Journal of Advanced Academics
jaa Volume 22 ✤ Number 3 ✤ Spring 2011 ✤ pp. 374–411

Can Stereotype Threat Be Measured? A Validation of the Social Identities and Attitudes Scale (SIAS)

Katherine Picho
Scott W. Brown
University of Connecticut

Years before teen talk Barbie uttered the phrase "Math class is tough!" scholars tried to understand and explain gender differences in math and science performance. Gender differences in visual-spatial ability (Baenninger & Newcombe, 1995; Gray, 1981; Terlecki, Newcombe, & Little, 2007), differential parent and teacher stereotype-based expectations (Jacobs, 1991; Jacobs & Eccles, 1992), and negative gender stereotypes regarding the mathematical ability of females are just some of the theories that have since been used to explain sex differences in mathematics. Regardless of which theory one might deem to be more plausible, there appears to be a systemic weeding out of females and ethnic minorities like African Americans and Hispanics in science, technology, engineering, and mathematics (STEM). This phenomenon begins in high school and continues through the doctoral level. Recent reports from the National Center for Education Statistics (NCES, 2007) showed that fewer girls than boys take Advanced Placement (AP) exams in STEM-related subjects like calculus and physics, and those who do generally earn lower scores

summary

This study reported the development and validation of the Social Identities and Attitudes Scale (SIAS)—a stereotype threat susceptibility measure. Exploratory and confirmatory factor analyses conducted with college students revealed that the scale, which constitutes six key ST moderators, possessed strong psychometric properties. The SIAS makes it possible for researchers to establish a baseline for measuring ST susceptibility and, subsequently, the impact of interventions attempting to reduce it. It provides researchers with the means to tease ST effects apart, differentiate between levels of ST risk (e.g., low, moderate, high), and facilitate the development of specialized interventions for different ST risk levels. Its use as a tool for identifying high-risk ST individuals might also be useful for mixed methods research seeking to understand contextual factors that exacerbate ST for these individuals and, importantly, how they respond to these environments.

Picho, K., & Brown, S. W. (2011). Can stereotype threat be measured? A validation of the Social Identities and Attitudes Scale (SIAS). *Journal of Advanced Academics, 22,* 374–411.

than their male counterparts. This trend also extends to African Americans and Hispanics. In 2005, for instance, 31% of Asian American and 16% of Caucasian high school graduates completed calculus, compared to only 6% and 7% of African American and Hispanic high school graduates, respectively (NCES, 2007). The downward trajectory of females and ethnic minorities continues through the undergraduate to graduate level and is even more pronounced for minority females. In 2007, African American women earned less than 2% of doctoral degrees awarded in engineering, the physical sciences, mathematics, and statistics (National Science Foundation [NSF], 2009). An even smaller proportion of Hispanic women and Native American women graduated from these fields (NSF, 2009). The American Association of University Women's most recent report *Why so Few?* (Hill, Corbett, & St. Rose, 2010) found that social and environmental factors like the societal beliefs about the nature of intelligence (i.e., fixed vs. incremental) and stereotype threat (ST) continue to pose serious barriers to the success of women in STEM.

ST is a psychological phenomenon that inhibits the academic performance of individuals in domains where negative ability stereotypes about their group are highlighted. The phenomenon affects females (Aronson, Quinn, & Spencer, 1998) and ethnic minorities (Spencer, Steele, & Quinn, 1999) in STEM on two levels: (1) It induces anxiety, which impairs short-term academic performance (Schmader, 2002; Spencer et al., 1999); and (2), chronic underperformance due to ST subsequently leads to disidentification—a process where individuals disengage from and lose interest in the quantitative domain (Steele, 1997). Steele (1997) demonstrated that females exposed to high levels of ST tended to disidentify with math and math-related careers more often than women who were not exposed to the threat. ST is therefore especially deleterious because it negatively affects the desire and career aspirations of bright, capable, ST individuals who value the stereotyped domains and have the capacity to be successful in these fields.

Empirical research documents negative ST effects on the math performance of females as early as elementary school

(Ambady, Shih, Kim, & Pittinsky, 2001) and middle school (Huguet & Regner, 2007). It is no surprise then that by high school, the critical point when students have to choose vocational paths, fewer females elect to pursue STEM education further. By compromising the performance of capable, math-identified females and minorities over time, ST weeds out a good number of qualified candidates through disidentification, which subsequently deters members of these groups from pursuing advanced education in STEM, consequently limiting career opportunities in these domains. Because disidentification is a slow process that might occur over several years and manifest itself when it is almost too late to remediate, identifying ST-susceptible individuals with the intent to intervene is critical.

Currently, no scale containing key ST dimensions exists. The purpose of this study is to develop and validate an integrated, psychometrically sound measure of ST; one that we hope will establish a baseline for measuring ST and the impact of interventions attempting to reduce it. Although ST effects are broad and can have an impact on non-STEM domains, the focus of this study is narrowed to ST as it relates to mathematics because mathematics remains a critical component and determinant of entry into careers in physical science domains like engineering. In addition, the majority of ST research has focused on ST in the mathematics domain.

Stereotype Threat

According to ST theory, mere knowledge that a negative stereotype exists about a social group is enough to inhibit one's performance on stereotype relevant tasks. However, in order for ST to occur one must: (a) believe the stereotype and also (b) have a high personal investment in the stereotyped domain (e.g., females and minorities who value mathematics; Steele, 1997). ST is offset by a cognitive imbalance between group and domain identification experienced by the individual (Schmader, Johns, & Forbes, 2008); the tension between these identities elicits negative affective responses like task-related worries (Beilock, Rydell, & McConnell, 2007) and anxiety (Steele & Aronson, 1995), which impair the working memory required for successful task completion, subsequently undermining performance.

Fortunately, ST does not impact all members of these stigmatized groups. For members belonging to these groups, ST is moderated by individual differences on other factors like group identification (Schmader, 2002), domain identification (Steele, 1997; Steele & Aronson, 1995), stigma consciousness (Brown & Pinel, 2003), and emotion regulation (Schmader et al., 2008). Hence individuals exhibiting high levels of the aforementioned factors are highly susceptible to ST effects and more likely to perform below their potential under ST conditions.

ST Mediator and Moderator Measures

There is no integrated ST measure at the moment—only instruments measuring individual ST moderators like math identification and self-efficacy (Brown & Josephs, 1999), gender identification (Luhtanen & Crocker, 1992), and stigma consciousness (Brown & Pinel, 2003). We surmise that the lack of an integrated ST scale to date has probably been because previous and current research had a different focus, and the tools used to answer those questions were sufficient given the research questions.

The first wave of ST research focused on establishing the existence and generalizability of the phenomenon. In these studies, experiments were conducted to examine whether ST effects had been observed in experimental groups that had been primed for negative stereotypes (Aronson et al., 1998; Steele & Aronson, 1995). The second wave of ST research focused on identifying moderators of the phenomenon. Research conducted during this wave focused on examining the effects of one or two moderators at a time on the performance of individuals exposed to ST (Cadinu, Maas, Rosabianca, Figerio, & Latinotti, 2003; Eriksson & Lindholm, 2007; Gonzales, Blanton, & Williams, 2002). Therefore, this area of research utilized subscales measuring specific moderator variables (e.g., gender identification and stigma consciousness). The Math Identification and Self-Efficacy Scale (Brown & Josephs, 1999) is an 11-item instrument assessing math identification and math self-efficacy; it has five math identification and six math self-efficacy items. Studies that have used the math identification subscale (Brown & Josephs, 1999; Brown & Pinel, 2003; Picho & Stephens, in press) have reported reliability estimates ranging from 0.65–0.83, while the math self-efficacy subscale has reported reliabilities of .96 (Brown & Pinel, 2003) and .86 (Picho & Stephens, in press). The Stigma Consciousness Questionnaire, a 10-item instrument developed by Pinel (1999), has reported reliabilities of .77 and .89 (Brown & Pinel, 2003). Most ST studies examining gender identification have used the gender subscale of Luhtanen and Crocker's (1992) collective self-esteem scale with reported reliability estimates ranging from 0.66–0.85 (Brown & Pinel, 2003; Eriksson & Lindholm, 2007; Keller & Molix, 2008; Schmader, 2002).

Although existing experimental research has established the individual contribution of each of the aforementioned variables to making one susceptible to ST, not much has been done to examine the extent to which these factors might collectively mitigate (or exacerbate) ST. For example, research demonstrates that under ST, women who identify strongly with their gender perform worse in mathematics compared to those for whom the reverse is true (Keller & Molix, 2008). Similar results have been found in the mathematics performance of women with high versus low stigma consciousness (Brown & Pinel, 2003) under ST. But are ST effects exacerbated in females who rank high on both (as opposed to just one) of these factors? Similarly, given deleterious effects of ST on African Americans for example, would ST have a uniform impact on Black and White females in mathematics, or would Black females be at a higher risk of suffering more adverse ST effects by virtue of group membership in two stereotyped groups (race and gender)? It is still not clear whether the effects of ST moderator variables are global or uniform for all at-risk individuals or whether the effects vary by degree for different ST-susceptible individuals depending on their group membership and individual differences on key moderator variables. Answering questions that speak to the potentially additive or interactive effects of ST would require the use of an integrated measure that incorporated key ST moderators, which currently does not exist.

The plethora of research uncovering key ST moderators ensures that the next wave in ST research explores how ST moderators interact with each other and what effect this has on the performance of ST-susceptible individuals. As current scales are not capable of addressing issues surrounding interactivity/additivity of key moderators and mediators, an integrated measure makes that possible. Such a measure is vital in advancing the ST research agenda because it would allow for the exploration of ST effects within a framework of degree as opposed to merely examining the absence or presence of ST effects on members of these stereotyped groups. Without an integrative measure, it is still unclear how these factors are likely to operate when they interact with one another and what impact, if any, these interactions would have in further distinguishing the degrees by which ST will affect the performance of females or minorities already ascertained to be at risk of performing poorly because of the threat.

The Social Identities and Attitudes Scale (SIAS) was conceptualized on the basis of the relationship between social identities related to race/ethnicity and gender, domain identification, and negative affect. In selecting factors for the SIAS, we focused on those moderators identified by the literature as being key to one's susceptibility to ST (see Table 1). Therefore, ST moderators that had undergone substantial investigation by several researchers were included. Moderator variables reported in only one or two ST studies in the literature were excluded because the number of investigations was too few for one to determine the relative consistency of the variable as a moderator. Thus, including them as part of the measure would have, in turn, had potentially serious implications for internal validity of the scale. We review the key individual differences factors forming the SIAS next.

Table 1

Stereotype Threat Factors

Construct	Conceptual Definition
Math Identification (MI)	Individuals who value math, have the skills to succeed in it, and perceive it as being useful to their future career (Steele, 1997)
Math Self-Concept (MSC)	One's beliefs about one's math abilities.
Gender Identification (GI)	The extent to which one's gender forms a central part of one's self-concept (Hoffman, 2006).
Gender Stigma Consciousness (GSC)	Extent to which one is chronically self-conscious of stigma attached to one's gender (Pinel, 1999)
Ethnic Identification (EI)	The extent to which one's ethnicity forms a central part of one's self-concept (Phinney, 1992)
Ethnicity Stigma Consciousness (ESC)	Extent to which one is chronically self-conscious of stigma attached to one's ethnicity (Pinel, 1999)
Negative Affect (NegAff)	Negative feelings of dejection experienced during math test taking (Marx & Stapel, 2006)

Key Stereotype Threat Factors

Mathematics identification. Domain identification is a key prerequisite for ST to occur (Steele, 1997). By definition, math-identified individuals have the intellectual ability and a vested interest in succeeding in mathematics (Steele, 1997). Math-identified individuals feel threatened when negative stereotypes about the intellectual ability of their social group regarding math performance are made salient in a testing situation (Aronson et al., 1998) because they value mathematics and want to succeed at it (Brown & Pinel, 2003). On the other hand, females who do not identify with math are less likely to be threatened because they lack interest in the domain, do not perceive it to be useful to them (Brown & Pinel, 2003), and as such, remain unaffected by negative ability stereotypes concerning their social groups in the field of mathematics.

Group identification. Social identity theory posits that group identity is an important part of self-concept and that people are motivated to maintain positive social identities when these identities are threatened by unfavorable comparisons with other groups (Tajfel, 1981). Because negative stereotypes define certain groups as inferior to others along a specific dimension, members of these groups are likely to experience a threat to their social identity. Under such circumstances, individuals who are highly identified with their group are more likely than the less identified group members to engage in behavioral and psychological strategies to protect and maintain their social identity (Schmader, 2002). Indeed, studies examining ST in females (Brown & Pinel, 2003; Eriksson & Lindholm, 2007; Keller & Molix, 2008; Schmader, 2002) and minorities (Gonzales et al., 2002; Steele & Aronson, 1995) indicated that when negative stereotypes about the ability of females or minorities (i.e., African Americans and Hispanics) are activated in academic contexts, certain members of these groups tend to experience a decline in performance on challenging tasks in the stereotype-relevant domain (e.g., mathematics). Hence, females tend to underperform relative to males on mathematics tests (Schmader, 2002), while African Americans (Steele, 1997; Steele & Aronson, 1995) and Hispanics (Gonzales et al., 2002) perform less well than Caucasians in verbal and quantitative tasks respectively.

The SIAS assessed group identification for both ethnicity and gender. Based on social identity theory, group identification was conceptualized as the degree to which an individual considered his or her membership to a given group (gender or race) to be central to his or her self-concept.

Gender identification. Individuals with strong gender identity attach great importance to their gender and perceive it to be central to their identity (Hoffman, 2006). Individual differences in gender identification affect the performance of individuals exposed to ST (Keller & Molix, 2008; Kiefer & Sekaquaptewa, 2007). Women exposed to ST perform worse than their male counterparts when they consider gender to be an important part of their self-definition and vice versa. Schmader's (2002) study showed that females tended to score lower than males on a mathematics test (a) when gender identity was linked to test performance, and (b) only if they strongly identified with their gender. Results from this study revealed that when exposed to ST, females who ranked highly on gender identification attempted significantly fewer questions than their male counterparts; by contrast, low gender-identified females performed comparably to males, regardless of whether they were exposed to ST or not. Hence, in conceptualizing items for the scale, gender identification was defined as the extent to which one considered one's femininity or masculinity to be an important part of one's self-definition (Hoffman, 2006).

Ethnic identification. Research shows that the activation of positive ethnic identities can counteract ST effects in mathematics performance for ST susceptible females. Ambady et al. (2001) illustrated the mitigating effects of other group positive identification on performance under ST using a sample of 81 Asian American girls. In this study, ethnic and gender identities were implicitly activated in grades K–2 students by asking pupils to color either a picture of two Asians holding chopsticks or a girl holding a doll, prior to administering a math test. The same identities were activated in grades 3–8 students by asking them ethnicity- and gender-related questions. Results showed that both lower elementary and middle school girls performed significantly worse on the math test when their gender identity was activated compared to when their ethnic identity or no identity was activated. They also found that with the middle school sample, Asian American girls in the ethnic identity activated group performed higher than the control (no identity activated) and the gender identity activated group. In this case, Asian American females were able to combat deleterious ST effects when a more favorable identity associated with a positive stereotype in the domain (i.e., Asians being good at math) was implicitly activated. These findings suggest that negative ST effects can be attenuated in ST susceptible individuals by invoking alternate, positive identity stereotypes associated with the stereotype-relevant domain. An ethnic identification subscale was therefore created on the basis of these findings as well as the implications they had for future ST research. That is, including an ethnic identification subscale would be useful in teasing out ST effects among males and females from the same negatively stereotyped ethnic group where one group had a positive alternate identity (e.g., for African American males vs. African American females, males would have a positive gender stereotype associated with mathematical ability). Examining the interaction between gender and ethnicity relative to ST and math performance would shed new light on the operation of ST. Ethnic identification was therefore defined as the extent to which one considered one's ethnicity to be an important part of one's self-definition.

Math self-concept. In ST contexts, individuals tend to act in ways that either confirm the stereotype (assimilation) or actively engage in counter-stereotypical behaviors that reject the stereotype (reactance; Kray, Galinksy, & Thompson, 2003). Studies show that for highly math-identified females, the ability to counter the negative effects of ST through reactance is moderated by self-efficacy (Hoyt, 2005).

Bandura (1986) described self-efficacy as one's judgments of one's capabilities to perform specific tasks. According to Bandura (2006), self-efficacy is domain-specific and thus any measures of the construct should be tailored as such. Although this task is easily accomplished for many subjects, it is not the case for a multifaceted subject like mathematics, which comprises different subcategories like geometry, algebra, and calculus. However, Bandura (2006) cautioned against creating general efficacy scales because they fail to account for the context surrounding performing specific tasks in the different branches of the domain, leading to ambiguity about the level of task and situational demands that one can achieve in a specific subdomain. Taken together, it is evident that a good math self-efficacy measure would have to assess math self-efficacy separately for each of these branches of mathematics.

However, the situational nature and stringent requirements for creating math self-efficacy scales makes it difficult for researchers to measure the construct because it would most likely vary from sample to sample. For this reason, previous ST studies assessing math-efficacy have used the general math self-concept items adapted from Marsh and O'Neil's (1984) self-description questionnaire in-lieu of math self-efficacy. Thus, a math self-concept factor was created as a factor for the SIAS instead of a general math efficacy scale or several math self-efficacy scales to accommodate each of the different math subdomains. Marsh and Craven (1997) conceptualized academic self-concept as a mental representation of an individual's aptitude in academics. Based on this, we defined math self-concept as one's perception of one's ability in mathematics.

Stigma consciousness. This refers to the extent to which individuals are chronically self- conscious of their stigmatized status (Pinel, 1999). High levels of habitual self-consciousness about negative ability stereotypes related to one's group in a particular domain (e.g., females in mathematics) is posited to make one more sensitive to environmental cues that trigger the stereotype and hence ST (Schmader et al., 2008) because individuals high on this construct are more likely to worry about how others might judge their performance on stereotype-relevant tasks (Brown & Pinel, 2003; Pinel, 2004). Studies have shown that when the negative stereotypes about women in mathematics were made explicit, females with high levels of stigma consciousness tended to perform more poorly than those with low levels of stigma consciousness (Brown & Pinel, 2003).

Negative affect. ST elicits negative emotions (Keller & Dauenheimer, 2003) like self-doubt (Steele & Aronson, 1995), negative expectancies (Stangor, Carr, & Kiang, 1998), dejection (Keller & Dauenheimer, 2003; Marx & Stapel, 2006), and task-related worries (Beilock et al., 2007). These emotions not only exert cognitive demands on working memory but also compete with the task at hand for limited cognitive resources required to perform the task successfully (Beilock et al., 2007; Schmader, 2010). Schmader and Johns (2003) tested the relationship between working memory and ST in women by activating negative stereotypes about women's quantitative ability before measuring working memory capacity. They found that working memory was significantly lower for women in the experimental group after receiving ST manipulation compared to women in the control group. Another ST study by Cadinu, Maas, Rosabianca, and Kiesner (2005) found that women taking a difficult math test reported having more negative thoughts under ST and that the number of negative thoughts they had during the first half of the test mediated the effect of ST (on performance) during the second half of the test. Hence, a negative affect factor was created to capture negative emotions triggered under ST conditions.

Theoretically, an individual is at risk of ST if he or she strongly identifies with the stereotyped domain and exhibits moderate to high self-efficacy in the stereotyped domain. Further, for those who meet these prerequisites, individual differences on group identification (i.e., gender or race), stigma consciousness, and negative affect determine whether or not they underperform on stereotype relevant tasks when subjected to ST. Thus, individuals who exhibit high levels of group identification, stigma consciousness, and negative affect are at greater risk of experiencing ST effects than their counterparts who exhibit low levels of the same.

Validating the Social Identities and Attitudes Scale (SIAS)

Content Validity

Based on the literature, items reflecting the following ST factors: math identification, math self-concept, gender identification, gender stigma consciousness, and negative affect were created. To gauge stereotype threat in minorities, group identification and stigma consciousness factors were created for ethnicity: ethnicity stigma consciousness and ethnic identification. Thus, seven scales in total were used to create the SIAS. The items went through two rounds of reviews. The initial pool of items totaled 80, with approximately 11 items per factor. A few stigma consciousness items like "Members of the opposite sex interpret my behavior based on my gender" and "People from other ethnic groups interpret my behavior based on my ethnicity" were adapted and modified from the stigma consciousness questionnaire by Pinel (1999). All items were placed onto a content validity form, which was distributed to 10 content validators. Five of the validators were expert professors in the fields of stereotypes, gender and math, and educational psychology, and the other five comprised graduate students in educational psychology.

The content validity form contained a qualitative and quantitative section. In the quantitative section, succinct definitions of the constructs were provided to the experts who were asked to place each item under the appropriate construct that it measured. A *none of the above* response option to the category that contained the constructs was added to eliminate forced choice in response options. Content validators were asked to indicate, on a 3-point scale, how confident they were that they had matched each of the 80 items to the correct construct. The response options for this were: *not confident*, *moderately confident*, and *very confident*. A similar response scale was included for relevance where experts checked either *not relevant*, *somewhat relevant*, or *very relevant* to gauge the relevance of each item as a measure for cognition, affect, or behavior. The qualitative section sought responses to scale improvement. The validation form also included four questions pertinent to the improvement of SIAS, such as suggestions for rewording any items of interest.

Content validation results. All forms were completed correctly and were used to compute the content validity indices (McKenzie, Wood, Kotecki, Clark, & Brey, 1999) based only on the percentages in the *very relevant* category. A priori, a cutoff of 80% was decided upon as appropriate for correct item placement. Any items placed in the correct category by less than 80% of the experts were deleted from the final instrument. In addition, any items that were not considered relevant to the scale were candidates for elimination. All items were placed correctly by content validity experts and had CVIs of 0.8 and above.

Item modification. Most of the qualitative feedback from the experts addressed the issue of the potential confound of variables measuring gender identification. Even though content validators correctly placed most items under their respective categories, a sizeable number were subsequently deleted on the basis of one or more of the following issues raised by content validators: item redundancy, ambiguity, items confounding factors not measured by the instrument, gender-biased indicators, and item variability (or lack thereof).

Item redundancy. Eight items lacking variation were considered redundant and subsequently removed for that reason. Sample items included: "I want to do well in math" and "Doing well in math is important to me," "I've always done well in math" and "I get good grades in math," and "I like my ethnicity" and "I love my ethnicity." The item "My gender forms a main part of my identity" appeared twice on the content validity form and one of the duplicate stems was also eliminated.

Double-barreled items. The indicator "I worry that my behaviors will be viewed as 'girly'" was initially created as an indicator for gender stigma consciousness. A few content validators pointed out that such an item might in fact have a double meaning for males completing the instrument. That is, such an item might be misconstrued as an indicator tapping into sexual orientation behaviors and confound responses on the gender stigma consciousness scale. This item was eliminated as a potential indicator for the scale.

Gender-specific stems. There were eight items that specifically referenced females that were omitted because they opposed the goal of creating an instrument that could be administered universally (i.e., to both males and females). Examples of such indicators were: "Most boys I know interpret my actions as 'girly,'" "Being female influences how males act with me," "I'm conscious about my femininity when I interact with males," "Stereotypes that females are poor in math affect me," "It is hard for males to take females seriously," and "Being female affects how people treat me." These items were female specific and would have complicated interpretation of responses from males, ultimately hindering the generalizability of the scale across both males and females.

Ambiguous items. Items like "I work hard at math" and "I need math skills to be successful in school," designed to measure math identification, were ambiguous in that that they could potentially measure other non-ST related concepts. Content validators pointed out that working hard at math and realizing its utility for success in school could underlie other factors besides math identification: Students might work hard in mathematics because they generally work hard in all other subjects, for example. In addition, respondents might understand the importance of doing math well in school but not necessarily identify with it. For these reasons, these four items were omitted from the final scale.

Broadly defined items. Ten items were eliminated because they were too broadly defined and left too much room for interpretation. Including these stems could have subsequently increased the risk of multidimensionality in the constructs being measured. For example, items like "I define myself in terms of my gender" and "My gender is central in defining who I am" were not only too broad, but they also lacked variance. The majority of these items were also problematic along the dimensions previously discussed.

Confounding factors not measured by the scale. Six items originally designed for stigma consciousness were confounding with stereotype endorsement, a factor not assessed by the SIAS. For example, "Math is a male subject," "It is unusual to find females who excel in math," "Males have a natural ability to excel in math," "Few females are interested in math," and "For most females getting good grades in math is a challenge" are very close to stereotype endorsement. Including these items could have potentially yielded an extra factor for the stereotype endorsement in the exploratory factor analysis. Also, items related to math self-concept, such as "I enjoy math," were noted as a possible confound for math enjoyment factors that exist and have been extensively studied. In total, 37 items were omitted. The final survey consisted of 43 items, which were rated on a 7-point Likert scale with responses ranging from (1) *strongly disagree* to (7) *strongly agree*.

Method

Sample and Procedures

The SIAS was administered online to students at a large Northeastern university. A total of 206 students responded and data from these were used to conduct an exploratory factor analysis. The sample was homogeneous with respect to gender and race and constituted: 79% females, 20.3% males, 82.3% Whites, 6.5% Asians, 2.7% Blacks, 4.3% Latinos, and 8% students who listed their race as "other." The sample also consisted of 49.6% undergraduate students: 21.8% of who were freshmen, 11.2% sophomores, 11.7% juniors, and 6.9% seniors. For 30.8% of the participants, most of their coursework involved the heavy use of mathematics compared to 69.2% whose courses were not mathematics-intensive.

SIAS: Factor Analysis

Principal axis factor (PAF) analysis with oblique rotation (direct oblimin) was used to determine the structure of the SIAS. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, eigenvalues greater than one, scree plot, parallel analysis (Horn, 1965), and a MAP test (Velicer, 1976; Velicer, Eaton, & Fava, 2000) were used to determine the factor structure. Both the MAP test and parallel analysis were used to inform factor retention because (a) they are the more accurate factor extraction methods and (b) they complement one another. In the worst-case scenario, parallel analysis tends to over extract and the MAP test tends to do just the opposite. Hence, the use of both methods in factor extraction reduces the risk of over- or underextraction.

Results. The KMO of the SIAS had a value of 0.862. The initial factor analysis converged in 13 iterations and explained 74% of the variance in the items. Three factors explained the 53.8% of the variance in stereotype threat: math identification (24.67%), ethnic stigma consciousness (19.71%), and negative affect (9.39%). The eigenvalues above 1.0 reported eight factors. An inspection of the scree plot, although subjective, seemed to suggest seven factors. Velicer's MAP test suggested eight factors, and the parallel analysis suggested seven factors. All seven hypothesized factors were supported in the initial EFA. The eighth factor contained one item. This item dealt with attitudes toward one's ability to do well in school, which, in retrospect, appeared unrelated to any of the existing factors. Hence a seven-factor solution was retained.

Seven factor solution. A second factor analysis specifying seven factors and suppressing values below .33 was run. Values below .33 were suppressed to simplify the patterns of item-factor loadings (Pett, Lackey, & Sullivan, 2003). The seven-factor solution, which converged in six iterations, explained 70.4% of the variance in the items and supported the seven ST factors.

All seven factors had indicators ranked on a 7-point Likert scale with values ranging from (1) *strongly disagree* to (7) *strongly agree*. Therefore, high values on each of the factors denoted high levels of math self-concept, identification (ethnic and gender), stigma consciousness (ethnic and gender), as well as negative affect. Theoretically, White males would be expected to have low scores on all factors but math identification and math self-concept because they suffer neither gender nor ethnicity/race negative stereotypes regarding math ability and hence have a very low risk of being subjected to ST. White females susceptible to ST are expected to be high on math identification, math self-concept, gender identification, negative affect, gender stigma consciousness, but low on ethnic identification and ethnic stigma consciousness. Low values on ethnic identification and ethnic stigma consciousness are expected for White males and females because unlike their Black or Hispanic counterparts, negative stereotypes about Whites' intellectual ability are either absent (Aronson & Disko, 1998) or exist in negligible proportions that they are not accessible enough to become part of a stigmatized personality (Crocker, 1999). Hence, for the White students, ethnic identification and stigma consciousness are not expected to have an impact on mathematics performance under ST. On the other hand, minority students (specifically African Americans and Hispanics) at risk of ST would be expected to be high on all seven factors.

Factor-correlations matrix. The factor correlations matrix in Table 2 showed math self-concept and math identification to be negatively related to negative affect and gender identification. Math identification was negatively correlated to negative affect and ethnic identity. Additionally, all social identity factors (ethnicity and gender) and stigma consciousness factors related to these identities reported moderate to strong positive correlations with each other. This seemed to suggest that for this sample, which was predominantly White, individuals high on math identification were low on ethnic identification and reported experiencing less negative affect during math test taking. Theoretically and conceptually this made sense because, as discussed in the previous paragraph, the lack of negative stereotypes about Whites relative to mathematics ensures that for White students, race as a factor is not negatively linked to performance in the stereotyped domain.

Communalities. Overall, the communalities were quite high—36 of 43 items had communalities above .6 (see Table 3). The remaining seven items had communalities that ranged between .33 and .58, values still considered to be moderate (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Pattern matrix. Initially SIAS had four gender identification, seven ethnic identification, seven ethnicity stigma consciousness, five gender stigma consciousness, six math identification, seven negative affect, and seven math self-concept items. Item-factor loadings on the pattern matrix in Table 4 were moderate to strong for math identification (.62–.93), ethnicity stigma consciousness (.37–.77), negative affect (.70–.86), ethnic identification (.49–.89), gender stigma consciousness (.57-.82), gender identification (.47-.62), and math self-concept (.49-.57). The moderate to strong range of values suggested a strong relationship between the items and their respective factors after controlling for other unrelated factors and variables (Pett et al., 2003). However, there were double loadings for two math self-concept items. These items double-loaded on math identification as well, with the factor

Table 2

Factor Correlations Matrix

Factor	MI	ESC	NegAff	EI	GSC	GI	MSC
MI	1.00	.14	-.31	-.07	.12	.05	.36
ESC		1.00	.03	.36	.41	.32	.08
NegAff			1.00	.01	.01	.02	-.25
EI				1.00	.14	.24	.01
GSC					1.00	.30	.000
GI						1.00	-.02
MSC							1.00

Table 3

Communalities for the Seven-Factor Solution

1. My gender influences how I feel about myself	.48
2. Math is important to me	.80
3. My ethnicity forms a major part of my identity	.66
4. My gender contributes to my self-confidence	.57
5. My gender influences how teachers interpret my behavior	.52
6. I have always done well in math	.88
7. I value my ethnic background	.71
8. I am good at math	.92
9. Most people judge me on the basis of my ethnicity	.70
10. My gender is central to defining who I am	.59
11. Being good at math will be useful to me in my future	.80
12. My ethnicity affects how I feel about myself	.75
13. Most people judge me on the basis of my gender	.67
14. My ethnicity influences how I feel about myself	.80
15. I learn things quickly in math	.87
16. My identity is strongly tied to my gender	.60
17. I feel a strong attachment to my ethnicity	.82
18. My gender affects how people treat me	.76
19. I am unable to do well in school	.33
20. Stereotypes about my ethnic groups bother me	.44
21. I have strong math skills	.90
22. My ethnicity is an important reflection of who I am	.78
23. I am connected to my ethnic heritage	.71
24. My gender affects how people act towards me	.78

Table 3, continued

25. Most people have unexpressed racist thoughts	.47
26.My math abilities are important to my academic success	.87
27. My ethnicity affects how my peers interact with me	.73
28. I can easily master advanced math concepts	.87
29. Doing well in math matters to me	.81
30. Members of the opposite sex interpret my behavior based on my gender	.75
31. My ethnicity influences how teachers interact with me	.69
32. I value math	.83
33. My ethnicity affects how I interact with people of other ethnicities	.56
34. I am capable of excelling in math	.78
35. Doing well in math is critical to my future success	.87
36. People from other ethnic groups interpret my behavior based on my ethnicity	.68
37. Experience doubt about my math abilities	.74
38. Experience feelings of frustration	.62
39. Feel like am letting myself down	.76
40. Start to lose confidence in my abilities	.80
41. Feel like a failure	.80
42. Feel hopeless	.80
43. Feel like giving up	.69

loadings for math self-concept being only slightly higher than for math identification. Two ethnic identification items also cross-loaded across ethnicity stigma consciousness and gender identification. This is probably because there was little wording variation between one of the items in both gender identification and ethnic identification. The multidimensionality of these items made them candidates for deletion.

Structure matrix. Similarly, the structure matrix shown in Table 5 revealed moderate to strong item-factor loadings for the factors in the SIAS, indicating moderate to strong correlations between items and their respective factors as well. In addition, there was little difference between the pattern and structure matrix with respect to the item-factor loadings, which provides support for low correlations between the factors.

Items deleted. Five items (two ethnic identification, one ethnicity stigma consciousness, and three math self-concept items) were removed because they were multidimensional. In addition,

Table 4

Pattern Matrix for Seven-Factor Solution

	MID	ESC	NegAff	EI	GSC	GI	MSC
26. My math abilities are important to my academic success	.93						
35. Doing well in math is critical to my future success	.92						
11. Being good at math will be useful to me in my future	.91						
29. Doing well in math matters to me	.85						
32. I value math	.81						
2. Math is important to me	.75						
34. I am capable of excelling in math	.62						
31. My ethnicity influences how teachers interact with me		.77					
27. My ethnicity affects how my peers interact with me		.76					
9. Most people judge me on the basis of my ethnicity		.73					
36. People from other ethnic groups interpret my behavior based on my ethnicity		.54			.39		
33. My ethnicity affects how I interact with people of other ethnicities		.51					
14. My ethnicity influences how I feel about myself		.48		.35		.34	
12. My ethnicity affects how I feel about myself		.44				.40	
25. Most people have unexpressed racist thoughts		.37					
39. Feel like am letting myself down			.86				
42. Feel hopeless			.86				
41. Feel like a failure			.86				
40. Start to lose confidence in my abilities			.83				
43. Feel like giving up			.72				

Table 4, continued

	MID	ESC	NegAff	EI	GSC	GI	MSC
37. Experience doubt about my math abilities			.71				
38. Experience feelings of frustration			.70				
19. I am unable to do well in school							
17. I feel a strong attachment to my ethnicity				.89			
23. I am connected to my ethnic heritage				.83			
7. I value my ethnic background				.78			
22. My ethnicity is an important reflection of who I am				.76			
3. My ethnicity forms a major part of my identity		.40		.49			
24. My gender affects how people act towards me					.82		
30. Members of the opposite sex interpret my behavior based on my gender					.79		
18. My gender affects how people treat me					.70		
13. Most people judge me on the basis of my gender					.57		
5. My gender influences how teachers interpret my behavior							
20. Stereotypes about my ethnic groups bother me							
1. My gender influences how I feel about myself						.62	
10. My gender is central to defining who I am						.57	
4. My gender contributes to my self-confidence						.54	
16. My identity is strongly tied to my gender						.47	
6. I have always done well in math	.45						.57
15. I learn things quickly in math	.39						.56
21. I have strong math skills	.47						.55
8. I am good at math	.49						.52
28. I can easily master advanced math concepts	.47						.49

***Note*. Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization.**
a. Rotation converged in 18 iterations. Suppressed values below .33.

Table 5

Structure Matrix for the Seven-factor Solution

	MID	ESC	NegAff	EI	GSC	GI	MSC
26. My math abilities are important to my academic success	.91						.34
35. Doing well in math is critical to my future success	.89						
11. Being good at math will be useful to me in my future	.86						
32. I value math	.86						.38
29. Doing well in math matters to me	.82						
2. Math is important to me	.82						.42
34. I am capable of excelling in math	.76		-.39				.48
8. I am good at math	.75		-.49				.74
28. I can easily master advanced math concepts	.73		-.47				.72
27. My ethnicity affects how my peers interact with me		.81			.38	.34	
9. Most people judge me on the basis of my ethnicity		.77		.35	.36		
31. My ethnicity influences how teachers interact with me		.77				.33	
14. My ethnicity influences how I feel about myself		.70		.60		.57	
36. People from other ethnic groups interpret my behavior based on my ethnicity		.67			.61		
12. My ethnicity affects how I feel about myself		.65		.52		.59	
33. My ethnicity affects how I interact with people of other ethnicities		.59			.37		
25. Most people have unexpressed racist thoughts		.45			.35		
41. Feel like a failure			.85				
40. Start to lose confidence in my abilities			.85				
42. Feel hopeless			.84				

Table 5, continued

	MID	ESC	NegAff	EI	GSC	GI	MSC
39. Feel like am letting myself down			.82				
37. Experience doubt about my math abilities	-.33		.77				-.43
43. Feel like giving up	-.36		.76				
38. Experience feelings of frustration			.71				
19. I am unable to do well in school							
17. I feel a strong attachment to my ethnicity		.38		.91			
22. My ethnicity is an important reflection of who I am		.47		.84			
23. I am connected to my ethnic heritage				.80			
7. I value my ethnic background				.78			
3. My ethnicity forms a major part of my identity		.56		.64		.39	
24. My gender affects how people act towards me		.41			.87	.41	
30. Members of the opposite sex interpret my behavior based on my gender		.37			.83		
18. My gender affects how people treat me		.46			.80	.43	
13. Most people judge me on the basis of my gender		.48			.69	.38	
5. My gender influences how teachers interpret my behavior		.45			.50	.39	
20. Stereotypes about my ethnic groups bother me				.33	.34		
10. My gender is central to defining who I am		.36			.38	.65	
4. My gender contributes to my self-confidence				.37	.37	.65	
1. My gender influences how I feel about myself						.64	
16. My identity is strongly tied to my gender				.35	.45	.59	
15. I learn things quickly in math	.68		-.51				.76
6. I have always done well in math	.71		-.44				.76
21. I have strong math skills	.73		-.46				.76

Note. **Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Suppressed values below .33.**

Table 6

Reliability Estimates for SIAS—Seven-Factor Solution

Subscale	No. Items	Alpha	CI95	Mean IIC	SD
MI	6	.95	.94–.96	.75	.05
ESC	6	.85	.81–.88	.47	.12
NegAff	7	.93	.91–.94	.65	.08
EI	5	.91	.89–.93	.68	.07
GI	4	.75	.68–.80	.43	.06
GSC	5	.88	.85–.90	.59	.11
MSC	4	.96	.94–.96	.84	.03

the lone item that had originally formed an eighth factor in the initial EFA was suppressed in the subsequent run of the seven-factor solution when values less than .33 were suppressed. In total, six items were removed, leaving a total of 37 items.

There were also concerns about whether to include math self-concept as a factor in future analyses because its indicators were multidimensional. Per the pattern matrix, two math self-concept items double loaded on math identification with factor loadings higher than .6. The structure matrix, however, revealed that all but one math self-concept item double loaded on the math identification factor with loadings ranging from .50 to .70. The evidence for multidimensionality of items measuring this factor was undeniably strong. The inter-item correlations between items in the math self-concept factor were very high (i.e., .84 and above)—signaling problems with multi-collinearity and redundancy.

Multi-collinearity tends to yield empirically underidentified model solutions (Kenny, 1979) in confirmatory factor analysis (CFA). Empirically underidentified solutions are theoretically identified but yield unstable, non-unique, and potentially invalid parameter estimates (Brown, 2006). Therefore, to preserve parsimony, unidimensionality, and to minimize risks of empirical underidentification, math self-concept was excluded as a factor of SIAS prior to conducting a CFA. This decision had no substantive bearing on the strength of SIAS because theoretically, math identification is the more significant factor central in discriminating ST-susceptible individuals from those who are not. An EFA excluding this factor revealed 65% of variance in the items was explained by the six factors. Therefore, in addition to the indicators representing math self-concept, four other multidimensional items (two each) from ethnic identification and ethnic stigma consciousness were removed. The revised 33-item scale was administered to a new sample of students and data were used to conduct a confirmatory factor analysis (CFA).

Confirmatory Factor Analysis

The modified version of SIAS was completed by a separate sample of 200 undergraduate college students. This sample, a little more heterogeneous compared to that used in the EFA, consisted of 50.8% males and 49.2% females. Of this, 68.4% were White, 13.4% Black, 8% Asian, 3.2% Latino, and 7% listed as "other."

Model specification. Prior to the CFA, normality for indicators were assessed and it was determined that the indicators were multivariate normal (skewness = .50, kurtosis =. 80). A six-factor model was specified for math identification, ethnic identification, gender identification, gender stigma consciousness, ethnic stigma consciousness, and negative affect. Indicators were specified to load onto their hypothesized factors, and latent variables representing the six aforementioned factors were scaled using marker variables, which were fixed to one. Variables that best represented the factor conceptually and had strongly correlations with their factors were selected as marker variables. All other variables were allowed to estimate freely. In sum, a total of 81 parameters were estimated: 33 error variances, 27 paths, 15 correlations, and 6 latent variances. The CFA was run in AMOS 17 using Maximum Likelihood (ML).

Results. Several guidelines for model fit have been proposed: For CFAs using maximum likelihood estimation, CFI and TLI near .95 or greater have been proposed (Hu & Bentler, 1999). For RMSEA, values < .05 suggest good fit, .08 suggests adequate model fit, and values greater than .1 should be rejected (Browne & Cudel, 1993).

The model produced reported a χ^2 of 1025(480), $p < .001$, CMIN/DF = 2.14, TLI = .87, CFI = .88, and RMSEA = .08. Based on the model fit guidelines above, the RMSEA suggested adequate model fit, while the CFI and TLI values did not. The low (.10–.20) to moderate (.35–.70) factor correlations reflected the achievement of acceptable levels of discriminant validity. The stigma consciousness factors (i.e., ethnicity and gender) were moderately correlated ($r = .62$), possibly because these factors were measuring different facets of the same construct—stigma consciousness. Similarly, gender identification and gender stigma consciousness appeared to be strongly correlated ($r = .70$), suggesting that individuals who strongly identified with their gender were also highly likely to be more sensitive to negative stereotypes about their gender with respect to mathematics. There were near-zero correlations between math identification and ethnic and gender stigma consciousness factors (.00 and .02 respectively). The same was true for the correlation between negative affect and ethnic identification, and these correlations were not statistically significant ($p > .05$). Correlations between math identification and negative affect, ethnic stigma consciousness and ethnic identification, ethnic and gender stigma consciousness, ethnic stigma consciousness and gender identification, as well as gender stigma consciousness and gender identification were all statistically significant at the $p < .001$ level.

Results also showed a statistically significant low-moderate, negative relationship between negative affect and math identification ($r = .38$) and a positive and relatively low-moderate relationship between ethnic stigma consciousness and gender identification ($r = .35$). Ethnic stigma consciousness and ethnic identification reported moderately positive correlations to each other, and the relationship between gender identification and gender stigma consciousness was relatively strong. Other factor correlations were not statistically significant ($p > .05$).

Table 7

SIAS Subscale Bivariate Correlations (CFA)

Factor	MI	ESC	NegAff	EI	GSC	GI
MI	1.00	.00	-.38***	.12	-.01	.02
ESC		1.00	.14	.52***	.62***	.35
NegAff			1.00	-.04	.17	.14
EI				1.00	.11	.20
GSC					1.00	.70***
GI						1.00

Note. *** $p < .001$.

Model Evaluation and Re-Specification

Localized areas of ill fit were examined using modification indices and standardized residuals. High values on a number of modification indices and high, standardized residuals suggested areas of misfit in the model, which were examined accordingly. Four items, one each from negative affect, ethnic identification, ethnic stigma consciousness, and gender identification, had high standardized residuals ranging from 2.30–3.40. Instead of eliminating all four items at once, two items, "I experience feelings of frustration during math test taking" (negative affect) and "Most people have unexpressed racist thoughts" (ethnic stigma consciousness) were first removed before rerunning the model. Both items were eliminated because of high standardized residuals with numerous items (i.e., 3.0 and 2.5 and above for negative affect and ethnic stigma consciousness, respectively). Additionally, the ethnic stigma consciousness item was also eliminated because it had a low communality (.08). Reanalysis of the model after deleting these items indicated that the issues of ill fit presented by the ethnicity identification item denoted earlier still persisted. With standardized residuals still above 2.50, "My ethnicity forms a major part of my identity" was also removed from the model. An examination of modification indices also suggested that the correlation of errors between pairs of negative affect (39 and 40), math identification (2 and 32), and gender stigma consciousness

(18 and 24) items would improve model fit. These errors were correlated on the basis of wording similarity (Brown, 2006). The final model eliminated a total of three items and three correlated error variances.

The revised model generated a χ^2 of 672.5(387), $p < .001$, CFI = .93, TLI = .93, RMSEA = .06. The results suggested adequate model fit to the data. Factor loadings remained high: math identification (.81–.91), ethnic stigma consciousness (.63–.83), negative affect (.63–.96), ethnic identification (.73–.91), gender stigma consciousness (.73–.83), and gender identification (.62–.81)—indicative of the psychometric properties of the indicators as both strong and stable across samples.

Reliability Analysis

A priori, a desirable alpha level of 0.8 was determined as acceptable for the subscales (Netemeyer, Bearden, & Sharma, 2003). Reliability estimates for the SIAS factors ranged from .81 to .95, and bivariate subscale correlations provided support for convergent and discriminant validity: Math identification was negatively correlated to feelings of negative affect experienced during math test taking. All social identities (ethnicity, gender) and stigma consciousness related to these identities reported statistically significant moderate to strong positive correlations with each other (see Table 8).

Discussion

Earlier, we pointed out the need for future research to expand its focus to examining ST as a phenomenon that occurs in degrees. We also argued that although this new perspective was necessary for the advancement of ST research in general, efforts to do so would require an integrated ST measure. The SIAS is the first integrated measure of ST, incorporating key moderators. Based on the EFA, the six factor SIAS explains 65% of the variance in the items measuring ST. The subscales have also demonstrated

Table 8

Reliability Estimates for SIAS (Post CFA)

Subscale	No. Items	Alpha	CI95	Mean IIC	IIC SD
MI	6	.95	.94–.96	.77	.14
ESC	5	.85	.82–.88	.54	.07
NegAff	6	.93	.91–.94	.69	.10
EI	4	.89	.86–.91	.67	.06
GI	4	.81	.76–.85	.51	.06
GSC	5	.88	.85–.91	.60	.08

high levels of reliability and discriminant and convergent validity. The CFA demonstrates adequate model fit to the data and the stability of the instrument and psychometric strength of the indicator variables in the instrument. The use of SIAS has the potential to help future research fill in the extant ST literature.

Examining Role of Context in ST in Authentic Educational Settings

ST is highly contextual. The test performance of individuals exposed to ST environments varies as a function of individual differences on the factors that make up the SIAS. Although preliminary evidence provides support for situational factors like numerical representation (Inzlicht & Ben-Zeev, 2003; Murphy, Steele, & Gross, 2007) as contextual moderators of ST, these inferences have been based on experimental or laboratory studies on ST. To date, few researchers have examined the contextual aspect of ST. Specifically, the SIAS makes it possible to investigate inquiry into the role of context in exacerbating ST in authentic learning environments. Its use as a tool for identifying highly ST-susceptible individuals may facilitate qualitative research that sheds more light on contextual factors that exacerbate ST for these individuals and, importantly, how they respond to these environments. Such data would be useful in examining school context and designing interventions that foster optimal (less threatening) learning environments that mitigate ST.

Structuring Focused Interventions

A number of strategies such as affirming domain belongingness (Steele, 1997), exposure to role models who have been successful in the domain (Aronson, 2004), and cognitive reappraisal (Schmader, 2010) have been proposed to minimize ST among highly identified individuals. Studies also show that teaching ST-susceptible students about the phenomenon and making them aware of its deleterious effects can be effective in greatly reducing ST effects (Aronson & Williams, 2004; Johns, Schmader, & Martens, 2005). More broadly, the SIAS may guide efforts to develop and curtail more effective interventions for ST individuals in STEM at different ST risk levels based on the aforementioned ideas. Specifically, differentiating between levels of ST risk (e.g., low, moderate, high) would make the SIAS useful in creating specialized interventions for females at different risk levels. Such a measure would also be crucial in measuring the effectiveness of interventions targeted at different ST risk groups.

Limitations and Recommendations for Future Research

We recognize that no single instrument can capture every construct related to a psychological phenomenon at any given time. The SIAS is not any different because it does not include all ST moderators. Although the SIAS constitutes key variables identified as moderators in ST research, other equally relevant but underinvestigated ST moderators like expectancy performance (Cadinu et al., 2003), locus of control (Cadinu, Maas, Rosabianca, Lombardo, & Figerio, 2006), self-affirmation (Martens, Johns, Greenberg, & Schimel, 2006), role models (Marx & Roman, 2002), and intergroup similarity (Rosenthal & Crisp, 2006) are not measured by the instrument. Researchers are therefore cautioned against construing the SIAS as a complete ST measure.

Secondly, the samples used to validate the SIAS were drawn from a population of college students. Because the SIAS has only been validated once, and with samples from one population, issues of external validity still abound. It is recommended that the SIAS undergo iterative validation processes with samples from different populations, as this would not only be useful in evaluating external validity of the scale, but also helpful in studying ST effects and the academic disengagement of students belonging to marginalized groups.

Thus far, the SIAS is limited in its application to ST groups in mathematics because it was developed to evaluate ST in STEM. As such, domain identification is represented by math identification in the SIAS. Nevertheless, the instrument is adaptable and can be used to capture ST across different domains. It is recommended that researchers wishing to use this instrument in other domains adapt the items assessing math identification to their domain of choice.

The development and validation of the SIAS is a good first step in a line of research that aspires to address the gender gap in STEM by targeting interventions for individuals at risk of disengaging from these domains because of ST. However, much more work remains on the SIAS before it can be used for these purposes. The next step in the iterative development of SIAS is to score it along a continuum and use it to classify students who are at medium to high risk of being susceptible to ST. By assessing whether ST moderators have interactive effects, and if these are significantly related to degrees of underperformance of ST females, researchers will be in a better position to evaluate whether differential effects exist for White versus Black females students in math and other STEM domains, for instance. An integrated ST measure like SIAS opens the avenue for researchers to further develop ST theory and bring us closer to fully understanding the extent to which she is affected when Barbie says, "Math class is tough!"

References

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12,* 385–390.

Aronson, J. (2004). The threat of stereotype. *Educational Leadership, 62*(3), 14–19.

Aronson, J., & Disko, D. (1998). *Frequency of stereotypes about Blacks, Whites, Asians and Latinos.* Unpublished raw data, University of Texas at Austin.

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

Aronson, J., & Williams, J. (2004). *Stereotype threat: Forewarned is forearmed*. Unpublished manuscript, New York University.

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences, 7,* 363–379.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4,* 359–373.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation and spillover. *Journal of Experimental Psychology, 136,* 256–276.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York, NY: The Guilford Press.

Brown, R. P., & Josephs, R. A. (1999). *The importance of importance: The Mathematics Identification Questionnaire*. Unpublished manuscript, University of Texas at Austin.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39,* 626–633.

Browne, M. W., & Cudel, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cadinu, M., Maas, A., Rosabianca, A., Figerio, S., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy performance. *European Journal of Social Psychology, 33,* 267–285.

Cadinu, M., Maas, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16,* 572–578.

Cadinu, M., Maas, A., Rosabianca, A., Lombardo, M., & Figerio, S. (2006). Stereotype threat: The moderating role of locus of control beliefs. *European Journal of Social Psychology, 36,* 183–197.

Crocker, J. (1999). Social stigma and self-esteem: Situational construction of self-worth. *Journal of Experimental Social Psychology, 35,* 89–107.

Eriksson, K., & Lindholm, T. (2007). Making gender matter: The role of gender-based expectancies and gender identification on women's and men's math performance in Sweden. *Scandinavian Journal of Psychology, 48,* 329–338.

Fabrigar, R. L., Wegener, T. D., MacCallum, C. R., & Strahan, J. E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Gonzales, P. M., Blanton, H., & Williams, K. J. (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.

Gray, J. A. (1981). A biological basis for the sex differences in achievement in science? In A. Kelly (Ed.), *The missing half: Girls and science education* (pp. 43–58). Manchester, England: Butler & Tanner Ltd.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few?: Women in science, technology, engineering and mathematics.* Washington, DC: American Association of University Women.

Hoffman, R. M. (2006). Gender self-definition and gender self-acceptance in women: Interception with feminist, womanist and ethnic identities. *Journal of Counseling and Development, 84,* 358–372.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32,* 179–185.

Hoyt, C. (2005). The role of leadership efficacy and stereotype activation in women's identification with leadership. *Journal of Leadership and Organizational Studies, 11,* 2–15.

Huguet, P., & Regner, I. (2007). Stereotype threat among school girls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99,* 545–560.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95,* 796–805.

Jacobs, J. E. (1991). The influence of gender stereotypes on parent and child mathematics attitudes. *Journal of Educational Psychology, 83,* 518–527.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and childrens' ability perceptions. *Journal of Personality and Social Psychology, 63,* 932–944.

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.

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

Keller, J., & Molix, L. (2008). When women can't do math: The interplay of self-construal, gender identification, and stereotypic performance standards. *Journal of Experimental Social Psychology, 44,* 437–444.

Kenny, D. A. (1979). *Correlation and causality.* New York, NY: Wiley-Interscience.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes. *Psychological Science, 18,* 13–18.

Kray, L. J., Galinsky, A., & Thompson, L. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology, 80,* 942–958.

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.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Orland, FL: Academic Press.

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21,* 153–174.

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

Marx, D. M., & Stapel, D. A. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology, 88,* 432–446.

McKenzie, J. F., Wood, L. M., Kotecki, E. J., Clark, K. J., & Brey, A. R. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health and Behavior, 23,* 311–318.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science and engineering settings. *Psychological Science, 18,* 879–885.

National Center for Education Statistics. (2007). *The condition of education, 2007.* Retrieved from http://nces.ed.gov/pubs2007/2007064.pdf

National Science Foundation. (2009). *Women, minorities, and persons with disabilities in science and engineering: 2009* (NSF 09-305). Arlington, VA: Hill.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications.* London, England: Sage.

Pett, M. A., Lackey, R. N., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in healthcare research.* London, England: Sage.

Phinney, S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research, 7,* 156–176.

Picho, K., & Stephens, J. M. (in press). Culture, context and stereotype threat: A comparative analysis of Ugandan females in coed and single sex schools. *Journal of Educational Research.*

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76,* 114–128.

Pinel, E. C. (2004). You're just saying that because I'm a woman: Stigma consciousness and attributions to discrimination. *Self and Identity, 3,* 39–51.

Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by reducing intergroup boundaries. *Personality and Social Psychology Bulletin, 32,* 501–511.

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

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science, 19*, 14–18.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 88*, 934–947.

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

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

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75,* 1191–1197.

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

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 68,* 797–811.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge, England: Cambridge University Press.

Terlecki, M. S., Newcombe, N., & Little, M. (2007). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22,* 996–1013.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 41–71). Boston, MA: Kluwer.