INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

SHEDDING LIGHT IN THE BLACK BOX OF STEREOTYPE THREAT: THE ROLE OF EMOTION.

by

DAVE I. COTTING

A dissertation submitted to the Graduate Faculty in Social-Personality Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2003

UMI Number: 3074642

UMI Microform 3074642

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/25/03 — Date

Chair of Examining Committee

1/27/03 — Date

Executive Officer

William E. Cross, Jr.

Joe Glick

Vita Rabinowitz

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

SHEDDING LIGHT IN THE BLACK BOX OF STEREOTYPE THREAT:

THE ROLE OF EMOTION

by

Dave I. Cotting

Adviser: Professor William E. Cross, Jr.

In this study, stereotype threat, a situational phenomenon that impairs the performance on ability or achievement tasks of members of stereotyped groups, is explored. To further our understanding of the phenomenon, an experimental manipulation of stereotype threat in students at an All-women College or University (AWCU) and at a Historically Black College or University (HBCU) was employed for the purpose of isolating the specific emotional structure and dynamics of the stereotype threat phenomenon. The paper begins with reviews of two frameworks that inform the current research: the literature on stereotype threat and the other on appraisal theories of emotion and emotion intensity. It was anticipated that a focus on emotions might reveal the hidden processes inside the "black box" of stereotype threat. The study did not find significant results and hence did not shed light on the black box of stereotype threat. However, post-hoc analyses suggested a causal model; stereotype threat causes changes in several affective dimensions, including sense of "power" (i.e., an emotion appraisal), irritation, dislike, and interest, which, in turn, cause impairment in performance. Future research should explore further this and other causal models that integrate stereotype threat theory and theories of emotion.

Acknowledgments

Because no one remembers anyone in my family graduating from high school, writing a dissertation may be considered a personal achievement. But the achievement must be shared, for this work would not have been possible without the labor of 3 women.

Maman: Mes succès font honneur à ta force de caractère et ton sens du sacrifice. Ton courage m'a toujours accompagné dans tout ce que j'ai entrepris. Parmis tant d'autres choses, tu m'as appris à ne jamais baisser la tête dans la poursuite de mes aspirations. Ce doctorat est le fruit de ton example et ton enseignement.

Mamy: Ton encouragement, ta douceur, et ta générosité envers moi ont toujours été sans limites. De mes premieres lectures sous forme de bandes dessinées jusqu'à la production de cette thèse, tu m'as toujours appuyé. Ce doctorat est le fruit de ton soutien.

Chandra: The rigor of your intellect, your unfailing support, and most importantly your love were instrumental at each step of this project. I am grateful for the multitude of experiences, conversations, and debates we have had together and the many more to come. You inspire me to be a better human being and a better researcher. This doctorate is the fruit of your contribution to every aspect of my life.

To Vernell and Lonnie, you have my most sincere gratitude for extending your love for your daughter to me. Also, many thanks for putting a roof over our heads for much longer than was originally planned and being most gracious about it.

I also want to extend my humble thanks in no particular order to:

Sean and Suzanne for their friendship and scholarship, and for their guidance through the meanderings of academic politics.

My cohort, Susan, Kenny, Anne, Rosemarie, Sarah, and Zay, for helping me get through my first two years in S/P.

My advisors, Klaus R. Scherer, Kay Deaux, Vita Rabinowitz, and William E. Cross, Jr., for your support and patient mentoring.

My dissertation committee, Vita Rabinowitz, Joseph Glick, Suzanne Ouellette, and Herb Saltzstein, for your scholarship.

The Chair of my dissertation committee, William E. Cross, Jr., for directing me over the last hurdle of graduate school.

Lastly, I do not believe that there is any reason to attempt such a feat of idiocy as the doctorate, other than the fact that some people, which is to mean some people like me, have a need to challenge themselves academically for self-definition. It's often a contest in purposeless suffering. But for reasons of my own, I had to have it.

Table of Contents

Table of Contents (continued)

List of Tables

List of Figures

I am the son of a Kabyle[1] father and a Swiss mother. On most days in the U.S., I pass for a dark-skinned White man. However, my multi-ethnic background has not always gone unquestioned. Growing up in Switzerland, I first was made aware of being different by children in school who often asked from where I came. In an attempt to distance myself from an abusive father repeatedly convicted of felonies, I always answered that I was Swiss, an answer that almost always was met with surprise and disbelief. The perception that I was different unfortunately was shared also by adults. For example, one day after school, I was stopped by the police while riding my bike home. As I stood on the side of the road and watched many lighter-skinned school children pass by on their bikes, I was requested to prove that the bike I was on was mine. Unable to do so in the same way as I would have been unable to prove that the clothes I was wearing were mine, I was given ten days to provide papers supporting my claim and pay a fine of 180 Swiss Francs (about 120 U.S. Dollars); I had been stereotyped as dishonest and thieving, two stereotypes about Arabs, because of the amount of melanin in my skin. Like W.E.B. Du Bois (1903) and many others before me, experiences such as these have helped me comprehend what it means to see oneself revealed through the eyes of others; it has affected how I feel, think, and behave today. Every time I go home to Switzerland, and more often in the U.S. since the events of September 11th 2001, the threat of being stereotyped hangs over my head, ready to strike.

The idea that stereotypes are detrimental to members of stereotyped groups, regardless of whether they are followed by overt discrimination, is at the center of the

stereotype threat paradigm (Steele, 1997; Steele & Aronson, 1995). Steele and Aronson (1995) have used this paradigm to describe the extra pressure that impairs members of stereotyped groups in stereotyped domains. To further our understanding of the phenomenon, an experimental manipulation of stereotype threat was employed for the purpose of isolating the specific emotional structure and dynamics of the stereotype threat phenomenon. I begin with a description of the theory of stereotype threat, a review of its empirical support, and a highlight of some of its limitations. A review of appraisal theories of emotion and emotion intensity follows. This literature, heretofore absent from stereotype threat research, provides information on processes that may take place inside the "black box" of stereotype threat.

A Psychosocial Explanation: Stereotype Threat

Stereotype threat: A theory

Stereotype threat is a situational phenomenon that occurs when members of groups about which negative stereotypes exist are reminded of the possibility of confirming these negative stereotypes (Aronson, Lustina, Good, Keough, Steele, & Brown, 1999; Aronson, Quinn, & Spencer, 1998; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995; Steele, 1997). According to the theory, stereotype threat creates anxiety or discomfort, which, in turn, may interfere with performance in the stereotyped domain. Hence, a negative stereotype creates a situation in which behavior from a member of a negatively stereotyped group risks being interpreted as confirming the stereotype (Aronson, Quinn, & Spencer, 1998; Goffman, 1963; Steele, 1997). Stereotype threat not only has a short-term impact on performance in the stereotyped domain, it also

has a long-term impact, in that members of the stereotyped group may eventually disidentify with the domain (Steele, 1997). Disidentification with the stereotyped domain is a strategy to prevent the domain from being used as a basis for self-esteem and self-evaluation. However, this self-protective strategy reduces the chance of future success in the domain by diminishing the individual's interest in the domain. Because stereotype threat is not related to any individual or group deficiency, it is not a phenomenon that follows a deficit model.

Stereotype threat has several characteristics. First, stereotype threat stems from a situational pressure (i.e., a threat in the air) rather than an internalized doubt about one's own ability in a domain or any other deficit; therefore, the response to this situation is a state and not a trait (Aronson, Quinn, & Spencer, 1998; Spencer, Steele, & Quinn, 1997; Steele, 1997; Steele & Aronson, 1995). Consequently, stereotype threat can be elicited in various populations, such as women (Brown & Josephs, 1999; Spencer, Steele, & Quinn, 1999), Blacks (Broadnax, Crocker, & Spencer, 1997; Steele & Aronson, 1995, 1998), Whites (Aronson, Lustina, Keough, Brown, & Steele, 1997; Stone, Sjomeling, Lynch, & Darley, 1999), Latinos/as (Aronson & Salinas, 1997), and students of low socioeconomic status (Croizet & Claire, 1998). Stereotype threat also can occur in different domains, such as verbal performance (Aronson & Salinas, 1997; Steele & Aronson, 1995, 1998), math performance (Brown & Josephs, 1999; Spencer, Steele, & Quinn, 1999), and athletic performance (Stone, Sjomeling, Lynch, & Darley, 1999). In this sense, the theory of stereotype threat differs from theories that suggest that members of negatively stereotyped groups internalize stereotypes (Clark & Clark, 1947; Howard & Hammond, 1985).

Second, stereotype threat affects performance only in domains for which negative stereotypes exist for some groups (e.g., academic domains for Blacks, math for women) (Inzlicht & Ben-Zeev, 2000). In this sense, it differs from Tokenism theory (Lord & Saenz, 1985; Saenz, 1994; Saenz & Lord, 1989), which suggests that a minority or token status can cause cognitive deficits in all domains. Saenz (1994) argued that such status increases self-consciousness and pressure to favorably represent one's group in any domain, thereby interfering with specific cognitive tasks (e.g., problem solving, memory).

A third characteristic of stereotype threat is the precondition that the negatively stereotyped domain must be considered self-relevant for a threat to be perceived. Therefore, to be threatening, a stereotype must be associated with a domain that is significant to one's self-concept. In other words, only those who are domain-identified and who care about the domain will be affected by stereotype threat (Steele, 1997). In terms of the specific topic of this paper, if sustained academic success requires identification with academic domains, then high academic achievers, whether female (in scientific domains) and/or Black or Latino/a (in any academic domain), should be more likely to be impaired by stereotype threat(s) than lower academic achievers (Steele, 1997), as low academic success does not require identification with academic domains. This hypothesis is congruent with other findings in psychology. For instance, Major and Schmader (1998) found that people who were psychologically engaged in a domain used social feedback about their performance in that domain as an important basis for self-evaluation. In other words, because what others think of one's performance may depend upon their expectations, beliefs, and/or stereotypes, a domain for which a stereotype

exists can be threatening to members of that stereotyped group under specific circumstances. Furthermore, threats to self-concept can be particularly debilitating for cognitive tasks. Pittner and Houston (1980), for instance, documented that threats to self-concept can have a negative impact on cognitive performance that was as strong as the impact of threats to physical well-being.

In past research, stereotype threat has been elicited by reminding participants of their membership in a stereotyped group (e.g., Steele and Aronson, 1995, study 4), by directly presenting the stereotype[2] (e.g., Aronson et al., 1998; Spencer et al., 1999), and by describing a test as diagnostic of ability (Steele & Aronson, 1995, study 1, 2, and 3; Steele, 1997). For example, when a test is described as diagnostic of intellectual ability, it makes "the racial stereotype about intellectual ability relevant to the performance of Black participants and [establishes] the threat that they might be seen stereotypically, or might fulfill the stereotype" (Steele & Aronson, 1998). According to Steele (personal communication, November 29, 2001), the "diagnosticity of the test" manipulation is the "cleanest," "surest," and "most portable into other contexts."

However, a review of the way stereotype threat has been elicited in past research would not be complete without the mention of an interesting new approach. This new approach, used by Inzlicht and Ben-Zeev (2000), demonstrated that stereotype threat could be elicited by any means used to increase the salience of stereotype. Building from Deaux and Major's (1987) interactive model of gender-related behavior, they found that an environment such as being outnumbered by members of the opposite sex while taking a math test elicited negative stereotypes in women.[3] Inzlicht and Ben-Zeev (2000) identified an environment that activates the threatening effects of stereotypes related to

academic performance as a "threatening intellectual environment" (p. 4).

In sum, stereotype threat stems from a situational pressure that can be reproduced in experiments by making salient a negative stereotype about one's group in a domain that is significant for one's self-concept. According to the theory, stereotype threat specifically refers to the acute anxiety experienced when members of groups about which negative stereotypes exist are reminded of the possibility of confirming these negative stereotypes (Aronson et al., 1999; Spencer et al., 1999; Steele & Aronson, 1995; Steele, 1997); acute anxiety, in turn, interferes with performance and identification in the stereotyped domain.

Stereotype threat: Empirical support

Empirical work on stereotype threat first focused on the stereotype that Blacks have an inferior intellectual ability. Steele and Aronson (1995) randomly assigned Black and White students to either a diagnostic, non-diagnostic only or non-diagnostic challenge condition. Participants in the first condition were asked to take a difficult verbal test described as diagnostic of intellectual ability, making the negative stereotype about Blacks' intellectual ability relevant to the task and establishing the stereotype threat. Participants in the non-diagnostic condition were asked to take a difficult verbal test not intended to evaluate ability but to help researchers understand psychological processes involved in solving verbal problems. In the third condition, non-diagnostic challenge, participants received the same description as the second condition, along with the information that the test was an attempt to provide "even highly verbal people with a mental challenge." Hence, a 2 (race: Black or White) X 3 (test description: diagnostic,

non-diagnostic only, and non-diagnostic challenge) analysis of covariance (ANCOVA) design with Scholastic Aptitude Test-Verbal (SAT-V) scores as a covariate was conducted. The results indicated that after controlling for SAT-V scores, Blacks' performance was significantly lower than that of Whites when the test was presented as one of intellectual ability (i.e., diagnostic condition), but matched that of Whites when the test was presented as non-diagnostic of intellectual ability (i.e., non-diagnostic only condition).

The results from Steele and Aronson's (1995) original study did not establish the activation of racial stereotypes, doubts about one's own abilities, or displays of disassociation from racial stereotypes in stereotype threat situations. Steele and Aronson (1995, study 3) tested these hypotheses in a follow-up study. Using a 2 (race: Black or White) X 3 (test description: diagnostic, non-diagnostic and control[4]) design, Steele and Aronson (1995, study 3) asked participants to complete a battery of measures before supposedly taking the verbal test.[5] These measures assessed the activation of racial stereotype, the activation of concerns about one's own ability, and the tendency to avoid racially stereotypic preferences.

The results revealed that Blacks in the diagnostic condition displayed significantly greater activation of racial stereotypes, more self-doubts, and a "greater tendency to avoid racially stereotypic preferences" (Steele & Aronson, 1998, p. 418) than Blacks in the other conditions and whites in all conditions. In addition, the proportion of Black participants who were willing to report their race on the questionnaire rose from 25% in the diagnostic condition to 100% in the other conditions.

Theoretically, stereotype threat is a phenomenon that occurs when the task is

challenging. With regard to women's stereotype threat in math, for instance, a test would need to be difficult enough to frustrate the skills and abilities of math identified students to induce a sense of threat (Aronson et al., 1998). This statement is congruent with the finding that unless the test is challenging, men do not outperform women in math (Benbow & Stanley, 1983; Hyde, Fennema, & Lamon, 1990; Spencer et al., 1999). Spencer et al. (1999, study 2) tested this hypothesis with strongly math-identified women and men. They randomly assigned participants to either a gender differences condition or a no gender differences condition. Participants in the first condition were told that the test they were about to take had shown gender differences in the past; this manipulation was expected to elicit a stereotype threat. In the second condition, participants were told that the test had never shown gender differences in the past; this manipulation was expected to eliminate (or at least reduce) the stereotype threat.

Participants in both conditions then were asked to take a challenging math test.[6] The results confirmed the hypothesis: women's performance in the gender differences (i.e., stereotype threat) condition was significantly lower than that of men, but matched that of men in the no gender differences (i.e., no stereotype threat) condition.

An empirical question that has been raised about stereotype threat is one of generalizability; is stereotype threat a phenomenon of the elite? Spencer et al.'s (1999, study 3) study provided an answer to this query. Participants were recruited from a non-selective university and were included on the basis of moderately selective criteria.[7] In addition, the test items came from a less advanced math test[8] than in study 2. As in a previous study (Spencer et al., 1999, study 2), women and men participants were randomly assigned to either a gender differences condition or a no gender differences

condition and were asked to take a math test. The results replicated the pattern found in more selective institutions: women's performance in the stereotype threat condition was significantly lower than that of men, but matched that of men when the stereotype threat was removed.

The previous study seems to indicate that some factor other than attending a selective institution[9] is necessary to experience academically related stereotype threat. Aronson et al. (1999, study 2) explored this question by testing the hypothesis that identification with the domain plays a role. Eliciting stereotype threat by directly presenting the stereotype, they compared the susceptibility to stereotype threat of participants who highly identified with math with that of participants who identified less. Results revealed that the performance of participants who highly identified with math was significantly lower in the stereotype threat condition than in the non-stereotype threat condition. Participants who identified less with math, on the other hand, performed better in the stereotype threat condition than in the non-stereotype threat condition, a finding that seems to provide support for the hypothesis that identification with the domain is necessary to experience stereotype threat. Yet, one important conceptual problem should be noted. Aronson et al. (1999, study 2) classified participants as highly identified with math when they indicated that they cared a great deal about being good at math, and participants who identified less or only moderately as those who cared less or moderately about being good in math. However, a student may care about being good in math but not identify with math (e.g., a pre-med student needing a high grade in math to be competitive for medical school). Thus, Aronson et al.'s study provided empirical support for the idea that caring about being good in the domain, rather than identification with the

domain, seems to be necessary to experience stereotype threat.[10]

The previous study constituted a follow-up to a study investigating whether a negative stereotype needed to be prevalent or frequently activated in order to generate a threat (Aronson et al., 1999, study 1). More specifically, the original study examined whether the stereotype about the superiority of Asian Americans in math could generate a threat for White male students who both identified with math (i.e., who cared about being good in math) and showed proficiency in math. Participants were randomly assigned to either a stereotype threat condition or a non-stereotype threat condition: in the first condition, participants were told that the purpose of the test was to evaluate their math ability and to understand why Asian Americans often outperform Whites on math tests; in the second condition, participants were told only that their math ability would be evaluated.

The results indicated that stereotype-threatened White males performed significantly lower on the math test than nonstereotype-threatened White males. In addition, this was the case despite the fact that participants in the stereotype threat condition reported making a significantly greater effort than those in the non-stereotype threat condition.[11] These findings suggest that chronically stigmatizing conditions or minority status are not necessary for stereotype threat to impair performance. Instead, being reminded of the possibility of confirming negative stereotypes about one's group is sufficient to cause impairment, regardless of the prevalence of the stereotype threat.

The causal link between stereotype threat-induced anxiety and performance

As noted in the last section, stereotype threat research has focused on the

identification of factors that make it possible to induce stereotype threat in lab settings and field experiments, and demonstration of the broad range of groups susceptible to a stereotype threat manipulation, inclusive of white men, white women, and Asian Americans. Now we turn to a handful of studies designed to isolate what is causing diminished performance in the stereotype threat condition. In a sense, these studies are trying to open what might be called the black box of stereotype threat. Much empirical evidence supports the effect of anxiety on test performance and suggests that anxiety in test-taking situations divert attention away from tests[12] (Baumeister & Showers, 1986; Le Gall, 1980; Lord & Saenz, 1985; Wine, 1971). Steele and Aronson (1995, study 2) followed up their original study to explore the possible role of anxiety in stereotype threat. Using nearly the same race by diagnosticity design (i.e., the non-diagnostic challenge condition was omitted) and a similar test[13] as in the original study, Steele and Aronson (1995, study 2) asked participants to complete a measure of anxiety induced by evaluation apprehension,[14] a measure of cognitive interference (Sarason, 1980), and to answer some questions about effort.

The data did not yield all the expected results. Although the 2 (race: Black or White) X 2 (test description: diagnostic or non-diagnostic) ANCOVA design (with scholastic SAT-V scores as a covariate) yielded a significant effect of race and a significant interaction effect of race by test description, no race or test description effects were found on the measures of test anxiety, cognitive interference, or effort. These results showed that the effects of stereotype threat were not mediated by anxiety, as stipulated by stereotype threat theory. This finding is supported by another study on anxiety and stereotype threat conducted by Spencer, Steele, and Quinn (1999). Their data showed

that, although anxiety was related to participants' performance, the stereotype threat manipulation did not predict levels of anxiety. In other words, anxiety appears to be a predictor of *performance* but not a mediator of stereotype threat.

In his article in the *Atlantic Monthly* (1999) as well as during a personal interview with the author of this study that followed Steele's recent lecture on stereotype threat at the Graduate Center-CUNY, it was conveyed that neither (a) observation reports on the behavior of participants in stereotype threat conditions nor (b) various self-reports provided by participants at the end of stereotype threat experiments, provide insight into participant affect states and experiences linked to stereotype threat. True, it has been determined that, during stereotype threat conditions, participants often report spending more time on task and being extra "careful" not to make mistakes [e.g., activities that slow and diminish performance], but, as to affect and anxiety, many participants state that they felt no different during the stereotype threat manipulation than they did in other test taking experiences.

Progress has been made by shifting attention from self-report and third-party observation data to physiological findings. In what might be the first reliable peek into the black box, Blascovich, Spencer, Quinn, & Steele (2001) report that Black participants exposed to stereotype threat showed a larger increase in *blood pressure* than both White and Black participants exposed to little or no stereotype threat. In effect, their data confirm a *physiological reactivity* on the part of participants in the stereotype threat condition, independent of whether the participants showed conscious awareness of their reactivity.

While the blood pressure study lays a general or global foundation for what might

be in the stereotype threat black box, there remains the question of whether it is possible to isolate a more refined picture of what is the emotional response to stereotype threat. The current study will try to extend the stereotype threat discourse through a more explicit exploration of the role of emotions in stereotype threat conditions.

Emotions: Flashlights in the stereotype threat black box?

Emotional experience

Although absent from the current stereotype threat literature, emotional reactions other than anxiety may be involved in stereotype threat. However, the decision of which framework to adopt and which emotions to investigate first is a challenging endeavor as the choice is vast and the risks to cast off the wrong framework and/or emotions are many. In regard to theories of emotion, Roseman and Smith (2001) explained that two-dimensional models have supplanted one-dimensional theories of emotion (i.e.. a conception of emotionality that corresponds to the degree of energy or activity in a behavior) inspired by behavioral theories that were dominant between the 1930's and 1950's. This suggests that one such two-dimensional model, like that of Watson and Tellegen (1985), which conceptualizes positive and negative affects as the primary dimensions of emotion, may need to be included in an exploration of the role of emotional experience in stereotype threat. However, some findings in emotion research cannot be accounted for by such one- or two-dimensional models. Universal facial expressions, characteristic action tendencies (i.e., approach, inaction, avoidance, and attack), and specific emotion appraisals, for example, support the conception that there are several distinct (or discrete) emotions. Among these emotions, anger, fear, and

sadness/disappointment appear to be the strongest candidates for stereotype threat; these emotions are mentioned often in the research on sports performance (e.g., Gould et al., 1987; Gould & Udry, 1994) and mental test performance (e.g., Baumeister & Showers, 1986; Le Gall, 1992). Another emotional response of interest for stereotype threat may be frustration. Frustration is experienced when a goal-directed behavior is blocked (Scherer, 2001). Stereotype threat may present such an obstruction. Yet, one potential limitation of the discrete emotion approach in regard to stereotype threat is that an emotion must be specified in order for the effect to be observed, in other words an emotion may be experienced, but if it is not tapped by the investigator, the emotion will go undetected.

The role of appraisals in theories of emotion

As certain emotional experiences have been overlooked in the literature on stereotype threat, so has the appraisal approach to studying emotions. A great amount of research on emotions in the last two decades has focused on its cognitive antecedents (i.e., appraisals). Informed by this research, an appraisal approach may provide an alternate view and a better understanding of the role of emotion in stereotype threat situations.

Emotion appraisal research originated in the work of Arnold (1960a, 1960b): she first suggested that emotions require thoughts or, more specifically, appraisals: cognitive processes that intervene between the perception of an event, its relevance for the individual, and the emotion. A number of theories born from this tradition (Frijda, 1986; Lazarus, 1991; Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1985), appraisal theories of emotion, claim that emotion is caused by cognitive, on-going processes of

evaluation of a stimulus event. Some of the basic assumptions of these approaches are described below, as they may contribute to a better understanding of the dimensions and determinants of emotion.

According to appraisal theories, an emotion results from an on-going evaluative process (i.e., appraisal) of a stimulus event (Arnold, 1960a). This process involves an evaluation of a situation in terms of one's concerns. As Arnold (1960a) pointed out, "To arouse an emotion, the object must be appraised as affecting me in some way, affecting me personally as an individual with my particular experience and my particular aims" (p. 171). Appraisals do not operate in the perceiver's conscious awareness. Rather, they are assumed to be direct, automatic "sense judgments" that Arnold described as resembling those of a baseball player adjusting to the speed and direction of a fly ball.

This approach considers emotion to be a multi-component, sequential phenomenon (Frijda, 1986; Lazarus, 1991; Scherer, 1984). For example, in his emotion model, Frijda (1986) proposed that perception and understanding of the emotion-provoking event occurs first, followed by evaluations (appraisals) of the relevance, seriousness, urgency, and importance of the event for one's concerns (i.e., dispositions to desire the occurrence or non-occurrence of a particular situation). These processes lead to a change in state of action readiness, affect (pleasantness/ unpleasantness), and arousal.

Other appraisal theories of emotion (Frijda, 1986; Roseman, 1984; Scherer, 1984; Smith & Ellsworth, 1985) share the basic assumption that emotions have distinctive patterns of appraisals, although the theories differ slightly on which dimensions of appraisal account for the differences among emotions. Smith and Ellsworth (1985) proposed the following list of eight appraisals -- attentional activity, pleasantness,

certainty, anticipated effort, perceived obstacle, responsibility, control, and legitimacy. In contrast to many studies in emotion research, they used real emotion-provoking situations to test the ability of these dimensions to discriminate among different emotions. In their study, students were asked to complete two questionnaires about their thoughts and emotions twenty minutes before a midterm exam and again immediately after getting the grade for that exam. The questionnaire was designed to identify which emotions the students felt, and to assess both how strongly the emotions were felt and to what degree different appraisal dimensions corresponded to how they were thinking about the exam. As expected, particular emotions were associated with specific patterns of emotional appraisals. For example, anger was associated with appraisals of high obstacle, low legitimacy, and high other-responsibility.

The appraisals investigated in Smith and Ellsworth's study (1985) were made in reference to emotion-provoking events (i.e., the exam and receiving a grade on the exam) that happened to the individual him/herself, and not to another. Appraisal theories stress that emotions occur when events are appraised as affecting the self, as in the case of stereotype threat.

An appraisal-based explanation: emotion intensity

One appraisal-based explanation that may prove useful in understanding stereotype threat is that of emotion intensity. In addition to bringing to the discussion on stereotype threat the most developed approach on emotion, emotion intensity protects against the risk of failing to detect specific discrete emotions or affective dimensions. Emotion intensity, which refers to the magnitude of one's feeling (Frijda, Orthony,

Sonnemans, & Clore, 1992; Sonnemans, 1991), is an understudied topic of research. One reason for this gap is the structural complexity of emotional intensity. First, there are low correlations among the physiological variables (e.g., heart rate, blood pressure, hormonal changes), which may be explained by the fact that each of them serves other individual functions than those related to emotions (for a review, see Frijda, 1986). Second, subjective (i.e., self-reported) emotional intensity, physiological variables, and behavioral acts appear to be only slightly related (Lang, 1984).

The structure of subjective emotion intensity is multidimensional. For example, a sudden, large change of water temperature during a shower and a poor grade on an exam may both elicit intense emotions; however, the intensities of both emotions are not necessarily comparable. While the former situation can cause an emotion that is short-lived, associated with very strong perceived bodily changes but not with belief changes or influence upon long-term behavior, a poor grade may cause a much longer emotional response, less significant bodily changes, and belief changes or influence upon long-term behavior. Further evidence of multidimensionality has been found in various studies in which responses to questions on aspects of emotional intensity of recalled experiences revealed different factors of emotion intensity (Frijda et al, 1992; Sonnemans, 1991; Sonnemans & Frijda, 1994). Sonnemans and Frijda (1994) found five factors: duration, drasticness of action tendency, bodily effects, recollection after the event, and changes in long-term behavior and beliefs. Therefore, the use of more than one subjective emotion intensity parameter (i.e., overall felt intensity) when attempting to measure emotion intensity is strongly recommended, as "it is unlikely that antecedents and consequences of emotions are consistently connected with overall felt intensity across all kinds of

emotions, or that they are consistently connected with all parameters" (p. 346).

Prior research suggests that any particular emotion can be determined by several different and independent factors (i.e., appraisal, affect, action readiness). Frijda's (1996) research suggests several such determinants, derived from his model, that are of particular significance to emotion intensity: concerns, appraisal, action repertoire, regulation, and mood. Concerns are viewed as goals, standards, norms, and wishes, and they can explain, for example, why teams may experience different emotions in games, or why students may experience different emotions when getting a grade of B. The strength of their concern influences the intensity of their emotions. Other evidence suggests that stronger concerns elicit more intense emotions. Frijda (1996) suggested that some of these concerns are related to a person's sense of group identity (e.g., concerns that one's group is treated fairly, or that people in one's group are not harmed).

Emotion intensity is also influenced by the relevance of an event or, in other words, the degree of relevance of an event for one's concerns. For example, the event of taking an important examination is relevant to a college student's concerns about academic standings. The intensity of one's emotion about the event will be influenced, however, by the strength of this concern. If the student cares about academic standing very much, or if the exam is in a core course in his or her major, the emotion will be more intense.

Emotion intensity also depends on prior mood state. An event such as being pushed on a crowded subway platform or train, for example, can elicit greater anger if one is already in an irritable mood. As Zillman's (1979) research on excitation transfer suggests, individuals have a tendency to attribute an affective state caused by a previous

event to current events. Experimental studies involving emotion intensity measurements, therefore, need to assess respondents' moods before the treatment.

The purpose of this research project

In general, the role of emotions in stereotype threat research has been limited to one variable, anxiety, and thereby the role of other affective variables has not been explored. This study investigates the importance of various affective dimensions on test score impairment in stereotype threatening situations in an attempt to shed light on the "black box" of stereotype threat. The affective dimensions of interest in this study are: emotion intensity, affects (i.e., positive and negative), and several discrete emotions (i.e., acute anxiety, anger, fear, disappointment, frustration). The proposed model is causal; stereotype threat causes changes in several affective dimensions, which, in turn, cause impairment in performance.

The specific ways in which affective dimensions may fit in the stereotype threat phenomenon are illustrated in the following example. Consider a student taking a test in a domain for which a negative stereotype exists about her group; she is, for example, African-American and the test is the SAT. She may be nervous because, just like anybody else, the score she will get on this test is important for her future. She provides information, including her ethnicity, which is requested on the answer form. This particular information makes negative stereotypes about African-Americans in this domain salient to her and affects the way she now appraises the situation; she is threatened by a stereotype and taking this test has now become relevant to other concerns (e.g., the way she will represent African-Americans, the way her score may be used to

increase or decrease academic funding, and/or the way the test may be ethnically biased). As a result, her emotions become more intense. She may also feel that taking this test carries a more positive value and feel more determined, proud, and/or enthusiastic. Simultaneously, she may also experience increased negative affect and feel somewhat more irritable, jittery, and/or afraid. In addition, the stereotype threat may elicit other specific emotions, such as acute anxiety, anger, fearfulness, disappointment, and frustration. As a result, these emotions and affects impair her performance and her test score does not reflect her true aptitude.

Two stereotypes were targeted in this study: the stereotype about women and mathematics and the stereotype about Blacks and academic performance, more specifically verbal test performance.

Hypotheses

The processes through which a stereotype threat impairs test performance will be investigated. From the most general to most specific, the following hypotheses are tested.

To the extent that stereotype threat activates one's concerns for oneself or one's group, stereotype threat increases emotion intensity. In other words, the effect of stereotype threat on performance may be mediated by emotion intensity, rather than by any specific emotion(s) (e.g., anxiety, shame, anger, and surprise) [15] (Model 1, Figure 1). Hence, it is hypothesized that:

H1a: Stereotype threat increases emotion intensity.

H1b: The higher emotion intensity, the lower the test performance.

To the extent that emotion processes mediate the effect of stereotype threat on

performance, other aspects of emotion processes that may be impacted by stereotype threat are primary dimensions of emotion, such as positive and negative affects. The effect of stereotype threat on performance may be mediated by positive and negative affects, rather than by specific emotions (Model 2, Figure 1). Hence, it is hypothesized that:

H2a: Stereotype threat increases positive affect.

H2b: The higher positive affect, the lower the test performance.

H2c: Stereotype threat increases negative affect.

H2d: The higher negative affect, the lower the test performance.

To the extent that emotion processes mediate the effect of stereotype threat on performance, a final measure of emotion processes that may be impacted by stereotype threat is the specific emotion(s) that are invoked. Stereotype-threatened individuals may experience the following emotions: a) acute anxiety and fear; b) anger and disappointment; and c) frustration, as suggested by literature on stereotype threat (e.g., Steele & Aronson, 1995; Steele, 1997), sport and mental test performance (e.g., Gould & Udry, 1994; Baumeister & Showers, 1986; Le Gall, 1992), and emotion (e.g., Scherer, 2001) respectively (Model 3, Figure 2). Hence, it is hypothesized that:

H3a: Stereotype threat increases acute anxiety.

H3b: The higher the level of acute anxiety, the lower the test performance.

H3c: Stereotype threat increases anger.

H3d: The higher the level of anger, the lower the test performance.

H3e: Stereotype threat increases fear.

H3f: The higher the level of fear, the lower the test performance.

H3g: Stereotype threat increases disappointment.

H3h: The higher the level of disappointment, the lower the test performance.

H3i: Stereotype threat increases frustration.

H3j: The higher the level of frustration, the lower the test performance.

Method

Participants. One hundred and six undergraduate students participated in this study for course credit. This sample included 94 women and 10 men, of multiple ethnicities[16] (4% Asian/Asian-American, 48% Black/African-American, 21% White/Caucasian, 1% Hispanic/Latino/a). These students were recruited from an All-women College or University (AWCU) to tap the stereotype about women and mathematics (N=51), and a Historically Black College or University (HBCU) to tap the stereotype about Blacks and academic performance (N=55). Demographic information pertaining to each sub-sample is reported in Table 1.

Instructions and procedure. The Principal investigator (PI), a dark-skinned multiracial male, introduced himself, and distributed a written description of the study and the participants' rights (Appendix A). Participants read the description and signed consent forms. Next, the PI informed the participants verbally that the study was investigating test-taking emotion and that they would be asked to answer a number of questions about how they feel before and after a test and how they felt during the test (i.e., post-hoc recall). The PI then asked the participants to fill out two affective state

questionnaires, the Positive and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988) and a discrete emotions scale (i.e., acute anxiety, anger, fear, disappointment, and frustration) (Appendix B), to provide a baseline assessment of their pre-test affective states. In addition, participants were asked to fill out a questionnaire to measure the importance of test scores to their self-concept (Appendix C). The participants then read a written description of the test (Appendix D and E). The specific description of the test, a slightly modified version of the description scripted by Aronson et al. (1999), served as the experimental manipulation. The PI remained blind to participants' condition assignments. That is, the front page of the study material distributed to participants in each condition was identical.

Following the "diagnosticity of the test" manipulation approach (Steele & Aronson, 1995, study 1, 2, and 3; 1998; Steele, 1997), two conditions were used: diagnostic and non-diagnostic. Participants were randomly assigned to one of these two conditions; a single pile of study material distributed to participants was arranged following the same pattern (i.e., diagnostic condition, non-diagnostic condition, diagnostic condition, etc.). On the last days of data collection, a pile containing only non-diagnostic condition material was used by mistake. Consequently, more participants were assigned to the non-diagnostic condition than to the diagnostic condition.

At both the HBCU and the AWCU, participants in the diagnostic condition read additional information about the study stating that the study is aimed at "better understanding what makes some people better at Math and English than others," making the racial stereotype about intellectual ability in verbal domains relevant to the performance of African American participants and the gender stereotype about

intellectual ability in mathematical domains relevant to the performance of women participants, and establishing the threat that they may be seen stereotypically or may fulfill the stereotype. In addition, the participants in the diagnostic condition read that they would receive feedback that "may be helpful to [them] by familiarizing [them] with some of [their] strengths and weaknesses" in mathematical and verbal problem solving. This manipulation can be found in Appendix D.

In the non-diagnostic condition, participants were told that the purpose of the research was to better understand the "psychological factors involved in problem solving" and that, after the test, they would receive feedback as a means of familiarizing them "with the kinds of problems that appear on tests [they] may encounter in the future" (Appendix E).

After participants read one these test descriptions, they took a test composed of 30 multiple-choice items (15 quantitative items and 15 verbal items) from various Graduate Record Examination (GRE) test guides (Appendix F) and some scratch paper. The PI informed the participants that they had 30 minutes to complete the test. After the testing period was over, participants were asked to complete four affective state questionnaires: two identical to those filled out before the manipulation (i.e., PANAS and discrete emotions scale), and two other questionnaires: a questionnaire based on the Emotion Intensity Questionnaire (Sonnemans, 1991) (Appendix G) and the Geneva Appraisal Questionnaire (GAQ) (Appendix H).

After the last emotion questionnaire (i.e., GAQ) was filled out, participants were asked to report their SAT scores and other demographic information to be used for describing the sample. The PI then debriefed the participants by describing the purposes

of the experiment, discussed issues related to the participants' experience, and answered any question the participants had. In addition, a debriefing document was distributed (Appendix I).

Dependent Measures.

The Positive and Negative Affect Scale

The PANAS (Appendix B) is arguably the most widely accepted measure of affect (see Watson & Clark, 1992, 1994) and consists of 20 words describing feelings and emotions. Using a five-point Likert scale ranging from 1 ("very slightly or not at all") to 5 ("extremely"), respondents indicated to what extent each word described their current feelings. The PANAS consists of two scales that capture two factors largely independent of one another (Diener & Iran-Nejad, 1986; Watson, 1988): positive affect (PANAS PA), and negative affect (PANAS NA). Persons experiencing high positive affect typically report various positive mood states, including feelings of cheerfulness, excitement, energy, alertness, self-confidence, and so on (Watson & Clark, 1994). Conversely, persons experiencing low levels of positive affect typically report feeling sluggish, unenergetic, and disinterested in their environment (Watson & Clark, 1994). Persons in a state of high negative affect typically experience a broad range of negative mood states, including feelings of nervousness, dissatisfaction, discouragement, and irritability (Watson & Clark, 1994), whereas persons in a state of low negative affect typically describe themselves as calm and relaxed (Watson & Clark, 1994). Adding the response scores on each ten items of the PANAS PA and dividing that number by ten gives the scores on the PANAS PA. The PANAS NA score is calculated in the same manner, but

using the ten PANAS NA items. High scores on each scale indicate high level of affect. This instrument is widely used. Reported internal consistency reliability (Cronbach's coefficient alpha) is high (ranging from .86 to .90 for PANAS PA, and from .84 to .87 for PANAS NA), and scale and item validity have been demonstrated with a sample comparable to the one in the present study (i.e., college students) (Watson et al., 1988).

The discrete emotions scale

Several specific emotions were added to the list of items from the PANAS[17] (items with asterisk in Appendix B). Some scholars of emotion (Ekman, 1992; Levenson, 1992) have postulated the existence of a number of discrete emotions, such as anger, fear, and sadness/disappointment. Although there is no apparent contradiction between a discrete emotions approach and one that emphasize emotional continua (reviewed in Ekman & Davidson, 1994), the importance of including these discrete emotions lies in the fact that they are not included in the PANAS items. Using solely PANAS scores thereby would overlook the possible effect of stereotype threat on discrete emotions. Using the same Likert scales as in the PANAS, respondents indicated to what extent they experienced the following discrete emotions: acute anxiety, anger, fear, disappointment, and frustration. As explained earlier, there is still disagreement among discrete emotion researchers about the number of emotions. However, these five emotions are often cited (Ekman, 1992; Izard, 1994, Levenson, 1992; Shaver, Schwartz, Kirson, & O'Connor, 1987) and are relevant to the emotion-inducing manipulation in this study.

The measure of importance of one's test scores[18]

How much participants care about being good in the domain (i.e., academic tests and exams) was measured for exploratory purposes with the 4-item identity sub-scale of

the collective self-esteem scale (Luhtanen & Crocker, 1992) and a question regarding the importance of one's test scores (i.e., "how important are your test scores to you?"). (Appendix C). Reported internal consistency reliability (Cronbach's coefficient alpha) is high (ranging from .90 to .96), and item validity has been demonstrated with a sample comparable to the one in the present study (i.e., college students) (Luhtanen & Crocker, 1992).

The emotion intensity measure

The measure, based on the Emotion Intensity Questionnaire (Sonnemans, 1991), includes items about two dimensions of emotion intensity: belief changes and influences upon long-term behavior (EI-BCIB), and overall felt intensity of emotion (EI-OFIE) (Sonnemans & Frijda, 1994). Respondents use a five-point Likert scale ranging from 1 ("little intense") to 5 ("very intense") to answer the questions (e.g., "how intense was your emotional feeling at its peak?"). The Emotion Intensity Questionnaire differs conceptually from the PANAS in that the questions on the Emotion Intensity Questionnaire concern the strength of emotion only, without making any reference to specific feelings or emotions (i.e., the quality of the emotion). The scores on the EI-BCIB and the EI-OFIE are both calculated by adding their related items (Appendix G) and dividing by the number of items. High scores indicate great emotional intensity. The internal consistency reliability of this instrument (Cronbach's alpha) is high, ranging from .91 (complete measure of emotion intensity and overall felt intensity dimension) to .89 (belief changes and influence upon long-term behavior). The Emotion Intensity questionnaire appears in Appendix G.

The Geneva Appraisal Questionnaire (GAQ)

This questionnaire, included for exploratory purposes, was developed by the Geneva Emotion Research Group. Based on Klaus Scherer's Component Process Model of Emotion (CPM),[19] the GAQ can be used to assess, through recall and verbal report, an individual's appraisal process in the case of a specific emotional episode. The GAQ consists of 34 questions that tap specific appraisal criteria. Respondents use a five point Likert scale ranging from "not at all" to "extremely" to answer the questions (e.g., "did you think that the actions that produced the event were morally and ethically acceptable?") for 30 of the items.

The internal consistency reliability of this instrument (Cronbach's alpha) cannot be computed as each question gathers information on a different facet of emotion appraisals (Scherer, personal communication, September 4, 2002). The Geneva Appraisal Questionnaire appears in Appendix H.

Pilot study. A pilot study was conducted with 15 participants to assess a) the effect of the manipulation on performance, b) whether participants understood the purpose of the study (i.e., manipulation check), and c) whether emotions other than those measured were experienced during the experiment. Given the small sample size, inferential statistical analyses were not appropriate. However, the general trends of the data were in the direction of performance impairment in stereotype threatened individuals compared to those in the control condition. In addition, discussions during the debriefing sessions revealed that participants understood the purpose of the study, and no emotion other than the one measured were reported. The discussions also encouraged the

inclusion of the Geneva Appraisal Questionnaire (GAQ).

Results

The dependent measures were mathematical and verbal test scores, differences in pre-post scores on the Positive and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988), differences in pre-post scores on a discrete emotions scales, a measure based on the Emotion Intensity Questionnaire (Sonnemans, 1991), and responses on the Geneva Appraisal Questionnaire (GAQ).

To test the assumptions that participants' test performance in the stereotype threat condition (i.e., diagnostic condition) differed from those in the control condition (i.e., non-diagnostic condition), I conducted independent samples t-tests on the main constructs (i.e., verbal test score, math test score, EI-BCIB, EI-OFIE, PANASD PA,[20] PANASD NA, anxiety, anger, fear, disappointment, and frustration). Results on these t-tests are presented in Table 2. As indicated by Levene's test for equality of variances and apart from one exception (i.e., anger in the AWCU sub-sample), the error variance across conditions was equal on all variables in both sub-samples.

In the HBCU sub-sample, both the verbal test performance and the change in the feeling of disappointment were lower for the participants in the stereotype threat condition. Thus, as expected, the stereotype threat manipulation had a significant negative impact on Black students' verbal test performance. In the AWCU sub-sample, participants in both conditions only differed on EI-BCIB. In contrast to my expectations, the stereotype threat manipulation did not have a significant negative impact on women's

math test performance. Therefore, the AWCU sub-sample was dropped from further hypothesis-testing analyses.

To assess measure reliability, I calculated Cronbach's alpha for each measure in the HBCU sub-sample (N=55) (see Tables 3a and 3b). Coefficient alphas for all scales, with the exception of PANASD NA, indicate that the measures were moderately reliable. One possible explanation for the low coefficient alphas for PANASD NA may be the lack of variance in the data (Crocker & Algina, 1986).

Bivariate correlations among the 10 main constructs for the HBCU sub-sample were calculated (see Table 4). Surprisingly, verbal test score was related only to condition (i.e., diagnostic-non-diagnostic). Condition was also related to disappointment (consistent with t-tests); participants in the diagnostic condition tended to experience less disappointment as a result of the stereotype threat than participants in the non-diagnostic condition. Contrary to my hypotheses, the data did not seem to unveil simple relationships among these variables.

Hypotheses testing

To test the fit of the hypothesized models, path analyses were conducted using LISREL and covariance matrices. One sample size recommendation for path analysis is 5-10 participants for each parameter (Klem, 1995). This requirement was met for each hypothesized model. Three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root Mean Residual (RMR), were used to evaluate the validity of the hypothesized models.

Hypothesis 1:

Hypotheses H1a (i.e., stereotype threat increases felt emotion intensity) and H1b (i.e., the higher the felt emotion intensity, the lower the test performance) together formed Model 1 presented in Figure 1. As indicated by the significance of the chi-square statistic, the patterns observed in the data were significantly different than the expected patterns. Thus, there was no need to look beyond the significant chi-square statistic to conclude that this model did not fit the data.

Hypothesis 2:

Hypotheses H2a (i.e., stereotype threat increases positive affect), H2b (i.e., the higher positive affect, the lower the test performance), H2c (i.e., stereotype threat increases negative affect), and H2d (i.e., the higher negative affect, the lower the test performance) together formed Model 2 presented in Figure 1. As indicated by the significance of the chi-square statistic, the patterns observed in the data were significantly different than the expected patterns. Again, there was no need to look beyond the significant chi-square statistic to conclude that this model did not fit the data.

Hypothesis 3:

Hypotheses H3a (i.e., stereotype threat increases felt acute anxiety), H3b (i.e., the higher the felt acute anxiety, the lower the test performance), H3c (i.e., stereotype threat increases the felt anger), H3d (i.e., the higher the felt anger, the lower the test performance), H3e (i.e., stereotype threat increases felt fear), H3f (i.e., the higher the felt fear, the lower the test performance), H3g (i.e., stereotype threat increases felt disappointment), H3h (i.e., the higher the felt disappointment, the lower the test

performance), H3i (i.e., stereotype threat increases felt frustration), and H3j (i.e., the higher the felt frustration, the lower the test performance) together formed Model 3 presented in Figure 2. As indicated by the significance of the chi-square statistic, the patterns observed in the data were significantly different than the expected patterns. Thus, I concluded that this model did not fit the data.

In search for an explanation for the lack of main effect of stereotype threat on performance at the AWCU, further statistical analyses were conducted to investigate any measured difference between the HBCU and the AWCU sub-samples. To test whether participants at the AWCU differed from those at the HBCU, I conducted independent samples t-tests on the main affective constructs (i.e., EI-BCIB, EI-OFIE, PANASD PA, PANASD NA, anxiety, anger, fear, disappointment, and frustration). As indicated by Levene's test for equality of variances and apart from two exceptions (i.e., PANAS NA (post-test) and anger (pre-test)), the error variance across sub-samples was equal on all variables.

Across sub-samples, participants did not significantly differ on EI-BCIB (t= -1.07, NS), EI-OFIE (t= .48, NS), PANASD PA[21] (t= -.09, NS), PANASD NA (t= .99, NS), anxiety (t= 1.04, NS), anger (t= -1.82, NS), fear (t= 1.19, NS), disappointment (t= -.16, NS), or frustration (t= .43, NS). In addition, similar t-tests on pre- and post- test measures of the PANAS PA, PANAS NA, and the discrete emotions items were conducted. They too did not indicate in finding any significant difference between the AWCU and HBCU sub-samples. Last, a t-test on an aggregate affect score comprised of all affect scores measured was conducted. No significant difference was observed either for the pre-test (t= 1.02, NS),

post-test (t= 1.44, NS), or the difference score (i.e., T2-T1) (t= .62, NS). Thus, no significant difference was observed between the AWCU and HBCU sub-samples on the main affective dimensions measures.

On a very exploratory basis, supplementary statistical analyses were conducted with data gathered from the Geneva Appraisal Questionnaire (GAQ) and items from the PANAS and Emotion Questionnaire. These analyses and their investigative rationale can be found in Appendices J and K.

Conclusions

According to stereotype threat theory, negative stereotypes create a situation in which behavior from a member of a negatively stereotyped group risks being interpreted as confirming the stereotype (Aronson, Quinn, & Spencer, 1998; Steele, 1997). Consequently, stereotype threat, a situational phenomenon, causes a conscious or unconscious sense of threat, which interferes with performance in the domain associated with the stereotype (Steele & Aronson, 1995; Steele, 1997). However, empirical evidence on the processes by which stereotype threat causes impairment in performance is equivocal, and, to a certain extent, self-reported acute anxiety has practically been ruled out as a mediating variable. On the other hand, it is now known that participants exposed to stereotype threat are subject to a higher level of physiological reactivity in the form of increased blood pressure (Blascovich et al, 2001). While the blood pressure findings are important, they do not point in the direction of a specific emotion or constellation of emotions.

The specific goal of this study was to examine the role of various affective dimensions on test score impairment in stereotype threatening situations in an attempt to shed light on the "black box" of stereotype threat. Two stereotypes were targeted in this study: the stereotype about women and mathematics (AWCU sub-sample) and the stereotype about Blacks and academic performance, with a focus on verbal test performance (HBCU sub-sample). The data indicated that the stereotype threat effect (i.e., the negative impact of stereotype threat on performance in the stereotyped domain) occurred in the HBCU sub-sample but not in the AWCU sub-sample. Consequently, the analyses about the importance of various affective dimensions on test score impairment in stereotype threatening situations were limited to the HBCU sub-sample. However, given the extensive use and the reliability of the "diagnosticity of the test" manipulation to induce stereotype threat (e.g., Steele & Aronson, 1995, study 1, 2, and 3; 1998; Steele, 1997), one may be left to wonder if some specific qualities of the AWCU participants and/or the AWCU environment may explain the lack of stereotype threat effect using this particular experimental manipulation or if these results were simply random.

One characteristic of scientific laws is generality; more often than not, such laws are assumed to be unrestricted in space and time. Yet, a scrutiny of typical research in psychology indicates that this condition is not always met. In experimental social psychology particularly, it is difficult to keep all but the manipulation equal across groups. Yet the notion of generality prevails. My expectation that the most typical stereotype threat manipulation would yield the predicted results was a consequence of this assumption. However, as statistical analyses reported earlier in this paper show, the phenomenon of stereotype threat in this study was not generalizable across time, space,

and population. Furthermore, the variables measured in this study cannot explain why the stereotype threat effect was not observed at the AWCU. Thus, any attempt to explain that observation is speculative. As such, a single explanation is proposed: the experimenter may have had a different effect on participants at the AWCU than at the HBCU. Specifically, the gender of the experimenter in a potentially stereotype threatening situation may have induced a stereotype threat in both stereotype threat and non-stereotype threat conditions at the AWCU (without an additive effect of manipulation and experimenter-related threat in the former). In contrast, the dark skin of the experimenter may have lessened any potential experimenter-related stereotype threat at the HBCU, so that the control condition was "clean;" participants may have perceived the experimenter as much as non-Black as non-White, thus minimizing any potential stereotype threat related to the skin color and ethnicity of the experimenter.

As the data indicated that the stereotype threat effect occurred in the HBCU sub-sample, analyses about the importance of various affective dimensions on test score impairment in stereotype threatening situations were conducted. The primary affective dimensions of interest in this study were: emotion intensity, positive and negative affects, and several discrete emotions (i.e., acute anxiety, anger, fear, disappointment, frustration). The proposed models were causal; stereotype threat triggers one or more affective dimensions, which, in turn, cause impairment in performance.

The findings did not support any of the hypothesized models that were suggested by the theoretical and empirical literature on stereotype threat. In effect, just as the search for the role of acute anxiety in stereotype threat episodes has lead to a dead in, so the current attempt to extend the stereotype threat discourse to include emotional valence,

emotional intensity, and a range of discrete emotions failed to throw light into the black box of stereotype threat. Other than the important findings about stereotype threat exposure and increased blood pressure (Blascovich et al, 2001), a finding that clearly establishes a rationale for searching and theorizing about reactivity to stereotype threat, this study was not able to link stereotype threat to specific emotions or emotional dynamics. Perhaps, specific emotions alone vary too much among individuals to account for the stereotype threat effect. Thus, we may need to investigate more sophisticated models (e.g., models that incorporate emotion antecedents) to understand the role of emotion in stereotype threat.

Post-hoc investigations regarding emotion appraisals and specific emotions in this study yielded support for such models. Specifically, two causal models appeared to fit the data for the HBCU sub-sample: Model A and B. Of these two models, Model B, which emerged as a better model than Model A, suggested that stereotype threat has an impact on performance in the stereotyped domain via two routes. By the first route, stereotype threat causes a decrease in interest, which in turn, causes a decrease in performance. The other route involves appraisal, irritation, and dislike for the task. Specifically, stereotype threat causes an increased sense of "power" appraisal, which causes greater irritation. Greater irritation, in turn, causes greater dislike for the task, which decreases performance. In Model B, all path coefficients were significant. Each path of Model B is discussed individually in Appendix K.

It cannot be overemphasized that these findings do not spark light in the black box of stereotype threat. Because of the post-hoc nature of the analyses used to reveal Model B, one should interpret this study cautiously. Nonetheless, the findings of this study

suggest that neither acute anxiety nor fear play a key role in explaining the effect of stereotype threat on performance in the stereotyped domain, as would be predicted by the theory of stereotype threat. Furthermore, post-hoc investigations suggest that one might conduct another study to test if individuals actually react to stereotype threat in a more proactive manner than simple anxiety or fear. For example, these individuals may distance themselves from the threatening situation by increasing their sense of "power" regarding the consequence of the event and lowering their interest in the task. These reactions, in turn, may interfere with performance in the stereotyped domain.

In sum, despite the lack of significant findings in the directions of the proposed hypotheses, this study can point up several general directions for future stereotype threat research. As noted throughout this discussion, individuals may respond with survival strategies that protect against the threat (i.e., the possible confirmation of the stereotype). Much like self-handicapping, these strategies may not be conducive to optimal performance but they may help individuals maintain how they feel about themselves and how they may relate to others. However, rather than infer a set of "disempowered" evaluations such as those involved in the experience of anxiety and fear, researchers of stereotype threat should test the model suggested by the post-hoc investigations in this study and explore the actual appraisals of stereotype-threatened individuals. Given the post-hoc results of this study, it would not be surprising if future research finds that individuals behave more like survivors of stereotype threat than like its victims. However, it is important to emphasize that the nature of post-hoc analyses and, to a certain extent, the operationalization of most constructs in the suggested model with single items, require that one does not jump to conclusions about the possible explanatory

power of the results. The work presented here simply suggests possible venues for future research on stereotype threat and emotion.

Aside from the failure of the data to reach significance in the test of the proposed hypotheses, two other limitations of this study should also be noted. First, this study addresses affective dimensions that may be of importance in stereotype threat from an appraisal theory of emotion (or cognitive) perspective. Other viable approaches to the study of emotions exist and should not be overlooked in the study of stereotype threat. One such approach is physiological in nature. Technological advances in instrumentation used to observe the brain, such as magnetic resonance imaging (MRI), now allow the measure of different activities in various parts of the brain. Using such instruments, we may be able to detect both the type of brain activity and the part/s of the brain that is/are activated during stereotype threat. Similarly, a facial expression approach to emotions may be used to explore stereotype threat. As an alternative to self-reports, facial expressions can be analyzed with tools such as the facial action coding system (FACS) to identify emotions in a non-intrusive manner.

Secondly, this study does not address the issue of "Black" and/or "woman" identity or how identification with the negatively stereotyped group may affect performance. Yet, identity is important for emotion appraisals, particularly in situations that involve groups with whom one identifies (Cotting & Deaux, 1999). Unfortunately, this limitation has also been the downfall of past studies on stereotype threat. Future studies on stereotype threat should address this gap.

We are only beginning to understand stereotype threat, but as our knowledge of the phenomenon increases, we may learn also how to counter stereotype threat and make

it a phenomenon of the past. It's a wiid dream, yet we cannot wait for stereotypes to disappear from society. This study is an expression of my hope to make that dream a reality.

Endnotes

[1] Kabyles in North Africa pre-date Arab invasions. They are part of the Berber people.

[2] Aronson et al., (1999) for instance, elicited stereotype threat by telling participants that the purpose of the experiment was to better understand why Asian Americans outperform European Americans on math tests.

[3] Based on McGuire, McGuire and Winton's (1979) finding that the spontaneous mention of gender increases proportionally to the increase in the relative number of opposite-sex individual, an interesting venue for future research may be to investigate if the effect of stereotype threat increases proportionally with salience of stereotypes.

[4] Participants in the control condition did not expect to take any verbal test, but were asked to complete a battery of measures.

[5] Participants did not actually take the verbal test.

[6] The items of the test were taken from the advanced GRE in math.

[7] Participants needed to have SAT-Math scores between 400 and 650, and no more than one year of calculus.

[8] Items from the Graduate management Test (GMAT) were used.

[9] It is worth noting that Aronson et al. (1999, study 2) likened high ability to attending a selective institution.

[10] However, note that Aronson et al. (1999) presented no evidence against the domain identification hypothesis.

[11] Note that an ANCOVA controlling for reported effort did not change the mean scores of participants in the two conditions, indicating that amount of effort could not fully account for differences in math scores (Aronson et al., 1999).

[12] Le Gall (1992) even argues that much of the lack academic success among students is induced by test-taking anxiety and not by lack of ability or motivation.

[13] Three difficult anagram questions were removed from the previously–used test, the test period was cut by nearly 17%, and the test was presented on a Macintosh computer.

[14] A version of the Spielberger State Anxiety Inventory (1970, 1980) was used.

[15] In addition, a mediating effect of overall emotion intensity may reveal the possibility that the impact of stereotype threat on performance is mediated by a single emotion. If more than one emotion is involved (e.g., stereotype threat increases anger but decreases surprise), overall emotion intensity may be equal.

[16] 26% of participants did not fill out the information about ethnicity.

[17] However, these items are analyzed separately and are not included in the analysis of PANAS PA and PANAS NA.

[18] Although data for this measure was collected, the hypotheses presented in this paper do not incorporate the importance of one's test score as a construct. Consequently, statistical analyses on this measure were not included in this paper.

[19] The most recent version of this model can be found in Scherer (2001).

[20] PANASD PA and PANASD NA represent the difference score on PANAS PA and PANAS NA.

[21] PANASD PA and PANASD NA represent the difference scores on PANAS PA and PANAS NA.

Table 1

Demographic Information

Demographic Characteristics		AWCU (N=51)	HBCU (N=55)
Sex			
	Male	0	10
	Female	51	43
	Missing	0	2
Ethnicity			
	White/Caucasian	22	0
	Black/African-American	5	46
	Asian/Asian-American	4	0
	Hispanic/Latino/a	1	0
	Missing	19	9

Table 2a
T-tests of study variables (HBCU sub-sample)

Variables		Means Diagnostic (N=26)	Non-diagnostic (N=29)	t
Math test performance		4.73 (SD=2.45)	5.59 (SD=2.39)	-1.26
Verbal test performance		5.23 (2.39)	6.93 (2.95)	-2.21*
EI-BCIB		2.22 (.71)	2.13 (.74)	.47
EI-OFIE		2.67 (.87)	2.72 (.71)	-.26
PANASD PA		-.72 (.61)	-.65 (.72)	-.35
	PANAS PA (pre)	2.65 (.87)	2.60 (1.87)	.20
	PANAS PA (post)	1.93 (.89)	1.94 (.68)	-.08
PANASD NA		.15 (.49)	.15 (.49)	.05
	PANAS NA (pre)	1.64 (.61)	1.53 (.50)	.72
	PANAS NA (post)	1.80 (.55)	1.68 (.58)	.74
Anxiety		.09 (1.27)	-.14 (1.43)	.59
	Anxiety (pre)	2.36 (1.22)	2.14 (1.19)	.67
	Anxiety (post)	2.45 (1.30)	2.00 (1.13)	1.33
Anger		.00 (1.07)	.55 (1.24)	-1.67
	Anger (pre)	1.68 (1.25)	1.34 (.72)	1.13a
	Anger (post)	1.68 (1.21)	1.90 (1.01)	-.69
Fear		-.05 (.38)	-.10 (.77)	.32
	Fear (pre)	1.32 (.57)	1.41 (.98)	-.41
	Fear (post)	1.27 (.63)	1.31 (.60)	-.22
Disappointment		.09 (1.48)	1.55 (1.62)	-3.32**
	Disappointment (pre)	1.82 (1.14)	1.41 (.83)	1.41a
	Disappointment (post)	1.91 (1.15)	2.97 (1.48)	-2.78
Frustration		.59 (1.50)	1.14 (1.46)	-1.31
	Frustration (pre)	2.32 (1.36)	1.93 (1.28)	1.04
	Frustration (post)	2.91 (1.23)	3.07 (1.44)	-.42

a As indicated by Levene's test of equality of error variances, the error variance of anger is not equal across groups. Therefore, the t-value when equal variances are not assumed is presented.
* $p<.05$; ** $p<.01$.

Table 2b
T-tests of study variables (AWCU sub-sample)

Variables		Means Diagnostic (N=20)	Non-diagnostic (N=31)	t
Math test performance		6.58 (SD=3.04)	6.94 (SD=2.53)	-.45
Verbal test performance		7.26 (2.47)	8.52 (2.74)	-1.63
EI-BCIB		2.29 (.83)	1.83 (.70)	2.07*
EI-OFIE		2.63 (1.06)	2.85 (.82)	-.82
PANASD PA		-.80 (.63)	-.63 (.57)	-1.00
	PANAS PA (pre)	2.68 (.69)	2.88 (.72)	-.96
	PANAS PA (post)	1.88 (.63)	2.25 (.71)	-1.85
PANASD NA		.23 (.61)	.28 (.51)	-.33
	PANAS NA (pre)	1.68 (.76)	1.57 (.47)	.60
	PANAS NA (post)	1.91 (.88)	1.86 (.73)	.23
Anxiety		.42 (1.30)	.45 (1.50)	-.07
	Anxiety (pre)	2.00 (1.20)	2.03 (1.02)	-.10
	Anxiety (post)	2.42 (1.26)	2.48 (1.29)	-.17
Anger		.11 (.57)	.39 (.92)	-1.34a
	Anger (pre)	1.26 (.81)	1.19 (.54)	.37
	Anger (post)	1.37 (.76)	1.58 (.87)	-.87
Fear		.05 (.71)	.00 (.45)	.32
	Fear (pre)	1.63 (1.12)	1.29 (.46)	1.27a
	Fear (post)	1.68 (1.20)	1.29 (.64)	1.32a
Disappointment		1.00 (1.56)	.94 (1.09)	.17
	Disappointment (pre)	1.58 (1.22)	1.35 (.61)	.75
	Disappointment (post)	2.58 (1.71)	2.29 (1.24)	.64a
Frustration		.68 (1.60)	1.32 (1.05)	-1.71
	Frustration (pre)	2.00 (1.33)	1.90 (1.01)	.29
	Frustration (post)	2.68 (1.70)	3.23 (1.12)	-1.23a

a As indicated by Levene's test of equality of error variances, the error variance of anger is not equal across groups. Therefore, the t-value when equal variances are not assumed is presented.
* $p<.05$; ** $p<.01$.

Table 3a

Reliability coefficients, means, and standard deviations of study variables

HBCU sub-sample (N=55)

Variables	No. Items	α	M	SD
EI-BCIB	8	.84	2.15	.73
EI-OFIE	3	.78	2.68	.78
PANASD PA	10	.76	-.67	.65
PANASD NA	10	.62	.16	.48

Table 3b

Reliability coefficients, means, and standard deviations of study variables

AWCU sub-sample (N=51)

Variables	No. Items	α	M	SD
EI-BCIB	8	.88	1.99	.79
EI-OFIE	3	.90	2.76	.91
PANASD PA	10	.76	-.68	.59
PANASD NA	10	.72	.26	.54

Table 4

Constructs correlation matrix (HBCU sub-sample)

Construct	1	2	3	4	5	6	7	8	9	10
Stereotype threata	-									
EI-BCIB	-.07	-								
EI-OFIE	.04	.43**	-							
PANASD PA	.05	-.06	-.15	-						
PANASD NA	-.01	.35**	.30*	-.21	-					
Anxiety	-.08	-.10	.10	.01	.37**	-				
Anger	.23	.38**	.28*	-.17	.36**	.16	-			
Fear	-.05	-.30*	.06	.04	.12	.23	.01	-		
Disapppointment	.43**	.27*	.24*	-.26*	.49**	.14	.42**	.18	-	
Frustration	.18	.39**	.28*	-.30*	.84**	.41**	.44**	.06	.64**	-
Verbal test Performance	.30*	-.13	.13	.15	-.19	-.11	-.08	-.06	.06	-.07

Listwise N= 45
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 5

Constructs correlation matrix (Model A, HBCU sub-sample)

Construct	1	2	3	4	5
Stereotype threata	-				
Interest	.31*	-			
Sense of Power	-.28*	.03	-		
Dislike	.03	-.09	.29*	-	
Verbal test Performance	.34*	.28*	-.18	-.33*	-

Listwise N= 45
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 6a

Constructs correlation matrix (Model A, AWCU sub-sample)

Construct	1	2	3	4	5
Stereotype threata	-				
Interest	.08	-			
Sense of Power	.08	.10	-		
Dislike	-.11	-.08	-.06	-	
Math test Performance	.05	.03	-.28*	-.49**	-

Listwise N= 49
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 6b

Constructs correlation matrix (Model A, overall sample)

Construct	1	2	3	4	5
Stereotype threata	-				
Interest	.20*	-			
Sense of Power	-.10	.06	-		
Dislike	.04	-.09	.11	-	
Total test Performance	.26**	.15	-.29**	-.41**	-

Listwise N= 94
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 7

Constructs correlation matrix (Model B, HBCU sub-sample)

Construct	1	2	3	4	5	6
Stereotype threata	-					
Interest	.31*	-				
Sense of Power	-.28*	.03	-			
Dislike	.03	-.09	.29*	-		
Verbal test Performance	.34*	.28*	-.18	-.33*	-	
Irritation	-.28*	-.09	.31*	.41**	-.17	-

Listwise N= 45
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 8a

Constructs correlation matrix (Model B, AWCU sub-sample)

Construct	1	2	3	4	5	6
Stereotype threata	-					
Interest	.08	-				
Sense of Power	.08	.10	-			
Dislike	-.11	-.08	-.06	-		
Math test Performance	.05 .03	-.28*	-.49**	-		
Irritation	.20	.16	-.25*	.20	.11	-

Listwise N= 49
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

Table 8b

Constructs correlation matrix (Model B, overall sample)

Construct	1	2	3	4	5	6
Stereotype threata	-					
Interest	.20*	-				
Sense of Power	-.10	.06	-			
Dislike	.04	-.09	.11	-		
Total test Performance	.26**	.15	-.29**	-.41**	-	
Irritation	-.03	.03	.01	-.30**	-.02	-

Listwise N= 94
a 0: diagnostic; 1: non-diagnostic
* p<.05; ** p<.01.

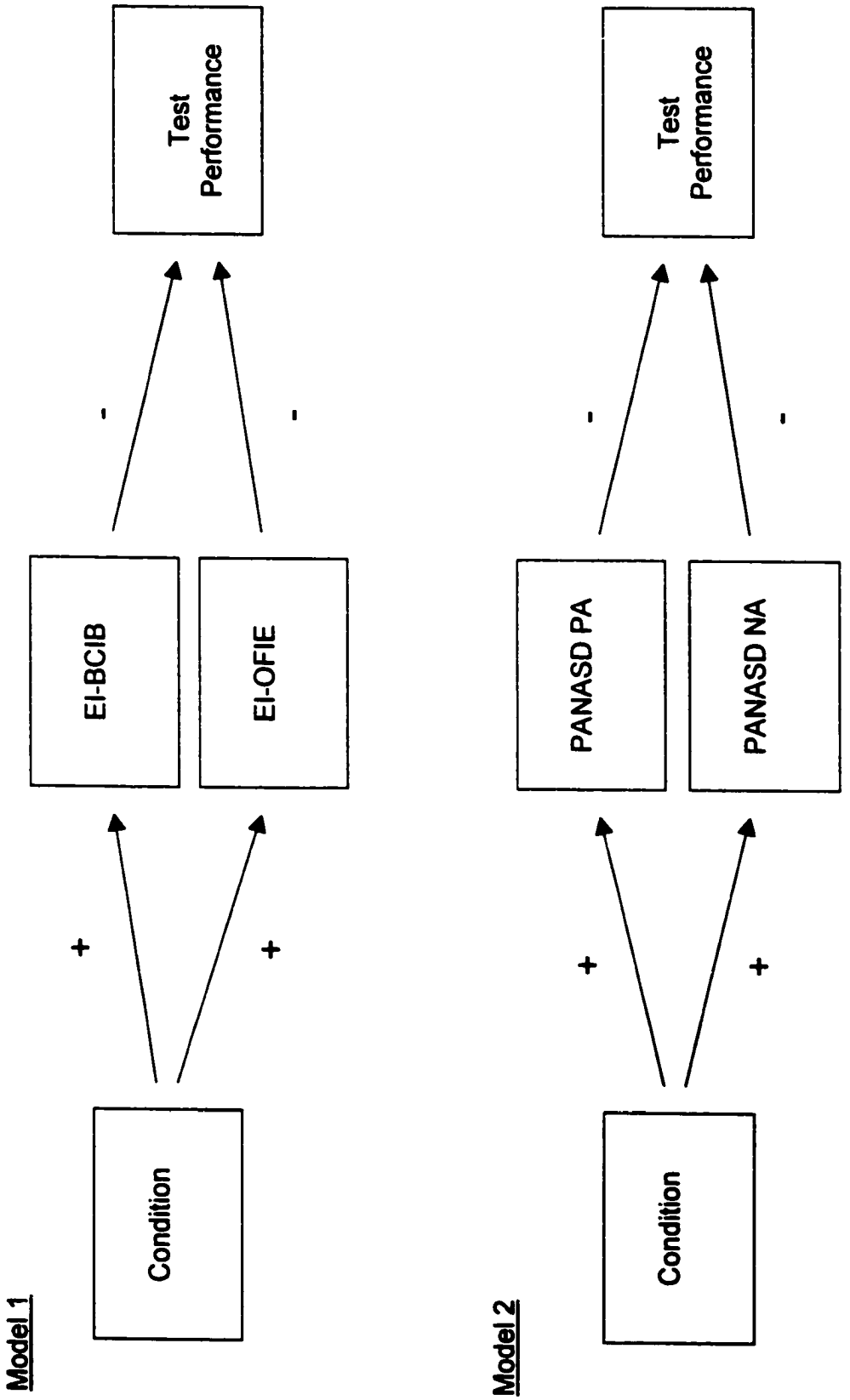


Figure 1.

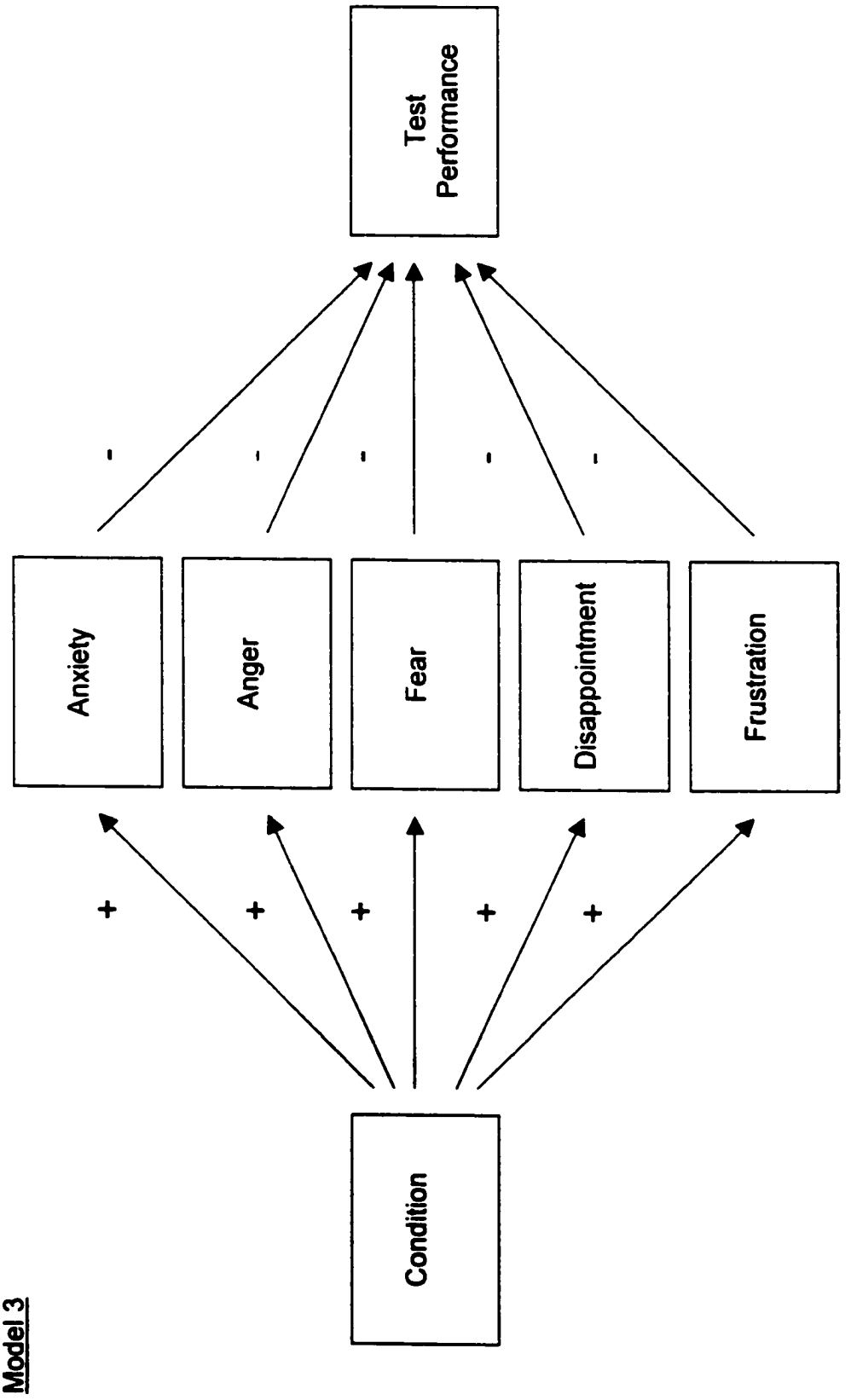


Figure 2.

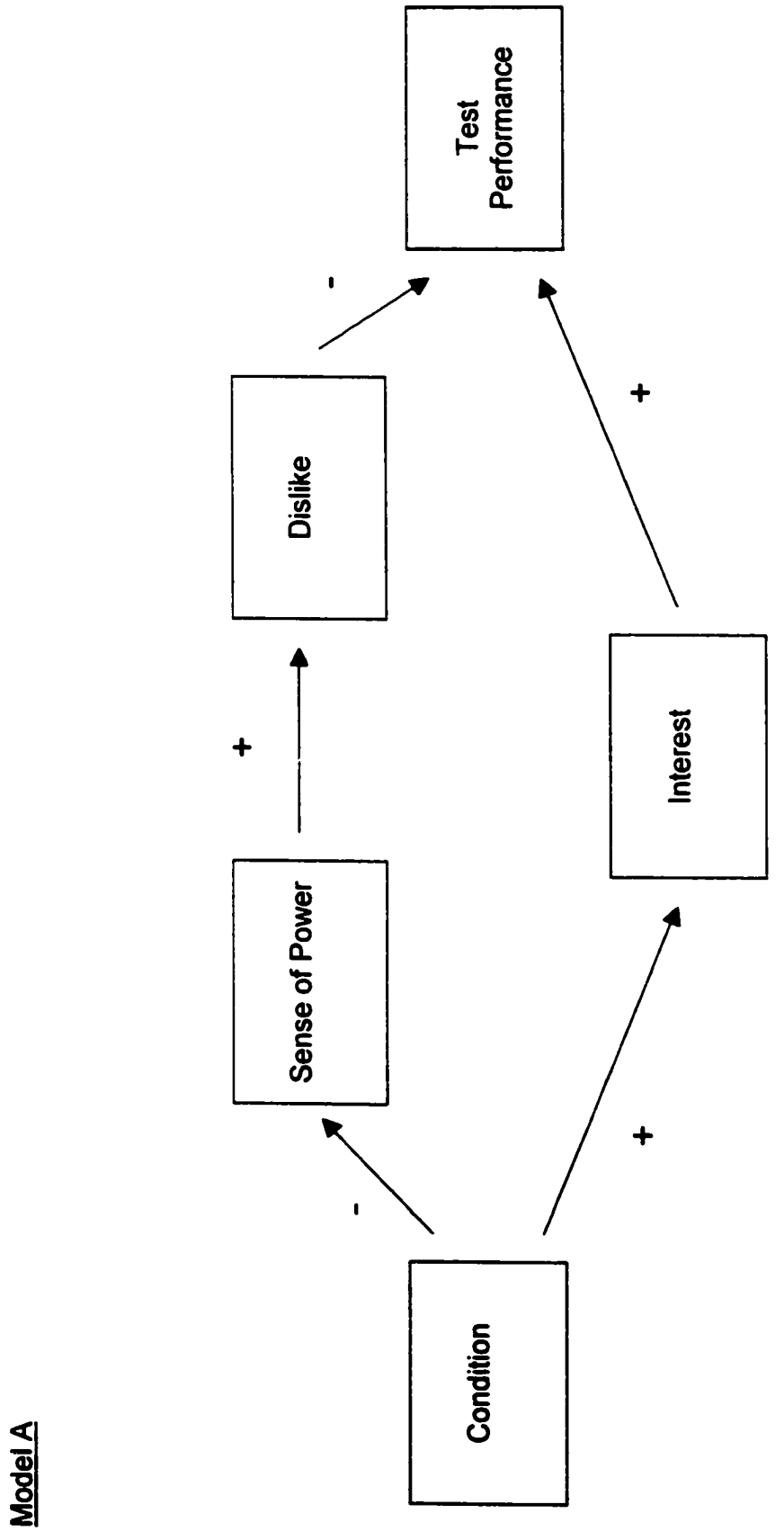


Figure 3.

Figure 4.

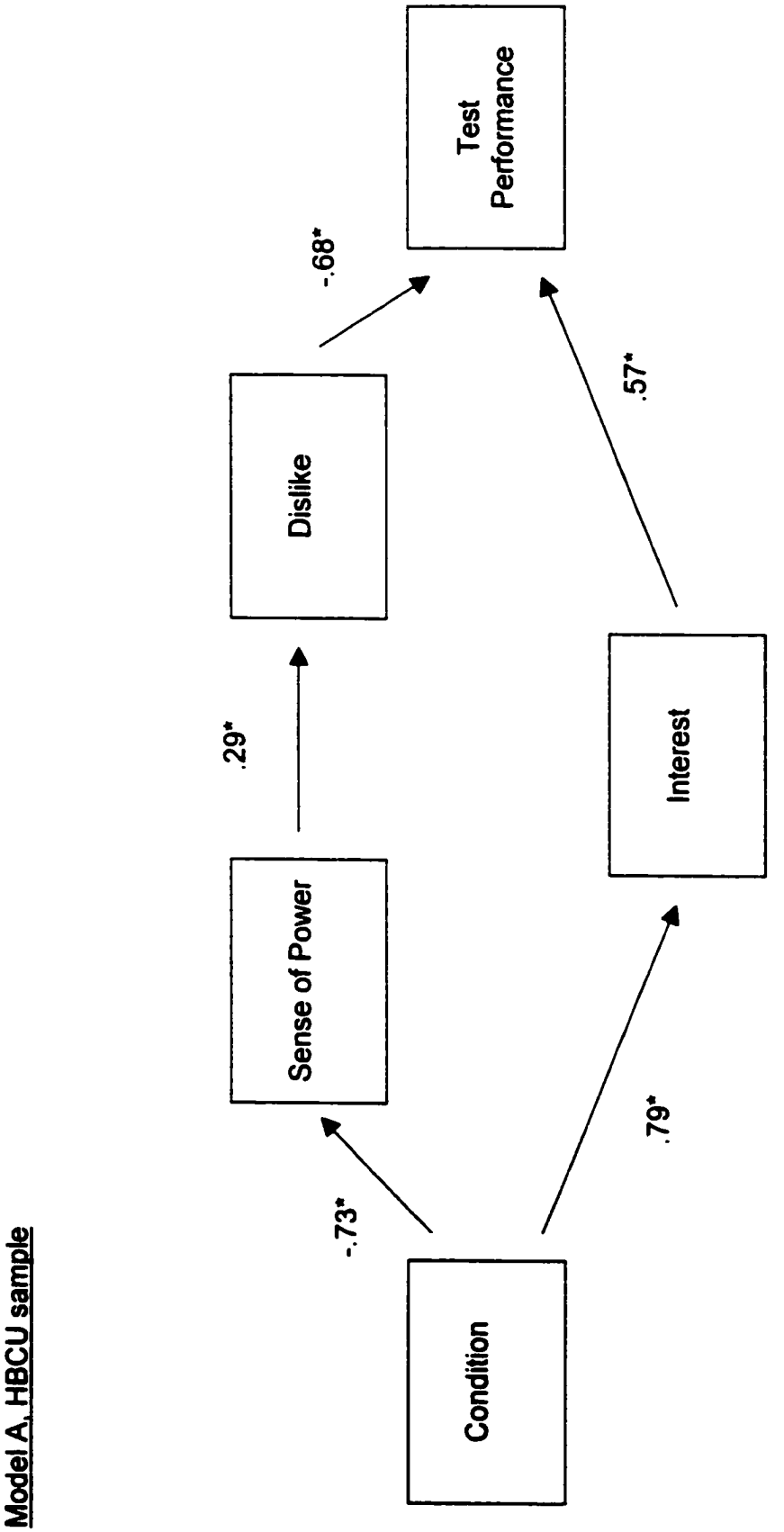
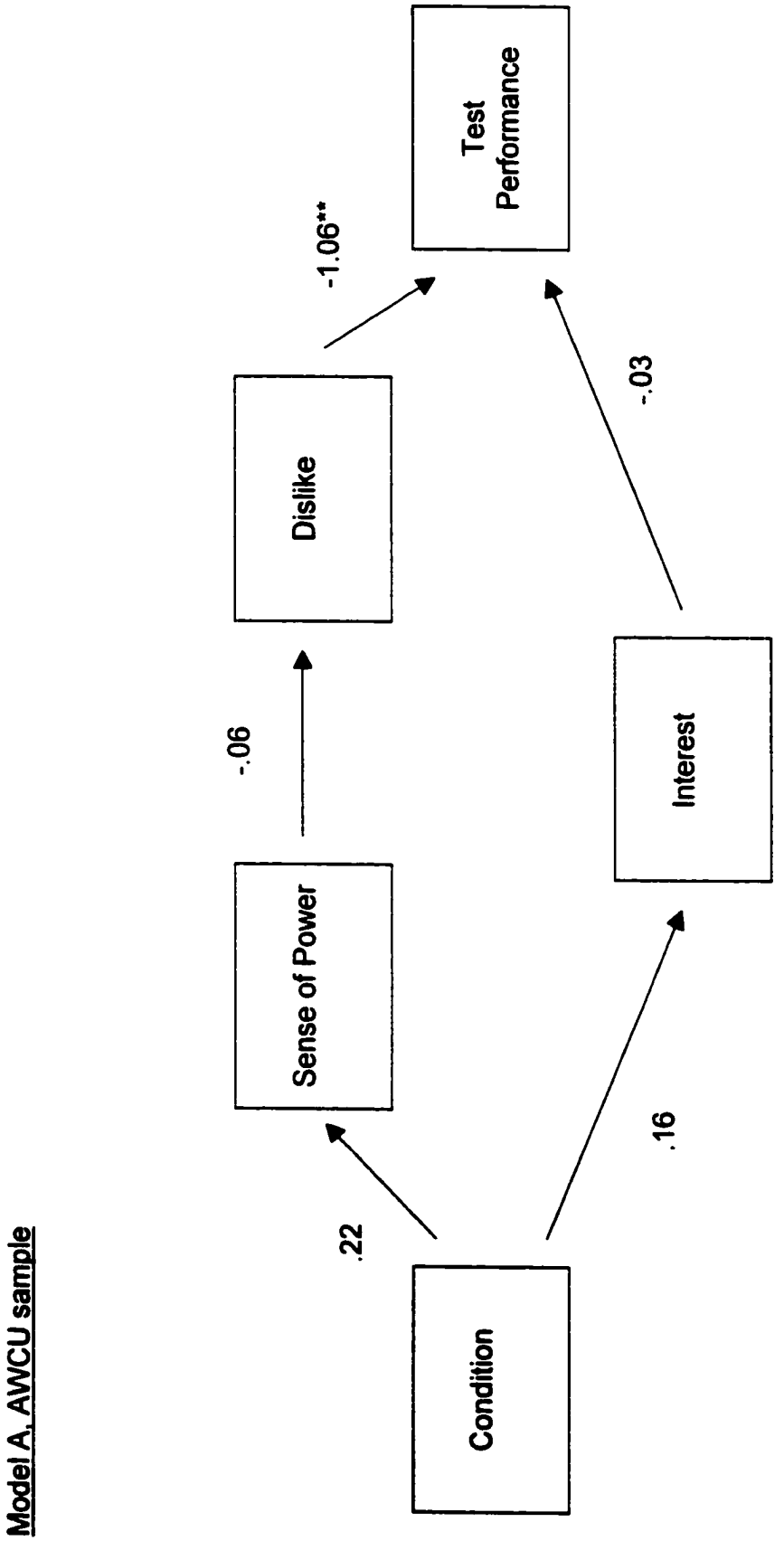
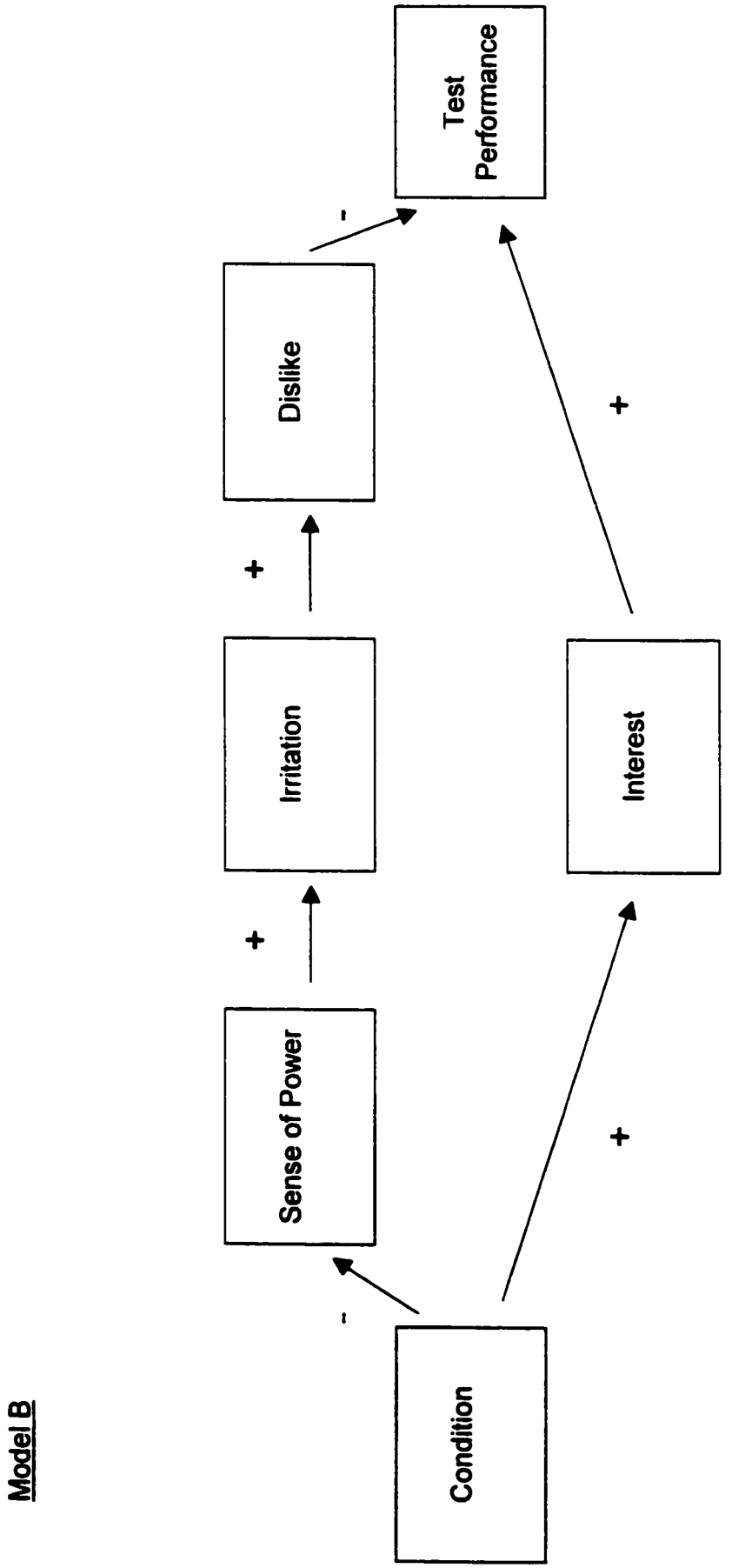


Figure 5.





Model B

Figure 6.

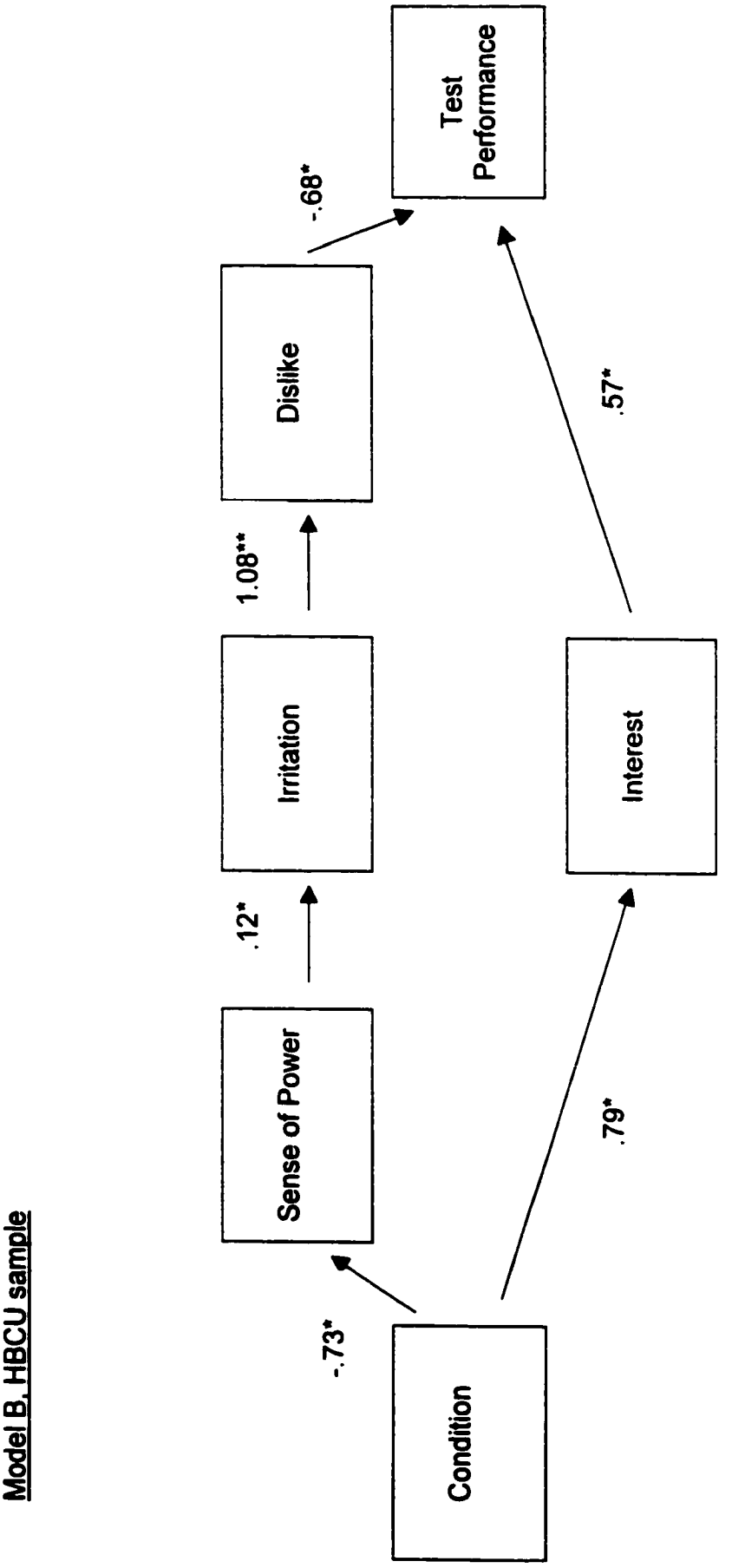
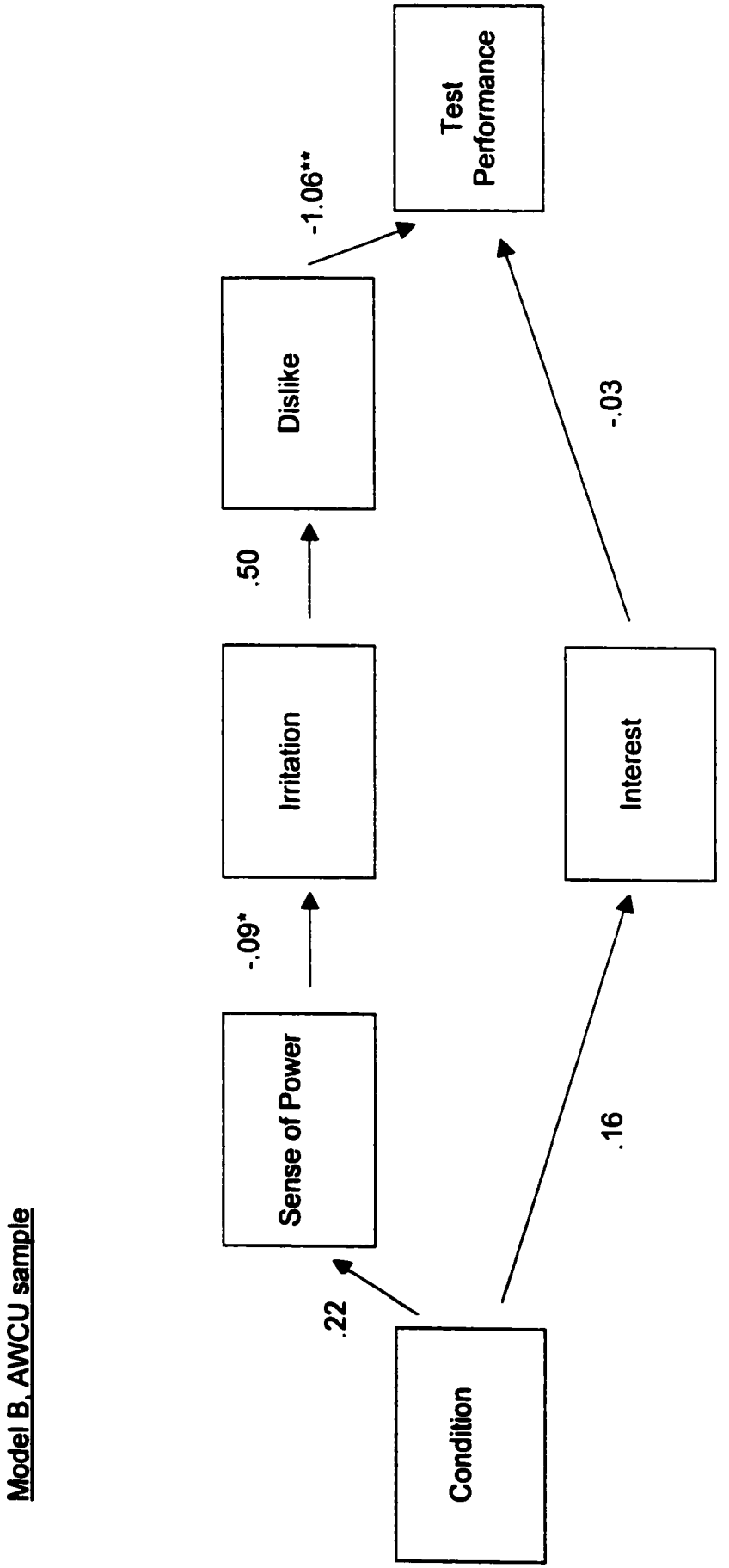


Figure 7.

Figure 8.



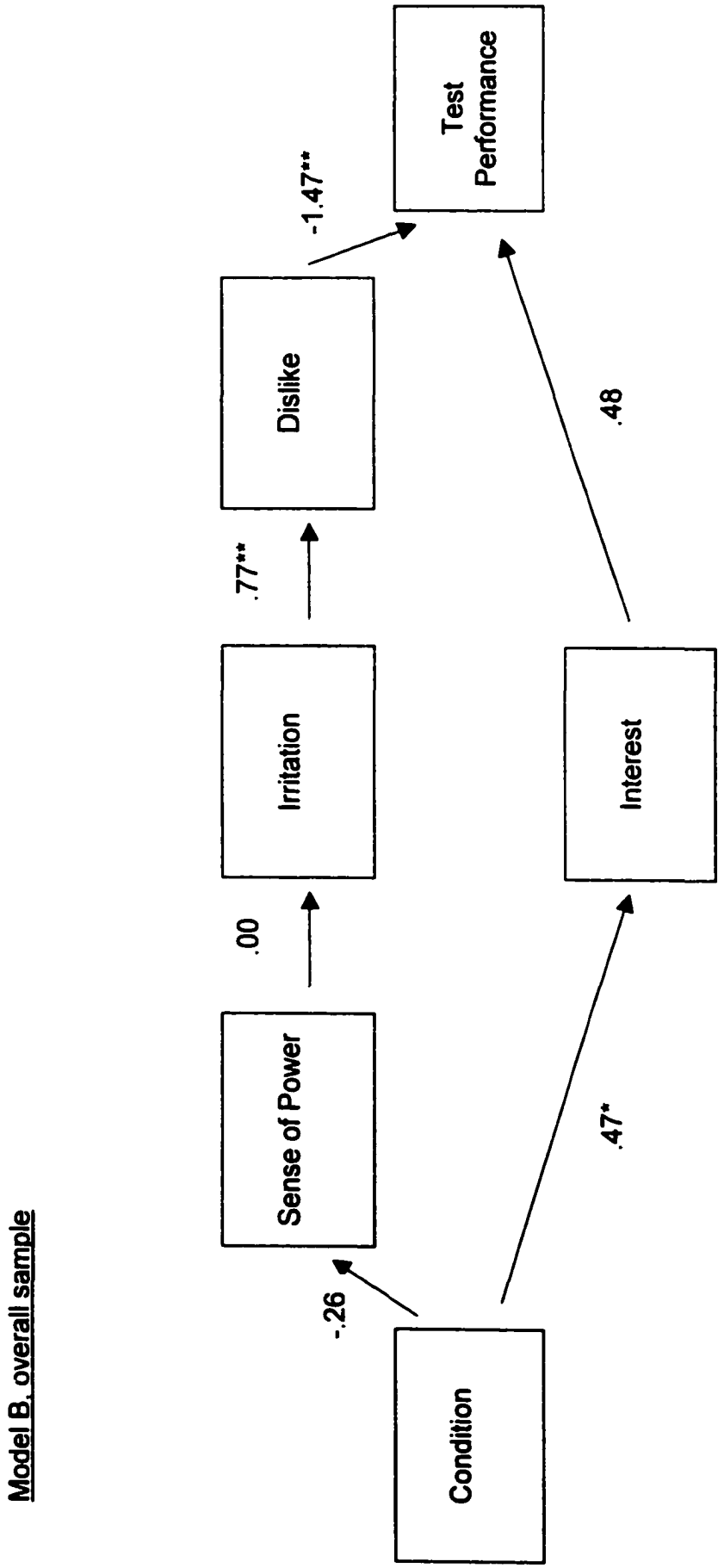


Figure 9.

Appendix A

Study description and my rights as a participant

I am a participant in a research study on performance in test taking and emotion. My participation in this study will take approximately 45 minutes and involves taking a short test and filling out a few questionnaires. I will receive credit in my Introduction to Psychology course for my participation. My responses to both the test questions and the questionnaires are confidential, and I will not write down my name or any identifier anywhere on the test or the questionnaires. Furthermore, I can refuse to answer any question and/or end my participation in this study at any time without penalty. Last, I can choose to not participate in this study and still earn the same amount of credit for my Introduction to Psychology course by a) reading one Psychology article and b) taking a short quiz on the reading.

I understand that I may have an emotional reaction to taking the quiz, especially if I have experienced emotions during tests in the past. These emotional effects, if experienced, should only last for the duration of the study. I understand that if these effects last beyond the duration of the study, I can contact the Student Counseling Center (Phone: 727-5697) to get information about how to cope with test-related emotions or get a reference of readings on this topic from the Principal Investigator (PI) of this study. My participation in this study is not expected to present any other risk for me or mine than possible emotional reactions. The study involves commonplace activities in educational settings (i.e., taking a test composed of multiple-choice items) and the use of questionnaires that present neither risks nor harm to me.

The results of the study may be published but my name will not be associated with this study. Furthermore, participants in this study will be referred to as Astudents from a historically Black college or university@ in any future publications based on these study data. Last, I will receive a receipt at the end of my participation in this study. This document, signed by the PI, will serve as a proof of participation, and I must provide this document to my professor in order to receive credit in my Introduction to Psychology course.

If I want a copy of the study, I can email the PI (dcotting@gc.cuny.edu) or write to him at: 739 Bexhill Drive, Vinton, VA 24179. In addition, I can call him at (540) 344-7400 or call the chair of his dissertation committee, William E. Cross, Jr., at (212) 817-8712 if I have any questions about this research. If I have questions about my rights as a participant in this study, I can contact Hilry Fisher, Sponsored Research, The Graduate School and University Center of The City University of New York, (212) 817-7523, hfisher@gc.cuny.edu .

_____ (participant) _____(Dave Cotting)

Note: This document was distributed to participants at the HBCU. At the AWCU, the telephone number of the student counseling center was different and the document read (line 23): "participants in this study will be referred to as "students from an All-women College or University".

Appendix B

The Positive and Negative Affect Scale (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now, that is, at the present moment. Use the following scale to record your answers.

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

_____interested	_____irritable	_____anxious*
_____distressed	_____alert	_____angry*
_____excited	_____ashamed	_____fearful*
_____upset	_____inspired	_____disappointed*
_____strong	_____nervous	_____frustrated*
_____guilty	_____determined	
_____scared	_____attentive	
_____hostile	_____jittery	
_____enthusiastic	_____active	
_____proud	_____afraid	

*These items have been added (discrete emotions) and are not part of the PANAS.

Appendix C

The Identity subscale of the Collective Self-Esteem

Please indicate the extent to which you agree or disagree with each of the following statements. If you strongly disagree with the statement, write 1 in the space provided next to the statement. If you strongly agree with the statement, write 7 in the space provided next to the statement. If you somewhat agree or disagree with the statement, use a number between 1 and 7 that best expresses your degree of agreement or disagreement with the statement and write it in the space provided next to the statement.

Please rate your agreement or disagreement with the following statements using the scale below:

1	2	3	4	5	6	7
strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree

_____ Overall, being a student with good scores on tests has very little to do with how I feel about myself.

_____ Being a student with good scores on tests is an important reflection of who I am.

_____ Being a student with good scores on tests is unimportant to my sense of what kind of person I am.

_____ In general, being a student with good scores on tests is an important part of my self-image.

The 1-item question regarding importance of one's test scores

How important are your test scores to you? *(Please circle one)*

1	2	3	4	5
not at all important	slightly important	somewhat important	important	very important

Appendix D

[Diagnostic condition]

As you probably know, mathematical and verbal skills are crucial to performance in many important subjects in college. Yet surprisingly little is known about the various personal factors involved in performance on problems requiring mathematical and verbal reasoning abilities. This research is aimed at better understanding what makes some people better at Math and English than others.

This test is difficult to provide a genuine test of your mathematical and verbal abilities and limitations so that we might better understand the factors involved in all of them. After the test, you will receive feedback that may be helpful to you by familiarizing you with some of your strengths and weaknesses in mathematical and verbal problem solving.

Please try to answer each question as well as you can to help us in our analysis of your mathematical and verbal abilities.

How to answer questions

Please read each item and then mark your response by circling the letter adjacent to the appropriate answer. To change an answer, cross out your previous one and circle your new answer.

Timing

You have 30 minutes to complete the entire test (30 multiple-choice items).

Material

A pen or a pencil. If you need scratch paper, please ask the investigator.

Appendix E

[Nondiagnostic Condition]

This test is a laboratory problem solving task that is not intended as diagnostic of ability. The purpose of the research is to better understand the psychological factors involved in solving mathematical and verbal problems.

This test is difficult because the focus of the research is on difficult mathematical and verbal problems. After the test, you will receive feedback that may be helpful to you by familiarizing you with the kinds of problems that appear on tests you may encounter in the future.

Even though we are not going to evaluate your ability, please try to answer each question as well as you can to help us in our analysis of the problem solving process.

How to answer questions

Please read each item and then mark your response by circling the letter adjacent to the appropriate answer. To change an answer, cross out your previous one and circle your new answer.

Timing

You have 30 minutes to complete the entire test (30 multiple-choice items).

Material

A pen or pencil. If you need scratch paper, please ask the investigator.

Appendix F

Mathematical problems

1) Oil costs $4 per gallon and diesel costs $5.24 per gallon. If the price of oil rises by 10% a month, and the price of diesel is unchanged, how many complete months will it take until a gallon of oil costs more than a gallon of diesel?

A) 2 B) 3 C) 4 D) 5 E) 6

2) If $x > 3$ and $y > -1$, then which of the following is always true?

A) $x < 3y$ B) $-x < 3y$ C) $xy < -3$ D) $-x > 3y$ E) $x < -3y$

3) A salesperson made a series of sales, 30% of sales were $80 each, 45% of sales were $40 each, and the remainder of his sales was $45 each. If he had 600 total sales, what was the total amount he sold?

A) $2,400 B) $9,600 C) $25,000 D) $31,950 E) $51,600

4)

Column A	Column B
1 - 1/27	8/9 + 1/81

A If the quantity in column A is greater than that in column B
B If the quantity in column B is greater than that in column A
C If the two quantities are equal
D If the relationship cannot be determined from the information given

5) If a stamp costs c cents, d stamps cost how many cents?

A) ad/cB) a/dcC) dc/aD) c/adE) acd

6) Bill has exactly enough money to buy 9 bottles of water. If each bottle costs 2 cents less, Bill has exactly enough money to buy 1 more bottle. How much money does Bill have?

A) $20 B) $2.00 C) $1.80 D) $0.18 E) $0.20

7) Three quarts of a bleaching chemical, Minum, contains 5 percent hydrogen peroxide and water. A different type of bleaching chemical, Maxim, which contains 20 percent hydrogen peroxide, will be mixed with the three quarts of Minum. How much of type Maxim should be added to the three quarts of Minum so that the resulting mixture contains 15 percent hydrogen peroxide?

A) 3 quarts B) 3.75 quarts C) 4.5 quarts D) 6 quarts E) 9 quarts

8)

Column A	Column B
The number of months in 7 years	The number of days in 12 weeks

A If the quantity in column A is greater than that in column B
B If the quantity in column B is greater than that in column A
C If the two quantities are equal
D If the relationship cannot be determined from the information given

9) In a certain year, Minnesota produced 2/3 and Michigan 1/6 of all the iron produced in the United States. If all the other states combined produced 18 million tons that year, how many million tons that year, how many million tons did Minnesota produce that year?

A) 27 B) 36 C) 54 D) 72 E) 162

10) The average (arithmetic mean) of x and y is 20. If $z = 5$, what is the average of x, y, z?

A) 81/3 B) 10 C) 12/2 D) 15 E) 17/2

11) N melons cost a total of T dollars, $N > 1$

Column A	Column B
Cost of M melons	M(T/N)

A If the quantity in column A is greater than that in column B
B If the quantity in column B is greater than that in column A
C If the two quantities are equal
D If the relationship cannot be determined from the information given

12) If $0 < s\,t < 1$, then which of the following can be true?

A $s < -1$ and $t > 0$
B $s < -1$ and $t < -1$
C $s > -1$ and $t < -1$
D $s > 1$ and $t < -1$
E $s > 1$ and $t > 1$

13) If n is an even integer, which of the following must be an odd integer?

A) $3n - 2$ B) $3(n + 1)$ C) $n - 2$ D) $n/3$ E) $n/2$

14) A group of 4 pumps are filling a tank. Each of the 3 smaller pumps works at 2/3rds the rate of the largest pump. If all 4 pumps work at the same time, they should fill the tank in what fraction of the time that it would have taken the largest pump if it operated alone?

A) 4/7 B) 1/3 C) 2/3 D) 3/4 E) 4/3

15) The price of a microchip declines by 66% every 6 months. At this rate, approximately how many years will it take for the price of an $81 microchip to reach $1 per chip?

A) 1.5 years B) 2 years C) 2.5 years D) 13 years E) 13.5 years

Verbal Problems

Analogies:

16) CURIOSITY : KNOW

A temptation : conquer
B starvation : eat
C wanderlust : travel
D humor : laugh
E survival : live

17) COLOR : SPECTRUM

A tone : scale
B sound : wave
C verse : poem
D dimension : space
E cell : organism

18) HEADLONG : FORETHOUGHT

A barefaced : shame
B mealymouthed : talent
C heartbroken : emotion
D levelheaded : resolve
E singlehanded : ambition

19) SEDATIVE : DROWSINESS

A epidemic : contagiousness
B vaccine : virus
C laxative : drug
D anesthetic : numbness
E therapy : psychosis

20) LAWYER : COURTROOM

A participant : team
B commuter : train
C gladiator : arena
D senator : caucus
E patient : ward

21) FRUGAL : MISERLY

A confident : arrogant
B courageous : pugnacious
C famous : aggressive
D rash : foolhardy
E quiet : timid

22) ANTIDOTE : POISON

A cure : recovery
B narcotic : sleep
C stimulant : relapse
D tonic : lethargy
E resuscitation : breathing

Antonyms:

23) MULTIFARIOUS

A deprived of freedom
B deprived of comfort
C lacking space
D lacking stability
E lacking diversity

24) DIFFUSE

A contend
B concentrate
C imply
D pretend
E rebel

25) FALLACY

A personal philosophy
B imaginative idea
C unconfirmed theory
D tentative opinion
E valid argument

26) ADULTERATION

A consternation
B purification
C normalization
D approximation
E rejuvenation

27) DEPOSITION

A process of congealing
B process of distilling
C process of eroding
D process of evolving
E process of condensing

Sentence completion:

28) Early _____ of hearing loss is _____ by the fact that the other senses are able to compensate for moderate amount of loss, so that people frequently do not know that their hearing is imperfect.

A discovery... indicated
B development... complicated
C detection... complicated
D treatment... facilitated
E incidence... corrected

29) The _____ science of seismology has grown just enough so that the first overly bold theories have never been _____ .

A magnetic... accepted
B fledgling... refuted
C tentative... analyzed
D predictive... protected
E exploratory... recalled

30) Since 1813 reaction to Jan Austen's novels have oscillated between _____ and condescension; but in general later writers have esteemed her works more highly than did most of her literary _____ .

A dismissed... admirers
B adoration... contemporaries
C disapproval... readers
D indifference... followers
E approbation... precursors

Appendix G

The Emotion Intensity questionnaire

Please, answer the following questions about the emotion you may have experienced while taking the test.

1) How intense was your emotional feeling at its peak?

*(Please circle one number)***

1	2	3	4	5
little intense	somewhat intense	rather intense	quite intense	very intense

2) How intense was your emotional feeling during the whole emotional episode (thus not only at the peak)? *(Please circle one number)***

1	2	3	4	5
little intense	somewhat intense	rather intense	quite intense	very intense

3) If you didn't like taking the test. How strong was that dislike?

(Please circle one number)

1	2	3	4	5
very weak	weak	rather strong	strong	very strong

4) If you enjoyed taking the test. How strong was that enjoyment?

(Please circle one number)

1	2	3	4	5
little intense	somewhat intense	rather intense	quite intense	very intense

5) To what extent did the emotion and the events change your opinion about or feelings towards test or exams you take? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

6) To what extent did the emotion and the events change your opinion about or feelings towards your performances in tests or exams?
*(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

7) To what extent did the emotion and the events change your opinion about or feelings towards test scores in general? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

8) To what extent did the emotion and the events change your opinion about or feelings towards psychology research? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

9) To what extent did the emotion and the events change your opinion about or feelings towards yourself? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

10) To what extent do you think the emotion and the events will change your long-term behavior when you take a test or an exam? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

11) To what extent do you think the emotion and the events will change your long-term behavior when you receive your score on a test or an exam? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

12) To what extent do you think the emotion and the events will change your everyday life? *(Please circle one number)**

1	2	3	4	5
not at all	slightly	somewhat	rather strongly	very strongly

13) On the whole, how intense was your emotion while taking the test? *(Please circle one number)***

1	2	3	4	5
or not at all				

* Items included in the belief changes and influence upon long-term behavior dimension of emotion intensity.
**Items included in the overall felt intensity of emotion dimension of emotion intensity.

Appendix H

GENEVA APPRAISAL QUESTIONNAIRE
Assessment of Emotion-Eliciting events
Version 2.1

Instructions

In the following questionnaire, please recall moments when you took the test in this study (this is referred to as Event X). Try to recall the details of the eliciting event(s) as clearly as possible.

Please respond to the questions on the following pages by placing a check mark in the appropriate space for the respective scale. If a particular question does not make sense in the specific situation in which you were (i.e., taking the test), please mark the circle "does not apply." It is extremely important that you answer all the questions and that you select only one alternative.

Event X

Please describe the event that produced your emotional experience in a few sentences, mentioning what happened and the consequences this had for you.

_____

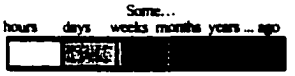
_____

_____

_____

Occurrence of the emotional experience

1. **How long ago** did this emotional experience occur?



2. **Where** were you when you experienced this emotion?

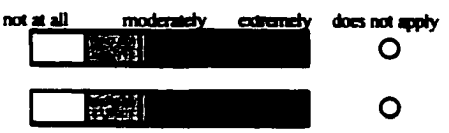
In my own home
In the home of friends or acquaintances
At work
In a public building or in a stranger's home
On a (motor)bike, in a car, bus, train, or plane
In the street or another public space
In a natural setting

3. **Who was present** when you experienced the emotion?

Nobody, I was alone
A partner or friend
Another person (acquaintance or colleague)
Several friends or acquaintances
One or more persons unknown to me
A large crowd

General evaluation of the event

How would you evaluate this type of event **in general**, independent of your specific needs and desires in the situation you reported above? *(Note: To allow assessing ambivalent situations, we ask you to respond to both scales.)*



4. **pleasant**

5. **unpleasant**

Characteristics of the event

At the time of experiencing the emotion, did you think that ...

		not at all — moderately — extremely	does not apply
6.	the event happened very **suddenly and abruptly**?		○
7.	you could have **predicted** the occurrence of the event?		○
8.	you were **familiar** with this type of event?		○
9.	the event would have **very important consequences** for you?		○
10.	the actions that produced the event were **morally and ethically acceptable**?		○
11.	the actions that produced the event **violated laws or social norms**?		○

Causation of the event *(Please note that, in many cases, several causes can be involved.)*

At the time of the event, to what extent did you think that one or more of the following factors caused the event?

		not at all — moderately — extremely	does not apply
12.	**chance, special circumstances**, or **natural forces**		○
13.	**your own** behavior		○
14.	– If so, did you cause the event **intentionally**?		○
15.	– If so, was your behavior **consistent with the image you have of yourself**?		○
16.	the behavior of **one or more other person(s)**		○
17.	– If so, did (this) these other person(s) cause the event **intentionally**?		○

Consequences of the event

At the time of experiencing the emotion, did you think that the real or potential consequences of the event ...

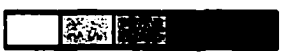
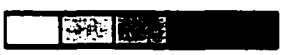
		not at all — moderately — extremely	does not apply
18.	had already been **felt** by you or were **completely predictable**?		○
19.	had been **expected** to occur at **that** time and in **that** specific form?		○
20.	could be **clearly envisaged** and might occur in the **near future** (with a fairly **high** probability)?		○
21.	were **rather unpredictable** but might occur in the **distant future** (with **uncertain** probability)?		○

Continued...At the time of experiencing the emotion, did you think that real or potential consequences of the event ...

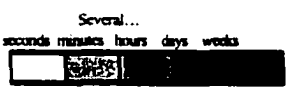
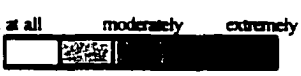
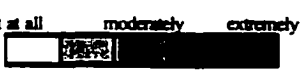
		not at all — moderately — extremely	does not apply
22.	did or would bring about **positive, desirable outcomes** for you (e.g., helping you to reach a goal, giving pleasure, or terminating an unpleasant situation)?		○
23.	did or would bring about **negative, undesirable outcomes** for you (e.g., preventing you from reaching a goal or satisfying a need, resulting in bodily harm, or producing unpleasant feelings)?		○
24.	were or would be **unjust** or **unfair**?		○
25.	could have been or could still be **avoided** or **modified** by **appropriate human action**?		○

Reactions with respect to the real or expected consequences

After you had a good idea of what the probable consequences of the event would be, did you think ...

		not at all — moderately — extremely	does not apply
26.	that it was **urgent** to act immediately?		○
27.	that you would be able to **avoid the consequences** or **modify** them to **your advantage** (through your own power or helped by others)?		○
28.	that you could **live with**, and **adjust to**, the consequences of the event that could not be avoided or modified?		○

Intensity and duration of the emotional experience

29.	How **intense** was the feeling that you experienced during the emotional episode described above?	weak — moderate — strong 	
30.	How **long** did the emotional experience last?	Several... seconds minutes hours days weeks 	
31.	To what extent did you try to **reduce the intensity** of your emotional experience and to **shorten its duration**?	not at all — moderately — extremely 	
32.	To what extent did you try to **control** or **mask the expression** of your feelings to keep them from being observed by others?	not at all — moderately — extremely 	

Verbal description of the emotional experience

33. How would you describe this emotional experience **in your own words**? Please write a word or a short expression in the box provided to the right.

34. Please decide which of the emotion terms listed below **corresponds best** to the emotional experience you reported above. Identify the term that comes closest to what you felt with a check mark. If you experienced an **"emotion blend"**, or two different emotions simultaneously, you can check two of the terms. In this case, please identify the stronger of the two emotions with two check marks.

Sadness	
Joy	
Rage	
Anxiety	
Surprise	
Fear	
Irritation	
Shame	
Contempt	
Guilt	
Disgust	
Pleasure	
Despair	
Pride	
None of the emotion terms above corresponds to what I felt during this emotion episode	

Appendix I
Debriefing

The purpose of the study in which you participated is to investigate the role emotion plays in test-taking situations that invoke a phenomenon known as stereotype threat. The idea that stereotypes are detrimental to members of stereotyped groups is at the center of the stereotype threat theory. Stereotype threat is a phenomenon that occurs when members of groups about which negative stereotypes exist are reminded of the possibility of confirming these negative stereotypes. As the thinking goes, the stereotype threat induces anxiety, which in turn interferes with performance and identification with the domain (e.g. school).

However, discussions about anxiety have been relatively few in the literature on stereotype threat, and some psychologists in the field even have used anxiety and fear interchangeably to describe this state. As a result, studies have not yielded the expected evidence about the role of anxiety in stereotype threat.

You were randomly assigned to one of two conditions. One condition (i.e., the experimental or, here, stereotype threat condition) described the study as diagnostic of one's ability. The other condition (i.e., control or no stereotype threat condition) described the study as non-diagnostic.

In fact, these descriptions are used as a way to manipulate stereotype threat. At times, and regardless of whether or not one believes stereotypes to be true, test/exam conditions have been shown to elicit stereotype threats. When that happens, one may feel different emotions and/or affects and his/her performance, as a result, may be impaired. The questionnaires about emotion intensity and a number of specific emotions you completed before and after the test will be used to determine whether there is support for this hypothesis.

I am available to answer any questions you may have about the study; I can be reached by email at dcotting@gc.cuny.edu. If you would like to receive an overview of the study's findings (expected availability Summer-Fall 2002), please contact me by email.

Thank you for your participation and effort.

Appendix J

Post-Hoc Analyses: Model A

Because the data did not support the study hypotheses, I turned to the specific emotion appraisals that participants reported on the Geneva Appraisal Questionnaire (GAQ) and notes from verbal exchanges I had with participants during the debriefing sessions to explore further the role of emotion in the stereotype threat phenomenon. However, it is important to note the following analyses are post-hoc in nature and should not be viewed as evidence for any particular explanation of stereotype threat. At most, the following analyses provide information for future studies on stereotype threat and emotion.

As noted, the GAQ can be used to assess, through recall and verbal report, an individual's appraisal process in the case of a specific emotional episode. Of particular interest in the case of a stereotype threat experience is coping potential, one of the appraisal criterion in Scherer's model. Based on observations made in past research on appraisals and identity (Cotting & Deaux, 1999), I suspected that an individual exposed to a stereotype threat would cope with the real or expected consequences of the situation by boosting one's sense of "power" (i.e., capability to avoid the consequences of the event or modify them to one's advantage and distancing him/herself from the threatening situation). "Power" depends on the individual's evaluation of the resources at his/her disposal to change contingencies and outcomes according to his/her interests (Scherer, 2001). In the case of stereotype threat, the individual may resort, for example, to believing that a sub-optimal performance in the test would not, despite the test description, accurately assess one's true abilities and limitations. I further suspected that a greater sense of "power" would increase dislike for the test; participants who relied on this strategy would dislike the test more than

those who didn't resort to such strategy. Indeed a greater sense of "power" may allow the individual to disapprove the task or the object of the task and translate into a greater dislike for the task or the object of the task (Lazarus & Folkman, 1984). Such dislike, in turn, may have an impact on one's test performance.

Next, I reviewed notes taken during debriefing discussions at the HBCU institution. Before describing the research methods and the details of the experiment, I started the debriefing sessions by asking how participants felt. Common responses to this inquiry suggested that some participants a) disliked taking the test and b) were not interested in hearing anything about the test, test taking, or test performance in general. However, whereas these participants did not alter their evaluation of their experience while taking the test, almost all expressed keen interest in stereotype threat research after hearing about it. Herein lies one possible element of the affective response to stereotype threat: stereotype threat may reduce interest; lower interest, in turn, may have a negative impact on test performance. According to the theory of stereotype threat, disidentification with the stereotyped domain is used as a strategy to prevent the domain from being used as a basis for self-esteem and self-evaluation. I propose that this strategy, in addition to having long-term consequences, may diminish the individual's interest in the domain at the time of evaluation and cause sub-optimal performance in the domain. To test these ideas, I calculated bivariate correlations among the following constructs (see Table 5):

a. Sense of "power" (GAQ item #27: "After you had a good idea of what the probable consequences of the event would be, did you think that you would be able to avoid the consequences or modify them to your advantage?").

b. Interest (i.e., interest at T2 minus interest at T1, PANAS item #1: "indicate to what extent you feel interested right now, that is, at the present moment").

c. Condition (diagnostic, coded as 0; non-diagnostic, coded as 1)

d. Verbal test score

e. Dislike (Emotion intensity item #3: "If you didn't like taking the test. How strong was that dislike?").

Condition was related to sense of "power" ($r=-.277$, $\alpha=.033$) and interest ($r=.307$, $\alpha=.020$); stereotype-threatened participants reported lower interest and a greater sense of "power" than non-threatened participants. Interest was positively related to verbal test performance ($r=.284$, $\alpha=.029$); the greater one's interest difference score (i.e., T2 minus T1), the better the verbal test performance. Further, sense of "power" was positively related to dislike for the test ($r=.292$, $\alpha=.026$); the greater one's sense of "power", the greater the dislike. Last, dislike was negatively related to verbal test performance ($r=-.330$, $\alpha=.013$); the greater the dislike, the lower the test performance.

To test the fit of the model in which stereotype threat causes change in verbal test performance through two paths involving sense of "power," dislike for the task, and interest (Figure 3), path analysis was conducted using LISREL and covariance matrices. Following the guidelines of 5-10 observations for each parameter (Klem, 1995), the recommended sample size for this path analysis is 25-50 participants; the model was evaluated using the covariance matrix based on 45 observations. Again, three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root

Mean Residual (RMR), were used to assess the fit of the model. The data, as indicated by the nonsignificance of the chi-square statistic (X^2=6.72, α=.24), suggest that the observed patterns are not significantly different than the expected patterns. In other words, the observed data fit the proposed model.

The second criterion, GFI, represents the degree of fit between the residuals of the expected data and the observed data without adjusting for the degrees of freedom. The range of values for GFI is 0 (poor fit) to 1 (perfect fit). The GFI for this model was .94, indicating that the fit between the observed and expected matrices was high. The third criterion, standardized RMR, is the average residual between the expected results and the observed data. A small (i.e., close to 0) RMR indicates that the difference between the matrices is small. For this model, the standardized RMR was .079. Although no guidelines have been established for the RMR, the index for this model suggests that the data and the model are similar.

Specific Path Analyses: Model A

To test the contribution of each path to the model, I calculated maximum likelihood estimates for each of them (Figure 4). All five paths were significant: the path from condition to interest, Beta = .79, z = 2.11, p < .017; the path from interest to verbal test performance, Beta = .57, z = 1.84, p < .033; the path from condition to sense of "power", Beta = -.73, z = -1.89, p < .029; the path from sense of "power" to dislike, Beta = .29, z = 2.00, p < .023; and the path from dislike to verbal test performance, Beta = -.68, z = -2.21, p < .014.

Generalizibility of Model A

To test whether Model A is applicable to threats related to other stereotypes (e.g., stereotype about women and math), I calculated bivariate correlations among the constructs of Model A for both the AWCU sub-sample (N=49) and the overall sample (combined HBCU and AWCU sub-samples, N=94) (see Tables 6a and 6b). The mathematical test score was used for the AWCU sub-sample to assess the impact of the threat of the stereotype about women and math on mathematical performance. Similarly, the total test score (i.e., combined verbal and mathematical test scores) was used to assess the impact of the threat of the stereotype about Black and academic performance and the stereotype about women and math for the overall sample in one variable. Even though stereotype threat was not significantly related to mathematical test performance in the AWCU sub-sample (r=.048, NS), stereotype-threatened individuals had on average slightly lower math scores than nonstereotype-threatened individuals. However, mathematical test performance was negatively related to dislike for the task (r=-.488, α=.000) and sense of "power" (r= -.276, α=.027).

In the overall sample, stereotype threat was related to total test score (r=.256, α=.006); stereotype-threatened participants had a lower test score than non-threatened participants. Condition was also significantly related to interest (r=.201, α=.026) but not significantly related to sense of "power" (r=-.095, NS); stereotype-threatened participants reported a lower interest than non-threatened participants. The relationship between interest and total test score was marginal (r=.153, α=.071). Further, sense of "power" was not significantly related to dislike for the test (r=.108, NS). Last, dislike was negatively related to total test score (r=-.407, α=.000); the greater the dislike, the lower the total test score.

Despite the above results, the fit of Model A (Figure 3) was tested for both the AWCU sub-sample (N=49) and the overall sample (N=94) conducting path analysis with LISREL and covariance matrices. Following the guidelines of 5-10 observations for each parameter (Klem, 1995), the recommended sample size for this path analysis is 25-50 participants; the model was evaluated using the covariance matrix based on 49 observations for the AWCU sub-sample and 94 observations for the overall sample. Again, three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root Mean Residual (RMR), were used to assess the fit of the model.

The data from the AWCU sub-sample, as indicated by the nonsignificance of the chi-square statistic (X^2=7.06, α=.22), suggest that the observed patterns are not significantly different than the expected patterns. In other words, the observed data fit the proposed model. GFI, the second criterion, was .94 for this model indicating that the fit between the observed and expected matrices was high. The third criterion, standardized RMR, was .091 indicating that the difference between the matrices was small. Although no guidelines have been established for the RMR, the index for this model suggests that the data and the model are similar. To test the contribution of each path to the model, I calculated maximum likelihood estimates for each of them (Figure 5). Only one path was significant: the path from dislike to math test performance, Beta = -1.06, z = -3.84, p < .000.

The data from the overall sample, as indicated by the significance of the chi-square statistic (X^2=13.32, α=.021), suggest that the observed patterns are significantly different than the expected patterns. In other words, the observed data did not fit the

Model A for the overall sample. Thus, there was no need to look beyond the significant chi-square statistic to conclude that this model did not fit the data for the overall sample. These findings suggest that Model A is generalizable to the AWCU sub-sample but not to the overall sample.

Post-Hoc Analyses: Model B

In search for any supplementary information provided by the GAQ that would strengthen the model further for the HBCU sub-sample, I analyzed participants' reports of the term that came closest to what they felt (i.e., item #34). Fourteen emotions (i.e., sadness, joy, rage, anxiety, surprise fear, irritation, shame, contempt, guilt, disgust, pleasure, despair, pride), alone or in combination (i.e., an "emotion blend"), offered participants 196 possibilities for any one emotion or two-emotions blend choice (e.g., contempt and guilt, joy and pride). I first examined the primary responses among the 15 choices (i.e., 14 emotions + "none of the above emotions" option) (Appendix H, item 34). 37.1% of the participants marked "irritation" as the primary emotion felt. As such, I created a dummy variable to assess the possible contribution of irritation in the phenomenon of stereotype threat. Next, I calculated bivariate correlations among the constructs involved in the model (i.e., condition, sense of "power", dislike, interest, and verbal test performance) and irritation. In addition to the correlations described above, irritation was positively related to sense of "power" ($r=.312$, $\alpha=.019$), dislike ($r=.409$, $\alpha=.003$); the greater the feeling of irritation, the greater the sense of "power" and the dislike for the test.

To test the fit of a model in which sense of "power" caused changes in dislike through irritation (Figure 6), path analysis was conducted using LISREL and covariance

matrices. The sample size recommended for this path analysis was 30-60 participants (i.e., 5-10 for each parameter) (Klem, 1995). This requirement was met for this model (N=45). Again, three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root Mean Residual (RMR), were used to assess the quality of the model. The data, as indicated by the nonsignificance of the chi-square statistic (Chi-Square=11.17, α=.26), suggest that the patterns observed are not significantly different than the expected patterns (i.e., the model).

The second criterion, GFI, represents the degree of fit between the residuals of the expected data and the observed data without adjusting for the degree of freedom. The GFI for this model was .92, indicating that the fit between the observed and expected matrices was high. The third criterion, standardized RMR, is the average residual between the expected results and the observed data. For this model, standardized RMR was .090. Although no guidelines have been established for the RMR, the index for this model suggests that the data and the model are similar.

Specific Path Analyses: Model B

To test the contribution of each path to the model, I calculated maximum likelihood estimates for each of them (Figure 7). All six paths were significant: the path from condition to interest, Beta = .79, z = 2.11, p < .017; the path from interest to verbal test performance, Beta = .57, z = 1.84, p < .033; the path from condition to sense of "power", Beta = -.73, z = -1.89, p < .029; the path from sense of "power" to irritation, Beta = .12, z = 2.16, p < .015; the path from irritation to dislike, Beta = 1.08, z = 2.95, p < .002; and the path from dislike to verbal test performance, Beta = -.68, z = -2.21, p < .014.

Generalizability of Model B

To test whether Model B is applicable to threats related to other stereotypes (i.e., stereotype about women and math), I calculated bivariate correlations among the constructs of Model B and mathematical test performance in the AWCU sub-sample (N=49), and the constructs of Model B and total test score (i.e., verbal and mathematical test scores) for the overall sample (N=94) (see Tables 8a and 8b).

In the AWCU sub-sample and in addition to the correlations described for Model A, sense of "power" was negatively related to irritation (r=-.248, α=.043). In the overall sample, irritation was positively related to dislike (r=.300, α=.002); the greater the feeling of irritation, the greater the dislike for the test.

To test the fit of Model B (Figure 6) for the AWCU sub-sample, path analysis was conducted using LISREL and covariance matrices. The sample size recommended for this path analysis was 30-60 participants (i.e., 5-10 for each parameter) (Klem, 1995). This requirement was met for this model (N=49). Again, three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root Mean Residual (RMR), were used to assess the quality of the model. The data, as indicated by the nonsignificance of the chi-square statistic (Chi-Square=13.23, α=.15), suggest that the patterns observed are not significantly different than the expected patterns (i.e., the model). The GFI for this model was .92, indicating that the fit between the observed and expected matrices was high. The third criterion, standardized RMR, was .11 for Model B, which suggests that the data and the model are similar.

To test the fit of Model B (Figure 6) for the overall sample, path analysis was conducted

using LISREL and covariance matrices. The sample size recommended for this path analysis was 30-60 participants (i.e., 5-10 for each parameter) (Klem, 1995). This requirement was met for this model (N=94). Again, three commonly used goodness-of-fit indices, chi-square statistic, Goodness-of-Fit Index (GFI), and Root Mean Residual (RMR), were used to assess the quality of the model. The data, as indicated by the nonsignificance of the chi-square statistic (Chi-Square=15.87, α=.070), suggest that the patterns observed are not significantly different than the expected patterns (i.e., the model). The GFI for this model was .95, indicating that the fit between the observed and expected matrices was high. The third criterion, standardized RMR, was .092 for Model B, which suggests that the data and the model are similar.

To test the contribution of each path to Model B for the AWCU sub-sample, I calculated maximum likelihood estimates for each of them (Figure 8). Of the six paths, two were significant: the path from sense of "power" to irritation, Beta = -.09, z = -1.75, p < .040; the path from dislike to verbal test performance, Beta = -1.06, z = -3.84, p < .000. In addition, the path from irritation to dislike was marginally significant, Beta = .50, z = 1.38, p < .083.

To test the contribution of each path to the model for the overall sample, I calculated maximum likelihood estimates for each of them (Figure 9). Of the six paths, three were significant: the path from condition to interest, Beta = .47, z = 1.96, p < .025; the path from irritation to dislike, Beta = .77, z = 3.02, p < .001; and the path from dislike to verbal test performance, Beta = -1.47, z = -4.20, p < .000. These findings suggest that Model B is generalizable to both the AWCU sub-sample and the overall sample.

Model comparison

One way to compare non-nested models is the use of the Bayesian Information Criterion (BIC)1 (Raftery, 1993; 1995). The difference between two BIC-values can be viewed as an approximation of the Bayes factor2, the ratio of (marginal) likelihoods of the data given the two competing models (Voeten, personal communication, September 27, 1997). As the BIC gets smaller (as a negative number), the fit increases. A difference of at least 6 between two models' BIC indices can be interpreted as evidence that the model with the smaller BIC value fits the data better than the other model3. Although the BIC may be considered problematic for selecting models because of its rationale in Bayesian statistics and not classical significance testing, it provides one of the best measures to evaluate competing non-nested models.

For the HBCU sub-sample (N=45), the BIC for Model A was –22.711 and the BIC for Model B was –44.541. For the AWCU sub-sample (N=49), the BIC for Model A was –22.796 and the BIC for Model B was –42.520. Using the BIC, it appears that Model B is better than Model A for the both the HBCU and AWCU sub-samples. This conclusion, as well as the lack of fit of Model A for overall sample, suggests that the contribution of irritation, the only distinction between Model A and B, may be important for the phenomenon.

---

1 BIC = Chi-square – df x ln (n (p + q)), where p is the number of y-variables, and q is the number of x-variables (Raftery (1993).
2 BIC only gives an approximation of the Bayes factor; computing the actual Bayes factor in the context of Structural Equation Modeling (SEM) is beyond the scope of this paper.
3 A difference of 6 means that the odds for the two competing models are 20:1 (i.e., 2 x ln (6)=20).

Appendix K

Stereotype threat to sense of "power"

In many situations, one's assessment of coping potential is socially grounded (Scherer, 2001); an individual's "power" is dependent upon others' real or imagined help or emotional support (expressed, for example, through empathy, compassion, encouragement, understanding). The condition (i.e., stereotype threat or control) was related to sense of "power" regarding the consequence of the event (i.e., the power to avoid the consequences of the event or modify them to one's advantage), suggesting that stereotype threat increased sense of "power". An interpretation of this finding is that one of the reactions to the situational phenomenon of stereotype threat is to distance oneself from the situation by boosting one's sense of "power"; rather than remain passive subjects of the consequences of the event, individuals seem to adopt an empowered outlook. Doing so may be an attempt on the part of the stereotype-threatened individual to maintain control over the possible interpretations of his/her performance in the stereotyped domain. In other words, the stereotype-threatened individual may be creating a safeguard for him/herself in the event that his/her performance confirms the negative stereotype about his/her group.

Sense of "power" to irritation

Sense of "power" was positively related to irritation, suggesting that as sense of "power" increases, the feeling of irritation increases. This path indicates that irritation is associated with a specific pattern of emotional appraisal involving high "power" or control and low threat (similar to Smith & Ellsworth, 1985). It is worth noting that a greater threat may have generated anger if the agent's (i.e., the person or group

responsible for the test) power had been perceived as low, or fear if the agent's power was perceived as high (Scherer, 2001).

Irritation to dislike

Irritation was positively related to dislike for the task (i.e., test), suggesting that as irritation increases, the dislike for the task increases. Not surprisingly, irritation negatively affects liking for the irritation-causing task. It is noteworthy to highlight that the involvement of dislike, whose appraisal belongs to the distancing family of appraisals (Roseman, 2001), confirms that this set of emotional reactions may be an expression of the stereotype-threatened individual's attempt to distance him/herself from the situation.

Dislike to verbal test performance

The path from dislike for the task to test performance was also supported, indicating that as dislike for the task increases, test performance decreased. This finding supports the theory that dislike, and its behavioral response component, a decrease in attention (Roseman, 2001), is not conducive to optimal performance in the task of test taking.

Stereotype threat to interest

Individuals usually do not experience a single emotion at a time, but react to emotion-provoking events with combinations of emotions (Ellsworth & Smith, 1988; Izard, 1972; Johnson-Laird & Oatley, 1992; Plutchik, 1991; Scherer & Ceschi, 1997). In addition to irritation, the condition (i.e., stereotype threat or control) appeared to cause a change in a co-occurring feeling: interest. Indeed, stereotype threat was related to interest, suggesting that individuals reacted to the stereotype-threatening situation by lowering their interest (or increasing their disinterest). In addition to boosting one's sense

of "power", lowering one's interest can be interpreted as yet another expression of how individuals seem to distance themselves from a stereotype-threatening situation. In other words, the stereotype-threatened individual may lower his/her interest to create another safeguard for him/herself in case his/her performance confirmed the negative stereotype about his/her group. This suggests that disidentification with the stereotyped domain is not only a long-term but also a short-term strategy to prevent the domain from being used as a basis for self-esteem and self-evaluation.

Interest to verbal test performance

Interest was positively related to test performance, suggesting that as interest decreases, verbal test performance decreases. Again, this finding indicates that low interest, perhaps by decreasing attention, is not conducive to optimal performance in a domain.

To assess the generalizability of Model B, post-hoc investigations were conducted with both the AWCU sub-sample and the overall sample. It is important to note that stereotype threat impacted test performance in the overall sample, but not in the AWCU sub-sample. The statistical analyses indicated that Model B fits the data for both the AWCU sub-sample and the overall sample. These results substantiate that Model B, if corroborated by future studies, may help us to understand stereotype threat as it appears in any stereotyped group.

References

Arnold, M. B. (1960a). Emotion and Personality: Vol. 1. Psychological aspects. New York, NY: Columbia University Press.

Arnold, M. B. (1960b). Emotion and Personality: Vol. 2. Physiological aspects. New York, NY: Columbia University Press.

Aronson, J., & Salinas, M. F. (1997). Stereotype threat, attributional ambiguity, and Latino performance. Unpublished manuscript, University of Texas.

Aronson, J., Lustina, M., Good, C., Keough, K., Steele, C. M., & Brown, J. L. (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. M., & 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.

Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. European Journal of Social Psychology, 16(4), 361-383.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. Science, 222, 1029-1031.

Blascovich, J., Spencer, S. J., Quinn, D., Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. Psychological science, 12 (3), 225-229.

Broadnax, S., Crocker, J., & Spencer, S. (1997). African-Americans and academic performance: The mediational role of stereotype vulnerability. Unpublished manuscript,

State University of New York at Buffalo.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. Journal of Personality and Social Psychology, 76(2), 246-257.

Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self and the emergence of racial identification in Negro preschool children. The Journal of Social Psychology, 10, 591-599.

Cotting, D., & Deaux, K. (1999). Affective consequences of social identification. Poster presented at the American Psychological Society Annual Convention, Denver, Colorado.

Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Orlando, FL: Hardcourt Brace Jovanovich College Publishers.

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

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. Psychological Review, Vol 94(3), 369-389.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5-18.

Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between various types of affect, Journal of Personality and Social Psychology, 47, 871-883.

Du Bois, W. E. B. (1903). The Souls of Black Folk. A. C. McClurg & Company.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6,

169-200.

Ekman, P., & Davidson, R. J. (1994) (Eds.). The nature of emotion, (pp. 146-149). New York, NY: Oxford University Press.

Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion,12, 271-302.

Forgas, J. P. (1991). Emotion and Social judgements. Oxford: Pergamon Press.

Frijda, N. H. (1986). The Emotions. Cambridge: Cambridge University Press.

Frijda, N. H. (1996). Passions: Emotion and socially consequential behavior. In R. D. Kavanaugh, B. Zimmerberg, S. Fein (Eds.). Emotion: Interdisciplinary perspectives, (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum.

Frijda, N. H., Orthony, A., Sonnemans, J., Clore, G. L. (1992). The complexity of intensity: Issues concerning the structure of emotion intensity. In M. S. Clark (Ed.), Emotion: Review of Personality and Social Psychology, (pp.60-89). Newbury Park, CA: Sage.

Goffman, I. (1963). Stigma. Simon & Shuster.

Gould, D., Petlichkoff, L., Simons, J., & Vevera, M. (1987). The relationship between Competitive State Anxiety Inventory-2 subscale scores and pistol shooting performance. Journal of Sport Psychology, 9, 33-42.

Gould, D., & Udry, E. (1994). Psychological skills for enhancing performance: Arousal regulation strategies. Medicine and Science in Sports and Exercise, 26, 478-485.

Howard, J., & Hammond R. (1985). Rumors of inferiority. New Republic, 72, 18-23.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in

mathematics performance: A meta-analysis. Psychological Bulletin, 107, 139-155.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychological Science, 11(5), 365-371.

Izard, C. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. Psychological Bulletin, 115, 288-299.

Johnson-Laird, P. N., Oakley, K. (1992). Basic emotions, rationality, and folk theory. In N. L. Stein & K. Oakley (Eds.), Basic Emotions (pp.201-223). Hove, UK: Erlbaum.

Klem, L. (1995). Path analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 65-97). Washington, DC: American Psychological Association.

Lang, P. J. (1984). Cognition in emotion: concept and action. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.). Emotions, cognitions and behavior, (pp. 192-226). Cambridge University Press.

Lazarus, R. S. (1966). Psychological Stress and the coping process. New York: McGraw-Hill.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. American Psychologist, 46, 819-834.

Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer.

Le Gall, A. (1980). Les insuccès scolaires (8th ed.) [The school failures]. Paris, France: Presses Universitaires de France.

Le Gall, A. (1992). L'anxiété et l'angoisse (4th ed.) [Anxiety and Anguish]. Paris, France: Presses Universitaires de France.

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. Psychological Science, 3, 23-27.

Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observer. Journal of Personality and Social Psychology, 49, 918-926.

Luhtanen, R., Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. Personality and Social Psychology Bulletin, 18 (3), 302-318.

McGuire, W. J., McGuire, C. V., & Winton, W. (1979). Effects of household sex composition on the salience of one's gender in the spontaneous self-concept. Journal of Experimental Social Psychology, 15, 77-90.

Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagment. In J. K. Swim & C. Stangor (Eds.), Prejudice: The target's perspective (pp. 219-241). San Diego, CA: Academic Press.

Manstead, A., & Fischer, A. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal Processes in Emotion (pp. 221-232). New York: Oxford University Press.

Mesquita, B., & Ellsworth, P. C. (2001). The role of culture in appraisal. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal Processes in Emotion (pp. 233-248). New York: Oxford University Press.

Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997) Culture and emotion. In J. E.

Berry, P. B. Dasen, & T. S. Saraswathi (Eds.), Handbook of cross-cultural psychology: Vol. 2. Basic processes and developmental psychology (pp. 255-297). Boston: Allyn & Bacon.

Pittner, M. S., & Houston, B. K. (1980). Response to stress, cognitive coping strategies, and the Type A behavior. Journal of Personality and social Psychology, 39(1), 147-157.

Plutchik, R. (1991). The emotions: Facts, theories, and a new model (rev. ed.) Lanham, MD: University Press of America.

Raftery, A. E. (1993). . Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models, (pp. 163-180). Newbury Park, CA: Sage.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), Sociological Methodology. Oxford: Blackwell.

Roseman, I. J. (1984). Cognitive determinants of Emotion: A structural theory. In P. Shaver (Ed.), Review of Personality and Social Psychology, vol. 5, (pp. 11-36). Beverly Hills, CA: Sage.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal Processes in Emotion (pp. 68-91). New York: Oxford University Press.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal Processes in Emotion (pp. 3-19). New York: Oxford University Press.

Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental

effects of distinctiveness and performance monitoring. Social Cognition, 12, 61-74.

Saenz, D. S., & Lord, C. G. (1989). Reversing roles: A cognitive strategy for undoing memory deficits associated with token status. Journal of Personality and Social Psychology, 56, 698-708.

Sarason, I. G. (1972).Introduction to the study of test anxiety. In I. G. Sarason (Ed.), Test Anxiety: Theory, Research, and Applications. Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1981). Wider die Vernachlässigung der Emotion in der Psychologie [On the neglect of emotion in psychology]. In W. Michaelis (Ed.), Bericht über den 32. Kongress der Deutschen Gesellschaft für Psychologie, Zürich, 1980 (pp. 304-317). Göttingen: Hogrefe Verlag.

Scherer, K. R. (1984). Emotion as a multi component process: A model and some cross-cultural data. In P. Shaver (Ed.), Review of Personality and Social Psychology, vol. 5, (pp. 37-63). Beverly Hills, CA: Sage.

Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. Geneva studied in Emotion and Communication, 1, 1-98.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal Processes in Emotion (pp. 92-120). New York: Oxford University Press.

Scherer, K. R., Ceschi, G. (1987). Lost luggage: A field study of emotion-antecedent appraisal. Motivation and Emotion, 21, 211-235.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. Journal of Personality and Social

Psychology, 52, 1061-1086.

Smith, T. W. (1990). Ethnic Images. (GSS Topical Report, 19). University of Chicago, National Opinion Center.

Smith, C. A., Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. Journal of Personality and Social Psychology, 48, 813-838.

Sonnemans, J. (1991). Structure and determinants of emotional intensity. Doctoral dissertation, Amsterdam University, Department of Psychology.

Sonnemans, J., Frijda, N. H. (1994). The structure of subjective emotional intensity. Cognition and Emotion, 8 (4), 329-350.

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

Spielberger, C. D., Gorsuch, R., Lushene, R. (1970). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Anton, E. D., Algaze, B., Ross, G. R., & Westberry, L. G. (1980). Test Anxiety Inventory. Sampler set: manual, test, scoring. Redwood City, CA: Mind Garden.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52(6), 613-629.

Steele, C. M. (1999). Thin ice: "stereotype threat" and Black college students. Atlantic Monthly, 282 (2), 44-54.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. Journal of Personality and Social Psychology, 68(11), 797-811.

Steele, C. M., & Aronson, J. (1998). Stereotype Threat and the Test Performance of Academically Successful African Americans. In C. Jencks & M. Phillips (Eds.), The Black-White Test Score Gap. Washington, D.C.: Brookings Institution Press.

Stone, J., Sjomerling, M., Lynch, C. I., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. Journal of Personality and Social Psychology, 77(6), 1213-1227.

van Reekum, C. M., & Scherer, K. R. (1997). Levels of processing in emotion-antecedent appraisal. In G. Matthews (Ed.), Cognitive science perspectives on personality and emotion (pp. 259-330). Amsterdam: Elsevier.

Watson, D. (1988). Intraindividual and interindividual analyses of Positive and Negative Affect: Their relation to health complaints, perceived stress, and daily activities. Journal of Personality and Social Psychology, vol. 54 (6), 1020-1030.

Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. Journal of Personality and Social Psychology, vol. 62, 489-505.

Watson, D., Clark, L. A., Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, vol. 54 (6), 1063-1070.

Watson, D., Tellegen, A. (1985). Toward a consensual structure of mood. Psychological Bulletin, 98, 219-235.

Wine, J. (1971). Test anxiety and direction of attention. Psychological Bulletin, 76(2), 92-104.

Zillman, D. (1979). Aggression and hostility. Hillsdale, NJ: Lawrence Erlbaum.