2011

Assessment of Implicit Attitudes Toward Women Faculty in Science, Technology, Engineering, and Math

Sarah Marie Jackson
Wright State University

ASSESSMENT OF IMPLICIT ATTITUDES TOWARD WOMEN FACULTY IN

SCIENCE, TECHNOLOGY, ENGINEERING, AND MATH

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science

By

SARAH MARIE JACKSON
B.A., The Ohio State University, 2002
M.B.A., University of Phoenix, 2005

2011
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

October 19, 2011

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Sarah Marie Jackson ENTITLED Assessment of Implicit Attitudes Toward Women in Science, Technology, Engineering, and Math Fields BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

_____
Tamera Schneider, Ph.D.
Thesis Director

_____
Scott Watamaniuk, Ph.D.
Graduate Program Director

_____
John Flach, Ph.D.
Chair, Department of Psychology

Committee on
Final Examination

_____
Tamera Schneider, Ph.D.

_____
Gary Burns, Ph.D.

_____
Peggy DesAutels, Ph.D.

_____
Andrew T. Hsu, Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Jackson, Sarah Marie. M.S., Department of Psychology, Wright State University, 2011.
Assessment of Implicit Attitudes Toward Women in Science, Technology, Engineering,
and Math Fields.

This study used two implicit attitude measures (a Go/No-Go Association Task; GNAT

and a personalized GNAT; PGNAT) and three explicit measures to assess attitude change

in faculty attending a diversity training session on women in STEM. It was hypothesized

that (1) pre- and post-training explicit scores would correlate more strongly with the

PGNAT than with the GNAT, (2) training would result in more positive attitudes toward

women in STEM, and (3) difference scores would be greatest in the explicit scales,

followed by the GNAT and PGNAT. Partial support was found for a stronger correlation

between the PGNAT and explicit scores, and the PGNAT revealed more positive implicit

attitudes following training. However, explicit scores did not change significantly, and

the GNAT and PGNAT change scores did not differ from one another. This study adds

support for use of a personalized GNAT and provides evidence that diversity training can

positively affect personal attitudes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I. INTRODUCTION AND PURPOSE

The link between attitudes and subsequent behavior has made the study of attitudes one of considerable interest to social psychologists. Traditionally, scientists have assessed attitudes using explicit measures such as self-reports and questionnaires (Antonak & Livneh, 2000; Wittenbrink & Schwarz, 2007). Explicit measures are vulnerable to threats to validity due to factors such as social desirability and experimenter demand effects (DeMaio, 1984; Wittenbrink & Schwarz, 2007). As a result, attitude researchers have focused on the use of implicit measures to assess attitudes (Fazio, Jackson, Dunton, & Williams, 1995; Fazio & Olson, 2003; Greenwald & Banaji, 1995; Greenwald, McGhee, & Schwartz, 1998). Indirect techniques, such as the Implicit Association Test (IAT), are less vulnerable to threats to validity (Antonak & Livneh, 2000; Greenwald, McGhee, & Schwartz, 1998; Thomas, Vaughn, & Doyle, 2007; Wilson, Lindsey, & Schooler, 2000). The Go/No-Go Association Task (GNAT) is a variation of the IAT that allows researchers to more precisely pinpoint the origin of biased response. The IAT does not correlate well with explicit measures, suggesting it might index normative associations (i.e., knowledge or awareness of stereotypes) rather than participants' personal attitudes (Blanton, Jaccard, Christie, & Gonzales, 2006; Han, Czellar, Olson, & Fazio, 2010; Olson & Fazio, 2004; Rudman, Ashmore, & Gary, 2001). Olson and Fazio (2004) developed a personalized version of the IAT (PIAT) that is more predictive of explicit attitudes under certain circumstances. Similarly, it may be possible

to personalize the GNAT to reduce the effects of extrapersonal associations. The purpose of this research was to use a traditional GNAT, a personalized GNAT (PGNAT), and three measures of explicit attitudes to discern training-induced attitude change about gender in science, technology, engineering, and math (STEM) fields.

Attitudes, Stereotypes, and the Effects of Training on Attitude Change

Social psychologists and sociologists have contributed considerable research to the field of attitude measurement in the past century (e.g., Allport, 1954; Cook & Selltiz, 2004; Devine, 1989; Eagly & Chaiken, 1993; Fazio & Olson, 2003; Thurstone, 1928). This area of research is of particular interest when studying stereotypes and examining which interventions aimed at changing negative attitudes are most effective. An attitude is commonly defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly and Chaiken, 1993, p. 1). A stereotype is an automatic, oversimplified image or idea of a particular group of people (Allport, 1954; Oxford American English Dictionary, 2005). Allport (1954) proposed that stereotyping is an intrinsic component of the cognitive system. People oversimplify their experiences by selectively attending to certain features in the environment and forming categories, concepts, and generalizations. It has been said that stereotypes are the cognitive components of attitudes, while discrimination comprises the behavioral component of attitudes (Devine, 1989; Hackney, 2005). Antonak and Livneh (2000) maintained that understanding a person's attitudes is necessary to be able to explain and ultimately predict that person's behavior toward a particular group.

The theory of implicit social cognition states that past experiences affect behavior, even when those experiences are not remembered or available to self-report or

introspection (Greenwald & Banaji, 1995; Rudman, 2004). This theory has spawned a significant body of research regarding implicit attitudes. Implicit attitudes reflect automatic psychological tendencies or social cognitions that are purported to be outside the control of the individual (Aberson & Haag, 2007; Greenwald & Banaji, 1995; Greenwald et al., 1998). Some researchers (e.g. Greenwald & Banaji, 1995) propose that an implicit attitude is one aspect of a person's true attitude toward a target. If this is the case, explicit and implicit measures both assess true attitudes, but through different mechanisms and often leading to differing results. Other researchers (e.g. Olson & Fazio, 2004) maintain that implicit attitudes might also reflect environmental associations, known as extrapersonal associations. This proposition suggests that implicit attitudes might measure associations that result from environmental experiences, rather than true attitudes toward a target. Either way, responses to implicit attitude measures are believed to reflect automatic activation of unconscious knowledge. Individuals responding to implicit attitude measures theoretically are unaware of this unconscious activation and therefore are not able to exert conscious control over their answers, reducing threats to validity such as social desirability bias. It is widely believed that stereotype activation, the automatic cognitive process whereby a stereotypical association is activated when one is exposed to a target group, operates on the same basic principle as implicit associations (Devine, 1989; Greenwald & Banaji, 1995; Rudman, Ashmore, & Gary, 2001). That is, stereotypes and implicit attitudes are automatic cognitive processes that occur outside the conscious control of the individual. Exposure to a stereotyped group or presentation with primes regarding such a group, even if the prime is subliminal, can lead to activation of stereotypes and implicit attitudes.

A number of researchers believe that attitudes are malleable (Blair, 2002; Rudman, Ashmore, & Gary, 2001; Thurstone, 1928). Experience, contact, and disconfirming information can lead to a change in attitudes and stereotypes. Whereas some researchers have shown that contact can reduce stereotypes, particularly if the contact is meaningful (Aberson & Haag, 2007; Aberson, Shoemaker, & Tomolillo, 2004; Pettigrew, 1997), others have shown that contact alone does not reduce stereotypes or bias (Aberson & Haag, 2007; Campbell, Gilmore, & Cuskelly, 2003; Jelenec & Steffins, 2002; Teachman & Brownell, 2001). Interventions that stress appreciation rather than elimination of group differences have been shown by some researchers to be effective at reducing stereotypes (Johnston & Hewstone, 1992; Wolsko, Park, Judd, & Wittenbrink, 2000). In order to change attitudes, people have to be presented with a combination of meaningful contact and information that disconfirms their existing stereotypes (Hunt & Hunt, 2004).

It is important to note that the nature of interventions must be taken into consideration when attempting to change attitudes. In situations where diversity training is forced, participants might perceive a threat to their freedom of expression or be offended by the implication that they are prejudiced (Brehm, 1966; Rudman, Ashmore, & Gary, 2001), reducing the effectiveness of training on attitude change. McCauley, Wright, & Harris (2000) found that 81% of U.S. colleges and universities use diversity training but, at the time of publication, none had evaluated the effectiveness of such training. Given the substantial amount of money that is invested annually in diversity training initiatives, it is prudent to begin evaluating these training sessions to determine whether attitudes change as a result of participation.

Stereotypes Toward Women in STEM Fields

Stereotypes are damaging to women scientists and engineers in several ways, as discrimination resulting from prejudice can impact education, hiring, promotion, retention, and availability of resources. For example, women faculty are paid less, promoted more slowly, receive fewer honors, and hold fewer leadership positions than men (National Academy of Science, 2006). The National Academy of Science (2006) has identified several common stereotypes toward women in STEM fields: (1) women are not as good in mathematics as men, (2) women are not as competitive as men and do not want jobs in academia, (3) women faculty are less productive than men, and (4) women are more interested in family than in careers. The implicit biases held by both men and women can significantly hinder the success of women who choose to enter STEM fields. Stereotypes can prevent women from initially entering STEM fields.

The belief that men are more inclined to participate and excel in math and science is widely held, even among women (National Academy of Science, 2006; Nosek, Banaji, & Greenwald, 2002b). Nosek et al. demonstrated that among college students, women's attitudes reflected negativity toward math and science relative to arts and language (2002b). Female deficits in problem-solving abilities increased relative to the number of males present in a group suggesting that situational cues can create a threatening intellectual environment for females (Inzlicht & Ben-Zeev, 2000; Sekaquaptewa & Thompson, 2003). However, two distinct empirical studies found that gender differences in performance on difficult math tests were eliminated when stereotype threat was lowered (Murphy, Steele, & Gross, 2007; Sekaquaptewa & Thompson, 2003). Stereotype threat is "being at risk of confirming, as self-characteristic, a negative

stereotype about one's group" (Steele & Aronson, 1995, p. 797). In other words, women perform more poorly on a task (difficult math task) when a relevant stereotype (women are not as good at math as men) is made salient. These findings suggest that gender differences in math performance might not reflect actual differences in ability, but rather that situational cues and priming can create a threatening environment that artificially reduces performance. In fact, contrary to the stereotype, female performance in high school mathematics now matches that of males (National Academy of Science, 2006).

Despite the fact that Canes and Rosen (1995) found that an increase in women faculty did not increase enrollment by females in a given department, Dasgupta and Asgari (2004) demonstrated that local environments, particularly exposure to biographical information about famous female leaders, exposure to women in leadership positions, and composition of social environment significantly affected women's nonconscious beliefs about their ingroup. Dasgupta and Asgari's findings suggest that stereotypic beliefs regarding women in fields typically thought to be dominated by men can be changed.

Attitude Measures

In 1928, Thurstone proposed one of the first widely-used explicit measures for assessing attitudes. Since that time, researchers have developed numerous implicit and explicit methods for the use of attitude measurement and debate has been heated over which techniques are most valid and reliable (e.g., Blanton & Jaccard, 2006; Blanton et al., 2009; Greenwald, Rudman, Nosek, & Zayas, 2006; Han et al., 2010; Nosek & Hansen, 2008). Cook and Selltiz (1964) identified five classes of attitude measures that are still in use today: self-report (explicit) measures, behavioral, physiological measures,

partially structured measures, and measures based on objective tasks. Partially structured measures use stimuli that can be interpreted in multiple ways. In measures based on performance on objective tasks, participants are presented with specific tasks to be performed. Behavioral, physiological, partially structured, and objective task measures are considered implicit measures because participants cannot consciously alter their responses to make them more desirable (Cook & Selltiz, 1964; Vargas, Sekaquaptewa, & von Hippel, 2007).

Direct, or explicit, measures have traditionally been the methods predominantly used in assessing attitudes (Antonak & Livneh, 2000; Wittenbrink & Schwarz, 2007). Explicit measures are subject to threats to validity resulting from a number of factors including experimenter demand effect, social desirability bias, acquiescence style, and the halo effect (Antonak & Livneh, 2000; Rudman et al., 2001; Wittenbrink & Schwarz, 2007). Even Thurstone (1928) acknowledged the uncertainty inherent in using one's explicit opinion to measure attitudes, citing the politician who "extends friendship and hospitality in overt action while hiding an attitude that he expresses more truthfully to an intimate friend" (p. 532). He further warned that one cannot automatically imply that a participant will act in accordance with the opinions he or she has explicitly expressed (Thurstone, 1928).

Another problem with explicit measures is founded in the implicit social cognition theory. If implicit attitudes are automatic cognitions that are outside the control of the individual and are not available to self-report or introspection, then the individual will not be able to explicitly report them (Aberson & Haag, 2007; Greenwald, & Banaji, 1995; Greenwald et al.,1998; Rudman, 2004; Rudman, Ashmore, & Gary, 2001). It is

therefore necessary to devise methods that are able to access these automatic and implicit associations.

In response to the previously mentioned threats to validity found in explicit measures, some researchers have endorsed the use of indirect, or implicit, measures for assessing attitudes. A number of researchers have claimed that implicit measures are less vulnerable to validity threats than explicit measures (Antonak & Livneh, 2000; Greenwald et al., 1998; Thomas et al., 2007; Wilson, Lindsey, & Schooler, 2000). Implicit measures fall under the following classes: behavioral, physiological measures, partially structured measures, and measures based on performance on objective tasks (Vargas, Sekaquaptewa, & von Hippel, 2007). In the past two decades, a number of researchers have proposed methods based on performance on objective tasks. The Implicit Association Test (IAT) is one widely-used tool that falls under this category.

The Implicit Association Test (IAT) and the Personalized IAT (PIAT)

Developed by Greenwald, McGhee, and Schwartz (1998), the Implicit Association Test (IAT) is a tool used to assess implicit attitudes and other automatic associations based on reaction times. It is purported to be more resistant to validity threats when compared with explicit measures. The IAT technique measures how quickly a participant can classify stimuli (e.g., words, names, or pictures) typically associated with a particular social group into target and attribute categories. Participants are asked to sort stimuli representing four concepts (i.e., attribute concepts "pleasant" or "unpleasant" and target concepts "insect" or "flower") into one of two response categories, each of which includes two of the four concepts (Greenwald et al., 1998). The target category contains the dichotomous aspects of the object attitude the researcher

is interested in studying. For example, the two concepts "flower" and "insect" would comprise the target category if a researcher was interested in studying participants' implicit attitudes toward flowers and insects. The attribute category contains the valence of the attitudes. Commonly used attribute category concepts include "good" and "bad" or "pleasant" and "unpleasant". In the traditional computer-based IAT, the response time indicates the relative strength of association by assessing how quickly a participant can pair a target category with the attribute dimension. If a target category is associated with an attribute dimension that reflects the participant's implicit attitude, he or she should respond more quickly (Greenwald et al., 1998; Pruett & Chan, 2006). For example, when "flower" is paired with "good", and "insect" is paired with "bad", one would expect to see faster response times, as the majority of participants most likely have more positive attitudes toward flowers than insects.

Because it is not always possible or practical to use computers to administer the IAT, a paper-based version of the test has been created (Lemm, Lane, Sattler, Khan, and Nosek, 2008). Lemm et al. validated the paper-based version through comparison with the computer-based IAT. In a paper-based IAT, participants are presented with a list of stimuli items from the target and attribute categories. Participants sort the list of items into the correct target or attribute categories by marking circles to the left or right of the each item, indicating to which category the item belongs. The test has two parts. In the first part, the attribute dimension of pleasant is paired with the first target group on one side of the column and the attribute dimension of unpleasant is paired with the second target group on the other side. In the second part, the pairings are switched (i.e., the pleasant category is now paired with the second target group). The dependent measure in

this test is the difference in the number of items correctly categorized in both parts in a specified amount of time, typically 20-30 seconds (Vargas, Sekaquaptewa, & von Hippel, 2007). The task has been designed so that participants are unable to sort all stimuli items in the time given. Participants must sort stimuli items as quickly as possible. This is intended to elicit as close to an instant, uncontrollable, and unconscious reaction as possible.

Potential problems with the IAT include the claims that the test does not predict discriminatory behavior (Blanton et al., 2009), that it does not correlate with explicit measures of attitudes (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), and that it measures normative associations rather than attitudes (Han et al., 2010). Olson and Fazio (2004) raised the argument that the ease with which participants sort stimuli into categories reflects the automatic *association* between a given attitude object and a given valence. As a result, this association does not necessarily reflect the individual's attitude due to the possible contamination of extrapersonal associations, associations to the category that may not be related to the individual's evaluation of the category (i.e., the individual's attitudes). Extrapersonal associations are associations present in one's memory that originate from sources outside one's personal attitudes. Social and cultural influences and norms contribute to these extrapersonal associations. As a result, the IAT score of an individual whose attitude contradicts a cultural norm might reflect attitudes that agree with the cultural norm because the IAT measures extrapersonal associations more so than personal preferences. For example, Nosek, Banaji, and Greenwald (2002) found that Blacks do not show an in-group preference on the IAT. Olson and Fazio (2004) argue that this finding does not measure a true attitude, but rather reflects the

prevalence of negative portrayals of Blacks in the media that contribute to extrapersonal associations.

To counter this possibility, Olson and Fazio developed a personalized version of the IAT (PIAT) that uses personalized attribute category labels ("I like" versus "I don't like") instead of normative attribute category labels ("good" versus "bad" or "pleasant" versus "unpleasant"). Olson and Fazio proposed that personalized attribute category labels lead to a more accurate measurement of attitudes, rather than associations because the "I like' and "I don't like" labels invoke personal associations as opposed to the traditional "Pleasant" and "Unpleasant" labels that allow for the effect of extrapersonal, normative associations. The personalized IAT also correlates more strongly with explicit attitudes when compared with the IAT (Han et al., 2010; Nosek & Hansen, 2008; Olson & Fazio, 2004).

The Go/No-Go Association Task (GNAT) and Personalized GNAT (PGNAT)

In a study examining the use of the IAT to measure self-esteem, Karpinski (2004) pointed out a potential problem with the IAT. He described the problem that arises from the use of dichotomous target variables (e.g. self versus other). Because participants are responding to two pairings of target variables and attribute variables, the IAT is not just a measure of attitudes toward a single target (e.g., self), but a combined measure that could result from attitudes toward one target variable (self) and attitudes toward the other target variable (other). It is possible that a high self-esteem score is the result of positive associations with the self or it could result from negative associations with others. Similarly, a positive score on an IAT measuring attitudes toward women scientists could therefore reflect a positive attitude toward women scientists or a negative attitude toward

11

men scientists. There is no way to definitively differentiate between the two since the response times or number of correct responses are a result of responding to both target variables simultaneously.

The Go/No-Go Association Task attempts to address this problem. The Go/No-Go Association Task (GNAT) is a variant of the IAT developed by Nosek and Banaji (2001). As with the IAT, the GNAT assesses the strength of association between a target category and an attribute dimension. The primary difference between the GNAT and the IAT is that the GNAT assesses association strength toward a single target concept as opposed to the dichotomous targets required in the IAT. Participants are presented with a single target variable and a single attribute dimension at the top of the screen. Stimuli items appear on the screen one at a time. Participants indicate whether each item fits in *either* category by pressing a button ("Go"). If an item does not fit in either category, known as a "distractor" item, the participant does nothing ("No Go"). The advantage of this design is that a researcher can parse out the source of variance in response times, allowing for a more direct investigation of a specific attitude object (Nosek & Banaji, 2001). For example, in a classic IAT, a researcher might determine that participants have faster response times when "White" is paired with "good" and "Black" is paired with "bad". It is impossible, however, to determine whether this difference in response latencies in the IAT is the result of an unconscious bias toward Whites being good or an unconscious bias toward Blacks being bad. With the GNAT, the researcher could administer a separate block for each possible pairing ("White" with "good", "White" with "bad", "Black" with "good", and "Black" with "bad").

Given the potential influence of extrapersonal associations on IAT scores, it is reasonable to assume that the GNAT will be similarly vulnerable, since it is merely a variation of the IAT. In order to correct for this possibility, the attribute variables "I like" and "I don't like" may be used in place of the traditional categories "Good" versus "Bad" or "Pleasant" versus "Unpleasant". The resulting personalized GNAT, or PGNAT, should effectively reduce the effects of extrapersonal associations.

Correlations Between Implicit and Explicit Measures

A number of empirical studies have revealed a lack of correlation between implicit and explicit attitudes. Lack of correlation between implicit and explicit measures could indicate dual processes (implicit attitudes and explicit attitudes), or it could indicate different ways of measuring a single process. If the latter is true, implicit and explicit measures both tap into a single attitude, but implicit measures could be more accurate because they are not subject to threats to validity the way explicit measures are (Greenwald & Nosek, 2008). The problem with this theory is that this lack of correlation has been detected when there is no reason to suspect that a participant would distort his or her responses on an explicit measure. For example, Nosek, Banaji, and Greewnwald (2002a) found that, despite the fact that explicit measures reveal an in-group bias for Blacks, the IAT failed to reveal such a bias. This finding is counterintuitive. Social psychology research has consistently shown that people respond more favorably to their own group. Olson and Fazio (2004) posited that the lack of in-group bias in IAT scores for Blacks may be a result of the influence of extrapersonal associations. Knowledge about stereotypes and bias toward Whites resulting from cultural and media exposure could influence IAT scores. In this case, an IAT score might not reflect an individual's

true attitude, but instead might reflect knowledge of stereotypes. The personalized IAT developed by Olson and Fazio (2004) corrected for the influence of extrapersonal associations, leading to a more accurate reflection of participants' true attitudes. The PIAT correlates more strongly with explicit attitude measures than the IAT (Olson & Fazio, 2004). Therefore, it is reasonable to assume that a personalized GNAT would correlate more strongly with explicit attitude scores than a traditional GNAT.

Hypothesis 1. Pre- and post-training explicit attitude scores will correlate more strongly with the PGNAT than with the GNAT.

Effective training has been shown to have an effect on attitudes, reflected in both explicit and implicit measures (Blair, 2002; Rudman, Ashmore, & Gary, 2001). Providing participants with information that disconfirms widely-held stereotypes about women in STEM should result in more positive attitudes, reflected in all three attitude measures administered. Attitude scores taken before and after the diversity training should reflect a significant increase in positive attitudes.

Hypothesis 2. Training (i.e. viewing a presentation on implicit bias) will result in more positive implicit and explicit attitudes toward women in STEM fields, reflected in more positive GNAT, PGNAT, and explicit measure scores.

Hypothesis 2a. The change from pre-training to post-training GNAT scores will be greater in the experimental group than in the control group.

Hypothesis 2b. The change from pre-training to post-training PGNAT scores will be greater in the experimental group than in the control group.

Hypothesis 2c. The change from pre-training to post-training explicit

attitude scores will be greater in the experimental group than in the control

group.

Han, Olson, and Fazio (2006) conducted a study wherein participants were

exposed to experimentally-created extrapersonal associations that conflicted with their

true attitudes. In the study, participants were given information about the children's

game Pokémon, and were given a choice of two Pokémon cards, one of which was

clearly superior in the context of the game. Participants expressed their card preference

through a semantic differential scale and by providing written statements explaining their

choice. They were then exposed to an artificially created extrapersonal association in the

form of a video recording depicting a young boy explaining which card he preferred and

providing reasons for his choice. Half of the participants watched a video that was

consistent with their explicit Pokémon card choice, and the other half watched a video

that was inconsistent with the selection. The researchers found that, despite the fact that

the reasons given in the video were viewed as not rational by the participants, exposure to

the extrapersonal association resulted in reduced implicit association scores reflected by

the IAT. The PIAT, however, was not affected, and scores correlated with explicit

attitudes.

Because the IAT has been shown to be more vulnerable to effects of extrapersonal

associations, it would be expected that exposure to disconfirming information in a

diversity training presentation would result in a greater change in attitude scores on an

IAT than on a personalized IAT. The GNAT should be similarly vulnerable to

extrapersonal associations, because it is a variant of the IAT. As a result, the GNAT

15

scores should reflect a more dramatic change following training than the PGNAT scores. Additionally, because it is easier for participants to consciously adjust their responses on explicit attitude measures, the change in explicit attitude scores should be greater than any changes detected in either the GNAT or the PGNAT.

Hypothesis 3. In the experimental group, the change in GNAT scores from pre- to post-training will be greater than the change in PGNAT scores, and the change in explicit attitude scores from pre- to post-training will be greater than the changes in both GNAT and PGNAT scores.

II. METHOD

Participants

Participants were 149 STEM faculty from four midwestern universities. Of these, 35 were female (26%) and 99 were male (74%) with a mean age of 49 years (age range: 31-75). Of those participants who provided ethnicity, 3.2% reported that they were African American, 6.5% selected Asian/Pacific Islander, 87.1% selected White, Non-Hispanic, and 3.2% selected other. Participants were randomly assigned to either a control group ($n = 80$) or an experimental group ($n = 69$) by department. A total of 18 departments were randomly assigned, with 8 departments receiving the control condition and 10 departments receiving the experimental condition. The research was supported in part by National Science Foundation ADVANCE Award (grant number HRD 0810989).

Error Rates

In their study utilizing a paper-based IAT, Teachman and Brownell (2001) omitted participants with an error rate of greater than 35%. This is somewhat more liberal than the standard error rate omission of 20% generally seen in computer-based IAT administration (Olson & Fazio, 2004). Given the fact that the paper-based implicit measure, by its design, does not provide error feedback, it is logical to use this less stringent recommendation. In the current study, 16 participants (10%) were omitted due to error rates greater than 35%, resulting in the final N of 140. This is less than the proportion omitted by Teachman and Brownell (18%).

Design. This study employed a mixed design. All participants completed the GNAT, the personalized GNAT, and three explicit attitude measures before and after viewing a diversity training presentation (semantic differential scale, stereotype scale, and feeling thermometer scale; within-subjects) and were randomly assigned to either a control or experimental group (between-subjects).

Materials

Individual trait measures. Prior to beginning the GNAT and PGNAT tasks, participants completed a brief survey comprised of 4 sub-scales designed to assess individual trait levels of egalitarianism (e.g. "There should be equality for everyone because we are all human beings;" adapted from Katz & Hass, 1988), reactance (e.g. "Advice and recommendations induce me to do just the opposite;" adapted from Hong & Faedda, 1996), social desirability ("I am always willing to admit it when I make a mistake;" adapted from Crowne & Marlowe, 1960), and self-discrepancy ("I should/would enjoy collaborating with a woman on a research project;" adapted from Monteith & Voils, 1998). A total of 13 items with 5-point agreement scales were presented. Participants rated each item using a Likert-type scale of 1 (strongly disagree) to 5 (strongly agree). One item from the reactance scale was eliminated due to low bivariate correlations. Cronbach's alpha for the egalitarianism, reactance, social desirability, and self-discrepancy scales were 0.64, 0.72, 0.58, and 0.71 respectively.

Egalitarianism is the belief that all people are equal and deserve equal rights and opportunities (Katz & Hass, 1988). Egalitarianism should correlate with more positive attitudes toward referent groups. Participants high in reactance will respond to situations that appear to threaten their freedom by resisting influence from others (Hong & Faedda,

1996). This could result in more negative attitudes on explicit measures. The Crowne-Marlowe social desirability scale was designed to assess individual differences in socially desirable responding (Crowne & Marlow, 1960). Participants scoring high on this scale should also explicitly report more positive attitudes toward referent groups. Self-discrepancy is a measure of prejudice-related inconsistencies to which people are prone and is calculating by looking at the difference between scores on paired should and would statements (Monteith & Voils, 1998). People may be prone to a conflict between their consciously-endorsed, non-prejudiced beliefs and their persisting stereotypical responses. The self-discrepancy scale is designed to assess this conflict.

GNAT and PGNAT. In the paper-based GNAT and PGNAT, a single target category ("female scientist") and a single attribute category ("pleasant" or "unpleasant" for the GNAT; "I like" or "I don't like" for the PGNAT) appeared at the top of the page. The stimuli items were listed in two columns. When a stimuli item belonged in the target category or the attribute category, the participants were instructed to circle the item ("Go"). If an item did not belong to either category, participants skipped the item ("No Go") and moved on to the next one in the list. Each pairing of a single target and attribute category was considered one block. Participants were timed for 15 seconds per block. At the end of the 15 seconds, they were instructed to draw a line below the last item they categorized.

In order to specifically examine attitudes toward women scientists, there were two blocks for the GNAT: "female scientist" paired with "pleasant" and "female scientist" paired with "unpleasant". A congruent target-attribute pairing is one that fits the participant's automatic association. For example, the target category "flower" paired

19

with the attribute category "pleasant" is generally considered to be a congruent pairing, whereas the target category of "flower" paired with the attribute category "unpleasant" is an incongruent pairing. Participants generally classify stimuli faster when the paired categories match their automatic attitudes toward the target. For the purpose of this study, congruent blocks were those that conformed to widely-held stereotypes regarding women in STEM. Specifically, because women scientists are generally viewed less favorably than men scientists, the congruent block was "female scientist" paired with "unpleasant", while the incongruent block was "female scientists" paired with "pleasant". The GNAT was personalized (PGNAT) by substituting "pleasant" with "I like" and "unpleasant" with "I don't like". Stimuli for the PGNAT were identical to stimuli used with the GNAT. There were two blocks for the PGNAT: "female scientist" paired with "I like" (incongruent block) and "female scientist" paired with "I don't like" (congruent block).

The number of items categorized in the time allowed reflects the speed at which participants sorted stimuli, while the number of items correctly categorized reflects accuracy. A ratio of correctly categorized items to total categorized items was obtained for both the congruent and incongruent conditions in the GNAT and PGNAT, and these ratios were transformed into a score for each implicit task using an algorithm suggested by Nosek & Lane (1999; see below). The resulting score is the difference in "sensitivity" between the congruent and incongruent conditions. This is believed to be a measure of automatic attitudes (Nosek & Banaji, 2001).

Stimuli. Stimuli items for both the GNAT and the PGNAT were a combination of words and symbols. Pleasant and unpleasant words were adapted from Greenwald,

McGhee, & Schwartz (1998), Nosek, Banaji, & Greenwald (2002a), and Olson and Fazio (2004). Symbols reflecting science and non-science icons overlaid on top of male and female forms were adapted from Nosek, Banaji, and Greenwald (2006). A pilot study was performed to validate the science and non-science symbols. 10 graduate students at Wright State University were presented with a list of ten symbols (5 science and 5 non-science) without the male and female figures and were asked to write down the first 2-3 words that came to mind when viewing each symbol. Out of the 10 original symbols, 4 symbols (2 science and 2 non-science) were selected based on consistency of responses. An additional symbol representing engineering was added for use with participants in engineering departments (see Appendix A for a complete list of stimuli items).

Scoring the implicit measures. The paper-based GNAT and PGNAT were scored using the following algorithm based on suggestions made by Nosek and Lane (1999) for scoring a paper-based IAT: $[(\pm \text{ max value } (A,B))/(\text{min value } (A,B))]$ x (square root of $|(A-B)|$). A is the number of items correctly categorized in the incongruent block ("female scientist" paired with "pleasant" or "I like") and B is the number of items correctly categorized in the congruent block ("female scientist" paired with "unpleasant" or "I don't like"). When A was greater than B, the score was multiplied by 1, and when B was greater than A, the score was multiplied by –1. Positive GNAT or PGNAT scores indicate more favorable attitudes toward women in STEM, whereas negative GNAT or PGNAT scores indicate more negative attitudes toward women in STEM.

Explicit attitude measures. Three explicit attitude scales were administered both before and after the presentation in the experimental group or before and after the department meeting in the control group. Higher scores on each explicit scale indicated

21

higher favorability toward women in STEM. The first two explicit attitude measures were adapted from Olson and Fazio (2004). The first part of the explicit measure contained 12 rating scales with semantic differential word choices at either end of the scale (e.g. "analytical-emotional" and "passive-proactive"). Participants circled a number from 1 to 5 on each scale which best reflected their beliefs regarding women scientists. A Principal Axis Factor analysis (PAF) with a Varimax rotation was conducted for the 12 semantic differential items. Guidelines proposed by Tabachnick and Fidell (1989) were employed to select items to be included in a factor: each item needed to load high (> .40) and have lower loadings on all other factors (< .30). A PAF was conducted for both the pre-training packet and the post-training packet. Of the 12 semantic differential items, 8 were included in the final scale. The final semantic differential scales had Cronbach's alpha of .92 for the pre-training packet and .93 for the post-training packet.

Part 2 of the explicit measure contained 22 stereotype statements derived in part from the National Academy of Science book *Beyond Bias and Barriers* (2006; e.g. "Women are worse at math than men" and "There are fewer women faculty because they are less qualified"). Participants rated each item using a Likert-type scale of 1 (strongly disagree) to 5 (strongly agree). Of the 22 statements, 5 were designated "filler" items which contained stereotype statements that were not related to attitudes regarding women in STEM (e.g. "Artists are more emotional than scientists"). A Principal Axis Factor analysis (PAF) with a Varimax rotation was conducted for these items, resulting in a final scale of 9 items with a Cronbach's alpha of .83 for the pre-training packet and .87 for the post-training packet.

Part 3 of the explicit measure consisted of 12 feeling thermometer items. Participants were asked to rate on a scale from 0 (very cold/unfavorable) to 100 (very warm/favorable) their feelings toward each item. Of these 12 items, three were specifically about women (e.g. "female scientists"), three were about men (e.g. "male engineers"), and the remaining six were filler items (e.g. "attending faculty meetings"). Only those items indexing favorability toward women were used in the final feeling thermometer scale. Cronbach's alpha for the pre-training female scale was .88 and the post-training female scale was .90.

Procedure

Participants were given a cover letter inviting them to participate in the study, along with an envelope containing the pre-training packet and the post-training packet. Participants were informed that their participation was completely voluntary, that their responses would remain confidential, and only aggregated data would be reported. The GNAT and PGNAT were presented as "brief, timed categorization tasks". Participants first completed the individual trait survey, followed by a practice GNAT task to ensure they understood the directions. Participants then completed a timed (15 seconds per block) GNAT (identifying stimuli as belonging to either "female scientists" and "pleasant" or "female scientists" and "unpleasant"), a timed (15 seconds per block) PGNAT (identifying stimuli as belong to either "female scientists" and "I like" or "female scientists" and "I don't like"), and the explicit attitude measures.

The order of presentation of the GNAT and PGNAT and the order of pairing of category labels within each task were counterbalanced across participants. A complete counterbalanced design was not possible given the small number of participants.

Therefore, an incomplete balanced square was used to ensure that each condition (i.e., each task) appeared equally often at each stage and each condition followed each other condition an equal number of times. Packets were randomly selected from one of eight possible combinations. In four of the packet designs, the GNAT came before the PGNAT and in the other four, the PGNAT came before the GNAT. Within the GNAT and PGNAT measures, the order of congruent ("female scientist" and "unpleasant" or "I don't like") and incongruent ("female scientist" and "pleasant" or "I like") blocks were completely counterbalanced.

In the experimental condition, participants first completed the pre-training packet, then took part in a half-hour presentation and discussion regarding women in STEM, diversity, and hiring. After the presentation, participants retrieved the post-training packet from the envelope and completed the series of tasks (GNAT, PGNAT, and explicit measures) as before along with a brief demographic questionnaire. In the control condition, researchers attended a previously scheduled department faculty meeting and administered the pre-training packet at the beginning of the meeting and the post-training packet at the end of the meeting, with no presentation given.

III. RESULTS

Correlations between Attitude Measures and Individual Trait Measures

None of the individual trait measures (social desirability, reactance, egalitarianism, and self-discrepancy) correlated significantly with any of the implicit attitude measures. There were several significant correlations between the individual trait measures and the explicit attitude measures (see Table 1). Social desirability correlated significantly with the stereotype scale ($r = .30$, $p < .01$) and with the feeling thermometer scale ($r = .22$, $p < .01$). Reactance correlated significantly with the stereotype scale ($r = .26$, $p < .01$). Egalitarianism correlated significantly with the stereotype scale ($r = .40$, $p < .01$) and the feeling thermometer scale ($r = .24$, $p < .01$). Self-discrepancy correlated significantly with the semantic differential scale ($r = -.26$, $p < .01$).

Correlations between Implicit and Explicit Measures

Hypothesis 1 stated that pre- and post-training explicit attitude scores would correlate more strongly with the PGNAT than with the GNAT. To test this hypothesis, bivariate correlation analyses were calculated between the GNAT, PGNAT, and three explicit scales. Table 2 shows correlations between the implicit and explicit measures. Steiger's (1980) z-tests of differences in dependent correlations were conducted to test differences between the GNAT scores and the explicit scales and between the PGNAT scores and the explicit scales.

Table 1

Bivariate correlations between pre-training attitude scores and individual trait measures.

	Attitude Measure	1	2	3	4	5	6	7	8
1	GNAT								
2	PGNAT	0.02							
3	Semantic Differential	0.02	-0.05						
4	Stereotype	-0.09	-0.11	0.31**					
5	Feeling Thermometer	-0.06	-0.04	0.44**	0.48**				
6	Social Desirability	-0.13	0.04	-0.02	0.30**	0.22**			
7	Reactance	0.05	0.02	-0.04	-0.26**	-0.08	-0.28**		
8	Egalitarianism	-0.13	-0.03	0.09	0.40**	0.24**	0.34**	-0.21*	
9	Self-Discrepancy	0.00	-0.02	-0.26**	0.01	0.02	0.10	-0.04	0.11

*Note. *$p < .05$, **$p < .01$.

The GNAT scores (both pre- and post-training) did not correlate significantly with any of the explicit attitude scales. The post-training PGNAT correlated significantly with the post-training feeling thermometer ($r = 0.32$). Post-training PGNAT correlations with the stereotype scale and feeling thermometer scale appeared to be larger than post-training GNAT correlations.

Semantic differential scale. The correlation between the pre-training GNAT and the pre-training semantic differential scale ($r = 0.07$) did not differ significantly from the correlation between the pre-training PGNAT and the pre-training semantic differential scale ($r = -0.16$), $z = 1.43$, $p > .05$. The correlation between the post-training GNAT and the post-training semantic differential scale ($r = -0.14$) did not differ significantly from the correlation between the post-training PGNAT and the post-training semantic differential scale ($r = 0.12$), $z = -1.50$, $p > .05$.

Stereotype scale. The correlation between the pre-training GNAT and the pre-training stereotype scale ($r = -0.07$) did not differ significantly from the correlation between the pre-training PGNAT and the pre-training stereotype scale ($r = -0.05$), $z = -0.17$, $p > .05$. The correlation between the post-training GNAT and the post-training stereotype scale ($r = -0.02$) did not differ significantly from the correlation between the post-training PGNAT and the post-training stereotype scale ($r = 0.13$), $z = -0.82$, $p > .05$.

Feeling thermometer scale. The correlation between the pre-training GNAT and the pre-training feeling thermometer scale ($r = -0.01$) did not differ significantly from the correlation between the pre-training PGNAT and the pre-training feeling thermometer scale ($r = -0.01$), $z = -0.01$, $p > .05$. The correlation between the post-training GNAT and the post-training feeling thermometer scale ($r = 0.06$) did not differ significantly from the

Table 2

Bivariate correlations between implicit attitude scores and explicit attitude scales in the experimental group.

	Attitude Measure	*M*	*SD*	1	2	3	4	5	6	7	8	9
Pre-training												
1	GNAT	0.79	3.34									
2	PGNAT	0.39	3.25	0.12								
3	Semantic Differential	3.84	0.77	0.07	-0.16							
4	Stereotype	4.09	0.61	-0.07	-0.05	0.25*						
5	Feeling Thermometer	75.30	17.71	-0.01	-0.01	0.59**	0.49**					
Post-training												
6	GNAT	1.06	2.74	-0.13	0.03	-0.11	0.80	0.05				
7	PGNAT	0.74	3.12	0.06	0.30*	0.02	0.13	0.27*	-0.01			
8	Semantic Differential	3.89	0.76	0.02	-0.07	0.82**	0.32**	0.61**	-0.14	0.12		
9	Stereotype	4.04	0.59	0.06	-0.07	0.39**	0.86**	0.57**	-0.02	0.13	0.50**	
10	Feeling Thermometer	74.50	17.84	-0.02	0.06	0.54**	0.45**	0.97**	0.06	0.32**	0.58**	0.55**

Note. *p < .05, **p < .01.

correlation between the post-training PGNAT and the post-training feeling thermometer scale ($r = 0.32$), $z = -1.54$, $p > .05$.

Effect of Training on Implicit and Explicit Attitudes

Hypothesis 2 stated that training would result in more positive implicit and explicit attitudes toward women in STEM fields. To test the assumption that participation in a diversity training session would result in more positive attitudes as reflected by the GNAT and PGNAT scores, a repeated measures ANOVA was conducted including GNAT (and PGNAT) scores at time 1 (pre-training) and time 2 (post-training), using group (experimental versus control) as a between-subjects variable.

GNAT scores. The experimental group appeared to have a higher baseline (pre-training) GNAT score ($M = 0.79$, $SD = 3.34$) than the control group ($M = 0.34$, $SD = 3.67$). The experimental post-training GNAT score ($M = 1.06$, $SD = 2.74$) appeared to be higher relative to baseline than the control post-training GNAT score ($M = 0.39$, $SD = 2.94$). The change in score from pre-training to post-training appeared to be greater in the experimental group (difference score = 0.27; see Figure 1) than in the control group (difference score = 0.05). However, the repeated measures ANOVA revealed no significant main effect for time, $F(1, 147) = 0.19$, *ns*, no significant main effect for group, $F(1, 147) = 2.24$, *ns*, and no significant time by group interaction, $F(1, 147) = 0.09$, *ns*. Furthermore, simple slopes analyses revealed no significant differences in the change in GNAT scores for the experimental or control groups, $\Delta R^2 = 0.01$, $\Delta F(1, 145) = 1.78$, *ns*.
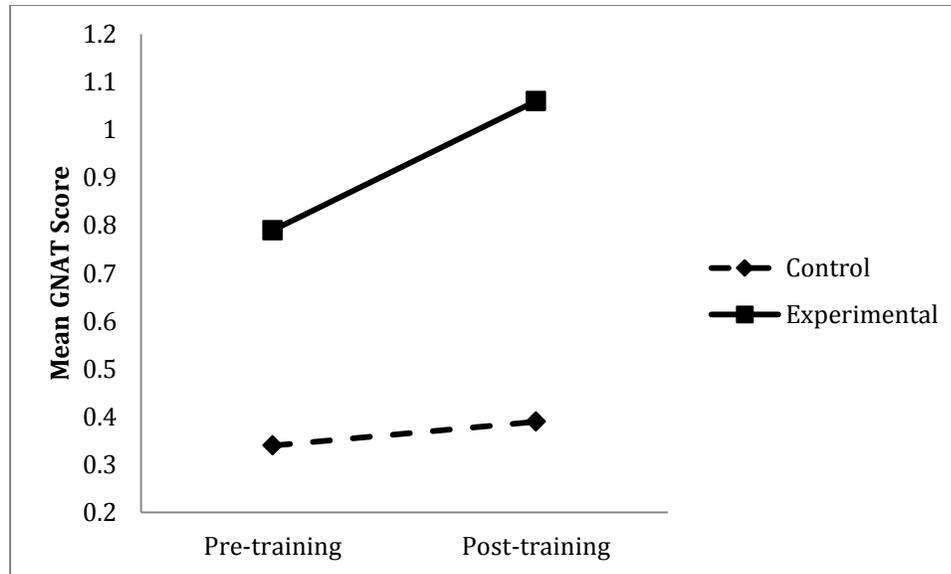
Figure 1. Mean GNAT scores for the pre-training and post-training control and experimental groups.

PGNAT scores. The baseline PGNAT scores for the experimental ($M = 0.45$, $SD = 3.23$) and the control ($M = 0.43$, $SD = 4.01$) groups were similar. The experimental post-training PGNAT score ($M = 0.74$, $SD = 3.12$) was higher relative to baseline than the control post-training PGNAT score ($M = 0.53$, $SD = 2.65$). As with the GNAT, the change in PGNAT score from pre-training to post-training appeared to be greater in the experimental group (difference score = 0.29) than in the control group (difference score = 0.10). The repeated measures ANOVA revealed no significant time by group interaction, $F(1, 143) = 0.07.$, *ns*. The main effects for time, $F(1, 143) = .28$, *ns* and group, $F(1, 143) = .08$, *ns* were not significant.

To conduct a simple slopes analysis, group (experimental versus control) was entered in the first step of a linear regression equation along with the pre-training PGNAT score. The interaction term of group by pre-training PGNAT score was added in step 2. Despite the fact that the ANOVA revealed a non-significant interaction, simple slopes analyses revealed that for the experimental group, the interaction term of group by

pre-training PGNAT score explained a significant amount of variance ($\Delta R^2 = 0.03$) in the post-training PGNAT score, $\Delta F(1, 141) = 5.04$, $p < .05$. This suggests that the slopes of the experimental and control lines differed significantly, indicating that the two groups changed differently over time.
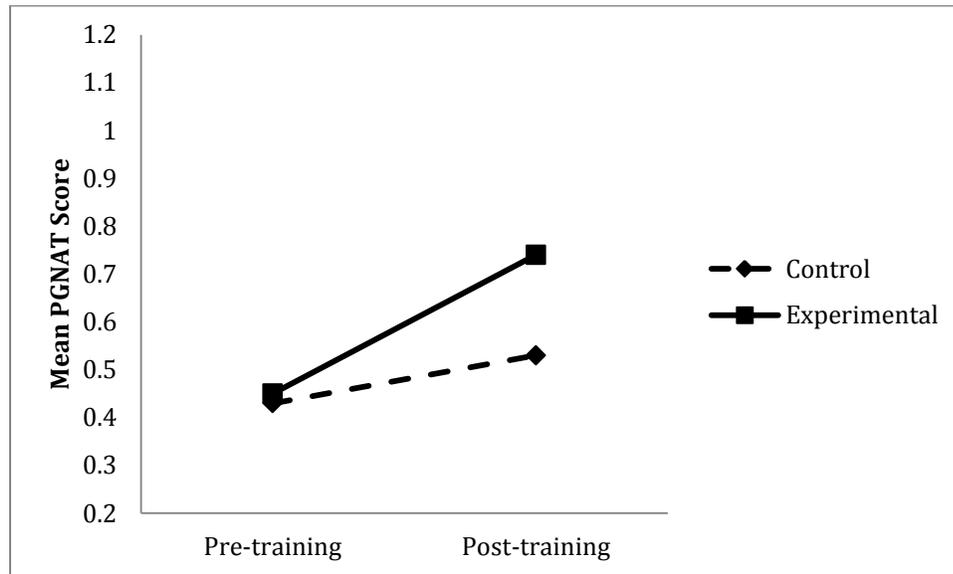
Figure 2. Mean PGNAT scores for the pre-training and post-training control and experimental groups.

Baseline differences between groups. A somewhat surprising trend was revealed when comparing the experimental and control group baseline implicit scores. The average baseline GNAT score in the experimental group was marginally greater ($p <$ .10) than the average baseline GNAT score in the control group. This difference was not observed in the PGNAT scores.

Explicit attitude measures. There were no significant differences between pre- and post-training scores for any of the explicit attitude measures (see Table 3). All explicit attitude measures revealed generally positive attitudes toward women in STEM, and these scores did not change significantly from pre- to post-training.

31

Table 3

Pre- and post-training scores for the semantic differential scale, stereotype scale, and feeling thermometer scale.

	Pre-training		Post-training	
Scale	*M*	*SD*	*M*	*SD*
Semantic Differential Control	3.94	0.81	3.94	0.79
Semantic Differential Experimental	3.86	0.76	3.89	0.75
Stereotype Control	4.11	0.49	4.05	0.53
Stereotype Experimental	4.12	0.58	4.06	0.57
Feeling Thermometer Control	74.25	15.83	73.54	16.77
Feeling Thermometer Experimental	75.73	17.69	74.79	17.83

Semantic differential scale. The control group semantic differential score did not change from pre-training ($M = 3.94$, $SD = 0.81$) to post-training ($M = 3.94$, $SD = 0.79$), $t(78) = -0.37$, *ns*. The experimental group semantic differential score did not change from pre-training ($M = 3.86$, $SD = 0.76$) to post-training ($M = 3.89$, $SD = 0.75$), $t(67) = -0.60$, *ns*. A repeated measures ANCOVA (controlling for individual trait measures) revealed a significant main effect for time, $F(1, 137) = 4.69$, $p < .05$. There was no significant main effect for group, $F(1, 137) = 0.29$, *ns*, and no significant interaction, $F(1, 137) = 0.12$, *ns*.

Stereotype scale. The control group stereotype score did not change from pre-training ($M = 4.11$, $SD = 0.49$) to post-training ($M = 4.05$, $SD = 0.53$), $t(79) = 1.82$, *ns*. The experimental group stereotype scale score did not change from pre-training ($M = 4.12$, $SD = 0.58$) to post-training (4.06, $SD = 0.57$), $t(68) = 1.31$, *ns*. A repeated measures

ANCOVA revealed no main effect for time, $F(1, 139) = 0.22$, *ns* or group, $F(1, 139) = 0.04$, *ns*, and no significant interaction, $F(1, 139) = 0.00$, *ns*.

Feeling thermometer scale. The control group feeling thermometer score did not change from pre-training ($M = 74.25$, $SD = 15.83$) to post-training ($M = 73.54$, $SD = 16.77$), $t(72) = 0.62$, *ns*. The experimental group feeling thermometer score did not change from pre-training ($M = 75.73$, $SD = 17.68$) to post-training ($M = 74.79$, $SD = 17.83$), $t(67) = 1.50$, *ns*. The repeated measures ANCOVA revealed no main effect for time, $F(1, 132) = 0.10$, *ns* or group, $F(1, 132) = 0.27$, *ns*, and no significant interaction, $F(1, 132) = 0.07$, *ns*.

Relative Change in Attitude Measure Scores

Hypothesis 3 stated that, among faculty receiving the presentation, the change in GNAT scores from pre- to post-training would be greater than the change in PGNAT scores, and the change in explicit attitude scores from pre- to post-training would be greater than the changes in both the GNAT and PGNAT scores. To test hypothesis 3, difference scores (post-training minus pre-training) for each of the implicit and explicit scores in the experimental group were calculated. Dependent-samples *t*-tests were conducted to compare each pairing of implicit and explicit scores.

There was no significant difference between the GNAT ($M = .28$, $SD = 4.58$) and PGNAT ($M = .29$, $SD = 3.76$) difference scores, $t(66) = -.82$, $p > .05$. The comparison of GNAT scores to explicit scales also revealed no significant differences: GNAT with semantic differential ($M = .03$, $SD = .46$), $t(67) = -.79$, $p > .05$; GNAT with stereotype scale ($M = -.05$, $SD = .32$), $t(68) = -56$, $p > .05$; GNAT with feeling thermometer ($M = -.83$, $SD = 4.55$), $t(67) = -.50$, $p > .05$. The comparison of PGNAT scores to explicit scales

were also non-significant: PGNAT with semantic differential, $t(66) = -.13$, $p > .05$;

PGNAT with stereotype scale, $t(66) = .09$, $p > .05$; PGNAT with feeling thermometer,

$t(65) = -.33$, $p > .05$.

IV. DISCUSSION

Stereotypes and implicit bias can negatively impact hiring, promotion, and retention of women in STEM fields. In order to evaluate diversity initiatives aimed at increasing the representation of women in these fields, accurate methods of assessing attitudes must be developed and validated. The paper-based IAT has been validated and compared with the computer-based IAT, but work is currently underway to validate a paper-based GNAT. To date, no prior research has examined the effect of personalizing the GNAT (PGNAT). The purpose of this study was to examine the effect of participation in a diversity training session on attitudes toward women in STEM, assessed using implicit and explicit measures of attitudes, and to compare the sensitivity of these measures in detecting attitude change.

There were several significant correlations between the individual trait measures and the explicit scales. As social desirability and egalitarianism increased, participants reported more positive attitudes toward women as reflected by the stereotype scale and feeling thermometer. As reactance increased, participants reported more negative attitudes on the stereotype scale, and as self-discrepancy increased, participants reported more negative attitudes on the semantic differential scale. These correlations make sense in light of previous research regarding these traits. However, given the fact that the individual trait measures are explicit scales by design, it stands to reason that there would be a stronger correlation with the explicit scales. The strong correlations between three

of the trait scales and the stereotype scale could also have resulted from similarity in scale design, as they all used 5-point Likert-type agreement scales.

Hypothesis 1 stated that explicit attitude scores would correlate more strongly with the PGNAT than with the GNAT. Pre-training correlations between the PGNAT and the explicit scales were similar to the pre-training correlations between the GNAT and the explicit scales. These correlations did not differ significantly or follow any discernible pattern. However, post-training correlations between the PGNAT and the stereotype scale and between the PGNAT and the feeling thermometer were larger than correlations between the GNAT and the stereotype and feeling thermometer scales. Most of the correlations between implicit and explicit measures were not significant. The one exception to this finding was the correlation between the post-training PGNAT score and the post-training feeling thermometer score. This correlation was significant and greater than the correlation between the post-training GNAT score and the post-training feeling thermometer score, partially supporting hypothesis 1.

Hypothesis 1 was based in part on the findings of Olson and Fazio (2004), who found that personalizing the IAT resulted in greater correlations between the PIAT and explicit attitude measures. The results of the current study appear to support this pattern at time 2 (post-training). This suggests that the GNAT might be vulnerable to contamination by extrapersonal association. The fact that this pattern emerged only post-training could indicate that participation in the training session resulted in increased salience of extrapersonal associations regarding women in STEM. Participants had greater awareness of social norms post-training, which might have resulted in attenuation of GNAT scores at time 2. These results also support the findings of Han, Olson, and

36

Fazio (2006), who found that exposure to an artificially-created extrapersonal association resulted in attenuation of the IAT, but not the PIAT.

Hunt and Hunt (2004) found that a one-hour presentation aimed at educating participants and providing stereotype-disconfirming information had a significant effect on attitudes toward people with disabilities in the workplace. Hypothesis 2 stated that training in the form of a one-half-hour presentation would result in more positive attitudes toward women in STEM. The GNAT scores for both the control and experimental groups indicated generally positive attitudes toward women in STEM as reflected by positive score values. The change in GNAT score from pre-training to post-training was greater in the experimental group than in the control group. The direction of the change and the larger difference score in the experimental group suggest that participation in the training session might have had some influence on attitudes. However, this difference was not significant; hypothesis 2a was not supported. As previously stated, this might be due to attenuation of the GNAT scores in the experimental group resulting from influence by extrapersonal associations.

The PGNAT scores for the control and experimental groups also indicated generally positive attitudes toward women in STEM. As with the GNAT, the change in PGNAT score from pre-training to post-training was greater in the experimental group than in the control group. The interaction between time and pre-training PGNAT score was significant, indicating support for the hypothesis that participation in the training session had a positive effect on attitudes as reflected by the PGNAT, supporting hypothesis 2b. This also suggests that the PGNAT might be more sensitive to detecting personal attitude change than the GNAT.

There were no significant differences in any explicit attitude scores between the experimental and control groups, and the only significant difference over time was observed in the semantic differential scale. For this scale, attitudes across groups increased slightly from pre-training to post-training. However, there was no difference between experimental and control groups. Hypothesis 2c was not supported. This may suggest that participants were able to recall the responses they provided on the pre-training packet and could therefore consciously attempt to respond similarly on the post-training packet. Because the social desirability pre-measure correlated significantly with the stereotype scale and the feeling thermometer, this could also suggest that participants were driven by social desirability bias, and did not want to appear to have experienced any change in attitude, despite the findings implied by the implicit attitude measures. If this is the case, the difference by group and time observed in the PGNAT might support the assertion that a personalized implicit measure of attitudes may be preferable in situations where social norms exert considerable pressure on participants, as is the case when the topic is highly sensitive or controversial (King & Bruner, 2000; van de Mortel, 2008).

Analyses looking at the effect of training on the GNAT and PGNAT also revealed a somewhat surprising trend. The experimental group had higher baseline GNAT scores than the control group. The pre-training difference between groups was not significant, but it was marginal. Since participants in the experimental group knew that the presentation they were attending was affiliated with the NSF ADVANCE LEADER grant, whose mission is to promote the advancement of women in STEM, it is possible that social norms regarding women in STEM were more salient for this group than the

control group. This suggestion supports previous research that has shown that the IAT (and therefore the GNAT) is more susceptible to influence by extrapersonal associations, such as awareness of social norms (Han et al., 2010). In contrast, the baseline PGNAT scores for the experimental and control groups were very close, which also supports previous research; the personalized implicit tasks are purported to be less vulnerable to extrapersonal associations, and the results of this study support this claim. The PGNAT may be less susceptible to influence by awareness of social norms and may better assess "true" attitudes when compared with the GNAT.

Hypothesis 3 predicted that, in the experimental group, explicit attitude scores would change more than GNAT scores, and the GNAT would change more than the PGNAT. The explicit scales did not change significantly from pre- to post-training. The PGNAT and GNAT difference scores were not significantly different from one another, but they appeared to be larger than the difference scores observed in the explicit scales, contrary to hypothesis 1. The lack of significant change in the explicit scales might be due to the fact that the pre-training scores started out relatively high, resulting in a ceiling effect. It could also be a result of the use of identical measures in both the pre-training and post-training packets, making it possible for participants to recall their responses from the pre-training packet and consciously attempt to duplicate these responses in the post-training packet. It is also worth noting that there was a relatively short amount of time between explicit scale administrations. Participants therefore might not have been given sufficient time to consciously reflect on the information given in the presentation. Such conscious reflection would be necessary to result in a change in conscious, explicit attitudes.

Limitations

One potential limitation in this study was the use of a paper-based GNAT and PGNAT, as opposed to using a computer-based GNAT/PGNAT. The computer-based implicit software would have allowed for more precision in measuring attitudes, based not only on number of stimuli items correctly categorized, but also on response latency. Furthermore, due to paper size constraints, we were limited to 28 items per page. Some participants were able to finish categorizing all 28 items within the 15 second time frame in the post-training packet, resulting in a ceiling effect. A practice effect was noted across the sample, resulting in a larger number of items categorized in the post-training packet. However, this was controlled for in the algorithm used to calculate GNAT and PGNAT scores by examining the ratio of correctly categorized items to number of items categorized. In a computer-based GNAT or PGNAT, response latency is measured on an item-by-item basis, which would have eliminated the ceiling effect observed in participants who displayed faster categorization abilities.

Another limitation in the design of this study was the use of a flower-insect GNAT and PGNAT in the practice tasks. We wanted to avoid biasing participants by suggesting that there was a "right" or "wrong" response, so we did not use the final stimuli pictures in the practice. An unforeseen result of this was that the number of items categorized on the first GNAT or PGNAT was lower than on subsequent blocks as participants simultaneously attempted to complete the task according to directions while trying to determine which picture items fit into the "female scientist" category. There was also a relatively high level of error in the first GNAT/PGNAT blocks as participants

attempted to learn the new stimuli. It is possible that a practice block using the final stimuli might have corrected for this, resulting in a more sensitive task.

There was no discernible change in explicit attitude measures from pre- to post-training. It is possible that this was due to the fact that the explicit measures were identical in the pre-training and post-training packets. There was a very short span of time between completing packets (about 30 minutes), which meant that participants could easily recall their responses from the pre-training and consciously attempt to replicate their responses in the post-training packet. Use of parallel forms, or a greater time lag between administrations might correct for this. Time between administration of pre- and post-training attitude measures is also a possible limitation. It is also possible that there was not enough time for participants to reflect on the information to which they were exposed, and therefore insufficient time for explicit attitudes to change in a meaningful way.

A final limitation results from the use of "Female scientist" as the target category. This might actually reflect two sets of attitudes, one toward females and one toward scientists. In essence, this creates the same problem Karpinski (2004) observed in the IAT, in that we are unable to parse out the source of variance resulting from attitudes toward these two groups.

Implications and Future Research

The paper-based GNAT and PGNAT developed for use in this study can be used to measure attitudes in future research settings, particularly when the use of a computer is not possible or practical. The findings of this study suggest that the PGNAT might be better suited for assessing personal attitudes regarding topics that are sensitive or

controversial in nature, particularly if there is concern that participants might be inclined to distort their responses on an explicit measure due to social desirability bias.

The findings of this study revealed that, while mean implicit attitude scores were positive rather than negative, the average response is closer to a neutral point. Furthermore, participation in a training session on implicit bias appeared to have a positive effect on implicit attitudes. This, combined with the fact that there continue to be lower representation of women in STEM fields, suggests that there is still work to be done to improve the climate for women in these fields. Education and conscious awareness of the biases that occur toward this group could help to create a warmer climate, resulting in greater numbers of women entering into and remaining in STEM fields.

Potential future research could include an attempt at duplicating these results using a computer-based GNAT and PGNAT. Increasing the number of participants might also increase significant findings when results are trending toward significant, but lacked the statistical power to achieve it. It would also be useful to administer the attitude measures used here in conjunction with a presentation on implicit bias to other populations, such as undergraduate STEM majors, in order to further validate the use of these measures.

Future use of the paper-based GNAT and PGNAT should include a practice task that uses the final stimuli to be used in the measured blocks to avoid confusion and reduce error rates. To avoid ceiling effects, the blocks themselves should also be adjusted to include a larger number of stimuli per page. The use of parallel forms on the explicit attitude measures could also increase sensitivity to attitude change. Follow-up

studies should be done using this population to determine whether any observed attitude change has a lasting effect.

Follow-up studies should be completed using this same population to determine whether the effect observed here is lasting. Future analyses of this data should also include a comparison between responses given by male and female participants as well as a comparison between responses in engineering and science departments.

Conclusion

Stereotypes toward women in STEM are pervasive and can negatively impact their hiring, promotion and tenure, and long-term success. In order to determine whether diversity initiatives are effective and to improve such initiatives, accurate attitude measures must be developed and validated. This study has shown that the PGNAT might be an effective tool for use in such initiatives. If measures such as the PGNAT are validated and shown to accurately detect attitude change, they can be used to evaluate and improve upon diversity training. Because attitudes can affect behavioral outcomes, these measures might also be useful in predicting and reducing stereotype-driven behaviors. The implications of such findings are wide-reaching, and can be applied to a variety of fields to improve training and reduce bias.

V. REFERENCES

Aberson, C. L., & Haag, S. C. (2007). Contact, perspective taking, and anxiety as predictors of stereotype endorsement, explicit attitudes, and implicit attitudes. *Group Processes*, *10*(2), 179-201. doi: 10.1177/1368430207074726.

Aberson, C. L., Shoemaker, C., & Tomolillo, C. M. (2004). Implicit bias and contact: The role of interethnic friendships. *Journal of Social Psychology*, 144, 335–347. Retrieved from http://www.heldref.org/pubs/soc/about.html.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Antonak, R. F. & Livneh, H. (2000). Measurement of attitudes towards persons with disabilities. *Disability and Rehabilitation*, *22*(5), 211-224.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, *6*(3), 242-261. doi: 10.1207/S15327957PSPR0603_8.

Blair, I.V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828-840. doi:10.1037//0022-3514.81.5.828

Blanton, H. & Jaccard, J. (2006). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *113*(1), 155-169. doi: 10.1037/0033-

295X.113.1.155.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*(3), 567-582. Retrieved from http://www.apa.org/pubs/journals/apl/index.aspx.

Brehm, J. W. (1966). *A Theory of Psychological Reactance.* New York, NY: Academic Press.

Campbell, J., Gilmore, L., & Cuskelly, M. (2003). Changing student teachers' attitudes towards disability and inclusion. *Journal of Intellectual & Developmental Disability*, *28*(4), 369-379. doi: 10.1080/13668250310001616407.

Canes, B. J., & Rosen, H. S. (1995). Following in her footsteps? Faculty gender composition and women's choices of college majors. *Industrial and Labor Relations Review, 48,* 486–504. Retrieved from http://www.ilr.cornell.edu/ilrreview/.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Cook, S. W., & Selltiz, C. (1964). A multiple-indicator approach to attitude measurement. *Psychological Bulletin*, *62*(1), 36-55. doi: 10.1037/h0040289.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping.

Journal of Experimental Social Psychology, 40, 642–658.

doi: 10.1016/j.jesp.2004.02.003.

DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F.

Turner & E. Martin (Eds.), *Surveying Subjective Phenomena* (Vol. 2, pp. 257-

281). New York: Russell Sage.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled

components. Journal of Personality and Social Psychology, 56, 5–18.

doi: 10.1037//0022-3514.56.1.5.

Eagly, A. E., & Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth, TX:

Harcourt Brace Jovanovich.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses

using G*Power 3.1: Tests for correlation and regression analyses. *Behavior*

Research Methods, 41, 1149-1160.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research:

Their meaning and use. *Annual Review of Psychology*, *54*(1), 297-327.

doi:10.1146/annurev.psych.54.101601.145225.

Fazio, R. H., Jackson, J. R., Dunton, B. C., Williams, C. J. (1995). Variability in

automatic activation as an unobtrusive measure of racial attitudes: A bona fide

pipeline? *Journal of Personality & Social Psychology*, *69*(6), 1013-1027.

doi: 10.1037//0022-3514.69.6.1013.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-

esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27. doi: 10.1037/0033-

295X.102.1.4.

Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K. (1998). Measuring individual

differences in implicit cognition: The Implicit Association Test. *Journal of*

Personality and Social Psychology, 74(6), 1464-1480.

Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean?

In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new*

implicit measures (pp. 65–82). Hillsdale, NJ: Erlbaum.

Greenwald, A. G., Rudman, L. A., Nosek, B. A., & Zayas, V. (2006). Why so little faith?

A reply to Blanton and Jaccard's (2006) skeptical view of testing pure

multiplicative theories. *Psychological Review*, *113*(1), 170-180. doi:

10.1037/0033-295X.113.1.170.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., Banaji, & M. R. (2009).

Understanding and Using the Implicit Association Test: III. Meta-Analysis of

Predictive Validity. *Journal of Personality & Social Psychology*, *97*(1), 17-41.

doi: 10.1037/a0015575.

Hackney, A. (2005). Teaching students about stereotypes, prejudice, and discrimination:

An interview with Susan Fiske. *Teaching of Psychology*, *32*(3), 196-199.

doi:10.1207/s15328023top3203_13

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or

malleability of the IAT? *Journal of Experimental Social Psychology*, *46*(2), 286-

298. doi: 10.1016/j.jesp.2009.11.011.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created

extrapersonal associations on the Implicit Association Test. *Journal of*

Experimental Social Psychology, *42*(3), 259-272. doi:10.1016/j.jesp.2005.04.006

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369-1385. doi: 10.1177/0146167205275613.

Hong, S. & Faedda, S. (1996). Refinement of the Hong Psychological Reactance Scale. *Educational & Psychological Measurement*, *56*(1), 173-182.

Hunt, C.S., & Hunt, B. (2004). Changing attitudes toward people with disabilities: Experimenting with an educational intervention. *Journal of Managerial Issues*, *16*(2), 266-280. Retrieved from http://www2.pittstate.edu/econ/jmi.html.

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,* 365–371. doi: 10.1111/1467-9280.00272.

Jelenec, P., & Steffens, M. C. (2002). Implicit attitudes toward elderly women and men. Current Research in Social Psychology, 7, 275–293. Retrieved from http://www.uiowa.edu/%7Egrpproc/crisp/crisp.html.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: III. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology*, *28*(4), 360-386. doi: 10.1016/0022-1031(92)90051-K.

Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin*, *30*(1), 22-34. doi:10.1177/0146167203258835

Karpinski, A. & Hilton, J.L. (2001). Attitudes and the Implicit Association Test. *Journal*

of Personality and Social Psychology, *81*(5), 774-787. doi:10.1037//0022-3514.81.5.774

Katz, I. & Hass, G. R. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality & Social Psychology*, *55*(6), 893-905.

King, M. & Bruner, G. 2000. Social desirability bias: a neglected aspect of validity testing. *Psychology and Marketing*, *17*(2), 79–103.

Lemm, K. M., Lane, K. A., Sattler, D. N., Khan, S. R., & Nosek, B. A. (2008). Assessing implicit cognitions with a paper-format Implicit Association Test. In T.G. Morrison & M. A. Morrison (Eds.), *The Psychology of Modern Prejudice*. Hauppauge, NY: Nova Science Publishers.

McCauley, C., Wright, M., & Harris, M. E. (2000). Diversity workshops on campus: A survey of current practice at U.S. colleges and universities. *College Student Journal*, *34*(1), 100-114. Retrieved from http://hwwilsonweb.com.

Monteith, M. J. & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality & Social Psychology*, *75*(4), 901-916.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18,* 879–885. doi: 10.1111/j.1467-9280.2007.01995.x.

National Academy of Science. (2006). *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering*. Washington, DC: National Academies Press.

Nosek, B.A. & Banaji, M.R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*(6), 625-668. doi:10.1521/soco.19.6.625.20886

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics, 6,* 101–115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83,* 44–59. doi: 10.1037//0022-3514.83.1.44.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2006). Website: http://implicit.harvard.edu/.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292). New York: Psychology Press.

Nosek, B. A., & Hansen, J. J. (2008). Personalizing the Implicit Association Test increases explicit evaluation of target concepts. *European Journal of Psychological Assessment, 24*(4), 226-236. doi: 10.1027/1015-5759.24.4.226.

Nosek, B. A. & Lane, K. (1999). *Analyzing paper-pencil IAT data.* Unpublished manuscript, Yale University.

Olson, M. A., & Fazio, R. H. (2004). The influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653-667. doi: 10.1037/0022-3514.86.5.653.

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality & Social Psychology Bulletin*, *23*(2), 173-185. doi: 10.1177/0146167297232006.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*(2), 79-82. doi 10.1111/j.0963-7214.2004.00279.x.

Rudman, L. A., Ashmore, R. D., Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*(5), 856-867. doi: 10.1037//0022-3514.81.5.856.

Schneider, T. R., Lyons, J. B., & Rivers, S. E. (2009). The biobehavioral model of persuasion: Generating challenge appraisals to promote health. *Journal of Applied Social Psychology*, *39*(8), 1928-1952. doi: 10.1111/j.1559-1816.2009.00510.x.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39,* 68–74. Retrieved from http://www.sciencedirect.com/science/journal/00221031.

Steele, C.M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality & Social Psychology*, *69*(5), 797-811.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

Stereotype. (2005). *The New Oxford American Dictionary* (2nd ed.). New York: Oxford Press.

Teachman, B. A. & Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: is anyone immune? *International Journal of Obesity & Related Metabolic Disorders*, *25*(10), 1525-1531.

Thomas, A., Doyle, A., & Vaughn, D. (2007). Implementation of a computer based
        Implicit Association Test as a measure of attitudes toward individuals with
        disabilities. *Journal of Rehabilitation*, *73*(2), 3-14.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*,
        529-554. Retrieved from
        http://www.brocku.ca/MeadProject/Thurstone/Thurstone_1928a.html.

Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report
        research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

Vargas, P., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and
        pencil: "Low-tech" measures of implicit attitudes. In B. Wittenbrink and N.
        Schwarz (Eds.) *Implicit Measures of Attitiudes*, pp.103-124.

Wilson, T. D., Lindsey, S., Schooler, T. Y. (2000). A Model of dual attitudes.
        Psychological Review, *107*(1), 101-126. doi: 10.1037/0033-295X.107.1.101.

Wittenbrink, B., & Schwarz, N. (Eds.). (2007). *Implicit Measures of Attitudes*. New
        York, NY: The Guilford Press.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic
        ideology: Effects of multicultural and color-blind perspectives on judgments of
        groups and individuals. *Journal of Personality and Social Psychology*, *78*(4), 635-
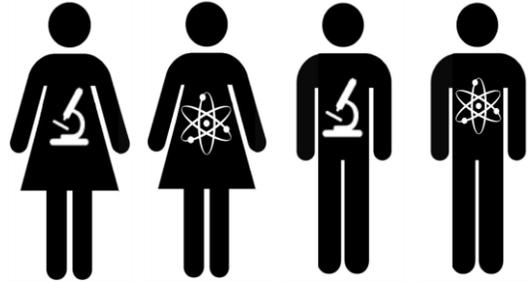   654. doi: 10.1037//0022-3514.78

List of Stimuli Items Used in the GNAT and PGNAT

Pleasant (I like) items:

joy

love

peace

friend
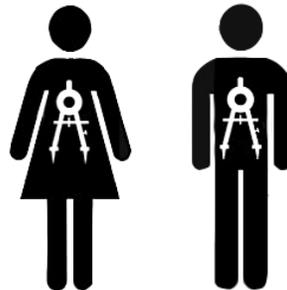
laughter

Male and Female Scientist Symbols:

Unpleasant (I don't like) items:

agony

poison

failure

evil

crash

Male and Female Non-scientist Symbols:

Male and Female Engineer Symbols:

Notes. Pleasant/Unpleasant (I like/I don't like) stimuli from: Greenwald, McGhee, & Schwartz (1998), Nosek, Banaji, & Greenwald (2002a), and Olson and Fazio (2004). Male and female figures from Nosek, Banaji, & Greenwald (2006).