THE ROLE OF ACTOR (EMPLOYEE) IDENTITY ON PERFORMANCE ASSESSMENT

by

BRIAN DAVID ROOTE

(Under the Direction of KECIA MONIQUE THOMAS)

ABSTRACT

Drawing from research on heterosexism and gender-role stereotypes, the systematic variations in performance evaluations for Gay and Straight workers were examined. Students in the lab and managers in the field viewed a ten-minute video clip of a target manager, presented as either Gay or Straight, engaged in several performance behaviors related to the sales industry in a high vs. average vs. low performance condition. Results demonstrated an overall aversive heterosexist (pro-Gay) effect in the high performance condition (lab) and average performance condition (field). Male students in the lab, and female managers in the field, assigned the highest ratings to the Gay manager, although their level of gender identity did not moderate this relationship. Finally, evaluations of the Straight manager in the field were more accurate than were evaluations of the Gay manager. Implications for theory and for practice as well as suggestions for future research are discussed.

INDEX WORDS: Aversive Heterosexism, Performance Appraisal, Personnel Evaluation, Rating Accuracy, Stereotypes, Gender Identity

THE ROLE OF ACTOR (EMPLOYEE) IDENTITY ON PERFORMANCE ASSESSMENT

by

BRIAN DAVID ROOTE

B. S., Georgia State University, 2003

M. S., University of Georgia, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

THE ROLE OF ACTOR (EMPLOYEE) IDENTITY ON PERFORMANCE ASSESSMENT

by

BRIAN DAVID ROOTE

Major Professor:	Kecia M. Thomas
Committee:	Brian J. Hoffman
	Victoria Plaut

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

DEDICATION

Others have said, and I firmly believe that I stand on the shoulders of those who have come before me. Their efforts continue to enrich my curiosity and inspire my contribution to the scientific community. I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Kecia Thomas, my committee chair, a mentor, and friend, and for her countless hours of reflecting, reading, encouraging, and most of all patience throughout the entire process. Her extensive career in diversity and inclusion continue to examine topics that are relevant, timely, and beneficial to a global workforce.

I would like to thank Dr. Brian Hoffman for agreeing to serve on my committee, and especially for the initial idea of examining the performance assessment and sexuality link, as well as his expertise and guidance along the way. I am grateful for the insight and support of my final committee member, Dr. Victoria Plaut, who was instrumental in the study design.

I am thankful for the love freely given by my family and friends, all of whom supported me during my educational journey. I look forward to the challenges ahead, and am profoundly grateful for the extensive opportunities afforded to me because of the care and investment of so many.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Lesbian, Gay, Bi and Transgendered (LGBT) individuals continue to stand on the margins of civil rights protection and can face discrimination in almost every aspect of their lives. Many Americans incorrectly assume that homosexual and transgendered employees have workplace protections against discrimination and harassment. Responses of self-identified Lesbian, Gay and Bi adults in a national sample indicate significant instances of violence and workplace discrimination, especially Gay men. For example, 63 percent of Gay men reported verbal abuse, 35 percent reported acts of violence, and 16 percent reported job discrimination at some point in their lives (Herek, 2008). The data obtained in a series of national health, social and census surveys, including self-reports of Gay men, found that Gay men earn anywhere from 10 to 32 percent less than otherwise similar Straight men. Within the same study, 10 to 28 percent of LGB respondents reported being denied a promotion or given negative performance evaluations (Badgett, Lau, Sears, & Ho, 2007). When the experiences of Transgendered individuals are included, the impact of discrimination is even more severe. More than 75 percent of the LGBT community has experienced or knows someone who has experienced discrimination in applying for college entrance, buying a house, applying for or keeping a job, getting insurance, or while serving in the military (The Kaiser Family Foundation, 2001). The widespread discrimination against the LGBT community demonstrates the urgent need to understand the workplace experiences of this under-researched group.

There is a gap in academic literature. Currently, no research has examined the assessments of the work performance of LGBT employees. Stereotypes and anti-Gay attitudes presumably impact performance evaluations of Gay male employees. This study seeks to make a

start on filling in the gap by exploring the extent to which stereotypes and attitudes toward Gay male managers influence subjective performance evaluations [1].

Overview of the Performance Evaluation Literature

By definition, performance evaluations continue to be widely utilized as an administrative tool for making critical decisions regarding promotions, layoffs, separations, and transfers. They are also commonly linked to important career development issues such as ownership, input, perceptions of being valued, and regarded as a contributor to the organizational team (Levy & Williams, 2004). Because evaluations by definition include forming an opinion of others by "discerning or comparing" (Hollenbeck, 2008), they are widely criticized for being subjective in nature and likely vulnerable to personal biases (Borman, Dunnette, & Hough1991; Dickinson, 1987; Feldman, 1981). Consequently, various forms of bias in performance evaluations can effectively provide some employees access to highly valued resources, such as promotions and pay raises, while blocking the opportunities of others. Typically, the careers of White men, assumed to be Straight, benefit through the maintenance of race, gender, and heterosexist based systems of privilege (e.g., the glass escalator; see Maume, 1999). On the other hand, these same systems can impede the upward mobility of racial minorities, women (e.g., the glass ceiling; Stroh, Langlands, & Simpson, 2004), and likely Gay managers as well (Button, 2001; Munoz & Thomas, 2005). Hence, the performance appraisal can serve as a persistent tool for maintaining systemic barriers to success when they include biases.

For these and other reasons, many psychologists regard work performance ratings as the Achilles' heel of the discipline of I-O psychology (Murphy, 2008). Rather than striving to evaluate subordinate employees' true performance, supervisors have been widely accused of promoting their own political agendas, and some have even openly admitted that this is the case.

In particular, supervisors' goals can include "moving" an employee up (down) and out of a particular work group, striving to motivate an employee by providing artificially inflated (deflated) ratings, or attempting to elicit perceptions of managerial effectiveness by increasing the ratings of subordinates (Longenecker, Sims, & Gioia, 1987).

Performance evaluation research conducted in a laboratory setting has also faced criticism. Landy (2008), recently criticized the usefulness of lab studies exploring stereotyping in work performance and other similar research paradigms, but he did not take into account the fact that in the applied setting, political goals, among other sources of error (e.g., social information processing errors [2]; (for a review, see Woehr, 2008), can seriously influence performance evaluations. It has been noted that undergraduate research participants only view and rate a thin slice of behavior. Moreover, they are untrained in assessing work performance, sometimes minimally invested in the lab study, not required to personally deliver feedback to the subordinate employee (Landy, 2008), and do not have the opportunity to interact with the subordinate employee (Dipboye, 1985).

In response, it is important first to establish the extent to which heterosexism exists in the observation and rating of Gay versus Straight managers. This initial attempt should take place in a carefully controlled laboratory study, in order to extend our current understanding of work performance theory and behavior. To increase the generalizability of the current study, a sample of managers currently working at least half time within an organizational setting will also be included.

CHAPTER 2

LITERATURE REVIEW

Stereotypes' Influence on Subjective Evaluations

The process of evaluating an employee's behavior has long been criticized as inherently subjective and open to personal biases (Borman, et al.1991; Dickinson, 1987; Feldman, 1981). Stereotypes, defined as beliefs or expectations about the status and characteristics of specific social groups, can influence these evaluations and judgments (Fiske & Taylor, 2008). More specifically, stereotypes are "…*everyday base rates*, or beliefs about the relative prevalence of traits, behaviors, or other attributes among a particular category of people, as compared to people in general" (Nelson, Acker, & Manis, 1996). To illustrate, the belief that Gay men are more likely to exhibit feminine characteristics than men are in general represents this sort of base rate overgeneralization.

A body of literature demonstrates that stereotypes direct our attention, influence both our social interactions and the behavior we store in memory (Jones, 2002). Empirical evidence suggests that information consistent with a group stereotype is better remembered than is information inconsistent with the stereotype[3]. For example, participants recalled specific "librarian like" information when the target was presented as a librarian (e.g., has glasses, books on the shelf) than when she was presented as a waitress (e.g., likes hamburgers), even though the video remained the same in both conditions. Further, participants' memory for the stereotype consistent information was not weakened by varying the time delay (i.e., immediate, 4 days, 7 days; (Cohen, 1981). Thus, memory for stereotype consistent information is robust and can influence future expectations.

Stereotypes also guide the way individuals make sense of their everyday lives. This process has been identified as a two-stage process. Once a stereotype is activated (stage one), individuals' transition to hypothesis confirmation mode (stage two); otherwise regarded as the "seeing is believing" phenomena. Consequently, individuals tend to better retain information that confirms their initial expectation and ignore refuting information. Darley and Gross (1983) demonstrated a biased information seeking process by asking participants to rate the academic ability of a fourth-grade girl. Half of the participants viewed a video of the target, Hannah, in a middle-class setting versus a lower-class setting. Next participants watched the target perform ambiguously on several achievement–test problems. That is, she answered both easy and hard questions accurately and inaccurately. Participants in the middle class condition rated her academic ability more favorably, attributing her mistakes to external conditions (i.e., a bad day), then did participants in the lower class condition, who were more likely to attribute her mistakes to internal conditions (i.e., less intelligent). It is likely that within the work setting, when encountering both effective and ineffective behaviors, raters may tend to focus on and recall ineffective behaviors of negatively stereotyped targets (i.e., Gay managers) more readily then than will for non-stereotyped targets.

Even with increasing tolerance for Gay men, cultural stereotypes, especially those propagated by the media, are criticized for being overly negative and can shape individuals' attitudes (GLAAD Media Reference Guide, 2007). Some of the most common negative stereotypes of Gay men characterize them as "mentally ill", "promiscuous", "lonely", "insecure", "dainty", "walk[ing] like girls", and "likely to be child molesters" (for a review see Herek, 1984; Madon, 1997). Stereotypes can also be positive. In a series of studies, researchers found that participants described Gay men as "more helpful to others", "more tactful", "warmer in

relationships", "fashionable," "good listeners," and "open with their feelings." (Madon 1997; & (Taylor, 1983).

Although stereotypes can be positive, they still reflect prejudice because they restrict and limit the lives of their targets. As an example, Wilson (1996) found that individuals holding positive stereotypes of Jews (e.g., financially successful, ambitious, and hardworking) also expressed blatant anti-Semitic attitudes. In fact, regardless of valence, stereotypes tend to be prescriptive in nature, and constrain the stereotyped individual's behavior to ways that people "ought" to act (Jones, 2002) and stereotypes of Gay men may thus limit employment opportunities and advancement to some of the highest levels within the organization.

In sum, both positive and negative stereotypes of Gay men serve to limit and define behavior. Individuals' memories are stronger for stereotype consistent information (Cohen, 1981). Additionally, regardless of valence, stereotypes ostensibly influence a rater's attention and recall of job relevant behavior to include some of the culturally shared aspects of stereotypes of Gay men. For the purpose of the study, the work performance of the manager will include behaviors that are both effective and ineffective (i.e., by manipulating performance level). As demonstrated in prior research, stereotypes should guide the interpretation of ambiguous information.

Heterosexism in the Workplace

If stereotypes comprise the cognitive function of prejudice, then what accounts for the negative attitudes and behaviors directed toward sexual minorities? A number of proposed psychological theories help explain the attitudes, stereotypes, and behaviors directed toward the LGBT community. Two of the most relevant perspectives, include heterosexism and the gender-role approach, which will serve as theoretical lenses for the proposed study.

Conceptualization of heterosexism includes both a structural and individual level of analysis. At the structural level, Herek (1990) defines heterosexism as "an ideological system that denies, denigrates, and stigmatizes any non-heterosexual form of behavior or identity" (p. 316). Further, Herek (2007) discusses heterosexism as a distribution of power in which sexual minorities, through heterosexist institutions (e.g., religion, the law, and medicine) assume an inferior status relative to heterosexuals. Two separate processes sustain the differentiation in power. First, the widespread assumption that everyone is heterosexual renders most LGBT people invisible in almost every aspect of society. Second, when LGBT individuals make their sexual identity known, heterosexist beliefs pathologize them. Specifically, through the lens of heterosexism, heterosexual people symbolize what is normal, prototypical and ideal. In contrast, non-heterosexual identities represent the "abnormal" and are denigrated as strange, unnatural and deviant.

The supposed invisibility of sexuality further maintains its stigma; however, current research demonstrates that Gay men may be more easily detected than was once believed. A sample of college students systematically selected the correct sexual orientation of stimulus targets after simply viewing photographs of Gay and Straight persons. The accuracy for correctly guessing sexuality increased when participants viewed either a short video clip (Rieger, Linsenmeier, Gygax, Garcia, & Bailey, 2007; Rieger, Linsenmeier, Gygax, & Bailey, 2008) or when a college sample viewed a computer simulation depicting subtle variations in the ways Gay versus Straight individuals walk (Johnson, Gill, Reichman, & Tassinary, 2007). In organizations, it is widely believed that Gay individuals can choose to disclose their identity or remain safe by keeping their sexuality closeted. However, it may be more realistic to view this phenomenon as the "myth of invisibility" because many Gay men are unsuccessful at hiding their identity (e.g.,

some Gay and Straight men possess characteristically gender nonconforming behaviors). In addition, Gay and Straight men alike can be "outted" by anyone who considers them to be sexually suspect. For example, unless a Gay man engages in a counterfeiting strategy (i.e., making up a Straight relationship), the level of interdependence in the typical work environment usually gives way to office gossip and assumptions about coworkers' personal lives and relationships (hypothesis confirmation discussed earlier). This may lead some employees to investigate the sexual identity of suspected Gay employees. Further, Gay employees attempting to manage a double life by concealing their identity at work, suffer in regards to career satisfaction, intent to remain at the job, and overall job satisfaction (Button, 2003). These types of outcomes reported by Gay men mimic the experiences reported by other minority employees based on race and gender.

According to the theory of heterosexism, once a Gay employee's sexuality becomes apparent, an implicit activation of assumptions and stereotypes operate. Individuals endorsing antigay attitudes, especially Straight men, may be more likely to engage in discriminatory behaviors (Herek, 2007). On the other hand, due in part to the rapid changes in social norms and the widespread need to maintain a positive self-image, some types of situations may be more likely to elicit discriminatory behaviors than others, but only when the behaviors can be rationalized.

Contemporary Discrimination

Prejudice has changed considerably over the last half of the 20th century. Before the civil rights movement, overt forms of discrimination were widely accepted. With the passage of explicit laws protecting minority group members, customs and social norms began to change. Individuals resisting minorities and diversity initiatives began exhibiting more subtle, hard to

detect forms of discriminatory behaviors instead. Two forms of contemporary prejudice emerged to help explain this change. The first perspective, modern racism, described by McConahay (1983; 1986), is a form of discrimination that can include overt behaviors but more likely includes covert and hard to detect behaviors. According to McConahay (1983; 1986), covert racists (or heterosexists) believe that discrimination is outdated, that Blacks (or Gays) push too hard in their fight for equal rights and that the demands are unfair, and any gains are undeserved.

Dovidio and Gaertner (1998; 2008; 1986) describe aversive racists as holding egalitarian values, and do not believe themselves to be prejudiced. Much like contemporary racists, aversive racists engage in subtle forms of discrimination rationalized based on some factor other than race, gender, or sexuality. The mechanism driving aversive racism is the simultaneous need to believe one is egalitarian and yet hold negative stereotypes about minority groups. Aversive racists are more likely to demonstrate discriminatory behaviors in ambiguous situations, where it would be easier to rationalize the situation as the cause for their reaction, as opposed to race, gender or sexuality. This process allows the aversive racist to maintain both an egalitarian belief system as well as protect his or her self-esteem.

Despite the lack of widespread workplace protection for Gay and Lesbian employees, overt forms of discrimination are still less likely to occur in workplaces. To illustrate, a study examining interviewer's reactions to a Gay versus Straight job applicant, found no difference in the hiring recommendations between the applicants. The differences surfaced however in the various micro aggressions of the interviewer, which included engaging in shorter verbal interactions, asking less follow up questions, and maintaining greater personal distance from the Gay applicant (Hebl, Foster, Mannix, & Dovidio, 2002). The implications of this study demonstrate the shift in discrimination from overt behaviors, to hard to detect ones. Therefore,

regardless of sexuality, when performance is strong, evaluations should reflect greater accuracy. Instead, when performance is ambiguous (i.e., average or low), and includes both positive and negative behaviors, discriminatory behaviors are more easily justifiable and a likely outcome. Within these types of situations, Gay managers in particular may suffer from antigay attitudes and stereotypes.

Systematic Bias in Performance Evaluations

Two forms of bias were examined in the present study. The first, psychometric bias, has been described as a statistical concept that refers to the systematic over or underestimation of true scores in the population (Jensen, 1980) due to variants such as respondent characteristics or testing situations. Tests of measurement equivalence between groups as well as evaluating accuracy indices typically provide an assessment of the existence of psychometric bias.

Psychosocial bias, on the other hand, reflects the negative attitudes and behaviors that benefit some groups at the expense of others. In order to examine the influence of psychosocial bias, researchers may determine mean level differences on outcome variables such as intent to hire, willingness to partner, and likeability between a majority and minority group of interest. Testing for psychosocial bias usually includes mean score tests such as analysis of variance and regression analyses.

Both psychometric and psychosocial biases continue to permeate the occupational landscape. Even minority groups who benefit from federal protection continue to experience discrimination at significant rates which are reported each year by the Equal Employment Opportunity Commission (EEOC). What has been described as "unintentional bias" (i.e., psychosocial bias) often obstructs the evaluations of racial and cultural minorities. For example, in a meta-analytic study, Kraiger and Ford (1985) found that on average supervisors rated

subordinates of the same race more favorably than they rated subordinates of a different race. In an attempt to replicate this finding, Sackett, DuBois and Noe (1991) examined the performance ratings of more than 13,000 military and civilian workers. Results demonstrated that both Whites *and* Black supervisors actually assigned higher ratings to White targets than to Black ones. This study thereby provides strong support that in many organizations, Whites consistently benefit from substantially higher ratings (Greenhaus, Parasuraman, & Wormley, 1990; Mount, Sytsma, Hazucha, & Holt, 1997).

The effects of biases also hold true for women. In a study that included over one hundred male middle managers examining potential explanations for the lack of women in upper levels of management (i.e., the glass ceiling), researchers demonstrated that female managers tended to be rated less favorably than their male counterparts on several measures of performance (Martell, Parker, Emrich, & Crawford, 1998). Although the overall effect was small, the authors established through a computer simulation of a hypothetical company that an effect accounting for only 1% of the variance dramatically limited the number of women who gained entry into senior management (Martell, Lane, & Emrich, 1996; Robison-Cox, Martell, & Emrich, 2007). Further, they stated that as employees proceed to the top levels within the organization, competition is greatly increased and even a small amount of bias can result in passing women over for promotions. The issue then in part is also one of discrimination.

Heterosexism in the workplace often consists of treating Gay employees less favorably than their Straight counterparts. Although no study has yet examined the relationship between sexual identity and work performance evaluations, one study has come reasonably close. When examining the evaluations of feminine (i.e., gentleness, understanding, and affection) and masculine (i.e., ambitious, self-reliant, independent, assertive) described men and women,

feminine acting men were rated lowest on various performance dimensions. Interestingly, the observed pattern of results applied to the high performance condition only. In the low performance condition, MBA graduate students provided the lowest ratings to the masculine described men (Moore, 1984). Unfortunately, the study utilized a within subjects design, and participants may have become aware of the hypotheses after rating 8 men and women who were presented as masculine or feminine.

In another laboratory study examining sexual orientation and selection recommendations demonstrated a differential pattern of ratings for Gay and Straight applicants. In the study, male undergraduate participants rated Gay men as less likely to experience career success and provided lower acceptance recommendations as elementary and fine art educators as compared to Straight men applying for the same position. There were no differences in ratings however, on measures of applicant qualifications and likelihood of graduate program success between the two groups (Jackson & Sullivan, 1989). Thus, Gay men were seen as qualified but were ultimately not selected for admission into the various graduate training programs. The results demonstrate the extent to which heterosexism and stereotypes can seriously limit the careers of Gay men.

The proliferation of antigay stereotypes within the work setting can impede Gay employees' ability to advance to some of the highest levels within some organizations. In a laboratory study, Golom and Liberman (2008) asked student participants to rate the extent to which homosexual and heterosexual, female and male managers were similar to a "successful manager" prototype. The results indicated that Gay male managers were viewed the least similar to the prototype. In fact, Gay male managers received lower mean ratings when compared to Straight male managers on various dimensions of instrumental performance, which included: *work competence, independence, activity/potency, rationality*. However, Gay male mangers

received higher ratings on only one dimension of <u>relational performance</u>: *concern for others*. In addition, the Gay manager received lower ratings on *emotional stability*. However, there were no differences found on the dimension of *hostility toward others* between the two groups. Based on that study, the proposed study will include several similar predictions on specific dimension level performance.

In sum, volumes of research have established a rating bias with regard to minority group members with varying levels of effect sizes (Greenhouse, Parasuraman, & Wormley, 1990; Martell, Parker, Emrich, & Crawford, 1998; Martell, Lane, & Emrich, 1996; Mount, Sytsma, Hazucha & Holt, 1997; Sackett and DuBois, 1991; Stroh, Langlands & Simpson, 2004). Like racism and sexism, the persistence of heterosexism in the workplace ostensibly accounts for both intentional and unintentional differences in assessing the performance of Gay Managers. It is likely that these same biasing effects spill over and impede the upward mobility of Gay managers when their performance includes both effective and ineffective behaviors. That is, at least some aspects of performance evaluations may include psychometric bias and include extraneous information related to sexual orientation but not related to actual performance. Based on the arguments above, the following predictions include:

Hypothesis 1: After controlling for social desirability, target sexuality will have a main effect on global performance ratings in the average and low performance conditions, such that the Straight manager will receive higher evaluations than will the Gay manager on a global evaluation of overall performance. No difference in performance evaluations are expected in the high performance condition.

Hypothesis 2: After controlling for social desirability, target sexuality will have a main effect on the dimensionality of performance ratings in the average and low performance conditions, such that the Straight manager will receive higher evaluations on the *Analysis, Decisiveness,* and *Leadership* evaluations as well as on the *Hiring Recommendation* dimension. Conversely, the Gay manager will receive higher evaluations on the *Interpersonal Sensitivity* evaluation. There is no anticipated difference expected on the *Conflict Management* assessment. No difference in performance evaluations are expected in the high performance condition.

Gender-role

The gender-role approach provides another framework from which to conceptualize negative attitudes towards Gay men and the LGBT community. Gender is socially constructed and the associated stereotypes are deeply entrenched to the extent that men are expected to possess agentic characteristics whereas women are expected to possess communal ones (Heilman & Okimoto, 2007). These prescriptive and descriptive expectations shape the landscape regarding how men and women ought to act. Men who emphatically endorse such restricted beliefs maintain their dominance and confirm their masculinity. Women, who lack dominance, receive benefits for complying with the gender role hierarchy, and rewarded for acting in ways that reinforce a patriarchal society. For example, in a business setting, the work performance of women who were described as possessing communal traits received higher evaluations from a sample of undergraduate students than when they were described as possessing masculine traits (Heilman & Okimoto, 2007). Conversely, greater hostility is directed towards homosexuals, especially Gay men, who presumably violate these traditional gender-role norms (Davies, 2004;

Kite, Whitley, & Herek, 1998). This may help explain why men endorse traditional gender norm

beliefs to a greater degree and endorse greater anti-Gay attitudes than do women (Kite, Whitley,

Garnets, & Kimmel, 2003; Whitley, 2001). Therefore, according to the gender-role approach,

Gay men will be viewed and rated negatively, especially by Straight men, because they are

thought to possess traits stereotypically associated with the opposite gender (e.g., Gay men are

effeminate; Herek, 1984; Kite & Deaux, 1987) and therefore violate gender norms.

Hypothesis 3: After controlling for social desirability, participant gender will have a main

effect on global and dimension level performance ratings, such that, regardless of

performance level, Straight male participants will rate the Gay manager lower on all

dimensions of performance than will Straight female participants.

Hypothesis 4: After controlling for social desirability, participant's level of gender

identity centrality is expected to moderate the target sexuality – participant performance

evaluations, such that:

a. Participants with lower gender identity centrality will evaluate targets similarly

   regardless of the sexual identity manipulation.

b. Participants with higher gender identity centrality will evaluate the Gay manager

   lower than they will evaluate the Straight manager on the global *and* dimension level

   performance ratings in the average and low performance conditions only.

c. No difference in performance evaluations are expected in the high performance

   condition.

Rating Accuracy

Rating accuracy is a fundamental indicator of performance evaluation effectiveness. Guion (1965) defines accuracy as a measurement used to describe both the strength and kind of relation between a set of measures and a corresponding set of measures (e.g., true scores) considered to be an accepted standard of comparison. Although there are no normative standards for acceptable and unacceptable levels of accuracy, there is an absolute standard of zero inaccuracy representing the highest level of performance (Cardy & Dobbins, 1994). Expert raters typically evaluate the target and compute pooled averaged scores across dimensions resulting in a true score rating (Sulsky & Balzer, 1988). There are varieties of methods researchers can employ when assessing rating accuracy, however Cronbach's Accuracy Component Scores are widely accepted and utilized in the performance appraisal literature (Sulsky & Balzer, 1988).

As has been discussed in previous sections of this paper, the effects of gender have resulted in a general pro-male bias. Of importance, is the extent to which evaluations of women depart from accuracy ratings. Research has demonstrated that ratings of (ostensibly White) women are influenced by the purpose of the appraisal and cultural stereotypes of women. Specifically, when making promotion decisions, raters who held traditional stereotypes of women evaluated female targets less accurately by deflating their ratings (Dobbins, Cardy, & Truxillo, 1988). This pattern of results was also established in an applied academic setting, researchers found that (ostensibly White) female professors were rated as less effective and were rated lower than their true level of performance when compared to (ostensibly White) male professors (Dobbins, Cardy, & Truxillo, 1986). Other studies have found liking to influence ratings such that participants inflated evaluations for those who were well liked, which distorted accuracy (Cardy & Dobbins, 1986). Thus, it is predicted that:

Hypothesis 5: The Gay manager will be evaluated less accurately than will the Straight manager in the low and average performing conditions. There are no expected differences in accuracy expected in the high performing condition.

CHAPTER 3

Method

Participants

Sample 1: Participants in this study included 482 undergraduate students recruited from an introductory psychology course. Only 437 (60% female; female participation number is representative of the institution) of those who completed the survey were included as 45 students (9% of the sample) missed at least one of the manipulation checks (see Table 1). The mean age of the sample was 19.55 years ($SD = 1.97$) and there were 87 (20%) students of color. All participating students received partial fulfillment of course experiment participation requirements.

Sample 2: In addition to obtaining ratings from a student sample, it is important to obtain ratings from professionals in the applied setting. Specific criticisms have been made questioning the usefulness of studies that include only undergraduate students, who have been described as less invested and not trained in rating performance (Landy, 2008). To increase the generalizability of the study findings, a sample of 288 managers and educators working at least part time in a variety of sectors across the world participated in the study. The mean age of the field sample was 42.47 years ($SD = 10.84$) and there were 55 (22.1%) managers of color. In the field sample, only 248 (55% female) of those who completed the survey were included as 40 participants (14% of the sample) missed at least one of the manipulation checks. The increased number of managers missing at least one of the manipulation checks include a large number of managers ($N = 16$), who were removed because they indicated that they knew the manager appearing in the video personally (see Table 1 for full manipulation check and condition details for both the lab and field samples). Managers and educators were recruited via a snowball

strategy to complete the online assessment. Additional demographic data for the lab and field samples can be found in Tables 2, 3 and 4.

Design

The proposed study was conducted using a 2 x 3 between subject factorial design, with sexuality of target manager (Gay vs. Straight) and performance level (low vs. average vs. high) as the independent variables. Participants were randomly assigned to one of the six levels of the between-subjects manipulations. The survey measured each participant's gender identity centrality and gender. The response variables of interest were a global and dimension level (i.e., *analysis, decisiveness, leadership, conflict management* and *interpersonal sensitivity*; all measured on a 5-point Likert scale from 1-very ineffective to 5-very effective) of performance.

Procedure

The procedure and stimulus materials were identical for the student and professional samples. The experimenter informed participants that the study concerned evaluating the performance of a potential new manager. Similar to the procedures in Heilman and Okimoto (2007), participants were told that they would evaluate a potential new manager applying for the position of regional manager of a sales and financial department within a consulting firm. Next, they were provided information about the new position, about the new manager, and then they watched a 10-minute video of the stimulus target[4]. Next participants were asked to complete a personality measure, which was included to increase forgetting and ensure the participants relied on their memory of the target manager. Next participants were asked to evaluate the performance of the target manager and provide their level of gender identity centrality on an online questionnaire. After the various questionnaires were completed, a written debriefing was provided revealing the purpose of the experiment and the manipulations used in the study.

Stimulus Materials

Similar to the materials used by Heilman and Okimoto (2007), the online stimulus survey began with a memorandum from the Chief Executive Officer (CEO) of a consulting firm to the company employees introducing an applicant who is applying for the position of regional manager, Mr. Phil Johnson. The memo described that the research participant was instrumental in the hiring process and asked the participant to rate a video that was originally taped during a hypothetical "role-play" scenario posed to Mr. Johnson during a formal interview. Analogous to Heilman and Okimoto (2007), the information contained in the memo was overwhelmingly male to convey the sex type of the industry (the memo indicated that 88% of these managers company-wide were male). Next, participants viewed an employee information form indicating either that Mr. Johnson has a domestic partner in the Gay condition or that he has a fiancée in the Straight condition. Participants were next presented with an interview schedule that described Mr. Johnson's leadership experience, reflecting either a Gay (Director of the Atlanta Gay men's chorus and Co-Chair of the Atlanta Gay Pride parade) or Straight (Director of an Atlanta Jazz Band and Co-Chair of the Atlanta Saint Patrick's Day parade) identity[5].

Performance Video

The performance stimulus consisted of a videotaped episode developed by Gorman and Rentsch (2008). The video depicts a "role-play" exercise in which an assessment center candidate takes on the role of a manager (portrayed as an applicant for the current study) and interacts with a direct report/subordinate (interviewer for the current study) while making an administrative decision (i.e., whether or not to fire a poor performing sales representative). Gorman and Rentsch (2008) originally designed the exercise to elicit behaviors from the candidate that could be grouped into several performance dimensions (high vs. average vs. low).

Due to a potential confound in the original video, that is, the target manager wore a visible

wedding band (although increasing, currently only five states recognize same-sex marriage),

which typically signifies a Straight marriage. Therefore, the videos were transcribed and re-shot.

The performance level was operationalized by the critical incidents presented in each of

the videos. For example, in the high performance condition, the target manager employed greater

analysis strategies and discovered that the low performing employee was actually going through

a difficult divorce. The manager further suggests that the divorce might be driving his poor

performance and decides to retain the employee. In the low and average performance videos, the

target manager employs lower analysis strategies, and does not discover the divorce scenario,

ultimately deciding to fire the low performing employee. Further, accuracy ratings provided by

subject matter experts (SME's) were used to distinguish between the three levels of performance.

Specifically, four SME's trained in performance assessment rated each of the three different 10-

minute videos. Inter-rater agreement was assessed according to the equations set forth in Lindell,

Brandt, and Whitney (1999) for a single target on a multiple item scale. The rater agreement

within each performance condition ($r_{wg}=.99$) demonstrated that there was excellent agreement

amongst the SME's.

Dependent Measures

Performance evaluations were assessed using a 5-dimension scale, all measured on a 5-

point Likert scale (1 *very ineffective* – 5 *very effective*). The scales were originally developed for

360-degree rating purposes in organizations and frame of reference training (FOR) (Gorman &

Rentsch, 2008). Interviews with three subject matter experts (SMEs), that were co-owners of a

consulting firm specializing in managerial assessment centers, led to the development of the

scales. The scales have been used effectively in examining the relationship between performance

standards (i.e., FOR training or no FOR training) and the corresponding level of accuracy in ratings (Gorman & Rentsch, 2008).

Performance Dimensions. Five dimensions of performance were measured. The first six items measured the target manager's level of (a) *analysis* ($\alpha = .89$; e.g., "In the video, Mr. Johnson recognized priorities among issues, material and data."). The next six items measured the target manager's level of (b) *decisiveness* ($\alpha = .88$; e.g., In the video, Mr. Johnson committed to a clear course of action). Nine items measured the target managers' (c) *leadership* ability ($\alpha = .91$; e.g., "In the video, Mr. Johnson attempted to motivate the District Manager."). Six items measured the target managers (d) *conflict management* ability ($\alpha = .93$; e.g., "In the video, Mr. Johnson defended himself when challenged by the District Manager."). The final six items measured the target managers' (e) *interpersonal sensitivity* ($\alpha = .85$; e.g., "In the video, Mr. Johnson tried to establish rapport with the District Manager (make small talk.)").

General Performance. A *general performance* composite was constructed by combining all performance items onto one composite ($\alpha = .96$). See appendix H for the full scale.

Hiring Recommendation. Hiring recommendations referred to the probability of hiring the applicant for the job. Four items were averaged to comprise the scale ($\alpha = .96$; e.g., "If I were in charge of filling the Region Manager position, I would select Mr. Johnson). See appendix I for the full scale.

Accuracy Rating. Analogous to the methods and descriptions provided by Gorman and Rentsch (2008), rating accuracy was assessed via Cronbach's (1955) four indexes of rating accuracy: (a) elevation (E), (b) differential elevation (DE), (c) differential accuracy (DA), and (d) stereotype accuracy (SA). Each index reflects a different portion of the distance between participants' ratings and the true scores provided by the SMEs. The specific formulas provided

by Sulsky and Balzer (1988) guided the accuracy analyses in the current study. Developed using an analysis of variance (ANOVA) framework, elevation represents the differential grand mean, differential elevation represents the differential main effect of ratees, stereotype accuracy refers to the differential main effect of dimensions, and differential accuracy refers to the differential Ratee x Dimension interaction (Sulsky & Balzer, 1988). Lower scores on these measures represent higher accuracy, whereas higher scores indicate lower levels of accuracy.

Borman's (1977) differential accuracy (BDA) was also assessed analogous to the procedures described by Gorman and Rentsch (2008). Borman's differential accuracy measures the correlation between ratings on each dimension and the corresponding target scores across ratees. Higher scores on the index reflect better rating accuracy. Borman's differential accuracy is closer conceptually to rating validity because it provides correlation information and is thus insensitive to distances between ratings and true scores (Sulsky & Day, 1994).

Independent Measure

Gender Identity Centrality. Cameron (2001) and Cameron and Lalonde (2001) developed a four-item scale to measure *Gender Identity Centrality* ($\alpha = .78$; e.g., "I often think about the fact that I am a (wo)man."). The four items were averaged to create the Gender Identity Centrality composite. All items were measured on a 5-point Likert scale (1 *strongly disagree* – 5 *strongly agree*). See appendix J for the full scale.

Control Measure

Social Desirability. Crowne and Marlowe's (1960) measure of social desirability has been used in psychological research for almost fifty years to control for the effects of biased responses. Reliability and validity analyses conducted by Fisher and Fick (1993) indicated that Straham and Gerbasi's (1972) 10-item short form is the best scale of choice (p. 423). Therefore,

it was used in the current study (α = .60; e.g., "I am always willing to admit a mistake. Note. Responding with "true" on this item indicates that the participant is answering in a socially desirable manner). A composite score was created by averaging the 10 items. Higher scores indicate greater levels of social desirability. See appendix K for the full scale.

Distracter Items

      Personality Measure. Costa and McCrae's (1992) forty-item NEO-PI personality inventory was included as a distracter task to increase the time between watching the stimulus performance video and making performance ratings. See appendix L for the full scale.

Demographic Information

      Lab participants were asked to provide demographic information as it pertains to their race/ethnicity, gender, sexual identity and parent's level of education. In addition to that, field participants were asked to respond to a series of questions related to their type of employment, work environment, job experience, and organizational job and position.

CHAPTER 4

RESULTS

The purpose of the study was to explore the impact of target (ratee) sexual identity on assessments of performance. This study utilized hierarchical multiple regressions to test for the main effects of sexuality and participant gender on both the global and dimension level, performance assessments while holding social desirability constant (Hypotheses 1, 2, and 3). Additionally, hierarchical multiple regressions were employed to test the moderating effect (James & Brett, 1984) of participant's level of gender identity on the sexuality – performance assessment relationship (Hypothesis 4). Finally, an analysis of variance (ANOVA) and correlational framework was adopted to assess the varying levels of rating accuracy (Sulsky & Balzer, 1988) in each of the experimental conditions (Hypothesis 5).

Preliminary Statistical Analysis

Descriptive analyses were employed and each variable was examined for adequate distributions. One of the interpersonal sensitivity items, "In the video, Mr. Johnson used humor" represented an unusually high level of skew (1.08) and kurtosis (1.36). In addition, one of the study's SME's recommended removing the item, primarily due to the lack of humor in all three of the stimulus videos. The item was removed from further analyses.

Performance Dimension Analysis

Confirmatory factor analyses (CFA) were conducted using LISREL 8.72 (Jöreskog & Sörbom, 1993) for the current data set to evaluate various conceptualizations of the latent structure of performance. Model fit was assessed by comparing multiple fit indices (Hu & Bentler, 1998, 1999; Hu, Bentler, & Hoyle, 1995; Vandenberg & Lance, 2000). The χ^2 goodness-of-fit test, the standardized root mean squared residual (SRMSR), root mean square error of

approximation (RMSEA), the Tucker-Lewis index (TLI), and the comparative fit index (CFI).
Hu and Bentler (1998; 1999) suggest that good model fit is indicated by a $\chi2 > .05$, SRMSR <
.08, RMSEA < .06, and TLI and CFI < .95.

The variance-covariance matrix among the performance subscales served as the input.
Previous empirical support for the taxonomy of differing performance structures was evaluated.
The first model, *general performance* (Campbell, Gasser, & Oswald, 1996; Guilford, 1954)
specifies a general factor that accounts for the shared variance among the measures of
performance. Therefore, a 1-Factor model was tested by loading all 32 performance items onto
one general factor, and it did not provide an adequate fit with the data (1 Factor; $\chi^2 = 4415.18$,
p<.01; RMSEA = .170; SRMSR = .093; TLI = .94; CFI = .95).

Second, a two factor structure of performance, which includes a task and interpersonal
factor was tested. The 2-Factor model was specified by loading the *analysis* and *decisiveness*
items onto the task factor, and by loading the *leadership, conflict management,* and *interpersonal
sensitivity* items to the interpersonal factor. The 2-Factor model did not fit the data well (2
Factor; $\chi^2 = 3947.68$, *p*<.01; RMSEA = .149; SRMSR = .097; TLI = .95; CFI = .95).

Next, guided by Borman and Brush's (1993) taxonomy of performance, a 3-Factor model
was specified based on three broad dimensions: (a) technical activities/mechanics of
management, (b) communication and interpersonal facilitation, and (c) leadership and
supervision. The 3-Factor model was tested by loading the *analysis* and *decisiveness* items onto
the task factor, the *leadership* items were loaded onto a leadership factor, and *conflict
management,* and *interpersonal sensitivity* was loaded onto a third interpersonal factor. The 3-
Factor model also did not fit the data (3 Factor; $\chi^2 = 3681.39$, *p*<.01; RMSEA = .139; SRMSR =
.110; TLI = .95; CFI = .96).

The next model, also guided by Borman and Brush's (1993) taxonomy, included a fourth dimension of performance: (d) useful personal behaviors and skills. Of the five dimensions of performance measured, *conflict management* was determined to fit the description provided by Borman and Brush (1993). Therefore, the 4-Factor model separated the interpersonal factor into a sensitivity (made up of the *interpersonal sensitivity* items) and a conflict management (made up of the *conflict management* items) factor. The 4-Factor model reasonably fit the data (4 Factor; $\chi^2 = 2342.76$, $p<.01$; RMSEA = .093; SRMSR = .069; TLI = .97; CFI = .97).

Next, based on the 5-dimension scale that the items were originally written to measure (Gorman & Rentsch, 2008), a 5-Factor model was tested and based on predetermined cutoff indices (Hu & Bentler, 1998; Hu & Bentler, 1999), it fit the data best (5 Factor; $\chi^2 = 2036.45$, $p<.01$; RMSEA = .086; SRMSR = .068; TLI = .97; CFI = .98). A final 6 Factor model was tested by placing the phi matrix to load onto one general factor and it, too, did not fit the data well (6 Factor; $\chi^2 = 2207.45$, $p<.01$; RMSEA = .094; SRMSR = .075; TLI = .97; CFI = .98). Therefore, the 5-factor measurement model fit was selected as a good fit to the data according to the cutoff criteria proposed by Hu and Bentler (1998; 1999). The results of the analyses are presented in Figure 2 and Table 5.

Level of Performance Manipulation Check

A one-way ANOVA verified whether the three levels of the target managers' performance were statistically distinguishable in the lab and field samples. Results demonstrated that student' ratings for the *general performance* composite indicated that they appropriately evaluated the target mangers' performance as either high ($X = 3.62$, $SD = .53$), average ($X = 2.88$, $SD = .57$), or low ($X = 2.35$, $SD = .76$), and each of the three means were significantly different from one another, $F(2, 436) = 192.79$, $p = .001$. In addition, managers' ratings for the *general*

performance composite indicated that they too, appropriately evaluated the target managers'

performance as either high (X = 3.44, SD = .57), average (X = 2.57, SD = .60), or low (X = 2.14,

SD = .48), and each of the three means were significantly different from one another, $F(2, 245)$ =

128.01, p = .001. Thus, the target manager's level of performance was effectively manipulated.

CFA of Measurement Equivalence/Invariance

   CFA's were conducted using LISREL 8.72 (Jöreskog & Sörbom, 1993) to determine

whether managers and students conceptualized the construct of performance differently across

sexuality conditions. The chi-square goodness-of-fit test is traditionally used for assessing model

goodness-of-fit. However, because the chi-square value is widely criticized for its sensitivity

both to minor differences between groups' factor patterns and sample size (Bentler & Bonett,

1980), it is supplemented with other fit indices in order to infer model goodness-of-fit (e.g.,

SRMR, RMSEA, TLI, etc.). The difference chi-square $(\Delta\chi_2)$ and difference CFI (ΔCFI) were

used as additional tests for ME/I. A nonsignificant difference chi-square as well as a change in

value of CFI less than or equal to –0.002 indicate that model fit is not worse with the additional

constraints, thus the null hypothesis of invariance should be retained (Meade, Johnson, &

Braddy, 2008).

   Fit indices for each model tested are presented in Tables 7 and 8. The omnibus test for the

equality of the variance-covariance matrices in both the lab and field data was conducted for

both groups (Gay vs. Straight). Failure to reject the null hypothesis indicates that overall

measurement equivalence exists and further tests are unnecessary, according to Vandenberg and

Lance (2000). Results indicated a good fit; however, tests for configural, metric and scalar

invariance were conducted to ensure full measurement equivalence across the sexuality

experimental conditions. The results of the field and lab sample demonstrated full metric

invariance primarily because the change in CFI failed to reach the recommended .002 cutoff (Meade, et al., 2008). Thus, the results of the CFA's supported full metric invariance in the way participants conceptualized performance in the lab and field samples regardless of the sexuality experimental manipulation. Therefore, it is appropriate to conduct substantive tests examining mean level differences primarily because comparisons between groups exhibiting full metric invariance may be tantamount to comparing "apples to apples" as opposed to comparing "apples to sparkplugs" in the absence of invariance (Vandenberg & Lance, 2000, p. 9).

Test of the Hypotheses

Prior to the substantive analyses of the hypotheses, steps were taken to examine the possible violations of the key assumptions for the proper use of regression analysis. Correlations among the predictor variables were assessed to ensure multicollinearity was not present (e.g., $r >$ .80). Investigation of the correlations among the independent variables did not reveal any evidence of multicollinearity (observed correlations ranged from .19 - .02). Cook's D (1977; 1979) is based on both "outlierness" (i.e. SRESID) and leverage characteristics of the observation. Values greater than 1 are considered problematic (Cook & Weisberg, 1980). Observed Cook's D were relatively low (ranging from .000 to .020), indicating the absence of influential observations. Table 6 illustrates the means, standard deviations, and correlations of the variables analyzed in the current study.

Mean Ratings across Performance Dimensions

A dimension rating (i.e., *general performance, analysis, decisiveness, leadership, conflict management, interpersonal sensitivity,* and *hiring recommendation*) by performance level (high vs. average vs. low) and sexual identity (Gay vs. Straight) mean rating examination was conducted in both the lab and field samples. Overall, student participants rated the Gay manager

slightly higher than they rated the Straight manager on fifteen of the twenty-one dimension level assessments (71%). Field managers, likewise, rated the Gay manager with greater positivity, rating him slightly more favorably on nineteen of the twenty-one dimension level assessments (90%). These results provide initial support against the predictions that the Gay manager will receive lower ratings on most of the performance dimensions. The following sections will test whether the mean ratings differed statistically.

Hypothesis 1: Global Performance Ratings

Hypothesis 1 predicted that the target manager's sexual identity would have a main effect on global performance evaluations, after controlling for the effects of social desirability. Specifically, Gay managers were expected to receive lower evaluations in the average and low performance conditions, with no difference expected in the high performance condition. To test this prediction, 6 hierarchical multiple regressions were conducted (one regression for each of the 3 levels of performance in both the lab and field samples). In step 1, the global performance composite was regressed onto the social desirability composite as a control variable. In step 2, manager sexuality was entered into the model and its standardized beta was examined in order to assess if it had a significant effect.

Three regressions were conducted for the lab sample and results demonstrated that none of the student generated mean ratings differed significantly. Three regressions were conducted for the field sample and only one of the resulting standardized coefficients for sexuality was significant. Professionals in the average performance condition evaluated the Gay manager (X = 2.73, SD = .72) significantly higher than they rated the Straight manager (X = 2.43, SD = .42; β = -.261, $t(63)$ = -2.17, p = .03) on the *general performance* dimension. The observed pattern of

results in the field sample was in the opposite direction from the prediction, providing partial discounting evidence for hypothesis 1 (see Table 9 & Figure 3).

Hypothesis 2: Dimension Level Performance Ratings

Hypothesis 2 predicted a pattern of main effects in the low and average performance conditions only. It was predicted that the Straight manager would receive higher evaluations on specific dimensions of performance, including *analysis, decisiveness,* and *leadership;* it was also predicted that he would receive higher ratings on the *hiring recommendation* measure. Whereas the Gay manager was expected to receive higher evaluations on the *interpersonal sensitivity* dimension of performance and no differences were expected on the *conflict management* dimension. Analogous to the steps employed to test hypothesis 1, the same steps were taken to explore hypothesis 2, with the exception that the outcome variable (e.g., *analysis, decisiveness, leadership,* etc.) changed for each hierarchical multiple regression analysis. Specifically, in step 1, each outcome variable (e.g., *analysis, decisiveness,* etc.) was regressed onto the social desirability composite as a control variable. In step 2, manager sexuality was entered into the model and its standardized beta was examined.

Analysis

Three regressions were conducted for the lab sample and results demonstrated that student participants assigned statistically more favorable ratings to the Gay manager ($X = 3.97$, $SD = .63$) than to the Straight manager ($X = 3.76$, $SD = .61$; $\beta = -.157$, $t(154) = -1.99$, $p = .05$) in the high performance condition only. Three further regressions were conducted for the field sample and results demonstrated that none of the means differed significantly. The observed pattern of results for the lab sample was in the opposite direction from the prediction, providing

initial evidence to discount hypothesis 2 on the *analysis* dimension of performance (see Table 10 & Figure 4).

Decisiveness

Three regressions were conducted for the lab sample and results demonstrated that participants assigned statistically more favorable ratings to the Gay manager (X = 3.44, SD = .66) than to the Straight manager (X = 3.16, SD = .68; β = -.195, $t(154)$ = -2.48, p = .01) in the high performance condition only. Three regressions were conducted for the field sample and results demonstrated that none of the means differed significantly. Taken together, results provided initial evidence to discount hypothesis 2 on the *decisiveness* dimension of performance (see Table 11 & Figure 5).

Leadership

Three regressions were conducted for the lab sample and results demonstrated that none of the means differed significantly. Three further regressions were conducted for the field sample and results demonstrated that managers assigned statistically more favorable ratings to the Gay manager (X = 2.64, SD = .81) than to the Straight manager (X = 2.29, SD = .68; β = -.261, $t(63)$ = -2.17, p = .03) in the average performing condition only. Again, the results provide evidence to discount hypothesis 2 on the *leadership* dimension of performance (see Table 12 & Figure 6).

Conflict Management

It was predicted that there would be no difference between the Gay and Straight manager in all three performance conditions. Three regressions were conducted for the lab sample and results demonstrated that none of the means differed significantly on the dimension of *conflict management*. Three further regressions were conducted for the field sample and the results demonstrated that managers assigned statistically more favorable ratings to the Gay manager (X

= 2.62, *SD* = .82) than to the Straight manager (*X* = 2.15, *SD* = .64; (β = -.318, *t*(63) = -2.70, *p* = .01) in the average performance condition. The results provide evidence to discount hypothesis 2 on the *conflict management* dimension of performance (see Table 13 & Figure 7).

Interpersonal Sensitivity

On the construct of *interpersonal sensitivity,* it was predicted that the Gay manager would be rated significantly higher than would the Straight manager in the average and low performance conditions. Three regressions were conducted for the lab sample and results demonstrated no significant differences. Three further regressions were conducted for the field sample and results demonstrated no significant differences. Thus, hypothesis 2 was not supported on the *interpersonal sensitivity* dimension (see Table 14 & Figure 8).

Hiring Recommendation

Three regressions were conducted for the lab sample and the results demonstrated that none of the means was significantly different. Three regressions were conducted for the field sample and results demonstrated that managers assigned statistically more favorable ratings to the Gay manager (*X* = 2.80, *SD* = 1.26) than to the Straight manager (*X* = 2.22, *SD* = .98; (β = -.259, *t*(63) = -2.14, *p* = .04) in the average performing condition only. Again, this was in the opposite direction than what was predicted. Both samples provided evidence to discount hypothesis 2 on the *hiring recommendation* dimension of performance (see Table 15 & Figure 9).

Hypothesis 3

Hypothesis 3 predicted that female participants would evaluate the Gay manager significantly higher than would their male counterparts (see Figures 10 – 15 for mean level ratings across gender, sample, and performance dimensions). In order to test for the main effect

of gender on the performance ratings of the Gay manager, the ensuing hierarchical multiple regressions followed the same two-step process conducted for hypotheses 1 and 2; however, only the data from the Gay manager conditions were analyzed. In step 1, the outcome measure (e.g., *general performance, analysis,* or *decisiveness*, etc.) was regressed onto the social desirability composite as a control variable. In step two, participant gender was entered into the model and the resulting standardized beta weight for gender was examined in order to assess if it had a significant effect.

Lab Sample

The mean level performance ratings were examined in the lab sample. Overall, of the twenty-one total dimensions of performance, female participants rated the Gay manager higher than did males on only nine dimensions (43%). When examining for significant differences between the mean level ratings, only one statistically significant difference was found. In the average performing condition, unexpectedly after controlling for social desirability, male students rated the Gay manager on *decisiveness* ($X = 2.95$, $SD = .81$) significantly higher than did female participants ($X = 2.58$, $SD = .78$; $\beta = .221$, $t(86) = 2.13$, $p = .04$). On average, male student participants provided higher mean level ratings for the Gay manager than did female student participants, which was in the opposite direction from the prediction.

Field Sample

The mean level performance ratings were examined in the field sample, as well. Overall, female managers rated the Gay manager higher than did male managers on fifteen of the twenty-one total dimension level assessments (71%). When examining the mean level ratings, two statistically significant differences were found. In the average performance condition, after controlling for social desirability, female managers rated the Gay manager on *leadership* ($X =$

2.93, $SD = .83$) significantly higher than did male managers ($X = 2.31$, $SD = .67$; $\beta = -.352$, $t(28)$ = 2.03, $p = .05$). Also in the average performance condition, female managers rated the Gay manager on *conflict management* ($X = 2.91$, $SD = .79$) significantly higher than did male managers ($X = 2.29$, $SD = .76$; $\beta = -.380$, $t(28) = 2.16$, $p = .04$). The obtained results provide partial support for hypothesis 3. In general, female managers rated the Gay manager higher on two of the constructs of performance as compared to male managers.

Hypothesis 4

Hypothesis 4 predicted that participant's level of gender identity would moderate the sexuality – performance assessment relationship. To test this relationship, a set of hierarchical multiple regressions were conducted according to the procedure outlined by James and Brett (1984). In step 1, the outcome measure (e.g., *general performance, analysis,* or *decisiveness,* etc.) was regressed onto the social desirability composite as a control variable. In step two, participant gender identity and manager sexuality were entered into the model. In step 3, the joint effect term (gender identity x sexuality) was added to the model and the resulting standardized beta weight was examined in order to assess if it had a significant effect. To test for possible moderation, regression analyses were run for each of the six dimensions of performance (e.g., *general performance, analysis, decisiveness,* etc.) by each of the performance levels (i.e., high, average, low) by sample (lab vs. field). Thus, thirty-six regressions were employed and none of the interaction terms resulted in significance (see Tables 16 – 21). Therefore, the data provided no support for hypothesis 4.

Hypothesis 5

Hypothesis 5 predicted that Gay managers' ratings would be less accurate than would Straight managers' in the average and low performance conditions. The results of the rating accuracy measures across the three performance levels are provided in Tables 22 – 24. No significant differences were found in the high or low performance conditions. In line with the accuracy score predictions, managers provided more accurate ratings of the Straight manager when compared to ratings of the Gay manager, but only on the elevation index (E= $t(64) = 3.44$, $p =.01$); providing partial support for hypothesis 5. See Table 25 for a full overview of the study results by each of the five hypotheses.

Post Hoc Analyses

Parallel analyses were conducted to explore the extent to which social desirability affected the overall and dimension level performance ratings. In addition to the higher mean level ratings of the Gay manager, demonstrated in each of the conditions of performance and in each sample, several additional dimensions of performance were also found to benefit the Gay manager. Students in the high performance conditions initially rated the Gay manager significantly higher than the Straight manager on the general performance ($\beta = -.156$, $t(155) = 1.97$, $p = .05$) and hiring recommendation ($\beta = -.155$, $t(155) = 1.96$, $p = .05$) composites. The significant beta weights dropped to non-significance after controlling for social desirability (see Tables 9 & 15). In a similar fashion, students in the low performance condition initially rated the Gay manager significantly higher on interpersonal sensitivity ($\beta = -.168$, $t(131) = 1.96$, $p = .05$), but the significant beta weight dropped to non-significance after controlling for social desirability (See Table 14).

In the field sample, managers in the average performance condition initially rated the Gay manager significantly higher on decisiveness ($\beta = -.236$, $t(64) = 1.96$, $p = .05$), but the significant beta weight dropped to non-significance after controlling for social desirability (see Table 11). In a similar fashion, managers in the low performance condition initially rated the Gay manager significantly higher on conflict management ($\beta = -.195$, $t(98) = 1.98$, $p = .05$), but once again, the significant beta weight dropped to non-significance after controlling for social desirability (see Table 13). The results of the post hoc analyses indicate that participants were often providing socially desirable responses in many of their evaluations, which indicate psychosocial bias.

CHAPTER 5

DISCUSSION

The purpose of this study was to examine the extent to which stereotypes about Gay managers affect subjective performance evaluations. In an effort to expand upon prior research, the influence of gender, gender identity, hiring recommendations and rater accuracy were explored. Drawing from research on heterosexism and gender-role centrality, it was predicted that the Gay manager would receive deflated ratings compared to the Straight manager, but only when the ratings were justifiable (in the low and average performance conditions). In the absence of ambiguity (in the high performance condition), no differences in ratings were expected.

The Influence of Sexual Identity

The study results indicate that students and managers' exhibited a general pro-Gay bias, which was unexpected, providing no support for hypothesis 1 and 2. There was, however, no clear pattern of results. The Gay manager received some of the strongest evaluations when exhibiting performance that was either high (lab) or average (field), with no difference in ratings when performance was low (lab and field). At the onset of the study, specific dimension level predictions were made based on Golom and Liberman's (2008) findings that stereotypical Gay managers were rated as less effective on instrumental performance measures (e.g., *work competence, independence, activity/potency, rationality*), yet higher on relational measures (e.g., *concern for others*). In the current study, the data did not support these predictions. Students evaluated the Gay manager as being better at analyzing data and making decisions, both of which are instrumental behaviors. In the field, managers did not evaluate the Gay manager higher on any instrumental behaviors. Instead, they evaluated the Gay manager as more effective overall, viewed him as more of a leader, possessing better conflict management strategies, and they

recommended him for hire at greater levels than the Straight manager. Unexpectedly, there were no evaluative differences between the Straight and Gay manager on the relational measure of interpersonal sensitivity despite the common stereotype that Gay men possess greater levels of sensitivity (Madon 1997; & Taylor, 1983).

The results were unanticipated given the extensive body of literature documenting discrimination against the LGBT community (see Badgett, Lau, Sears, & Ho, 2007; Herek, 2008). However, a critical examination of within group differences (i.e., the Gay community) may assist with understanding the current study's outcomes. Often times, researchers artificially combine groups by observable demographic characteristics (e.g., gender, race, and sexual identity) and avoid examining within group differences. Thomas, Phillips, & Brown (1998) discuss the lack of acknowledgement from the field of industrial-organizational (I/O) psychology to the differences within groups. They challenge researchers interested in understanding the organizational realities for *all* workers, to consider the issue of identities of both minority and majority group members. Thus, when examining responses to the Gay community, a greater number of targets must be presented (e.g., an effeminate Gay White man, a Black Gay man, a Black Lesbian woman, etc.). Given the vast diversity within the Gay community, the current study is an initial step in understanding the performance evaluation realities of the LGBT community. Moreover, the study has only explored the reactions toward a White, masculine, Gay man. Given the numerous negative cultural stereotypes of Gay men as possessing effeminate characteristics, the target manager in the study likely counteracted these stereotypes and received benefits for doing so.

Ambivalent Amplification

Katz and Hass (1988) have named this phenomenon ambivalent amplification, which posits that dominant group members hold both negative and positive opinions of stigmatized groups, and will amplify positive responses in favorable situations but will also amplify negative responses in unfavorable ones. In other words, a socially stigmatized target gives off signals that can lead to either a positive or a negative response. The cues of a masculine, articulate, and successful White manager in the stimulus video may have led participants to respond by assigning favorable ratings to the Gay manager who they may have perceived as someone who could "pass" as straight. The observed ratings may not have been quite this favorable had he fallen within the negative stereotypical expectation.

Aversive Heterosexism

Another theory that partially accounts for the study findings is the theory of aversive racism. Gaertner and Dovidio's (1986) theory of aversive racism (or heterosexism) similarly emphasizes ambivalence as the underlying mechanism driving dominant group member's reactions. In other words, Straight individuals hold egalitarian beliefs and a nonprejudiced self-image. Nevertheless, society continually reinforces negative portrayals of stigmatized groups, especially Gay men. Ambivalence ensues when Straight individuals encounter a Gay person and respond in an overly nondiscriminatory manner in order to preserve their nonprejudiced self-image. Although not demonstrated in the current study, according to the theory of aversive heterosexism, it is predicted that Straight persons would behave in a discriminatory manner in more ambiguous situations when the discrimination can be rationalized.

In this study, the target Gay manager seemingly overcame the negative images of a stereotypically effeminate Gay man. Even though the Gay manager was highly committed to his

identity as evidenced by his Gay Men's Chorus and Gay Pride Parade affiliation, managers and students responded by amplifying positive responses in either the high (lab) or average (field) performance conditions. The fact that there were no penalties for the Gay manager in the low performance condition may signal that participants in the study were still responding in an overly nondiscriminatory manner, even when his low performance would have provided justification for low evaluations. Perhaps the strongest support for the ambivalent amplification and aversive heterosexism argument lies in the fact that when controlling for the effects of social desirability, many of the positive ratings provided by both students and managers dropped to non-significance. It would appear that at least some of the evaluations of the Gay manager in this study were made with the conscious intention of portraying a positive self-image.

Changing Social Norms

Another theoretically plausible explanation for the study results indicates that social norms are changing to be more inclusive of differences. Perhaps the biased responses favoring the Gay manager represent a shift in cultural thinking, which is exemplified by the mounting support for same sex marriage across the nation. In fact, over the past several weeks Iowa, Vermont and Maine have joined the list of states allowing same sex couples the right to enter into marriage. Fierce opposition from opponents however, continues to block the federal recognition of marriage through legislation such as the defense of marriage act and the federal marriage amendment, which both serve to define marriage as a union between one man and one woman. Thus, further support for this explanation is necessary by examining the relationships between the study variables in a greater number of contexts and situations. Until now, there has been little support in the literature demonstrating a pro-Gay bias. On the other hand, perhaps attitudes towards the LGBT community may remain generally negative, but openly expressing

these viewpoints continue to elicit social sanctions, which lead to a motivation to provide socially acceptable responses. In sum, the results of the study likely support a combination of the proposed theoretical explanations, warranting further examination, and emphasize the complexity of contemporary discrimination.

The Influence of Gender

Participants in the laboratory responded very differently from managers in the field. Overall, male students provided inflated evaluations of the Gay manager more often than did male managers in the field. When assessing statistical significance, the laboratory sample of male students provided higher decisiveness ratings than did female students when the Gay manager exhibited average performance. In the field study, managers responded more consistently with the study prediction that females would provide the highest ratings. Female managers evaluated the Gay manager as more of a leader and better at employing conflict management strategies when his performance was average, thus providing partial support for hypothesis 3 in the field sample only. The findings from the laboratory sample however provided support against hypothesis 3.

The inability to find consistent ratings between the field and laboratory samples might suggest generational differences, especially given that after controlling for social desirability, the male students provided some of the highest evaluations of the Gay manager. Meta analytic research has demonstrated that Straight women possess more positive attitudes towards homosexuality than do Straight men (Kite, et al., 1998). The results provide another indication that social norms are changing and the gap between the generations may be growing. To illustrate, a study examining the demographics of California voters on a 2008 ballot to ban same-

sex marriage found that 67 percent of voters 65 or older supported the ban, while majorities

under 65 opposed it (Egan & Sherrill, 2009).

The inconsistencies may also be a result of the participants' varying level of experience

in both the work setting and at evaluating performance. Furthermore, the frame of reference

invoked by the students may have been quite different from that of the managers. For instance,

student participants, many possessing limited work experience, may have viewed the target

manager as a potential boss, which led to amplifying ratings when the Gay manager exhibited

high performance. Yet, the more experienced managers may have viewed the target as a potential

colleague or direct report which led to amplifying ratings when the Gay manager exhibited

average performance. A current debated exists in the multi-source ratings (MSR) literature (i.e.,

360-degree assessments), arguing whether differences in ratings between the various rater levels

(e.g., subordinate, peer, supervisor) is due largely to either (a) psychometric bias, and should be

treated as error, or (b) substantive differences in the targets' behavior between raters, and should

be treated as substantive method variance (Lance, Hoffman, Gentry, & Baranik, 2008). To

illustrate the latter point, an employee may strive to appear intellectual to a supervisor but take a

very different (and perhaps more relaxed approach) when surrounded by a group of peers. Thus,

differences in performance ratings between a supervisor and peer may not be the result of error,

but rather differences in the target's true performance levels. It would be premature to attempt to

draw conclusions regarding the difference in assessments observed in the study at this point.

Future studies might explore the extent to which participant's frame of reference influence

performance ratings.

It was further predicted that a strong identification with a gender identity would help to

identify the type of person more likely to respond in discriminatory ways. The data provided no

support for hypothesis 4. In retrospect, the target Gay manager in the video may have counteracted stereotypes that depict Gay men as effeminate. Thus, regardless of the participant's level of gender identity centrality, observing a "Straight-acting" masculine Gay manager may have overridden any negative associations. Taken together, these results may be yet another indicator that social norms continue to change to be more inclusive of masculine Gay men.

Accuracy Ratings

The results indicated that rater accuracy was better for the Straight manager in only one condition. In other words, managers more accurately assessed the Straight manager's performance in the average performance condition according to the *elevation* index, providing partial support for hypothesis 5. In hindsight, the rating accuracy indices should have been more discrepant and sensitive to rating inaccuracy, especially given that the Gay manager was rated more favorably on 84 percent of the performance dimensions. One of the drawbacks however was the inability to assess the way raters tend to evaluate in general (i.e., the average of multiple ratees). Previous research evaluates a rater's accuracy by assessing variability across multiple ratees (Cardy & Dobbins, 1986; 1994; Dobbins, Cardy, & Truxillo, 1986; & Gorman & Rentsch, 2008). In the current study, participants watched and assessed only one target across several dimensions of performance. As a result, the rating accuracy equations were less sensitive to rating nuances and did not identify inaccuracies that may have been detected otherwise.

A finding that was not specified at the onset of the study was that the accuracy scores for the laboratory participants were generally similar to scores obtained from the field. The generalizability of student samples is often called into question (Landy, 2008). The data from the current study would suggest that student participants were invested in the study and provided relatively similar evaluations to the ones provided by experienced managers.

In sum, the data support that, (a) regardless of performance appraisal experience, evaluations provided by students were generally similar to the ones provided by managers; (b) managers evaluated the Straight manager more accurately than they evaluated the Gay manager when his performance was average, (c) both students and managers exhibited a pro-Gay bias when performance was either strong (lab) or average (field); and (d) difference in evaluations were apparent when examining gender, such that male students and female managers provided some of the highest evaluations of the Gay manager, indicating a possible generational shift in acceptance for homosexuality.

Limitations

It is important to note several limitations of this laboratory and field study. The first limitation is the nature of using an experimental design. In organizational settings, it is unreasonable for managers to conduct evaluations after observing only 10-minutes of performance. In an attempt to increase realism in the study, the target manager was presented as a job applicant engaged in a "role play" interview scenario. Therefore, the performance ratings provided might realistically follow the format of a new hire selection system of an actual organization. The extent to which the relationship between sexuality and performance evaluations changes longitudinally remains unanswered. Participants may have felt less threatened by the target manager, knowing that these ratings would ultimately change neither his life nor their own. It further remains unclear how the evaluations would have changed had this been an actual applicant applying for the position of a boss, coworker, or subordinate.

The inclusion of undergraduate students could pose potential problems that limit the generalizability of the findings. Most of the students had limited real work experience in a corporate setting in which the scenario was based. The accuracy score indices specifically

addressed this issue; however, as the student population provided relatively accurate performance evaluations. Another limitation noted is the geographical differences in students' motivation to utilize or overcome the use of rigid stereotypes, or the socio-economic differences in the sample that may not generalize to other parts of the country. Again, the inclusion of the field sample helps circumvent many of the drawbacks of using a student sample and increase its generalizability. For instance, most of the managers came from across the United States, but a significant portion of the sample (11 percent) also came from a global population, where norms and strength of Gay stereotypes might be different. Adding to its strength, the data represents a cross-section of industries, drawing largely from education (14.5 percent), consulting (12.9 percent), and technology (12 percent) related industries.

Another limitation is the level of education of the study participants, which may have led to a restriction in range in the sample. Past research demonstrates that education is highly correlated with discrimination against homosexuality, such that the higher the level of education, the more favorable is the attitudes towards homosexuality (Herek & Glunt, 1993). The managers in the sample were highly educated (93 percent held a bachelors degree or higher). The lab participants, however, may have provided a proxy for education, as many of the students are only in their first year of college training. Nevertheless, the observed pro-Gay bias may be largely due to the participant's increased level of education.

Finally, the snowball approach to the data collection in the field sample may have led to including participants differing from the norm. Many of the networks accessed primarily included a network of socially conscious and sensitive professionals. In addition, the modest number of participants included in the field sample likely reduced the statistical power necessary to detect all of the possible effects influenced by the experimental manipulation. In the

laboratory, participants' exhibited a pro-Gay bias in the high performance condition but the field managers demonstrated it in the average condition. Perhaps the inclusion of a greater number of managers may result in a pro-Gay bias in the high performance condition as well. As it stands, the observed pro-Gay bias was not necessarily a robust effect, and differed according to the sample and performance level.

Future Research

An obvious extension to the current study is the exploration of reactions to a Gay manager who exhibits effeminate characteristics. Much of the past research has examined responses to a generic target, usually presented in a vignette, described simply as "a Gay man." As a result, participants likely relied on their stereotypes and conceptualizations of Gay men or Lesbian women and provided deflated ratings (e.g., Golom & Liberman, 2008). Given the mounting research demonstrating heterosexism in society (Herek, 2007) as well as organizational settings (Ragins & Martocchio, 2004), it is expected that responses to an effeminate Gay man would follow the pattern of hypotheses proposed in the current study.

Future studies should also include greater racial, ethnic, and gender diversity amongst the target managers. The LGBT community is vastly diverse and the experiences of minorities who exhibit a double (or triple) minority status (Black, Gay, and Female) tend to report greater levels of discrimination (Berdahl & Moore, 2006). Studies such as these present specific challenges, especially when accounting for the effects of negative (or positive) responses resulting from one of several potentially stigmatizing identities.

The effects of sexuality should also be explored in a greater number of contexts and real world settings. This study examined sexuality using a laboratory experimental manipulation. A new research paradigm might strive to compare Gay and Lesbian managers' performance

evaluations to similarly situated Straight counterparts in actual organizations. Methodologically, this introduces unique challenges primarily because a number of sexual minorities avoid disclosing their sexual identity out of the fear of retaliation (Ragins & Martocchio, 2004). In addition, sexuality is a socially constructed, and as such difficult to measure. For instance, a person may self identify as heterosexual even though he or she engages in homosexual behaviors. Despite these real-world challenges, researchers should continue to strive to examine the sexuality – performance evaluation link directly in corporate settings.

The use of MSR's, especially for management, are quickly replacing single rater performance evaluations, which include the ratings of individuals at all levels of the organization (i.e., coworker, direct reports, and supervisors). Future studies should examine the relationship between sexuality and multi-source ratings. In the current study, participants were told to rate the target manager as a peer, direct report, or supervisor. Student participants may have been more likely to evaluate the target manager as a potential supervisor, but in the field, managers may have viewed him as a potential peer or direct report. The way that participants viewed the target manager may have affected their performance evaluations. Thus, the current results might guide initial MSR predictions regarding the impact of performance level, or position of the rater, may guide ratings.

Research objectives may examine the boundary conditions beyond the proposed study that likely elicit anti-Gay biases in evaluating employee performance. Social norms continually change to support a more inclusive, demographically diverse workforce but it is often context specific. For instance, the context might affect the ratings of Gay men and Lesbian women depending on whether one works in a feminine versus masculine dominated industry. To that

end, a Gay manager may have very different experiences in a transportation industry (i.e., negative) than he would experience in a fashion industry (i.e., positive).

Implications

Despite the limitations of this study, the findings reveal several implications for organizations as well as Gay managers. The results indicate an initial pro-Gay bias in performance evaluations. This could be attributed to one, or to a combination of several explanations: (a) attitudes towards homosexuality are changing and improving, (b) attitudes toward masculine Gay men are generally favorable but there is a potential offsetting discriminatory response toward effeminate Gay men, or (c) attitudes toward the LGBT community generally remain negative, but acting on negative attitudes is becoming less socially acceptable, leading to a motivation to sound politically correct.

Stereotypes may have originated to help speed up mental processing, but when applied to individuals, they can limit individual lives. Moreover, regardless of valence, stereotypes, even when positive, serve to restrict the identity of its targets by defining acceptable standards of behavior (Jones, 2002). When looking at favoritism, McIntosh and Plous (2003) warns that privilege (unearned advantage) afforded to one group (e.g., Whites, males), usually comes at the expense of other groups (e.g., ethnic minorities, women). Although the results of the current study may be initially good news, research demonstrates that individuals holding positive stereotypes of stigmatized groups also have been found to endorse blatant negative stereotypes. To illustrate, students possessing benign Jewish stereotypes were also found to endorse anti-Semitic stereotypes (Wilson, 1996). In another study, students describing Asian students as being highly competent were also less willing to roommate with an Asian American student (Lin, Kwan, Cheung, & Fiske, 2005). In yet another study, Walls (2008) demonstrated student's

endorsement of positive stereotypes of Gays and Lesbians were associated with increased levels of social dominance orientation and decreased support for Gay and Lesbian rights. Thus, an increasing number of scholars are beginning to examine the role of positive stereotypes in the maintenance of systems of stratification (Czopp, 2008).

It is important to note the implications for organizations. Performance evaluations continue to serve as the criteria for many important administrative decisions that can advance and limit talent when biased. The outcome of the current study would suggest that talented Straight White managers initially might be at a disadvantage as compared to masculine Gay White managers. What's more, inflated performance evaluations could be detrimental to the career of a Gay manager, limiting his access to the same developmental opportunities necessary for improving performance his Straight counterpart likely receives. Thus, the challenge for organizations is to create performance evaluation systems that assist with the observation and recall of specific job performance behaviors, offer incentives for ratings that accurately reflect the evaluation of each subordinate's individual performance, and protect supervisors from the negative consequences of providing honest ratings (Murphy, 2008). Otherwise, biased evaluations could potentially result in a loss of talent, greater levels of attrition, and disengagement amongst managers whose efforts are not fully valued, Gay or Straight. This finding reiterates that discrimination is complex and subtle preferences can still be detrimental to a person's career. In an ideal world, the evaluations of a Gay and Straight manager exhibiting equal performance would not differ in any appreciable way.

REFERENCES

Badgett, L., Lau, H., Sears, B., & Ho, D. (2007). Bias in the workplace: Consistent evidence of sexual orientation and gender identity discrimination. *The University of California School of Law: The Williams Institute*, 1-31.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Borman, W. C. (1977). Consistency of rating accuracy and rating error in the judgment of human performance. *Organizational Behavior and Human Performance, 20*, 238-252.

Borman, W. C., & Brush, D. H. (1993). More Progress Toward a Taxonomy of Managerial Performance Requirements. *Human Performance, 6*(1), 1.

Borman, W. C., Dunnette, M. D., & Hough, L. M. (1991). Job behavior, performance, and effectiveness *Handbook of industrial and organizational psychology, Vol. 2 (2nd ed.).* (pp. 271-326). Palo Alto, CA, US: Consulting Psychologists Press.

Button, S. B. (2001). Organizational efforts to affirm sexual diversity: A cross-level examination. *Journal of Applied Psychology, 86*(1), 17-28.

Cameron, J. E. (2001). Social identity, modern sexism, and perceptions of personal and group discrimination by women and men. *Sex Roles, 45*(11), 743-766.

Cameron, J. E., & Lalonde, R. N. (2001). Social identification and gender-related ideology in women and men. *British Journal of Social Psychology, 40*(1), 59-77.

Campbell, J. P., Gasser, M., & , & Oswald, F. (1996). *The substantive nature of job performance variability.*: Hillsdale, NJ: Erlbaum.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology, 71*(4), 672-678.

Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal : alternative perspectives.* Cincinnati: South-Western Pub. Co.

Cohen, C. E. (1981). Person categories and social perception: Testing some boundaries of the processing effect of prior knowledge. *Journal of Personality and Social Psychology, 40*(3), 441-452.

Cook, & Weisberg, S. (1980). Characterizations of an emperical influence function for detecting influential cases. *Technometrics, 22*, 495-508.

Cook, R. D. (1977). Detection of influencial observations in linear regression. *Technometrics, 19*, 15-18.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Association, 70*, 428-434.

Costa, P. T., Jr.,, & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Resources, Inc.

Cronbach, L. (1955). Processes affecting scores on 'understanding of others' and 'assumed similarity.'. *Psychological Bulletin, 52*(3), 177-193.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349-354.

Czopp, A. M. (2008). When is a compliment not a compliment? Evaluating expressions of positive stereotypes. *Journal of Experimental Social Psychology, 44*(2), 413-420.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*(1), 20-33.

Davies, M. (2004). Correlates of Negative Attitudes Toward Gay Men: Sexism, Male Role
Norms, and Male Sexuality. *Journal of Sex Research, 41*(3), 259-266.

Dickinson, T. L. (1987). Designs for evaluating the validity and accuracy of performance ratings.
Organizational Behavior and Human Decision Processes, 40(1), 1-21.

Dipboye, R. L. (Writer) (1985). Some Neglected Variables in Research on Discrimination in
Appraisals [Article], *Academy of Management Review*: Academy of Management.

Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. (1988). The effects of purpose of appraisal and
individual differences in stereotypes of women on sex differences in performance ratings:
A laboratory and field study. *Journal of Applied Psychology, 73*(3), 551-558.

Dovidio, J. F., Gaertner, S. L., Eberhardt, J. L., & Fiske, S. T. (1998). On the nature of
contemporary prejudice: The causes, consequences, and challenges of aversive racism
Confronting racism: The problem and the response. (pp. 3-32). Thousand Oaks, CA US:
Sage Publications, Inc.

Dovidio, J. F., Gaertner, S. L., & Willis-Esqueda, C. (2008). New directions in aversive racism
research: Persistence and pervasiveness *Motivational aspects of prejudice and racism.*
(pp. 43-67). New York, NY US: Springer Science + Business Media.

Egan, P. L., & Sherrill, K. (2009). California's Proposition 8: What Happened, and What Does
the Future Hold? *National Gay and Lesbian Task Force Policy Institute*, 1-16.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal.
Journal of Applied Psychology, 66(2), 127-148.

Fisher, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-
Crowne social desirability scale *Educational and Psychological Measurement, 53*(2),
417-424.

Fiske, S. T., & Taylor, S. E. (2008). *Social cognition : from brains to culture* (1st ed.). Boston: McGraw-Hill Higher Education.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism *Prejudice, discrimination, and racism.* (pp. 61-89). San Diego, CA US: Academic Press.

GLAAD Media Reference Guide (2007). Retrieved August 12, 2008, from http://www.glaad.org/Page.aspx?pid=373

Golom, F. D., & Liberman, B. E. (2008, April). Think manager, think male? Stereotypes of gay and lesbian managers. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Gorman, C. A., & Rentsch, J. R. (2008). Evaluating the Effectiveness of Frame-of-Reference Scales for Improving Rating Accuracy

Paper presented at the 23rd Annual Meeting of the Society for Industrial/Organizational Psychologists, San Francisco California.

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33*(1), 64-86.

Guilford, J. P. (1954). *Psychometric methods* (2d ed.). New York,: McGraw-Hill.

Guion, R. M. (1965). *Personnel Testing*. New York: McGraw-Hill.

Hastie, R. (1981). Schematic principles in human memory. In E. T. Higgins, C. P. Herman & M. P. Zanna (Eds.), *Social cognition : the Ontario Symposium, volume 1* (pp. x, 437 p.). Hillsdale, N.J.: L. Erlbaum Associates.

Hebl, M. R., Foster, J. B., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin, 28*(6), 815-825.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology, 92*(1), 81-92.

Herek, G. M. (1984). Beyond 'homophobia': A social psychological perspective on attitudes toward lesbians and gay men. *Journal of Homosexuality, 10*(1), 1-21.

Herek, G. M. (1986). The social psychology of homophobia: Toward a practical theory. *Review of Law & Social Change, 14*(4), 923-934.

Herek, G. M. (1990). The context of anti-Gay violence: Notes on cultural and psychological heterosexism. *Journal of Interpersonal Violance, 5*(3), 316-333.

Herek, G. M. (2007). Confronting sexual stigma and prejudice: Theory and practice. *Journal of Social Issues, 63*(4), 905-925.

Herek, G. M. (2008). Hate crimes and stigma-related experiences among sexual minority adults in the United States: Prevalence estimates from a national probability sample. *Journal of Interpersonal Violance*(In press).

Herek, G. M., & Glunt, E. K. (1993). Interpersonal Contact and Heterosexuals' Attitudes toward Gay Men: Results from a National Survey. *The Journal of Sex Research, 30*(3), 239-244.

Hollenbeck, G. T. (2008). "When I use a word ...". *Industrial and Organizational Psychology, 1*(2), 183-184.

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hu, L.-T., Bentler, P. M., & Hoyle, R. H. (1995). Evaluating model fit *Structural equation modeling: Concepts, issues, and applications.* (pp. 76-99). Thousand Oaks, CA US: Sage Publications, Inc.

Jackson, L. A., & Sullivan, L. A. (1989). Cognition and affect in evaluations of stereotyped members. *Journal of Social Psychology, 129*(5), 659-672.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*(2), 307-321.

Jensen, A. R. (1980). *Bias in Mental Testing*.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology, 93*(3), 321-334.

Jones, M. (2002). *Social psychology of prejudice*. Upper Saddle River, N.J.: Prentice Hall.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago, IL Hillsdale, NJ USEngland: Scientific Software International

Lawrence Erlbaum Associates, Inc.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The Forgotten Ones? The Validity of Consideration and Initiating Structure in Leadership Research. *Journal of Applied Psychology, 89*(1), 36-51.

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

Kite, M. E., Whitley, B. E., Jr., Garnets, L. D., & Kimmel, D. C. (2003). Do heterosexual women and men differ in their attitudes toward homosexuality? A conceptual and methodological analysis *Psychological perspectives on lesbian, gay, and bisexual experiences (2nd ed.).* (pp. 165-187). New York, NY US: Columbia University Press.

Kite, M. E., Whitley, B. E., Jr., & Herek, G. M. (1998). Do heterosexual women and men differ in their attitudes toward homosexuality? A conceptual and methodological analysis *Stigma and sexual orientation: Understanding prejudice against lesbians, gay men, and bisexuals.* (pp. 39-61). Thousand Oaks, CA US: Sage Publications, Inc.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70*(1), 56-65.

Lance, C. E., Hoffman, B. J., Gentry, W. A., & Baranik, L. E. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resource Management Review, 18*(4), 223-232.

Landy, F. J. (2008). Stereotypes, Bias, and Personnel Decisions: Strange and Stranger. *Industrial and Organizational Psychology, 1*(4), 379-392.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30*(6), 881-905.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype Content Model Explains Prejudice for an Envied Outgroup: Scale of Anti-Asian American Stereotypes. *Personality and Social Psychology Bulletin, 31*(1), 34-47.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23*(2), 127-135.

Longenecker, C. O., Sims, J. H. P., & Gioia, D. A. (1987). Behind the Mask: The Politics of Employee Appraisal. *Academy of Management Executive, 1*(3), 183-193.

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles, 37*(9), 663-685.

Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist, 51*(2), 157-158.

Martell, R. F., Parker, C., Emrich, C. G., & Crawford, M. S. (1998). Sex stereotyping in the executive suite: "Much ado about something". [Article]. *Journal of Social Behavior and Personality, 13*(1), 127-138.

Maume, D. J., JR. (1999). Glass Ceilings and Glass Escalators: Occupational Segregation and Race and Sex Differences in Managerial Promotions. *Work and Occupations, 26*(4), 483-509.

McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin, 9*(4), 551-558.

McConahay, J. B., Dovidio, J. F., & Gaertner, S. L. (1986). Modern racism, ambivalence, and the Modern Racism Scale *Prejudice, discrimination, and racism.* (pp. 91-125). San Diego, CA US: Academic Press.

McIntosh, P., & Plous, S. (2003). White privilege: Unpacking the invisible knapsack *Understanding prejudice and discrimination.* (pp. 191-196). New York, NY US: McGraw-Hill.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592.

Moore, D. P. (1984). Evaluating in-role and out-of-role performers. *Academy of Management Journal, 27*(3), 603-618.

Mount, M. K., Sytsma, M. R., Hazucha, J. F., & Holt, K. E. (1997). Rater-ratee race effects in developmental performance ratings of managers. *Personnel Psychology, 50*(1), 51-69.

Munoz, C. S., & Thomas, K. M. (2005). A multi-level examination of career barriers for sexual minorities employees. *Unpublished Dissertation.*

Murphy, K. R. (2008). Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance. *Industrial and Organizational Psychology, 1*(2), 148-160.

Nelson, T. E., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology, 32*(1), 13-38.

Ragins, B. R., & Martocchio, J. J. (2004). Sexual orientation in the workplace: The unique work and career experiences of gay, lesbian and bisexual workers *Research in personnel and human resources management, Vol 23.* (pp. 35-129). US: Elsevier Science/JAI Press.

Rieger, G., Linsenmeier, J., Gygax, L., Garcia, S., & Bailey, J. (Writer) (2007). Dissecting Gaydar: Accuracy and the Role of Masculinity and Femininity.

Rieger, G., Linsenmeier, J. A. W., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology, 44*(1), 46-58.

Robison-Cox, J. F., Martell, R. F., & Emrich, C. G. (2007). Simulating Gender Stratification. *Journal of Artificial Societies & Social Simulation, 10*(3), 1-18.

Sackett, P. R., DuBois, C. L., & Noe, A. W. (1991). Tokenism in performance evaluation: The effects of work group representation on male-female and White-Black differences in performance ratings. *Journal of Applied Psychology, 76*(2), 263-267.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*(5), 885-904.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111*(1), 42-61.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne social desirability scale. *Journal of Clinical Psychology, 28*, 191-193.

Stroh, L. K., Langlands, C. L., & Simpson, P. A. (2004). Shattering the galss ceiling in the new millenium. In M. S. Stockdale & F. J. Crosby (Eds.), *The psychology and management of workplace diversity* (pp. xix, 377 p.). Malden, MA: Blackwell Pub.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*(3), 497-506.

Sulsky, L. M., & Day, D. V. (1994). Effects of frame-of-reference training on rater accuracy under alternative time delays. *Journal of Applied Psychology, 79*(4), 535-543.

Taylor, A. (1983). Conceptions of masculinity and femininity as a basis for stereotypes of male and female homosexuals. *Journal of Homosexuality, 9*(1), 37-53.

The Kaiser Family Foundation (2001). A Report on the Experiences of Lesbians, Gays and Bisexuals in America and the Public's Views on Issues and Policies Related to Sexual Orientation. Retrieved August 15, 2008 from http://www.kff.org/kaiserpolls/3193-index.cfm

Thomas, K. M., Phillips, L. D., & Brown, S. (1998). Redefining race in the workplace: Insights from ethnic identity theory. *Journal of Black Psychology, 24*(1), 76-92.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69.

Walls, N. E. (2008). Modern heterosexism and social dominance orientation: Do subdomains of heterosexism function as heirarchy-enhancing legitimizing myths? In T. G. Morrison & M. A. Morrison (Eds.), *The Psychology of Modern Prejudice* (pp. 225-259). Hauppauge, NY: Nova Science Publishers.

Whitley, B. E., Jr. (2001). Gender-role variables and attitudes toward homosexuality. *Sex Roles, 45*(11), 691-721.

Wilson, T. C. (1996). Compliments will get you nowhere: Benign stereotypes, prejudice, and anti-Semitism. *Sociological Quarterly, 37*, 465-479.

Woehr, D. J. (2008). On the Relationship Between Job Performance and Ratings of Job

Performance: What Do We Really Know? *Industrial and Organizational Psychology,*

1(2), 161-166.

Table 1

Lab and Field Manipulation Checks

Number	Condition	Manipulation Check Item	Sample	Number P's missing the item	Total
1		Whose Performance will you rate?			
	Both	Chose: The Direct Report	Lab	9	
	Both	Chose: The Direct Report	Field	4	13
2		Item was removed from analysis	N/A	N/A	N/A
3		Mr. Johnson has a / is:			
	Gay	Engaged	Lab	2	
		Single	Lab	2	
		Girlfriend	Lab	7	
	Straight	Domestic Partner	Lab	14	
		Single	Lab	2	
		Girlfriend	Lab	4	31
3		Mr. Johnson has a / is:			
	Gay	Engaged	Field	3	
		Single	Field	1	
		Girlfriend	Field	6	
	Straight	Domestic Partner	Field	1	
		Single	Field	1	
		Girlfriend	Field	0	12
4		It appears Mr. Johnson's affinity Orientation is:			
	Gay	Straight	Lab	1	
	Straight	Gay	Lab	1	2
4		It appears Mr. Johnson's affinity Orientation is:			
	Gay	Straight	Field	4	
	Straight	Gay	Field	3	7
5		Mr. Johnson belonged to:			
	Gay	St. Patrick's Day Parade	Lab	0	
	Straight	Gay Pride Parade	Lab	1	1
5		Mr. Johnson belonged to:			
	Gay	St. Patrick's Day Parade	Field	0	
	Straight	Gay Pride Parade	Field	1	1
6		Do you know the person in the video?			
	Both	I know him well	Lab	1	
	Both	I know him well	Field	16	17
				Total Lab	**45**
				Total Field	**40**

Table 2

Demographic Characteristics of Lab Sample

Variables	Percentage	Frequency	Total
Gender	Female = 60%	266	
	Male = 40%	173	**439**
Age	>18 = 28.5%	127	
	19 = 28.5%	125	
	20 = 21.5%	94	
	<21 = 17.8%	79	
	Skipped item = 3.2%	14	**439**
Race	Native American = .7%	3	
	Caucasian = 80%	352	
	Asian/Pacific Islander = 7.8%	34	
	Latino/a = 1.6%	7	
	African American = 7.1%	31	
	Multi-Racial = 2.7%	12	**439**
Sexuality	Gay = 2.1%	9	
	Bi = 1.6%	7	
	Straight = 96%	421	
	Skipped item = .4%	2	**439**
Mother's Level of Education	Elementary/Middle School = .9%	4	
	High School or GED = 26.3%	115	
	Technical Training = 5%	22	
	Associates Degree 7.5%	33	
	Bachelor's Degree = 39%	171	
	Master's Degree = 13.2%	58	
	Professional Degree 3.4%	15	
	Doctoral Degree = 2.3%	10	
	Skipped item = 2.5%	11	**439**
Fathers' Level of Education	Elementary/Middle School = 1.4%	6	
	High School or GED = 16.4%	72	
	Technical Training = 6.6%	29	
	Associates Degree = 7.8	34	
	Bachelor's Degree = 37.4	164	
	Master's Degree = 15.3	67	
	Professional Degree = 6.6%	29	
	Skipped item = 2.5%	11	
	Doctoral Degree = 6.2%	27	**439**

	Descriptive Statistics				
	Mean	*SD*	*Min*	*Max*	*Range*
Age	19.55	1.97	17.00	39.00	22.00

Table 3

Demographic Characteristics of Field Sample

Variables	Percentage	Frequency	Total
Gender	Female = 54.4%	135	
	Male = 45.6%	113	
Age	>30 = 15%	37	
	31-35 = 16.6%	41	
	36-40 = 16.5%	41	
	41-45 = 10.5%	26	
	46-50 = 12.5%	36	
	51-55 = 13.3%	39	
	56-60 = 6.1%	19	
	<61 = 3.2%	9	**248**
Race	Native American = 0%	0	
	Caucasian = 77.9%	193	
	Asian/Pacific Islander = 4.8%	12	
	Latino/a = 3.2%	8	
	African American =10.4%	26	
	Multi-Racial = 2.4%	6	
	Other = 1.2%	3	**248**
Sexuality	Gay = 6.9%	17	
	Bi = 2%	5	
	Straight = 89%	221	
	Skipped Item = 2%	5	**248**
Level of Education	Elementary/Middle School = 0%	0	
	High School or GED = 1.6%	3	
	Technical Training = 1.2%	3	
	Associates Degree 4%	10	
	Bachelor's Degree = 27.3%	68	
	Master's Degree = 45%	112	
	Professional Degree = 4%	10	
	Doctoral Degree = 16.1%	40	
	Skipped item = .8%	2	**248**
Marital Status	Single (Never Married) = 20.9%	51	
	Married = 67.1%	167	
	Divorced/Separated 11.2%	28	
	Widowed = .4%	1	
	Skipped item = .4%	1	**248**

State of Residence	Georgia = 39%	97	
	California = 4.8%	12	
	Florida = 4%	10	
	New York = 3.6%	9	
	Texas = 3.2%	8	
	All other states = 38%	84	
	Outside the US = 11.3%	28	**248**
Country of Residence	Argentina = .4%	1	
	Australia = .85	2	
	Belgium = .4%	1	
	Canada = 1.2%	3	
	Germany = .8%	2	
	India = .8%	2	
	Israel = .4%	1	
	Malaysia = .4%	1	
	Netherlands = .4%	1	
	Puerto Rico = .4%	1	
	Romania = .4%	1	
	Scotland = .4%	1	
	South Africa = .4%	1	
	UK = 2.8%	7	
	USA = 89%	221	
	Vietnam = .4%	1	
	Skipped item = .4%	1	**248**
Employment Status	40 Hours a week = 78.7%	196	
	20 Hours a week = 6%	15	
	Other (e.g., Business Owner/Consultant/Retired) = 8.1%	20	
	Laid off/Recently unemployed = 6.9%	17	**248**
Number of Employees	less than 100 = 27.8%	69	
	101 - 499 = 15.3%	38	
	500 - 999 = 5.6%	14	
	1,000 - 10,000 = 24.2%	60	
	over 10,001 = 20.2%	50	
	Skipped Item = 6.9%	17	**248**
Industry	Advertising/Marketing/PR = 1.6%	4	
	Arts/Entertainment/Media = 1.6%	4	
	Banking/Financial Svcs/Accounting = 6.4%	16	
	Consulting Services 12.9%	32	
	Education = 14.5%	36	
	Engineering = 2.8%	7	

	Government and Policy = 4%	10
	Medical/Healthcare = 7.6%	19
	Human Resources/Recruiting = 9.6%	24
	Information Technology/Computers = 12%	30
	Internet/e-Commerce = 2%	5
	Legal = 1.2%	3
	Non-profit = 4.8%	12
	Publishing = .8%	2
	Real Estate = .4%	1
	Retail/Wholesale = 3.6%	9
	Sales = 1.6%	4
	Science/Biotechnology/Pharmaceuticals 1.6%	4
	Telecommunications = 2%	5
	Other (e.g., Program coordinator, Talent Mgmt) = 8.5%	21	**248**
Corporate Position	Administrative/Support = 2.8%	7
	Technical = 3.6%	9
	Supervisor = 6.0%	15
	Manager = 25.7%	64
	Director = 11.6%	29
	Senior Director = 3.2%	8
	Vice President = 6.0%	15
	Senior/Executive Vice President = 2.4%	6
	CEO/Executive = 5.6%	14
	N/A = 9.2%	23
	Skipped item = 23.4%	58	**248**
Education Position	Administrative/Support = 4.4%	11
	College Level Professor = 9.3%	23
	Grade School Teacher = 1.2%	3
	Grade School Director = .8%	2
	College Level Faculty/Administration = 3.2%	8
	Skipped item = 81%	201	**248**
Yearly Individual Salary	$10-19,999 = 2.4%	6
	$20-39,999 = 8.8%	22
	$40-59,999 = 14.5%	36
	$60-79,999 = 18.9%	47
	$80-99,999 = 13.3%	33
	$100K+ = 33.7%	84
	Skipped item = 8.1%	20	**248**

	Descriptive Statistics				
	Mean	*SD*	*Min*	*Max*	*Range*
Age	42.57	10.75	27.00	75.00	52.00
Years in Org	6.46	6.63	0.10	31.00	31.10

Table 4

Demographics by Experimental Condition

	Race by Condition												
	Lab A Gay	Lab A Straight	Lab B Gay	Lab B Straight	Lab C Gay	Lab C Straight	Field A Gay	Field A Straight	Field B Gay	Field B Straight	Field C Gay	Field C Straight	Total
Other	0	0	0	0	0	0	0	0	0	0	0	3	3
Native American	2	0	1	0	0	0	0	0	0	0	0	0	3
White	72	56	75	41	59	49	35	34	20	32	37	35	545
Asian Pac-Island	8	5	4	8	2	7	1	2	4	0	1	4	46
Latino/a	0	0	2	1	1	3	0	0	2	0	5	1	15
Black	5	6	8	4	5	3	3	2	4	3	10	4	57
Multi-Race	3	1	0	3	2	3	2	1	2	0	1	0	18
Total	90	68	90	57	69	65	41	39	32	35	54	47	687

	Gender by Condition												
	Lab A Gay	Lab A Straight	Lab B Gay	Lab B Straight	Lab C Gay	Lab C Straight	Field A Gay	Field A Straight	Field B Gay	Field B Straight	Field C Gay	Field C Straight	Total
Female	55	45	49	33	44	40	18	18	17	23	36	23	401
Male	35	23	41	24	25	25	23	21	15	12	18	24	286
Total	90	68	90	57	69	65	41	39	32	35	54	47	687

	Sexuality by Condition												
	Lab A Gay	Lab A Straight	Lab B Gay	Lab B Straight	Lab C Gay	Lab C Straight	Field A Gay	Field A Straight	Field B Gay	Field B Straight	Field C Gay	Field C Straight	Total
Skipped	0	0	1	0	0	0	0	2	0	1	0	2	6
Bi	1	1	1	1	0	3	0	1	2	0	2	0	12
Gay	2	0	4	0	1	2	6	1	4	2	2	2	26
Straight	87	67	84	56	68	60	35	35	26	32	50	43	643
Total	90	68	90	57	69	65	41	39	32	35	54	47	687

Table 5

Model Fit Statistics for Performance Ratings

	df	χ^2	RMSEA	SRMSR	TLI	CFI
1. 1 Factor	464	4415.18	.170	.093	.94	.95
2. 2 Factor	463	3947.68	.149	.097	.95	.95
3. 3 Factor	461	3681.39	.139	.110	.95	.96
4. 4 Factor	458	2342.76	.093	.069	.97	.97
5. 5 Factor	454	2036.45	.086	.068	.97	.98
6. 6 Factor	459	2207.33	.094	.075	.97	.98

Note. N = 121–248. *df* = model degrees of freedom; SRMSR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; TLI = Tucker-Lewis Index; CFI = comparative fit index.

Table 6

Correlation Analysis

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Gender [a]	--																
2. Race [b]	.06	--															
3. Age	-.06	-.02	--														
4. E	-.03	.00	-.09*	--													
5. DE	-.01	-.02	-.08*	.23**	--												
6. SA	-.01	-.02	-.10*	.22**	.79**	--											
7. DA	.05	.14**	-.18**	.20	-.01	-.01	--										
8. BDA	.02	-.03	-.04	-.15**	-.04	-.11**	-.18**	--									
9. SDES	.03	-.03	-.29**	-.04	-.06	-.03	.06	.03	(.60)								
10. GID	.27**	.07	-.17**	-.03	-.05	-.07	.02	-.03	.21*	(.79)							
11. GPERF	-.01	-.01	-.17**	.26**	.01	-.01	.05	-.01	.09*	.10**	(.96)						
12. ANAL	-.02	.00	-.16**	.22**	-.01	-.08*	.01	.14**	.10*	.11**	.89**	(.89)					
13. DECIS	-.06	.01	-.06	.25**	-.02	-.01	.04	-.08*	.02	.08	.88**	.75**	(.88)				
14. LEAD	.02	-.02	-.19**	.24**	.05	.02	.09*	-.07	.09*	.07	.94**	.79**	.78**	(.91)			
15. CONF	.03	-.04	-.14**	.19**	.07	-.01	.03	.04	.09*	.10*	.82**	.66**	.65**	.77**	(.93)		
16. SENS	-.04	.00	-.14**	.20**	-.10**	.03	.01	-.09*	.06	.07	.69**	.57**	.58**	.58**	.30**	(.85)	
17. HIRE	-.01	.01	-.18**	.19**	.01	-.00	.00	.02	.07*	.09*	.85**	.78**	.73**	.81**	.66**	.63**	(.96)
M	1.59	2.45	28.03	0.54	0.23	0.17	0.19	0.58	1.54	3.28	2.87	3.15	2.66	2.91	2.94	2.65	2.89
SD	0.49	1.06	12.98	0.63	0.29	0.21	0.15	0.28	0.21	0.97	0.78	0.90	0.89	0.88	0.97	0.93	1.32

Note. $N = 671-687$. E = elevation. DE = differential elevation. SA = stereotype accuracy. DA = differential accuracy. BDA = Borman's differential accuracy. SDES = Social Desirability. GID = Gender Identity. GPERF = General Performance. ANAL = Analysis. DECIS = Decisiveness. LEAD = Leadership. CONF = Conflict Management. SENS = Sensitivity. HIRE = Hiring Recommendation.
[a] 1 = male, 2 = female. [b] 0 = Other, 1 = Native American, 2 = Caucasian, 3 = Asian/Pacific Islander, 4 = Latino/a, 5 = African American, 6 = Multi-racial..
*p < .05. **p < .01.

Table 7

CFA Model Goodness-of-Fit Indices: Omnibus Performance Model by Sample

Group	*df*	χ^2	SRMSR	RMSEA	TLI	CFI
All Samples	528	1010.05**	.08	.04	.99	.99
Lab	528	614.14**	.07	.02	.99	1.00
Field	528	950.98**	.07	.06	.98	.99

Note. n = 122-437; *df* = model degrees of freedom; SRMSR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; TLI = Tucker-Lewis Index; CFI = comparative fit index; *p* < .05; **p* < .01.

Table 8

CFA Model Goodness-of-Fit Indices: 5-Factor Performance Model for Gay and Straight Manager

Model	*df*	χ^2	SRMSR	RMSEA	TLI	CFI	Δdf	Δχ^2	ΔCFI
Lab									
1. Configural Invariance	908	2054.20**	.086	.088	.951	.972	----	----	----
1 versus 2	----	----	----	----	----	----	27	46.06*	.001
2. Metric Invariance	935	2100.26**	.092	.088	.950	.971	----	----	----
2 versus 3	----	----	----	----	----	----	32	59.40**	.000
3. Scalar Invariance	967	2159.66**	.094	.087	.949	.971	----	----	----
Field									
1. Configural Invariance	908	1666.11**	.088	.085	.947	.975	----	----	----
1 versus 2	----	----	----	----	----	----	27	63.80**	.001
2. Metric Invariance	935	1729.91**	.099	.086	.945	.974	----	----	----
2 versus 3	----	----	----	----	----	----	32	40.19**	.001
3. Scalar Invariance	967	1770.10**	.099	.084	.943	.973	----	----	----

Note. N = 122-248; *df* = model degrees of freedom; SRMSR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; TLI = Tucker-Lewis Index; CFI = comparative fit index; *p* < .05; **p* < .01.

Table 9
Test of Hypothesis 1: General Performance Ratings

	General Performance	
	Lab	Field
High Performance		
Sexuality β	-.148ψ	-.102
Average Performance		
Sexuality β	.020	-.261*
Low Performance		
Sexuality β	.019	-.157

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta ψ = The Standardized Beta dropped from a level of significance to non-significant when controlling for social desirability. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 10

Test of Hypothesis 2: Analysis Dimension Ratings

	Analysis	
	Lab	Field
High Performance		
Sexuality β	-.157*	-.041
Average Performance		
Sexuality β	.005	-.208
Low Performance		
Sexuality β	-.082	-.087

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 11

Test of Hypothesis 2: Decisiveness Dimension Ratings

	Decisiveness	
	Lab	Field
High Performance		
Sexuality β	-.195**	-.187
Average Performance		
Sexuality β	-.081	-.232ψ
Low Performance		
Sexuality β	-.128	-.086

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. ψ = The Standardized Beta changed from a level of significance to non-significance when controlling for social desirability. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 12

Test of Hypothesis 2: Leadership Dimension Ratings

	Leadership	
	Lab	Field
High Performance		
Sexuality β	-.080	-.059
Average Performance		
Sexuality β	.054	-.277*
Low Performance		
Sexuality β	-.084	-.132

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 13

Test of Hypothesis 2: Conflict Management Dimension Ratings

	Conflict Management	
	Lab	Field
High Performance		
Sexuality β	-.064	.000
Average Performance		
Sexuality β	.002	-.318**
Low Performance		
Sexuality β	.007	-.189ψ

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. ψ = The Standardized Beta changed from a level of significance to non-significance when controlling for social desirability. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 14

Test of Hypothesis 2: Interpersonal Sensitivity Dimension Ratings

	Interpersonal Sensitivity	
	Lab	Field
High Performance		
Sexuality β	-.110	-.099
Average Performance		
Sexuality β	.095	.127
Low Performance		
Sexuality β	-.162ψ	-.108

Note: *N* = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. ψ = The Standardized Beta changed from a level of significance to non-significance when controlling for social desirability. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 15

Test of Hypothesis 2: Hiring Recommendation Ratings

	Hiring Recommendation	
	Lab	Field
High Performance		
Sexuality β	-.148ψ	-.003
Average Performance		
Sexuality β	-.090	-.259*
Low Performance		
Sexuality β	-.033	-.116

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. β = Standardized Beta. ψ = The Standardized Beta changed from a level of significance to non-significance when controlling for social desirability. *p < .05. **p < .01. In step 1, Social Desirability was entered into the model as a control variable. In step 2, Sexuality was entered as a predictor and the Standardized Betas for Sexuality are reported above.

Table 16
Test of Hypothesis 4: General Performance Ratings

	General Performance	
	Lab	Field
High Performance	-.133	.334
Average Performance	.400	-.602
Low Performance	.335	.184

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.
In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.
The Standardized Betas from step 3 are reported in this table.

Table 17

Test of Hypothesis 4: Analysis Ratings

	Analysis	
	Lab	Field
High Performance	-.140	.361
Average Performance	.453	-.429
Low Performance	.425	.028

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.

In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.

The Standardized Betas from step 3 are reported in this table.

Table 18
Test of Hypothesis 4: Decisiveness Ratings

	Decisiveness	
	Lab	Field
High Performance	-.395	.132
Average Performance	.389	-.286
Low Performance	.130	.374

Note: $N = 32 - 90$. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.

In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.

The Standardized Betas from step 3 are reported in this table.

Table 19

Test of Hypothesis 4: Leadership Ratings

	Leadership	
	Lab	Field
High Performance	-.018	.404
Average Performance	.294	-.614
Low Performance	.458	.092

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.

In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.

The Standardized Betas from step 3 are reported in this table.

Table 20

Test of Hypothesis 4: Conflict Management Ratings

	Conflict Management	
	Lab	Field
High Performance	-.290	.480
Average Performance	.301	-.188
Low Performance	.390	.402

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.

In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.

The Standardized Betas from step 3 are reported in this table.

Table 21
Test of Hypothesis 4: Interpersonal Sensitivity Ratings

	Interpersonal Sensitivity	
	Lab	Field
High Performance	.285	-.081
Average Performance	.103	.352
Low Performance	-.352	-.180

Note: N = 32 - 90. Sexuality: Gay = 0, Straight = 1. Participant Gender: Female = 0, Male = 1. B = Standardized Beta. *p < .05. **p < .01.
In step 1, Social Desirability was entered into the model as a control, in step 2, Sexuality and Participant Gender were entered as predictors. In step 3, the joint effect term was entered as a moderator.
The Standardized Betas from step 3 are reported in this table.

Table 22

Hypothesis 5: Means, Standard Deviations, and Rating Accuracy Test Results Across High Performance Condition by Sample

	Lab High Performance		Field High Performance	
	Gay	Straight	Gay	Straight
Accuracy Index	(*n* = 90)	(*n*=68)	(*n* = 41)	(*n*=39)
Elevation				
M	.38	.47	.50	.61
SD	.39	.43	.54	.76
Differential Elevation				
M	.36	.48	.34	.34
SD	.37	.40	.34	.36
Stereotype Accuracy				
M	.30	.26	.25	.21
SD	.22	.22	.24	.21
Differential Accuracy				
M	.18	.10	.14	.16
SD	.11	.01	.09	.12
Borman's Differential Accuracy				
M	.62	.56	.55	.60
SD	.26	.31	.32	.26

Note. Lower values on elevation, differential elevation, stereotype accuracy, and differential accuracy denote greater accuracy. Higher values on Borman's differential accuracy represent greater accuracy. Means with different subscripts are significantly different at *p* < .05.

Table 23

Hypothesis 5: Means, Standard Deviations, and Rating Accuracy Test Results Across Average Performance Condition by Sample

	Lab Average Performance		Field Average Performance	
	Gay	Straight	Gay	Straight
Accuracy Index	($n = 90$)	($n=57$)	($n = 32$)	($n=35$)
Elevation				
M	.87	.70	.79$_a$	.29 $_b$
SD	.92	.58	.81	.32
Differential Elevation				
M	.13	.06	.16	.05
SD	.22	.08	.25	.05
Stereotype Accuracy				
M	.13	.07	.16	.05
SD	.23	.09	.25	.11
Differential Accuracy				
M	.21	.22	.15	.12
SD	.15	.13	.11	.13
Borman's Differential Accuracy				
M	.61	.60	.48	.64
SD	.25	.28	.26	.30

Note. Lower values on elevation, differential elevation, stereotype accuracy, and differential accuracy denote greater accuracy. Higher values on Borman's differential accuracy represent greater accuracy. Means with different subscripts are significantly different at $p < .05$.

Table 24

Hypothesis 5: Means, Standard Deviations, and Rating Accuracy Test Results Across Low Performance Condition by Sample

	Lab Low Performance		Field Low Performance	
	Gay	Straight	Gay	Straight
Accuracy Index	(*n* = 69)	(*n*=65)	(*n* = 54)	(*n*=47)
Elevation				
M	.57	.49	.35	.36
SD	.61	.58	.37	.55
Differential Elevation				
M	.19	.22	.16	.13
SD	.22	.20	.15	.17
Stereotype Accuracy				
M	.15	.15	.08	.10
SD	.22	.18	.09	.15
Differential Accuracy				
M	.26	.22	.15	.12
SD	.24	.15	.22	.02
Borman's Differential Accuracy				
M	.56	.55	.56	.56
SD	.27	.29	.28	.29

Note. Lower values on elevation, differential elevation, stereotype accuracy, and differential accuracy denote greater accuracy. Higher values on Borman's differential accuracy represent greater accuracy. Means with different subscripts are significantly different at $p < .05$.

Table 25

Overview of Hypotheses

Hypothesis	Supported	Partially Supported	Not Supported
Hypothesis 1:			
Main effect of Sexuality on General Performance			X
Hypothesis 2:			
Main effect of Sexuality on Dimension Level Performance			X
Hypothesis 3:			
Main effect of Participant Gender on General and Dimension Level Performance		X	
Hypothesis 4:			
Moderation of Gender Identity on Performance			X
Hypothesis 5:			
Straight Manager rated with Greater Accuracy than Gay Manager		X	
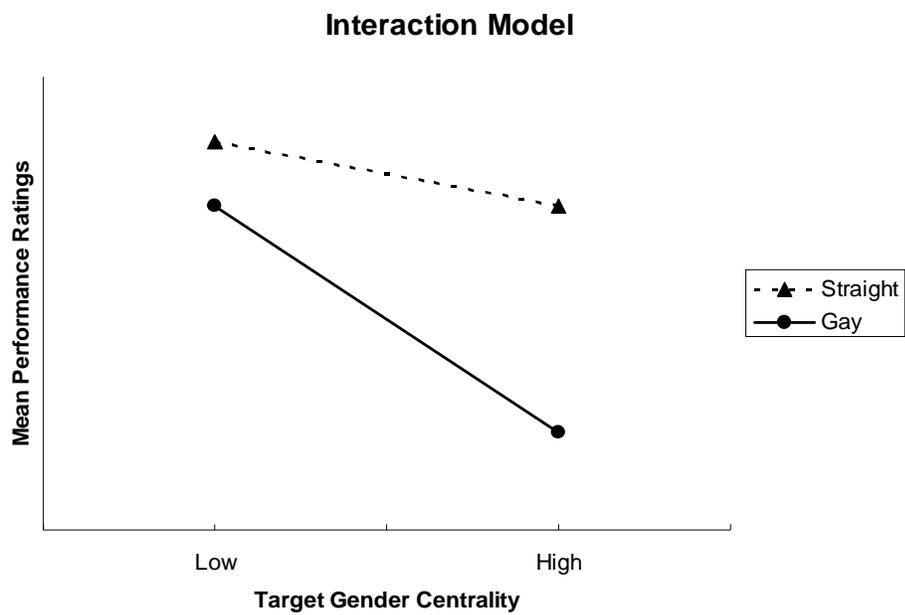
Figure 1: Expected interaction between gender centrality and mean performance ratings in the average and low performance conditions.

One- Factor Model
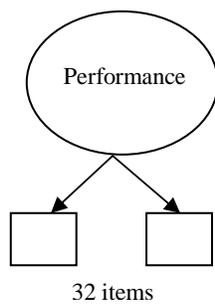
One-Factor Model

Performance

32 items

Two-Factor Model

Task

Inter Personal

12 items

11 items

Three-Factor Model

Technical

Leadership

Inter-personal

12 items

9 items

11 items

Four-Factor Model

12 items   9 items   6 items   5 items

Five-Factor Model

6 items   6 items   9 items   6 items   5 items

Six-Factor Model

6 items   6 items   9 items   6 items   5 items
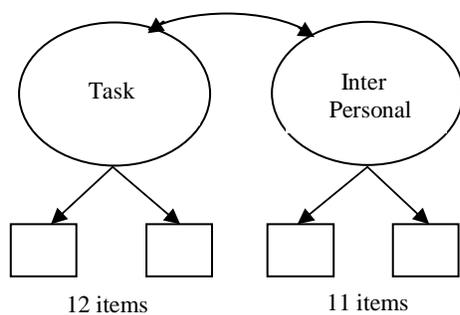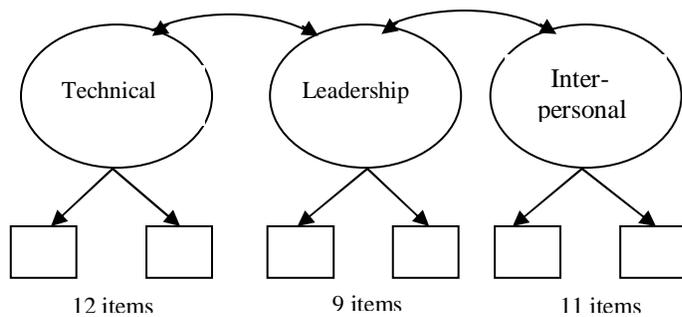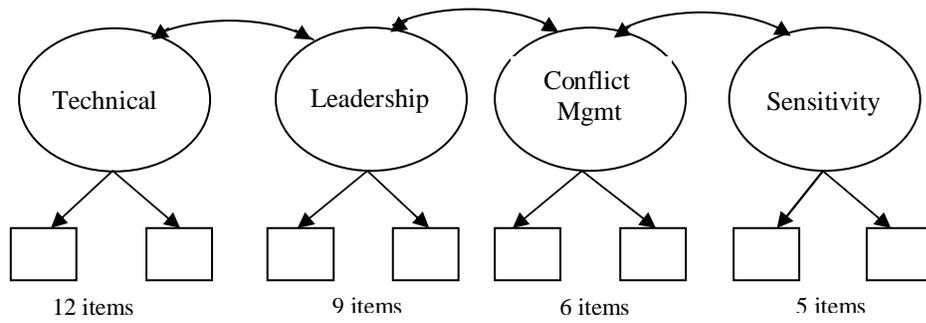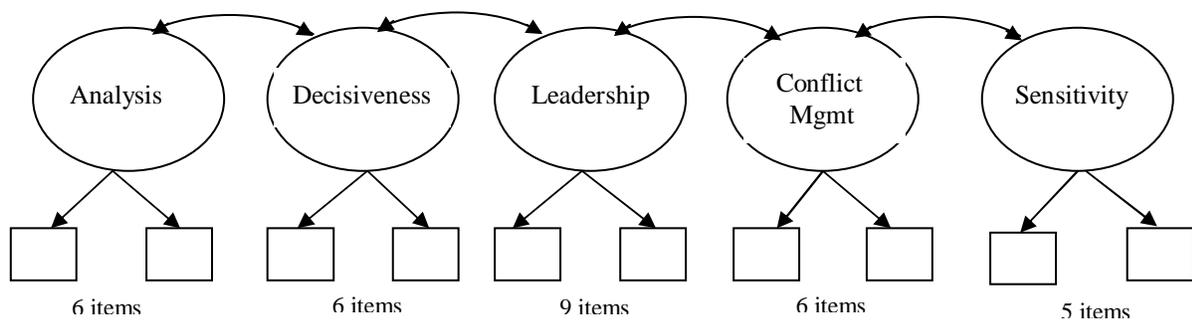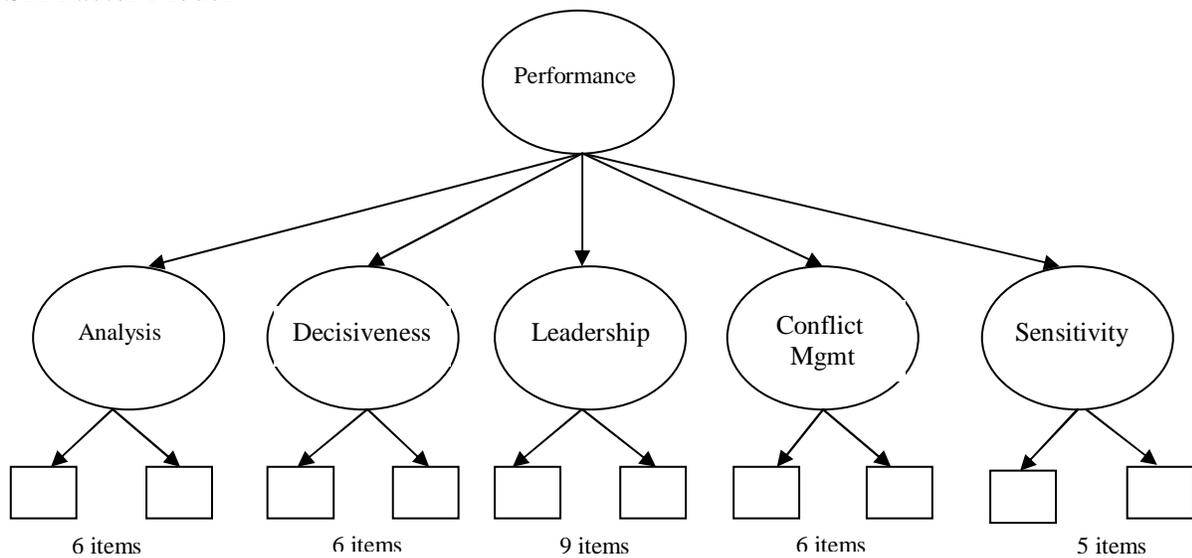
Figure 2. Graphic illustrations of conceptual performance models. For simplification, only subsets of manifest indicators are displayed. To further simplify visual presentation, Two, Three, Four, Five, and Six-Factor variables display correlations between adjacent factors, but all factor intercorrelations were estimated in the analyses.
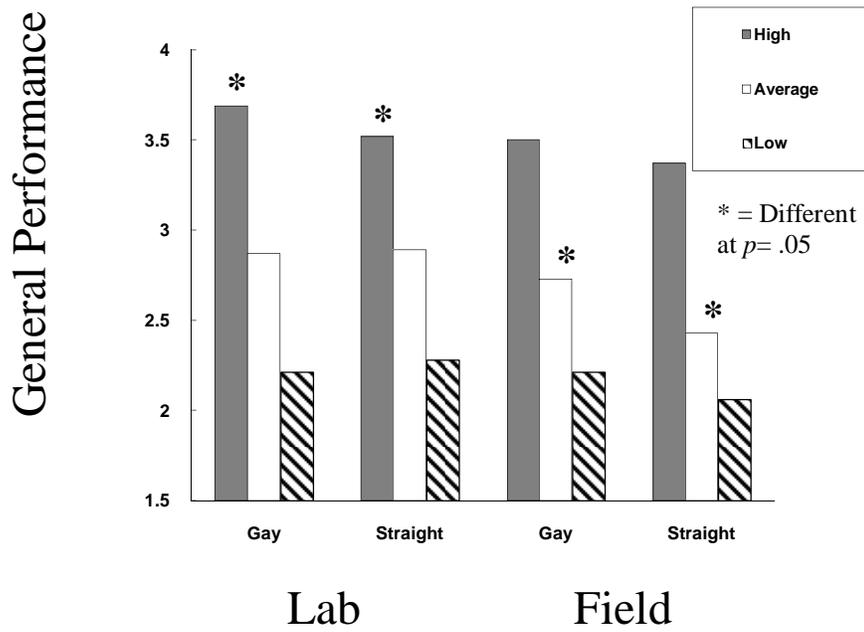
Figure 3. Mean general performance ratings across level of performance and sexuality in lab and field samples.
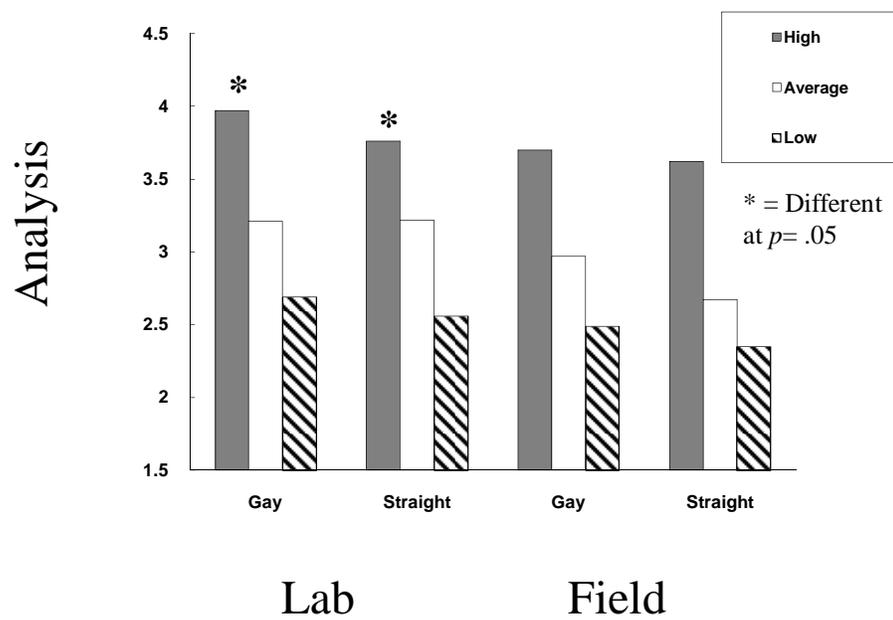
Figure 4. Mean analysis dimension ratings across level of performance and sexuality in lab and field samples.
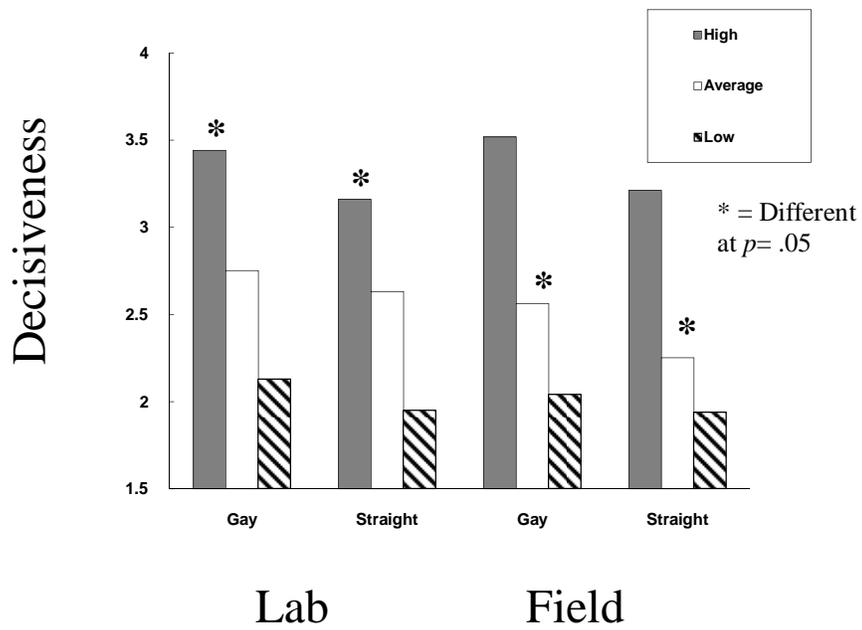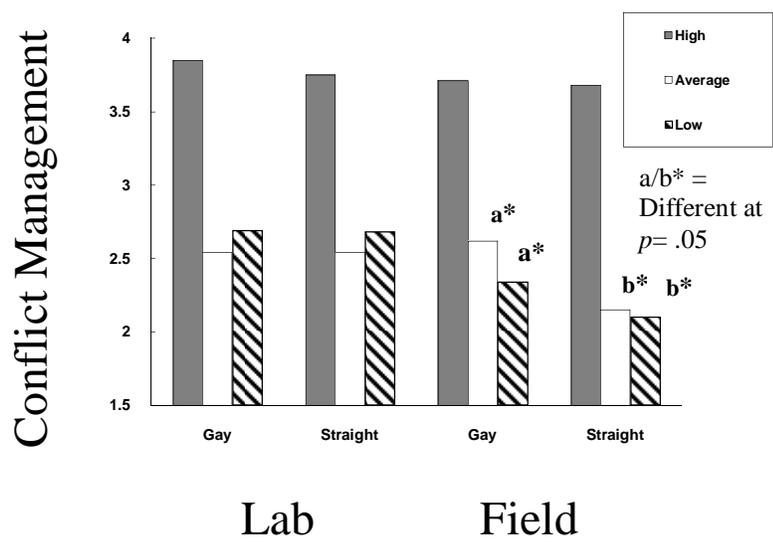
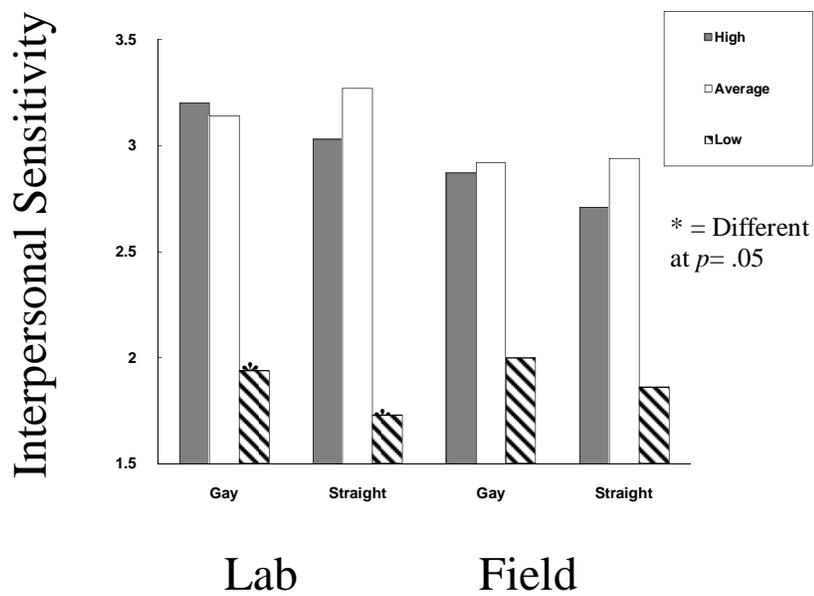*Figure 5.* Mean decisiveness dimension ratings across level of performance and sexuality in lab and field samples.
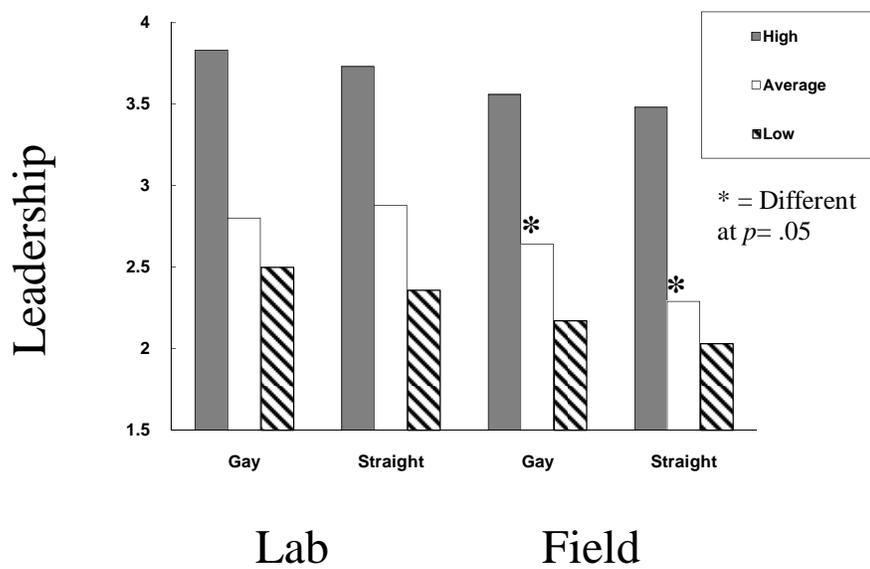
Figure 6. Mean leadership dimension ratings across level of performance and sexuality in lab and field samples.

Figure 7. Mean conflict management dimension ratings across level of performance and sexuality in lab and field samples.

Figure 8. Mean interpersonal sensitivity dimension ratings across level of performance and
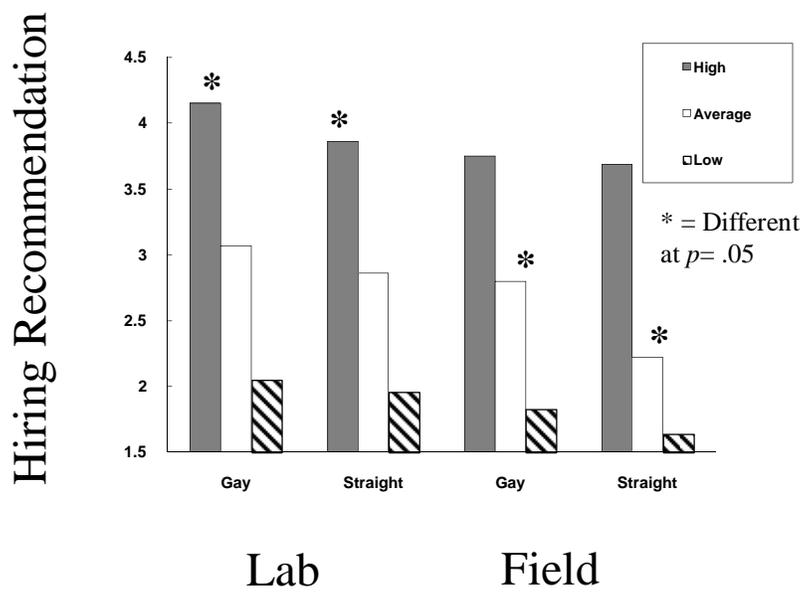sexuality in lab and field samples.

Figure 9. Mean hiring recommendation ratings across level of performance and sexuality in lab and field samples.
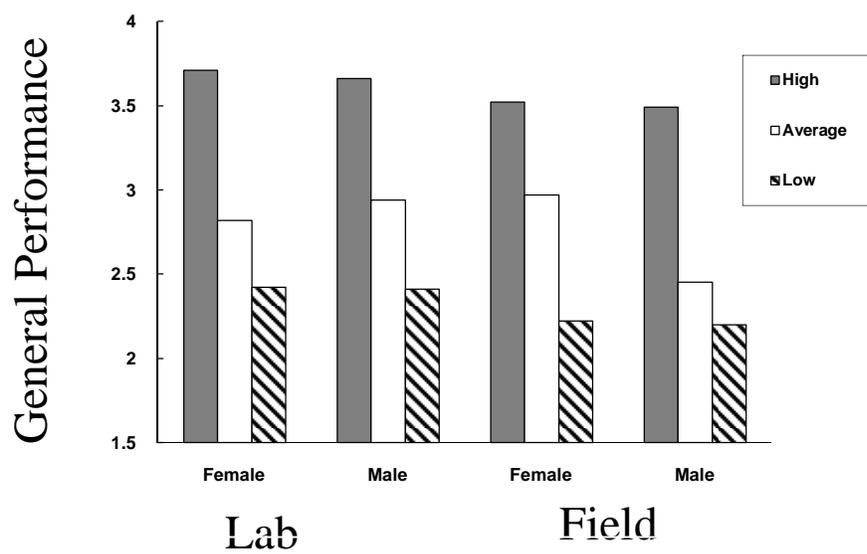
Figure 10. Mean general performance ratings for the Gay manager across level of performance and gender in lab and field samples.
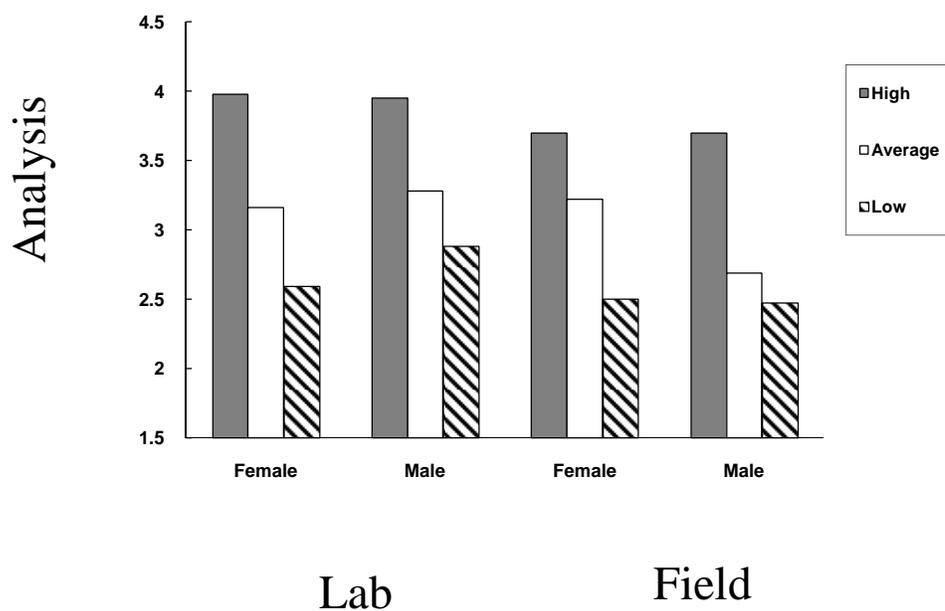
Figure 11. Mean analysis dimension ratings for the Gay manager across level of performance and gender in lab and field samples.
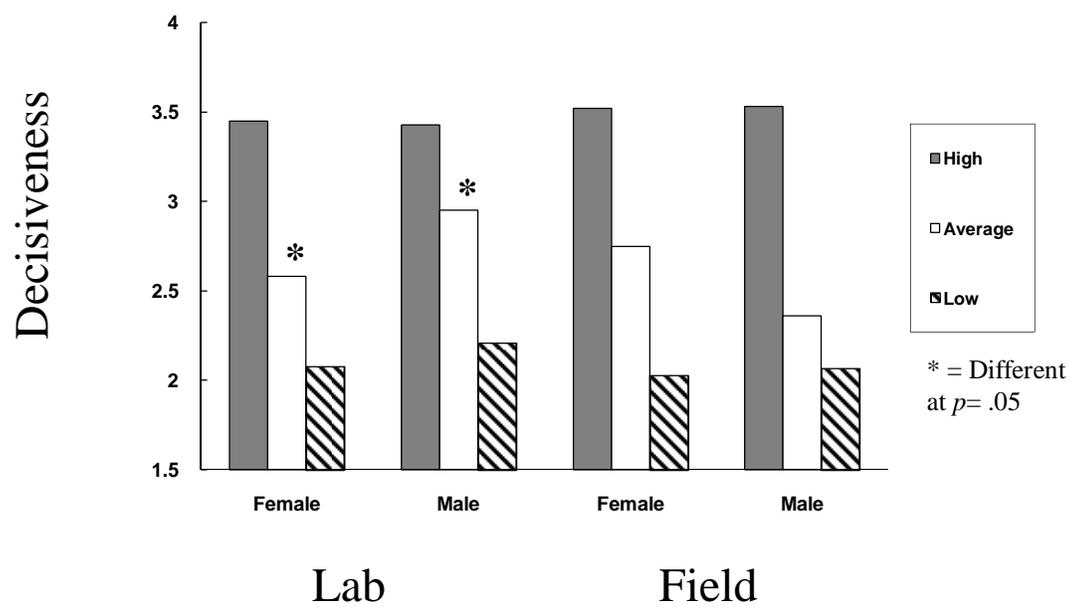
Figure 12. Mean decisiveness dimension ratings for the Gay manager across level of

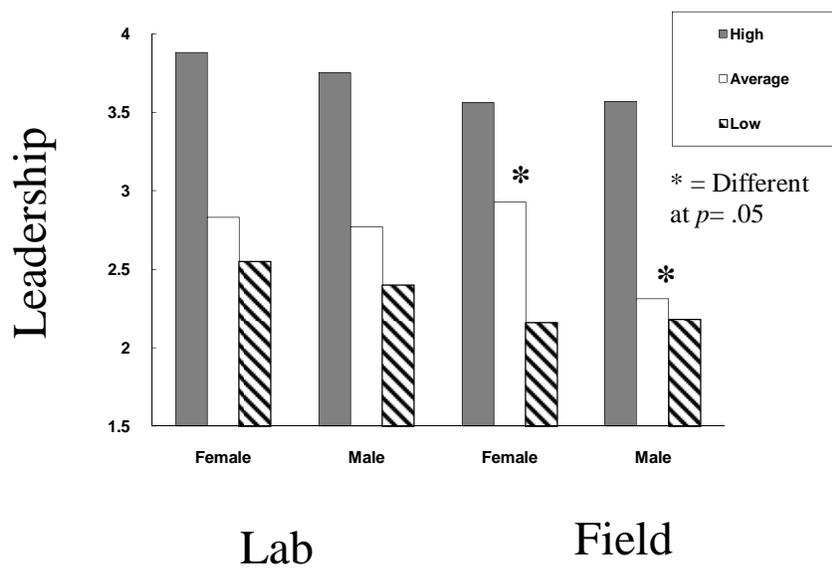performance and gender in lab and field samples.

Figure 13. Mean leadership dimension ratings for the Gay manager across level of performance

and gender in lab and field samples.

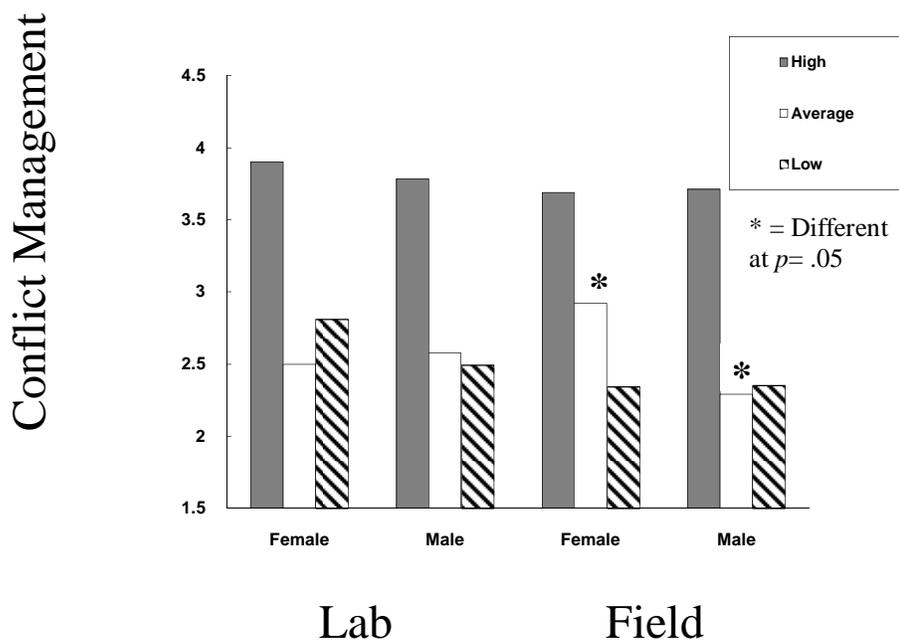*Figure 14.* Mean conflict management dimension ratings for the Gay manager across level of
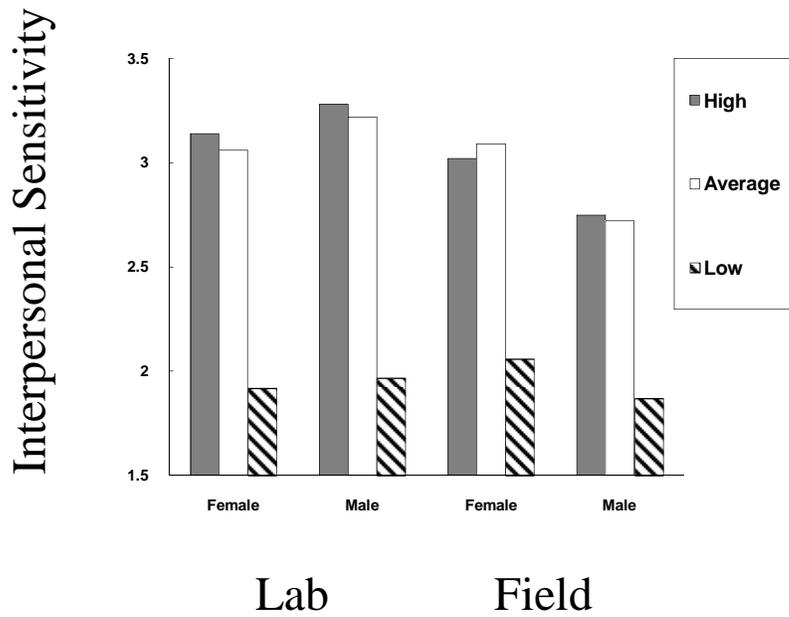
performance and gender in lab and field samples.

Figure 15. Mean interpersonal sensitivity dimension ratings for the Gay manager across level of performance and gender in lab and field samples.
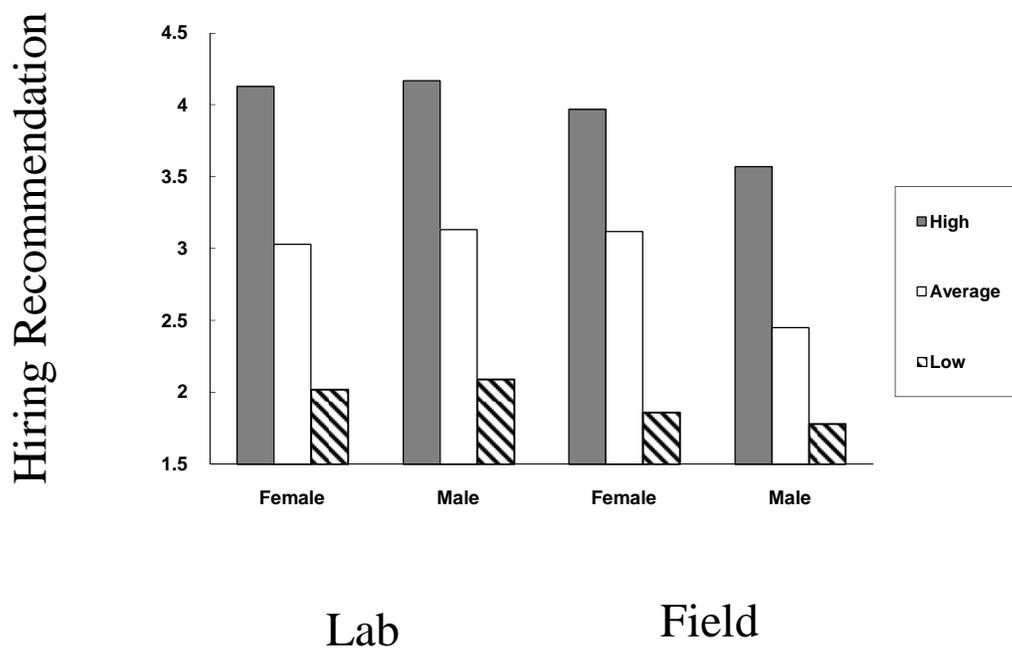
Figure 16. Mean hiring recommendation ratings for the Gay manager across level of performance and gender in lab and field samples.

ENDNOTES

[1] The need to understand the experiences of transgender, bi and lesbian employees of all different races is crucial. At this preliminary stage however, the proposed research will first explore the relationship between work performance and White Gay males, and propose a follow up study examining this relationship for LBT employees.

[2] The extent to which information-processing errors demonstrated in the laboratory actually reduce the relationship between ratings and actual job performance in the applied setting has not been empirically supported.

[3] Although not as prevalent, there is research to support that in some instances, stereotype inconsistent information is better recalled when individuals have enough time to make sense of the inconsistency (Hastie, 1981) and are motivated to form accurate impressions (Stangor & McMillan, 1992).

[4] Research demonstrates that individuating information can reduce the use of stereotypes when making evaluations. Sinclair and Kunda (1999) demonstrated that when White student participants were motivated to see someone with positive regard, as little as 10 minutes of observation with Black individuals caused negative stereotypes to dissipate.

[5] Although the specific documents utilized in the study would not typically make up the interview protocol in a typical organization, direct conversations with a person with a high level of a Gay identity would be more likely to disclose the type of information contained in the documents (Button, 2001; Ragins & Marocchio, 2004).

[6] The effects of participant gender and sexual identity were examined and did not affect any of the consecutive analyses, thus the participants' gender and sexuality variables were collapsed.

APPENDICES

APPENDIX A: CHIEF EXECUTIVE OFFICER MEMORANDUM

Memorandum

To: All ACG Employees

CC: Mr. Phil Johnson

From: The Chief Executive Officer

Re: Hiring for the position of **Southeast Region Manager**

American Consulting Group (ACG)

As a valued employee of your input is extremely important as we consider candidates to fill the position of Southeast Region Manager. We have interviewed many qualified candidates, with Finance and Sales experience, who will lead this group. Currently about 88% of our Finance and Sales Managers company-wide are male.

Our upper management team has narrowed the field down to two extremely competitive candidates. Mr. Phil Johnson is the first of the two external candidates. **Importantly, the performance evaluations you provide will influence our final hiring decision.** If hired, you will be in close contact with Mr. Johnson who will serve as your supervisor, peer, or subordinate.

You will soon view a short video of a hypothetical "role-play" scenario posed to Mr. Johnson as part of his initial interview. **Mr. Johnson will appear on camera** and Brenda Mitchell (the subordinate/direct report) will be heard speaking off camera. It is important that you pay attention to and rate only Mr. Johnson's performance.

Your input is highly valued and I would like to personally thank you for your assistance in this important initiative.

APPENDIX B: EMPLOYEE INFORMATION FORM GAY CONDITION

American Consulting Group

Employee Information

Personal Information		

Full Name: **Johnson** / **Phil** / **R.**
Last / First / M.I.

Address: **2124 Cheshire Bridge Road** — **Apt 8310**
Street Address — Apartment Unit #

Atlanta — **GA** — **30324**
City — State — ZIP Code

Home Phone: **404-638-4776** — Alternate Phone: **( )**

E-mail Address: **pjohnson@gmail.com**

Social Security Number or Government ID: **xxx-xx-xxxx**

Birth Date: **xx/xx/xxxx** — Marital Status: **Domestic Partner**

Spouse's Name: **Mark Owens**

Spouse's Employer: **Corporate Accounting** — Spouse's Work Phone: **404-227-2232**

Emergency Contact Information		

Full Name: **Owens** / **Mark** /
Last / First / M.I.

Address: **2124 Cheshire Bridge Road** — **Apt 8310**
Street Address — Apartment Unit #

Atlanta — **GA** — **30324**
City — State — ZIP Code

Primary Phone: **(404-638-4776)** — Alternate Phone: **( )**

Relationship: **Domestic Partner**

APPENDIX C: EMPLOYEE INFORMATION STRAIGHT CONDITION

American Consulting Group

Employee Information

Personal Information			
Full Name:	Johnson	Phil	R.
	Last	First	M.I.
Address:	2124 Cheshire Bridge Road		Apt 8310
	Street Address		Apartment Unit #
	Atlanta	GA	30324
	City	State	ZIP Code

Home Phone:	404-638-4776	Alternate Phone:	( )
E-mail Address:	pjohnson@gmail.com		
Social Security Number or Government ID:		xxx-xx-xxxx	
Birth Date:	xx/xx/xxxx	Marital Status:	Engaged
Spouse's Name:	Mary Owens		
Spouse's Employer:	Corporate Accounting	Spouse's Work Phone:	404-227-2232

Emergency Contact Information			
Full Name:	Owens	Mary	
	Last	First	M.I.
Address:	2124 Cheshire Bridge Road		Apt 8310
			Apartment Unit #
	Street Address		
	Atlanta	GA	30324
	City	State	ZIP Code
Primary Phone:	(404-638-4776)	Alternate Phone:	( )
Relationship:	Engaged		

APPENDIX D: INTERVIEW SCHEDULE GAY CONDITION

American Consulting Group

Interview Schedule

Applicant
Name: Phil Johnson

Position Applied
for: Southeast Region Manager – Finance and Sales

Recruiter's   Recruiter's
Name: John Wright   Phone: (404) 587-6699

Start Time	End Time	Interviewer	Location
10:00am	10:30am	Mark Shae: Welcome	Roswell Office
10:30am	11:00pm	Elizabeth Porter: Finance	Roswell Office
11:00am	11:30am	Allen Hoffman: Sales	Roswell Office
11:30am	12:00pm	Assessment Center	Roswell Office
12:00pm	1:00pm	Lunch with Sales Force	Roswell Office

Comments

- Some graduate training
    - Has BS. in Finance

- Leadership Experience
    - Director of the Gay Men's Chorus
    - VP of the Atlanta United Way
    - Co-Chair of the Atlanta Gay Pride Parade

American Consulting Group
Roswell, Georgia

APPENDIX E: INTERVIEW SCHEDULE STRAIGHT CONDITION

American Consulting Group

Interview Schedule

Applicant Name: Phil Johnson

Position Applied for: Southeast Region Manager – Finance and Sales

Recruiter's Name: John Wright

Recruiter's Phone: (404) 587-6699

Start Time	End Time	Interviewer	Location
10:00am	10:30am	Mark Shae: Welcome	Roswell Office
10:30am	11:00pm	Elizabeth Porter: Finance	Roswell Office
11:00am	11:30am	Allen Hoffman: Sales	Roswell Office
11:30am	12:00pm	Assessment Center	Roswell Office
12:00pm	1:00pm	Lunch with Sales Force	Roswell Office

Comments

- Some graduate training
  - Has BS in Finance
- Leadership Experience
  - Director of the Atlanta Jazz Band
  - VP of the Atlanta United Way
  - Co-chair of the Atlanta St. Patricks Day Parade

American Consulting Group
Roswell, Georgia

APPENDIX F: CONSENT FORM: FIELD

Thanks for taking the time to participate in this survey-based study that should take only 25-30 minutes of your time. The title of this research is 'Evaluating the Performance of New Managers'. Specifically, you will view a short 10-minute video and then answer several questions. For each survey completed, you may choose from four different charities to donate 1.00$, which will be made by the principle investigator OR you may place your email address into a random drawing to receive a $50.00 American Express gift card. (Note: participation in the research is not required in order to enter the raffle.) In addition, by completing this web survey, you are making this important project possible.

Please be assured that your participation in this study will remain confidential. Any answers that you provide will not be traced back to you, and data collected on this website will be kept in a secured site. Although the site is secure, should you prefer an alternative means of completing the survey, you may print a copy of the survey from the website and mail the completed survey to the principle investigator: Dr. Kecia M. Thomas, or Mr. Brian Roote, Department of Psychology, The University of Georgia, Athens, GA 30602-3013.

This web survey is voluntary however, you may refuse to participate or stop taking part at anytime without penalty. By completing the web survey, you are agreeing to participate in this research. Please complete this web survey as soon as possible, but no later than two weeks from today in order to ensure that your response is included in this study. A follow-up e-mail message will be sent in one week as a reminder of this deadline. No discomfort or risks are foreseen in participating in this study.

We realize that your time is very valuable and thank you in advance for your help with this important study. If you have any questions or comments about this study, now or in the future, or if you would like to receive a copy of the survey results, please feel free to contact the principle investigator, Dr. Kecia M. Thomas or Mr. Brian Roote, Department of Psychology, The University of Georgia, at roote@uga.edu. Questions or concerns about your rights as a research participant should be directed to: The Chairperson, University of Georgia Institutional Review Board, 612 Boyd Graduate Studies Research Center, University of Georgia, Athens, GA 30602-7411. Telephone (706) 542-3199; e-mail address: irb@uga.edu.

Thanks again for your participation!

APPENDIX G: CONSENT FORM: LAB

Thanks for taking the time to participate in this survey-based study. The title of this research is 'Evaluating the Performance of New Managers'. Specifically, you will view a short 10-minute video and then answer several questions. By completing this web survey, you are making this important project possible.

Your participation in the study will help you gain experience at evaluating the performance of others. The researcher also hopes to learn more about the process of evaluating employees work performance in the real world setting.

As a participant of the study, you will receive one credit hour toward your research participation hours. You must provide your email address at the end of the survey in order to receive your credit hour. (PLEASE NOTE: Your e-mail address will not be connected with your responses). Please be assured that your participation in this study WILL REMAIN CONFIDENTIAL. Any answers that you provide will not be traced back to you, and data collected on this website will be kept in a secured site. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once the materials are received by the researcher, standard confidentiality procedures will be employed.

This web survey is voluntary however, you may refuse to participate or stop taking part at anytime without penalty or loss of benefits to which you are otherwise entitled. By completing the web survey, you are agreeing to participate in this research. No discomfort or risks are foreseen in participating in this study. However, in order to make this study a valid one, some information about my participation will be withheld until after the study.

We realize that your time is very valuable and thank you in advance for your help with this important study. If you have any questions or comments about this study, now or in the future, or if you would like to receive a copy of the survey results, please feel free to contact the principle investigator, Dr. Kecia M. Thomas or Mr. Brian Roote, Department of Psychology, The University of Georgia, at roote@uga.edu. Questions or concerns about your rights as a research participant should be directed to: The Chairperson, University of Georgia Institutional Review Board, 612 Boyd Graduate Studies Research Center, University of Georgia, Athens, GA 30602-7411. Telephone (706) 542-3199; e-mail address: irb@uga.edu.

Thanks again for your participation!

APPENDIX H: WORK PERFORMANCE MEASURES

1	2	3	4	5
Very Ineffective	**Ineffective**	**Satisfactory**	**Effective**	**Very Effective**

Leadership:
In the video, the manager….

1) Stated the goals and the purpose of the meeting.
2) Maintained control of the meeting.
3) Provided direction/redirected the discussion.
4) Asked for input from the direct report.
5) Established multiple agendas.
6) Articulated smooth transitions between topics.
7) Clarified roles.
8) Resisted the manipulations of the District Manager.
9) Attempted to motivate the District Manager.

Conflict Management
In the video, the manager….

1) Confronted the District Manager about her ideas.
2) Defended himself when challenged by the District Manager.
3) Corrected the District Manager.
4) Voiced a dissenting opinion.
5) Challenged the District Manager's ideas.
6) Asserted an uncommon or unpopular position.

Interpersonal Sensitivity
In the video, the manager…

1) Displayed attentive behaviors (e.g., eye contact, nodding).
2) Tried to establish rapport with the District Manager (make small talk).
3) Used humor.
4) Exchanged social pleasantries.
5) Acknowledged the contributions of the District Manager.
6) Did not interrupt the District Manager.

1	2	3	4	5
Very Ineffective	Ineffective	Satisfactory	Effective	Very Effective

Decisiveness
In the video, the manager…

1) Made specific recommendations.
2) Used emphatic speech.
3) Committed to a clear course of action.
4) Delineated clear action plans.
5) Strongly expressed beliefs.
6) Recognized the need for immediate action.

Analysis
In the video, the manager…

1) Searched for additional information by asking probing questions.
2) Recognized priorities among issues, material, and data.
3) Identified possible solutions.
4) Integrated information and ideas from different sources.
5) Correctly identified basic issues, including: data, facts, names/titles of people.
6) Correctly identified relationships among: data, people, and problems.

APPENDIX I: HIRING RECOMMENDATION

1	2	3	4	5
Strongly disagree	Disagree a little	Neither agree or disagree	Agree a little	Strongly agree

1. I would recommend Mr. Johnson for hire.

2. I am confident that Mr. Johnson can do the job.

3. If I were in charge of filling the Region Manager position, I would select Mr. Johnson.

4. Mr. Johnson is a very likeable applicant. (Note: This item was not used in the hiring recommendation composite.)

5. Overall, Mr. Johnson has a lot of potential to be a successful Region Manager.

APPENDIX J: GENDER CENTRALITY SCALE

1	2	3	4	5
Strongly disagree	**Disagree a little**	**Neither agree or disagree**	**Agree a little**	**Strongly agree**

1. I often think about the fact that I am a (wo)man

2. Overall, being a (wo)man has very little to do with how I feel about myself

3. In general, being a (wo)man is an important part of my self-image

4. The fact that I am a (wo)man rarely enters my mind

APPENDIX K: SOCIAL DESIRABILITY SHORT FORM

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you personally.

1. I'm always willing to admit it when I make a mistake. (T)

2. I always try to practice what I preach. (T)

3. I never resent being asked to return a favor. (T)

4. I have never been irked when people expressed ideas very different from my own. (T)

5. I have never deliberately said something that hurt someone's feelings. (F)

6. I like to gossip at times. (F)

7. There have been occasions when I took advantage of someone. (F)

8. I sometimes try to get even rather than forgive and forget. (F)

9. At times I have really insisted on having things my own way. (F)

10. There have been occasions when I felt like smashing things. (F)

APPENDIX L: TEN-ITEM PERSONALITY INVENTORY-(TIPI)

The Big Five Inventory (BFI)
Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1	2	3	4	5
Strongly disagree	Disagree a little	Neither agree or disagree	Agree a little	Strongly agree

I see myself as:

1. Practical
2. Rude
3. Warm
4. Sympathetic
5. Cold
6. Harsh
7. Efficient
8. Cooperative
9. Unsympathetic
10. Kind
11. Bashful
12. Withdrawn
13. Inefficient
14. Sloppy
15. Organized
16. Bold
17. Systematic
18. Extroverted
19. Disorganized
20. Careless
21. Creative
22. Uncreative
23. Complex
24. Philosophical
25. Jealous
26. Unintellectual
27. Temperamental
28. Deep
29. Intellectual
30. Imaginative
31. Fretful
32. Relaxed
33. Energetic
34. Shy
35. Moody
36. Talkative
37. Unenvious
38. Touchy
39. Envious
40. Quiet

APPENDIX M: RECALL FORM

Please answer the following questions to the best of your ability.

1. Whom did you rate on the various dimensions of performance (e.g., leadership, interpersonal sensitivity, conflict management, etc.)?
a. Mr. Johnson, the man I saw ON camera
b. Brenda, the subordinate/direct report I heard OFF camera

*2. Numerically speaking, most of the finance and sales manager's at the American Consulting Group are:
a. Female
b. Male
c. Not sure

3. On the Employee Information form, Mr. Johnson indicated that he was/had a:
a. Engaged
b. Domestic Partner
c. Girlfriend
d. Single

4. It appears Mr. Johnson's affinity orientation is:
a. Straight (Heterosexual)
b. Gay (Homosexual)
c. Bisexual
d. Not sure

5. Mr. Johnson belonged to which of the following professional organizations:
a. Co-Chair of the Atlanta Gay Pride Parade
b. Co-Chair of the Athens Habitat for Humanity
c. Co-Chair of the Atlanta NAACP
d. Co-Chair of the Atlanta St. Patrick's Day Parade

6. Have you ever seen or do you know the man appearing in the video (Mr. Johnson)?
a. I do not know the person in the video
b. I am familiar with or I know the person in the video
c. Although I have briefly seen the person in the video before, I do not know him personally

*Removed from analysis

APPENDIX N: DEMOGRAPHIC INFORMATION: LAB

Please tell us a little about yourself by responding to the questions below.

1. Gender
Male
Female

2. Race
Native American
Caucasian / White
Asian / Pacific Islander
Hispanic or Latino/a
African American / Black
Multi-Racial
Other:

3. What is the highest level of formal education that your Mother has completed?
a. Elementary/Middle School (grades 1-8)
b. High school or GED (General Equivalency Diploma)
c. Technical training or apprenticeship
d. Associate's degree
e. Bachelor's degree (e.g., B.A., B.S.)
f. Master's degree (e.g., M.S., M.Ed., M.A., M.B.A)
g. Professional degree (e.g., J.D., M.Div., M.D., D.V.M)
h. Doctoral degree (e.g., Ph.D., Ed.D.)

3. What is the highest level of formal education that your Father has completed?
a. Elementary/Middle School (grades 1-8)
b. High school or GED (General Equivalency Diploma)
c. Technical training or apprenticeship
d. Associate's degree
e. Bachelor's degree (e.g., B.A., B.S.)
f. Master's degree (e.g., M.S., M.Ed., M.A., M.B.A)
g. Professional degree (e.g., J.D., M.Div., M.D., D.V.M)
h. Doctoral degree (e.g., Ph.D., Ed.D.)

4. What is your age?

5. What is your sexual orientation / identity?
Bi
Lesbian or Gay (Homosexual)
Straight (Heterosexual)
Other

APPENDIX O: DEMOGRAPHIC INFORMATION: FIELD

1. Gender
Male
Female

2. Race
Native American
Caucasian / White
Asian / Pacific Islander
Hispanic or Latino/a
African American / Black
Multi-Racial
Other:

3. What is your age?

4. What is your affinity orientation / identity?
Bi
Lesbian or Gay (Homosexual)
Straight (Heterosexual)
Other

5. Indicate your employment status
Employed Full Time (at least 40 hours per week)
Employed Part Time (at least 20 hour per week)
Employed (under 20 hours per week)
Unemployed

6. What is your marital status?
Single (Never Married)
Married
Divorced or Separated
Widowed

7. What is the highest level of formal education that you have completed?
Elementary/Middle School (grades 1-8)
High school or GED (General Equivalency Diploma)
Technical training or apprenticeship
Associate's degree
Bachelor's degree (e.g., B.A., B.S.)
Master's degree (e.g., M.S., M.Ed., M.A., M.B.A)
Professional degree (e.g., J.D., M.Div., M.D., D.V.M)
Doctoral degree (e.g., Ph.D., Ed.D.)

8. What state do you live in? (If international, please provide the country)
Choose from 50 states or write in country.

9. What industry do you work in?
Advertising/Marketing/Public Relations
Arts/Entertainment/Media
Banking/Financial Services/Accounting/Auditing
Consulting Services
Education
Engineering
Government and Policy
Medical/Healthcare
Human Resources/Recruiting
Information Technology/Computers
Internet/e-Commerce
Legal
Non-profit
Publishing
Real Estate
Retail/Wholesale
Sales
Science/Biotechnology/Pharmaceuticals
Telecommunications

10. If you work in education, please indicate the level of your position.
Administrative/Support
College Level Professor (Asst, Assoc, Full)
Grade School Teacher (Elem, Middle, High)
Grade School Director (Principal, Vice Principal, Assist. Principal)
College Level Faculty/Administration (Dean, Assist. Dean, Vice President)

11. How many people work in your organization? Please estimate the approximate number of employees.
less than 100
101 - 499
500 - 999
1,000 - 10,000
over 10,001

12. How many years have you been in your current position?

13. How many years have you been with your organization?

14. What is your yearly individual salary (NOT household combined income)?
$10-19,999
$20-39,999
$40-59,999
$60-79,999
$80-99,999
$100K+

APPENDIX P: DONATE TO A CHARITY: FIELD

I would like the principle investigator to donate 1.00$ on my behalf to the following organization:

a. Act Up: The AIDS Cure Project
b. Juvenile Diabetes Research Foundation
c. The United Way
d. National Society for Hispanic Master of Business Administration (NSHMBA)
e. Cinderblock Foundation (helping underprivileged children)
g. National Breast Cancer Foundation, Inc.
h. Enter the $50.00 American Express card drawing. Note: please provide your email address (your email address will not be linked to your responses and will remain confidential)

APPENDIX Q: DEBRIEF: FIELD AND LAB

Thank you for taking part in the research study titled "Antecedents that Influence the Accuracy of Work Performance Evaluations" which is being conducted by Brian Roote, roote@uga.edu, under the direction of Dr. Kecia M. Thomas, Department of Psychology, University of Georgia (706-542-0057). The purpose of the study was to see how sexual orientation influences job performance evaluations in the workplace. Performance appraisals continue to serve as an administrative tool for making decisions regarding promotions, layoffs, separations, and transfers. In addition, they are sometimes criticized for being subjective in nature and likely vulnerable to personal biases. This research will aid future researchers in better understanding the role that sexual orientation can have on performance evaluations.

APPENDIX R: SOLICITATION EMAIL: FIELD

Dear Prospective Participant,

Have you ever wondered if the performance evaluations you make of your employees were accurate?

This study seeks to examine this link. By participating in the current study, you can obtain the study results and gain a better understanding of the evaluations you make and their general degree of accuracy. If interested in obtaining the results, please submit your email address at the end of the study.

The principal researcher, Brian Roote (PhD candidate at the University of Georgia) is seeking Managers/Educators employed at least half time. The study specifically examines the accuracy of the performance evaluations manager's make. Employed working Managers and/or Educators are invited to participate in this 35-minute online study. Specifically, participants will view a 10-minute video (embedded in the survey) and then answer several questions about themselves and the person appearing in the video.

At the studies completion, participant's will have the opportunity to select one of six charitable organizations that the principle investigator will donate 1.00$ on your behalf. Your information will remain confidential and the University of Georgia Institutional Review Board (Project # 2009-10118-1) has approved the study.

I would like to thank you for participating in this research project.

To participate, click on the "working professionals click here" link.
http://www.uga.edu/psychology/grads/research/broote/

Sincerely,
Brian Roote
Email:
roote@uga.edu

APPENDIX S: PILOT STUDY

The researcher conducted a pilot study to gather information on varying aspects of this current study. One hundred and seventy-five undergraduate students participated in a pilot-testing phase of the study. This pilot test helped to clarify my understanding of the specific manipulations in the study such as sexuality and performance level. In addition, the researcher asked the participants to guess the hypotheses of the study. These results initiated several appropriate changes to the stimuli and measures before data collection began.