of the test.
5. I was worried when I was completing the second trial of the test.
6. I enjoyed the second trial of the test.

Performance Attributions

"What factors do you think influenced your performance on the anagram tasks IN GENERAL." Responses from 1 (*strongly disagree*) to 7 (s*trongly agree*).

1. My natural abilities
2. The testing environment
3. My effort on the tasks
4. The way the tasks were structured
5. *Internal* motivation to do well
6. *The way the tasks were administered*
7. The type of tasks

8. Other:_____

Self-Handicapping

"Please answer the following questions about the anagram tasks IN GENERAL."

Responses from 1 (*not at all*) to 7 (*very much*).

1. How distracted did you feel during the tasks?
2. How much effort did you expend on the tasks?
3. How motivated were you to succeed at the tasks?
4. How much experience do you have with similar types of tasks as the ones today?
5. How many hours of sleep did you get last night?
6. How stressed out have you been in the past few days?

Note. None of these scales or items were used in the final analysis.