

4-8-2003

## Gender Stereotypes of Citizenship Performance

Lisa Wilkinson  
*University of South Florida*

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

---

### Scholar Commons Citation

Wilkinson, Lisa, "Gender Stereotypes of Citizenship Performance" (2003). *USF Tampa Graduate Theses and Dissertations*.

<https://digitalcommons.usf.edu/etd/1507>

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact [digitalcommons@usf.edu](mailto:digitalcommons@usf.edu).

GENDER STEREOTYPES OF  
CITIZENSHIP PERFORMANCE

by

LISA WILKINSON

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts  
Department of Psychology  
College of Arts and Sciences  
University of South Florida

Major Professor: Tammy Allen, Ph.D.  
Louis Penner, Ph.D.  
Judith Bryant, Ph.D.

Date of Approval:  
April 8, 2003

Keywords: Organizational Citizenship Performance, Shifting Standards Model, Objective Scales, Reward Recommendations, Overall Performance

© Copyright 2003, Lisa Wilkinson

## Table of Contents

List of Tables	iii
List of Figures	v
Abstract	vi
Chapter 1 – Introduction	1
Citizenship Performance	3
Construct Development	3
Dimensions of Citizenship Performance	6
Consequences of Citizenship Performance	8
Stereotypes of Citizenship Performance	9
Shifting Standards Model	12
Overall Performance	17
Reward Recommendations	25
Citizenship Performance Dimensions	31
Chapter 2 – Method	32
Participants	32
Design	32
Procedure	33
Materials	34
Measures	38
Citizenship Performance	38
Overall Performance	39
Reward Recommendations	40
Chapter 3 – Results	41
Pilot Study	41
Manipulation Check	41
Order Effects	42
Rater Gender Effects	43
Descriptive Statistics	44
Hypothesis Testing	47
Exploratory Questions	53
Post Hoc Analysis	57

Chapter 4 – Discussion	60
Gender and Citizenship Performance	61
Overall Performance	64
Reward Recommendations	66
Citizenship Performance Dimensions	67
Post Hoc	67
General Limitations	68
Implications and Future Research	71
References	74
Appendices	85
Appendix A: Consent Form	86
Appendix B: Cover Story	88
Appendix C: Performance Log	89
Appendix D: Citizenship Performance Scale	91
Appendix E: Overall Performance Scale	94
Appendix F: Reward Recommendations	96
Appendix G: Demographic Information	98
Appendix H: Participants Debriefing	99

## List of Tables

Table 1	Distribution of Participants into 4 Conditions	32
Table 2	Perceived Number of Males and Females in 14 Occupations	34
Table 3	Descriptive Statistics (z-scores) by Gender and Scale Type	43
Table 4	Descriptive Statistics (raw) by Gender and Scale Type	43
Table 5	Zero-Order Correlation Coefficients Among Variables with Objective Citizenship Performance	45
Table 6	Zero-Order Correlation Coefficients Among Variables with Subjective Citizenship Performance	45
Table 7	Zero-Order Correlation Coefficients Among Variables with Objective Citizenship Performance, Split by Professor Gender	47
Table 8	Zero-Order Correlation Coefficients Among Variables with Subjective Citizenship Performance, Split by Professor Gender	47
Table 9	Hierarchical Regression Analysis with Objective Overall Performance as the Dependent Variable	48
Table 10	Hierarchical Regression Analysis with Subjective Overall Performance as the Dependent Variable	50
Table 11	Hierarchical Regression Analysis with Objective Reward Recommendations as the Dependent Variable	51
Table 12	Hierarchical Regression Analysis with Subjective Reward Recommendations as the Dependent Variable	52
Table 13.	Hierarchical Regression Analysis with Subjective Measures of Both Overall and Citizenship Performance.	53
Table 14.	Hierarchical Regression Analysis with Subjective Measures of Both Reward Recommendations and Citizenship Performance.	54
Table 15	Promax Oblique Factor Rotation Pattern Matrix	56

Table 16	Mean Factor Scores by Gender	56
Table 17	Outliers and their Influence on the Variables	57

## List of Figures

Figure 1.	Predicted Results for Hypothesis 1	17
Figure 2.	Predicted Results for Hypothesis 2a	23
Figure 3.	Predicted Results for Hypothesis 2b	24
Figure 4.	Predicted Results for Hypothesis 3	24
Figure 5.	Predicted Results for Hypothesis 4a	28
Figure 6.	Predicted Results for Hypothesis 4b	29
Figure 7.	Predicted Results for Hypothesis 5	30
Figure 8.	Scree Plot of Eigenvalues for Factor Analysis of Citizenship Performance	55
Figure 9.	Interaction Plotted According to Aiken and West's procedure	58

## Gender Stereotypes of Citizenship Performance

Lisa Wilkinson

### ABSTRACT

The relationship between citizenship performance and overall performance and the relationship between citizenship performance and reward recommendations were investigated, with gender and scale type as possible moderators. Two hundred and fifty-four University of South Florida students were used in this study. The majority of these participants were undergraduate, psychology majors, female, and between the ages of 17 and 23. Participants were given statements describing a teacher's performance and were asked to evaluate the professor on citizenship and overall performance and recommend them for rewards. No support was found for the hypothesis that men would have a stronger relationship than women between citizenship and overall performance. No support was found for the hypothesis that men would have a stronger relationship than women between citizenship performance and reward recommendations. Scale type was not found to influence these relationships. These results are not consistent with the shifting standards model. Numerous suggestions are made for changes to the experiment, including performing a field study instead of a lab study.



## Chapter 1

### Introduction

Relative to men, women have been historically undervalued in the workplace with regard to promotions (Rosen & Jerdee, 1973), compensation (Judge, Cable, Boudreau, & Bretz, 1995) and performance evaluations (Hamner, Kim, Baird, & Bigoness, 1974). Stereotypes of performance held by raters have been studied as a possible explanation for why women's contributions are devalued. Gender stereotypes (beliefs about male or female behaviors that cause generalizations to all members of that sex) can dictate raters' expectations of male and female behavior within the workplace. Consequently, these expectations may influence how raters evaluate performance and make promotion decisions.

The focus of the current study is to examine the consequences of gender stereotypes of citizenship performance for men and women. Citizenship performance represents behaviors that are important in shaping the work environment and in supporting task performance (Borman, Penner, Allen, & Motowidlo, 2001). There is some evidence to suggest that female stereotypes that characterize women to be unselfish and supportive (Eagly & Crowley, 1986) may place higher expectations on women to be organizational citizens (Allen & Rush, 2001). The shifting standards model provides a theory of how raters' stereotypes effect ratings (Biernat, Manis, & Nelson, 1991). The theory predicts that we judge people's behaviors based on expectations we have for the group (race, gender, etc.) they belong to and we shift our standards accordingly. For

example, when rating a woman's athleticism, raters will compare a female to other females, not to males. Therefore, even if the best female athlete is not as athletic as the average male athlete, she will still be rated as an above average athlete because raters will compare her to the lower athletic standard they have for women.

The shifting standards model provides the framework used in the present study for investigating whether men and women are equally evaluated and rewarded for citizenship performance. Considering that citizenship performance has been found to contribute variance to both performance evaluations (Allen & Rush, 1998; Borman, White, & Dorsey, 1995; Conway, 1999; MacKenzie, Podsakoff, & Fetter, 1991; Motowidlo & Van Scotter, 1994) and different types of organizational rewards (Allen, 2000; Allen & Rush, 1998; Chen & Heilman, 2001; Hui, Lam, & Law, 2000; Van Scotter et al., 2000), it seems important to study gender as a possible moderator of the relationships between citizenship performance and evaluations and rewards.

In what follows, a review on the development of the citizenship performance construct will be provided. Following the construct review, the development of dimensions and consequences of citizenship performance will be detailed. Next, there will be a discussion on gender stereotypes of citizenship performance. A detailed description of the shifting standards theory will follow and then a section on the relative contribution of the personal support dimension. Finally, the consequences of differential rewards and evaluations of citizenship performance will be discussed in the context of both performance evaluations and promotions.

### *Citizenship Performance*

*Construct development.* References of helping behaviors in the work place can be found as far back as 1939 in Chester Barnard's book, "The Functions of the Executive." In his book, Barnard described individuals who were responsible for giving the office a collective atmosphere (Barnard, 1939). The topic of altruistic behaviors in the workplace, however, was only sparsely mentioned throughout the literature for the next forty years (e.g., Katz, 1964; Roethlisberger & Dickson, 1939). In 1977, Organ resumed interest in helping behaviors in a discussion about the causal relationship between job performance and job satisfaction. In his article, Organ attempted to provide support for the dying theory that a satisfied employee is a productive employee. Organ claimed that if theorists considered a broader definition of performance, one that included behaviors that assist in the ease of organizational functioning, then the model of satisfaction causing performance would find more empirical and theoretical support. Some examples of the behaviors described by Organ are, arriving to work on time, not breaking the rules, and going along with company decisions and action without raising objections.

Beginning in the 1980's, discussion of citizenship behaviors began to permeate the organizational literature. Throughout its evolution, the citizenship performance construct has been given many titles and different variations on its definition. In 1983, Smith, Organ, and Near introduced the construct of Organizational Citizenship Behavior (OCB). The authors described OCB as a "myriad of acts of cooperation, helpfulness, suggestions, gestures of goodwill, altruism, and other instances of what we might call citizenship behavior" (p. 653). In 1988, Organ modified his original definition of OCB by defining it as a behavior that does not receive formal rewards, but that helps in the

functioning of the organization. As Organ worked on the construct of OCB, other authors developed related constructs.

In 1986, Brief and Motowidlo introduced a similar construct called Prosocial Organizational Behaviors (POB). These authors defined the construct as “positive social acts carried out to produce and maintain the well-being and integrity of others” (p. 710). In their article, Brief and Motowidlo examined different ways to break the concept of prosocial behaviors into parts and came up with the in-role and extra-role distinction. In-role behaviors are defined as behaviors that are written in an employee’s job requirements, whereas extra-role behaviors are defined as acts that are not included in an employee’s job tasks or duties. However, Morrison (1994) found the in-role and extra-role division to be overlapping and hard to distinguish. In a study of 317 clerical workers, Morrison found that employees were more likely to perform OCB if they defined the behaviors as in-role rather than extra-role. They also found that employees and supervisors didn’t agree, above chance, on what behaviors would be defined as extra-role and which would be defined as in-role. Subsequently, Borman and Motowidlo (1997) derived another distinction that avoided the problems of in-role versus extra-role by focusing on whether the behavior represented a “core” task.

Borman and Motowidlo (1997) defined contextual performance as behaviors that improve or maintain the environment of the organization. Behaviors that typify contextual performance are important because they surround and support the core tasks being performed. The term task performance is used in contrast to contextual performance and represents behaviors that have a direct contribution to the productivity of the company. Since job requirements vary between jobs and companies, this

distinction can be easily generalized across situations. Results from Motowidlo and Van Scotter (1994) supported the contextual performance and task performance distinction, showing that each type of performance contributed uniquely to performance evaluations. Task performance contributed 13% of the variance in overall performance and contextual performance contributed 11% above the variance accounted for by task performance. Further, the researchers found that the two constructs had their own unique antecedents. Van Scotter, Motowidlo, and Cross (2000) similarly found support for the division of performance into task and contextual performance. In this study, the authors were interested in accounting for the variance found in systematic rewards. The authors found that the two different types of performance each differentially predicted variance in medals, promotability ratings, rewards, and rank.

In light of the recent research on citizenship performance, Organ (1997) wrote a review in order to, in part, consolidate the different terms for helping behaviors into one concept and one definition. Organ agreed with the usefulness of the contextual and task performance distinction, but argued that the name, contextual performance, does not provide the reader with ready knowledge of what the concept means. He described the concept of contextual performance as “cold, gray, and bloodless” (p. 91).

In a recent article by Borman et al. (2001), the authors use the term citizenship performance in place of the term contextual performance. Citizenship performance is defined as behaviors that contribute positively to the work place environment and enhance workers' task performance. Citizenship performance takes on the same definition as contextual performance, but its meaning can more readily be understood from the name alone than contextual performance. Examples of these behaviors are

assisting co-workers with their jobs, following the rules, being friendly and having a positive attitude, staying late to finish one's work, showing support of the organization, and giving extra effort on the job. In light of the empirical findings supporting the contextual and task distinction, the definition of citizenship performance given will be used in the present study.

*Dimensions of citizenship performance.* Beyond the problems of naming the construct citizenship performance, there have been a number of different dimensions proposed. Smith et al. (1983) used interviews with managers from different organizations to create 16-items that operationalized citizenship behaviors. The managers or supervisors were asked to identify behaviors that contributed to the organization, but that were not explicit requirements of the job. These 16 items were factor analyzed and loaded on two factors: altruism and generalized compliance. Altruism was defined as help given to other individuals in the organization. Generalized compliance was defined as “a more impersonal sort of conscientiousness, more of a ‘good soldier’ or ‘good citizen’ syndrome of doing things that are ‘right and proper’ but for the sake of the system rather than for specific persons” (p. 662). This two-factor model has also been referred to as OCBO and OCBI (e.g., Skarlicki & Latham, 1996). The distinction between OCBO and OCBI lies in the object to which the behavior is being directed. OCBO is organizational citizenship behavior directed at the organization and OCBI is OCB directed at the employee's co-workers.

Organ (1988) proposed 3 new dimensions to add to the previous 2-dimension model of altruism and generalized compliance: sportsmanship, civic virtue, and courtesy. Organ felt that these five dimensions more adequately covered the entire breadth of the

OCB concept. The following are definitions of these constructs according to Organ (1988). Sportsmanship is the willingness of an employee to ignore the small problems that arise and not complain. Civic virtue is when an employee takes personal responsibility for how the organization performs. Courtesy is the amount one tries to ameliorate the situation when conflict arises between co-workers or personal problems develop for co-workers.

Podsakoff and Mackenzie (1997) used a 3-dimension model and introduced the new dimension of helping behavior. The authors felt that raters in their studies were unable to distinguish between the altruism and courtesy dimensions provided by Organ. The authors chose to use the term helping behaviors to encompass both altruism and courtesies. They define helping behaviors as behaviors that help co-workers solve work-related dilemmas (Podsakoff & MacKenzie, 1994).

Recently, Coleman and Borman (2000) conducted a study for the purpose of bringing together the different dimensions of OCB used throughout the literature and conducting several types of analyses to find the best factor structure. These authors found the most consistent and successful model to be a 3-factor model. The three dimensions of this model were personal support, organizational support, and conscientious initiative. Personal support encompasses the previous concepts of altruism and helping behaviors and involves assisting co-workers with problems, being considerate of co-workers needs, and cooperating with co-workers. Organizational support includes the earlier concept of compliance and represents behaviors that demonstrate support for the organization, both by following the rules and making one's organizational commitment evident with fellow co-workers and people outside the

company. Conscientious initiative involves a desire and perseverance to fulfill one's job duties and create the best opportunities for self and company. The current study will utilize the three dimensions (personal support, conscientious initiative, and organizational support) derived by Borman and Coleman (2000).

*Consequences of citizenship behavior.* One reason for the growing research on citizenship performance in the past decade is the consequences it holds for employees. For example, it has consistently been found that employees are evaluated for their citizenship performance as well as for their task performance (Allen & Rush, 1998; Borman, White, & Dorsey, 1995; Conway, 1999; MacKenzie, Podsakoff, & Fetter, 1991; Motowidlo & Van Scotter, 1994). More specifically, organizational citizens receive higher performance appraisals than do employees who don't perform citizenship behaviors. For instance, MacKenzie et al. (1991) found that OCB accounted for about 30% of the variance in manager's performance evaluations of salespeople. Podsakoff and MacKenzie (1994) found similar results in a field study of insurance salesmen. These researchers found that OCB accounted for 48% of the variance in employee evaluations.

Further demonstration of the impact of citizenship performance on employees is illustrated with findings that supervisors consider citizenship performance when making various reward recommendations (Allen, 2000; Allen & Rush, 1998; Chen & Heilman, 2001; Hui, Lam, & Law, 2000; Van Scotter, 2000; Van Scotter et al., 2000). In a study of military personnel, Van Scotter et al. (2000) found that contextual performance predicted promotability ratings above the variance explained by task performance. Further, the authors found that contextual performance explained variance in informal rewards, whereas task performance did not. Informal rewards are rewards given to



employees that are not recorded in an employee's personnel file. Examples of these rewards include special assignments, new positions, training, and aiding in career advancement. Similarly, in a field study, Allen (2000) tested the relationships between OCB and promotions and OCB and salary. The results showed that both salaries and promotions were significantly correlated with the amount of OCB exhibited by the employee.

Allen and Rush (1998) used a five-item measure of reward recommendations that was created to reflect common organizational rewards. The five items were increases in salary, promotions, public recognition, high profile project, and opportunities for professional development. The researchers found that OCB correlated with both reward recommendations and performance evaluations. The growing evidence demonstrating the importance of citizenship performance to individual evaluations and rewards, underscores the need to accurately evaluate citizenship performance. For example, it has been found, in the past, that stereotypes held by the raters can impact the accuracy of overall performance evaluations (Dobbins, Cardy, & Truxillo, 1988).

*Stereotypes of citizenship performance.* Stereotypes can have a crucial impact on evaluations of employee performance in the workplace. They can be especially harmful because as Heilman (1995) states, once we have associated particular behaviors with a group of people, we generalize that behavior to all individual group members. For example, Dobbins et al. (1988) found that raters with more traditional gender stereotypes rated women less favorably than men on overall performance evaluations. In her review on the effects of sex stereotypes in the workplace, Heilman describes four factors that work to maintain and reinforce stereotypes. These four influences are perceptions,

interpretation, memory, and inferences. First, perceptions are affected when we focus our attention on information that is consistent with stereotypes and ignore inconsistent information. For example, observers are more likely to recognize a man that is acting aggressively than a woman that is acting aggressively. Second, people interpret information in a way that is consistent with stereotypes. For instance, if a woman is observed acting aggressively, the witness may interpret a woman's motives differently from a man's motives. The viewer may assume that the woman was provoked but assume that the man started the conflict.

Third, people tend to remember what is consistent with previous stereotypes or even remember events that did not occur because they are consistent with stereotypes. For example, when someone witnesses a female acting aggressively toward a male, the observer might falsely remember that the man provoked the attack. Finally, inferences are made when there is little or no information given about an individual. When lacking information people tend to rely on shortcuts based on superficial characteristics of someone. For example, when meeting a man and a woman for the first time, it might be assumed that the man is more aggressive than the female, based on their sex alone.

Resilience of stereotypes was demonstrated in a study by Nelson, Biernat, and Manis (1990), in which the researchers attempted to reduce various gender stereotypes. In an effort to increase accuracy of judgments, participants were placed in one of three groups. They were either told the truth that there were no gender differences between the rates, were given monetary incentives for accuracy, or were educated on gender stereotypes. Only in the condition where the participants were told the truth, that men and women rates were matched for height, was there a significant decrease of stereotype

effects. In this condition men were still rated significantly taller than were women, however, there was a significant decrease in the rating differences found between men and women. The cash incentives and the training on stereotypes had no significant impact on the ratings made by the participants.

In a business setting, the persistence of the stereotypes held by supervisors could have implications for how they view their subordinates' performance. Allen and Rush (2001) found support for the theory that people possess gender stereotypes of citizenship performance. In this study, participants were given a list of 10 citizenship behaviors and 10 task behaviors, and were asked to determine how likely a male or a female employee would be to perform the behavior, to predict the percentage of males or females who would perform these behaviors, and to provide an expected salary for the job. Results showed that expectations for ratees to engage in citizenship behavior were greater for women than for men.

Allen and Rush's (2001) findings were not surprising considering the past findings on gender and helping behavior stereotypes. In their meta-analysis on helping behaviors Eagly and Crowley (1986) discuss common stereotypes that are attributed to women regarding their altruistic or social role. The authors assert that "women are expected to care for the personal and emotional needs of others, to deliver routine forms of personal service, and, more generally, to facilitate the progress of others toward their goals" (p. 284). The results supported the contention that thoughtfulness and nurturing traits are considered female characteristics. Similarly, Eagly, Makhijani and Klonsky (1992) described women as having "communal qualities, such as being friendly, unselfish, concerned with others, and emotionally expressive" (p. 6). These behaviors are

similar to the behaviors of an organizational citizen. Examples of citizenship behaviors that resemble female descriptions are helping co-workers when they need emotional support, demonstrating consideration for co-workers, encouraging co-workers successes, and putting the needs of the team first.

The current research is an attempt to extend the findings of Allen and Rush by investigating if women and men are rated the same for their citizenship performance. The shifting standards model, introduced in the following section, provides a theory of how gender expectancies can create differential evaluation of men's and women's citizenship performance.

#### *Shifting Standards Model*

The shifting standards model, borne out of the social psychology literature, describes how stereotypes can be hidden when a particular type of measurement is used. In the previously described study by Nelson et al. (1990), the authors developed the shifting standards model while studying the resilience of the stereotype that men are taller than are women. The authors found that men were consistently rated taller than were women, even though height was controlled and there were no actual differences, on average, between men and women.

Nelson et al. (1990) attempted to explain the large and consistent difference found between men and women on height. The authors suggested that scale type could have an impact on how raters make differential ratings for men and women. Feet and inches were used to measure height in this particular study. The authors claimed that objective scales “have a special virtue, in that there is universal agreement that a man of 5’7” and a women of 5’7” are in fact equal in height” (p. 673). Objective scales are familiar and

quantifiable and therefore provide a more accurate description of rater stereotypes. When height is provided in feet and inches, there is no need for interpretation because the meaning of 5'7" is the same for all raters.

The objective scale, just described, was considered in contrast to the Likert scale. When a Likert scale is used, Nelson et al. (1990) hypothesized that the endpoints of the scale could have a different meaning depending on the rater's comparison group. Likert or continuum scales were considered subjective, because the meaning of these ratings could vary depending on the rater's standard of comparison. For example, there is a common stereotype that men are better at math than women. Due to this stereotype, when raters judge a woman to be an above average math performer, this judgment may not be equivalent to a man receiving an above average rating. Because of the stereotype that men are better mathematicians than are women, male performance may be judged against a higher standard. Therefore, the rater shifts the standard down when rating a female as compared to when rating a male. Consequently, since women are compared against a lower standard, they would not need the same math expertise in order to receive an equivalent evaluation, as would men.

These different rating patterns, based on scale type, led to the development of the shifting standards theory. The premise of the theory is that when subjective ratings are used, a rater's standard of comparison is unknown. However, when objective scales are used, raters are forced to compare the men to the women.

In a test of the shifting standards theory, Biernat, Manis, and Nelson (1991) examined ratings of height, weight, and income across gender. Participants looked at 44 pictures and were asked to judge the height, weight, and income of the person in the

photograph. The participants rated (subjective scale) and ranked (objective scale) the people in the pictures. Specifically, participants rated the pictures on a Likert scale with seven choices (e.g., very tall to very short) and ranked them in order from most to least (e.g., tallest to shortest). The authors expected to find that when using the Likert scale, there would be less evidence of a stereotype influence on ratings. However, the results were not the same for the three independent variables. For height and weight, men were rated significantly taller and heavier than were women on both subjective and objective measures. When the subjects used objective measures to rate the people in the pictures, the difference between men and women was larger than when rating on the subjective scale.

For income, the results demonstrated, what the authors refer to as a reversed stereotyping effect. When rating on a subjective scale of financial success, women were rated as more financially successful than were men, even though the only information provided was a picture. When providing salary amounts using an objective measure, raters gave men higher salaries than they gave to women. A woman who makes \$40,000 a year is considered successful. However, a man who earns \$40,000 may receive an average rating because the raters are comparing the man to the standard they have for men, which is higher than the standard they have for women. That is to say, even though the man and the woman have the same salary, the value is being rated against a different standard depending upon the sex of the ratee. These findings support the shifting standards model, because with the subjective ratings, "for a man to be labeled financially very successful, he had to earn much more money than a woman who was similarly labeled" (p. 5).

Biernat and Manis (1994) tested the shifting standards theory with more variables including aggressiveness, assertiveness, verbal ability, and mathematical ability. The researchers found similar results across all stereotyped variables (Beirnat & Manis, 1994). Recently, the shifting standards model was introduced into the industrial/organizational psychology literature by Martel and DeSmet (2000). The authors studied gender stereotypes in leadership. In order to avoid subjective scales that might mask stereotypes, the authors asked participants to rank the ratees. The study was designed for participants to identify the abilities that are expected of a good leader and to discover if people have the same expectations of male and female leaders. Further, the authors wanted to determine if gender stereotypes are the reason why women are not promoted into the managerial positions at equal rates as men. Martel and DeSmet found that some behaviors were not subjected to gender stereotyping, but that there were several behaviors that were considered to be more characteristically male or female. For example, behaviors such as mentoring, supporting, and rewarding were considered to be more likely performed by women. Behaviors considered characteristic of males were delegating and upward influence. According to the shifting standards model, these stereotyped behaviors were evident because an objective scale was used that forced the raters to compare the ratee to everyone, not just the members of the ratee's gender.

The proposed study will test the applicability of the shifting standards model to gender stereotypes of citizenship performance. The stereotype that women are more likely than men to perform as organizational citizens is expected to increase expectations for women to perform citizenship behaviors. In their meta-analysis, Organ and Ryan (1995) found no evidence that gender related to ratings of OCB ( $r(1300)=.03$ , ns).

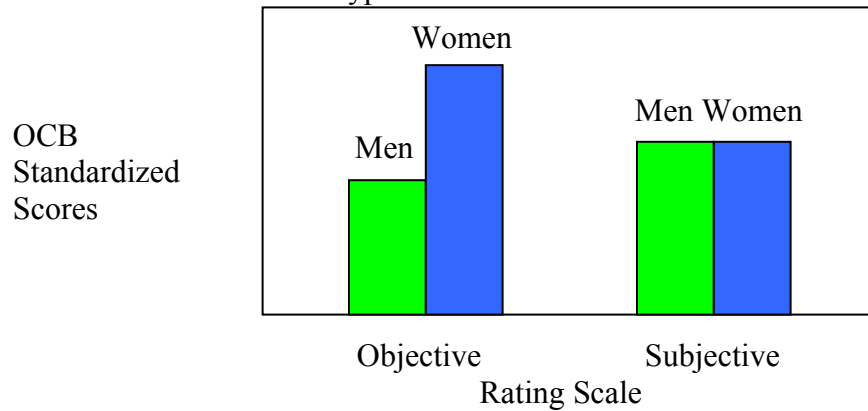
Podsakoff, Mackenzie, Paine, and Bachrach (2000) discussed researchers' surprise with the null findings from the meta-analysis. They provided examples of characteristics of citizenship performance that share features with stereotypes of men and women. Further, most OCB measures are Likert scales (subjective). It is reasonable that raters could be using a different reference group when rating men than when rating women, which would mask perceived differences between men and women. Therefore, even though women may be perceived to be greater organizational citizens than men, they will not receive higher ratings than will men on subjective scales.

For the present study, it is predicted that when rating citizenship performance on an objective scale, raters will compare men and women and will consequently rate women higher because they are expected to be better organizational citizens. However, when using subjective scales, raters are expected to compare women against other women, thereby, hiding the gender stereotype. As a consequence, when rating on a subjective scale, there will be no significant difference between men and women on citizenship performance. Figure 1 displays the expected results for Hypothesis 1.

*Hypothesis 1:* When participants rate employees on a subjective citizenship performance scale there will be no significant difference between men and women. However, when participants rate employees using an objective citizenship performance scale, women will be given significantly higher ratings than will men.



Figure 1. Predicted Results for Hypothesis 1.



### *Overall Performance Evaluation*

In an organizational setting, it is important to understand the way citizenship behavior may affect overall performance appraisals and how the relationship may differ across gender. Research studying gender bias in performance evaluation has yielded inconsistent results (Landy & Farr, 1980). Studies have produced results supporting male favored bias, (Hamner et al., 1974), female favored bias (Atwater & Van Fleet, 1997), and no bias (Schwab & Grams, 1985). Considering that performance is generally measured using subjective scales, the shifting standards theory would postulate that men and women are not being rated against each other, but against members of their own group. In other words, women are rated relative to other women and men relative to other men. According to the shifting standards model, in order to explore raters' stereotypes and biases, an objective scale should be used (Biernat et al., 1991).

Biernat, Crandall, Young, Kobrynowicz, and Halpin (1998) tested the shifting standards model using peer and self-ratings of officer performance in a leadership-training course. For the objective scale, the participants ranked their groupmates from most capable to least capable. According to the shifting standards model, rankings are objective because raters must compare the target groups, in this case men and women.

For the subjective scale, subjects rated their groupmates on a Likert 5-point scale of capability from “excellent” to “needs much improvement.” Men were rated higher than women were on both rankings and the Likert scales. However, when subjects ranked their peers, the difference between men and women was greater than when a Likert scale was used. In other words, when raters were forced to compare men and women by ranking them, sex differences were even more pronounced. Further, these findings provide support for the application of the shifting standards model in the realm of performance evaluations.

These findings also support the hypothesis that when objective scales are used to evaluate overall performance, the difference raters perceive between men and women are more apparent. An important consideration, for the proposed study, is how citizenship performance influences overall performance evaluations. As discussed earlier, both task performance and citizenship performance have been found to contribute to the variance associated with overall ratings of performance (e.g. Allen & Rush, 1998). Therefore, gender stereotypes of citizenship performance could subsequently impact overall performance evaluations. As will be described in the following paragraphs, only a few studies (Allen & Rush, 2001; Chen & Heilman, 2001; Lovell, et al., 1999) have attempted to investigate how gender influences the relationship between citizenship performance and overall performance evaluations.

Chen and Heilman (2001) tested differences in performance evaluations and reward recommendations for men and women while manipulating the level of OCB in the ratee performance descriptions provided. Vignettes were used to describe the performance of either a male or a female employee, who either performed OCB, who

chose not to perform OCB, or who only exhibited task behaviors. When subjects were rating vignettes with only task performance information, no significant differences were found between men and women. However, in the two groups where the ratee either performed OCB or did not perform OCB, men were rated significantly higher than were women on both overall performance evaluations and reward recommendations. The results also showed that the female ratees were rated significantly lower when they chose not to perform OCB than when only task behaviors were provided, but no differences were found for men. Furthermore, women who did perform OCB were not rated significantly higher than were women with no OCB information provided. This was in contrast to men who were rated higher when performing OCB than when no OCB information was provided. In brief, men were positively evaluated for performing OCB and women were given lower evaluations for not performing OCB.

Research has consistently demonstrated that citizenship performance contributes to overall performance appraisals (e.g., Allen & Rush, 1998). In the Chen and Heilman (2001) study, citizenship performance did not contribute to overall performance ratings the same way for men as for women. The notion that women engage in citizenship behavior more frequently than do men is predicted to create higher expectations for women to perform citizenship behaviors relative to men. As a consequence, women may need to perform more citizenship performance in order to be equally recognized for their citizenship performance. In order to understand how the citizenship performance ratings contribute to overall performance differently for men and for women, both overall and citizenship performances need to be evaluated. Further, Chen and Heilman used subjective ratings to evaluate overall performance. As uncovered by Biernat, Manis, and

Nelson (1991), when objective ratings are used, a rater's comparison group is more easily distinguished than when subjective ratings are used. Even though Chen and Heilman found a significant difference between men and women on overall performance, if an objective scale had been used the difference may have been greater.

Lovell et al. (1999) tested the effect of gender on the relationship between OCB and performance evaluations using resident advisors (RA) in college dorms. A survey of the five dimensions of OCB provided by Organ (1988) was created by the researchers and then factor analyzed. The factor analysis produced a three-factor solution of altruism, sportsmanship, and mediation. The experimenters asked the RAs to rate each other on OCB and asked the dorm directors to rate the RAs on overall performance. There was a significant correlation of .38 found between OCB and performance evaluations. In other words, higher levels of OCB were associated with higher scores on performance appraisals. Further, women were given marginally higher OCB ratings than were men ( $p = .063$ ). However, even though women received higher OCB ratings than did men, there were no differences found between women and men on overall performance evaluations. Although the pattern of results concerning overall OCB was not statistically significant, they suggest that the effect of gender on ratings of citizenship performance is worth further investigation.

A weakness of Lovell et al.'s (1999) study is that no task ratings were provided (i.e., the men could have been significantly better task performers than were the women). True levels of citizenship performance were also not available. The proposed study is designed to address this limitation by controlling for task and citizenship performance in a laboratory setting. Further, subjective ratings were used to measure both OCB and

overall performance. If the experimenters had used an objective scale, it is possible that they would have found a significant difference between men and women. Lastly, peers provided the ratings of citizenship performance and supervisors provided the overall performance ratings. Consequently, there is no way of knowing if the supervisors and peers perceive the same behaviors and if the supervisor would have given the resident advisors the same ratings on OCB as did the peers.

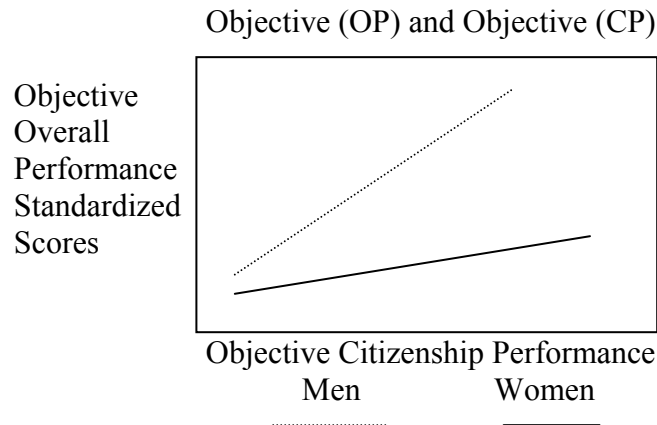
Allen and Rush (2001) provided a third study that looked at gender, OCB, and overall performance evaluations. The authors manipulated levels of task performance and OCB and manipulated the gender of the ratee. Results indicated no significant differences on ratings of overall performance or reward recommendations between men and women. However, for both dependent measures subjective scales were used. This could be the reason why there were no significant differences found. Perhaps raters were rating women against women and men against men as predicted by the shifting standards theory. In the present study, objective scales will be used in order to see how raters compare men and women without masking perceived differences with a subjective scale. Further, participants will evaluate the ratee's citizenship performance as well as overall performance, in order to look for a moderating effect of gender.

In two of the three studies described in this section (Lovell et al., 1999; Chen & Heilman, 2001), men and women were not equally evaluated for performing citizenship behaviors. The current study plans to add to this line of research by using an objective scale as well as a subjective scale to measure overall performance. In the previous hypothesis, citizenship performance was measured using subjective and objective scales. Therefore, the following hypotheses will be divided by the type of scale used to measure

citizenship performance to better understand under what conditions gender differences may emerge. First, when an objective scale is used to measure overall performance, gender is expected to moderate the relationship between citizenship performance and overall performance. The relationship between citizenship performance and overall performance is expected to be greater for men than for women. When we break down the results by the scale used to measure citizenship performance, the nature of the interaction is expected to vary depending on the scale used. When an objective scale is used to measure citizenship performance, the relationship between citizenship performance and overall performance is predicted to be weakest for women and the strongest for the men. Figure 2 displays the expected results of Hypothesis 2a.

*Hypothesis 2a:* Gender is predicted to moderate the relationship between citizenship performance and overall performance when both are measured using an objective scale. It is predicted that citizenship performed will be more highly related with overall performance for men than it will for women. The difference for men and women, in the strength of the relation between citizenship and overall performance, is proposed to be greatest when both types of performance are measured using objective scales.

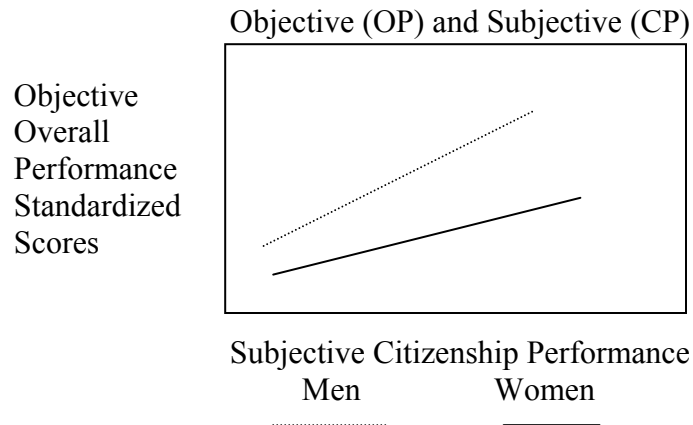
Figure 2. Predicted Results for Hypothesis 2a.



When overall performance is measured using an objective scale and citizenship performance is measured using a subjective scale, gender is expected to moderate the relationship between the two types of performance. However, the difference in the strength of the relationship between overall and citizenship performance for men and for women is not expected to be as great as when objective scales are used to measure both types of performance. Figure 3 displays the expected results of Hypothesis 2b.

*Hypothesis 2b:* Gender is predicted to moderate the relationship between citizenship performance and overall performance when citizenship performance is measured using a subjective scale and overall performance is measured using an objective scale. The relationship between citizenship and overall performance is expected to be stronger for men than it is for women.

Figure 3. Predicted Results for Hypothesis 2b.

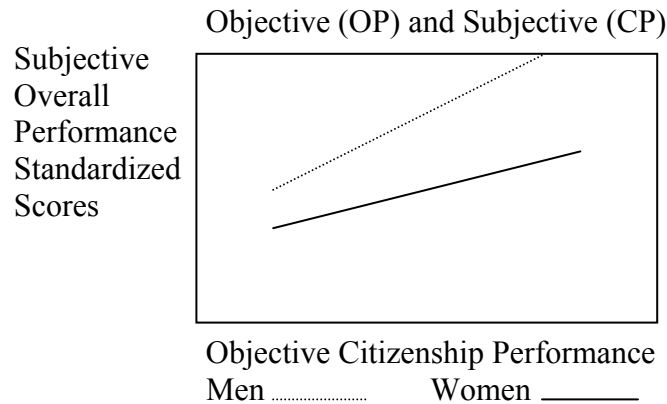


Next, how gender influences the relationship between citizenship and overall performance when a subjective measure of overall performance is used will be discussed. When citizenship performance is measured using an objective scale and overall performance is measured using a subjective scale, gender is predicted to moderate the relationship between the two types of performance. As predicted with Hypothesis 2b, citizenship performance for men is expected to have a stronger relationship with overall performance than it does for women. Figure 4 displays the expected results for Hypothesis 3.

*Hypothesis 3:* When citizenship performance is measured using an objective scale and overall performance is measured using a subjective one, gender is predicted to moderate the relationship. Citizenship performance is predicted to have a stronger relationship with overall performance for men than it does for women. The difference between men and women in this case is not predicted to be as great as when both types of performance are measured using objective scales.



Figure 4. Predicted Results for Hypothesis 3.



When both citizenship performance and overall performance are measured using a subjective scale, gender is not expected to impact the relationship between citizenship performance and overall performance. Therefore, the following research question will be an exploratory one.

*Exploratory question 1:* Does gender moderate the relationship between citizenship performance and overall performance when both are measured using subjective scales?

#### *Reward Recommendations*

Also of significance is the question of how citizenship performance ratings relate to reward recommendations. As described earlier, Allen and Rush (1998) used the concept reward recommendations to encompass five organizational rewards: salary, promotions, public recognition, high profile project, and opportunities for professional development. Citizenship performance has been found to contribute to reward recommendations (e.g., Allen & Rush, 1998), as well as to promotions (Hui, et al., 2000), informal rewards (Van Scotter et al., 2000), salary (Allen, 2000), and recommendations

for a fast-track development program (Kiker & Motowidlo, 1999). However, only a few studies (Allen, 2000; Allen & Rush, 2001; Chen & Heilman, 2001) have looked at the influence of gender on the relationship between citizenship performance and organizational rewards.

The previously discussed articles by Chen and Heilman (2001) and Allen and Rush (2001) tested the effect of gender on the relationship between OCB and reward recommendations. Even though these two studies used a similar dependent measure, their results were quite different. Allen and Rush (2001) found no effect of gender on reward recommendations, whereas, Chen and Heilman (2001) found that men were rated significantly higher on reward recommendations than were women. Both of these studies used subjective scales. Chen and Heilman (2001) used a different design than did Allen and Rush (2001). Chen and Heilman provided participants with quantitative values of the ratees task performance and wrote a paragraph describing a time when the ratee either performed OCB or decided not to perform OCB. Therefore, the written, story style of the OCB information may have been more salient than the task information. The salience of the OCB information may have given it a larger impact on the ratings of overall performance than was the case with the study by Allen and Rush (2001) who provided both task behaviors and OCB together in a videotape of job performance. Providing task and citizenship behaviors together seems to be more applicable to an organizational setting. These study design differences may explain the discrepant results between the two studies and will be addressed in the following way, task and citizenship performance will be provided together in a performance log in an attempt to capture a more realistic

setting. Further, both subjective and objective measures will be used, in order to more accurately gauge raters' perceptions of men and women (Biernat et al., 1998).

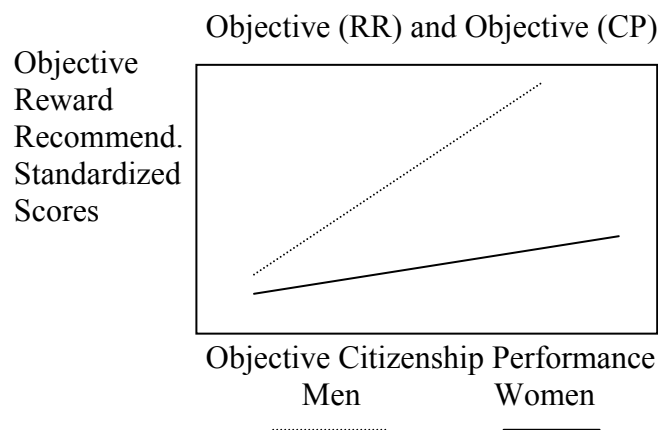
Allen (2000) tested the hypotheses that gender moderates the relationship between OCB and promotions and OCB and salary. In a field study in which self-reports of OCB and promotions were used, Allen found that OCB was correlated with employee salary and number of promotions. More importantly, gender moderated the relationship between OCB and promotions, but not OCB and salary. The correlation for men between OCB and promotions was .21 and between OCB and salary was .20. For women, neither type of OCB significantly correlated with promotions. Objective measures of promotions were used in this study by asking the participants to include the number of promotions they have received to date. This study supports the claim that women are not rewarded for their citizenship behavior to the same extent as are men and further supports the use of objective scales to identify these differences. One feature of the Allen study that should be noted is that self-ratings were used for OCB, therefore, the rater evaluating OCB did not make the salary or promotion decisions. To better understand how OCB correlates with reward recommendations differently for men than for women, the same rater needs to be used for both OCB and reward recommendations. This is important because the rater recommending rewards could have perceived the employee's OCB level differently than the rater providing information about the employee's OCB. The present study will ask participants to rate the employee on citizenship performance and also decide on reward recommendations for the employee.

It is predicted that when raters are using a subjective scale to measure reward recommendations, there will be no difference between men and women, but when using

an objective scale, there will be a significant difference. As with Hypothesis 2, the following hypotheses will be divided based on the scale type used to measure citizenship performance and reward recommendations. First, when reward recommendations and citizenship performance are measured with objective scales, gender is expected to moderate the relationship with the relationship being stronger for men than for women. Further, when both reward recommendations and citizenship performance are measured with objective scales, the difference between men and women is predicted to be greatest. Figure 5 displays the expected results for Hypothesis 4a.

*Hypothesis 4a:* Gender is predicted to moderate the relationship between citizenship performance and reward recommendations when both are measured with objective scales. The relationship between reward recommendations and citizenship performance is predicted to be stronger for men than for women.

Figure 5. Predicted results for Hypothesis 4a.

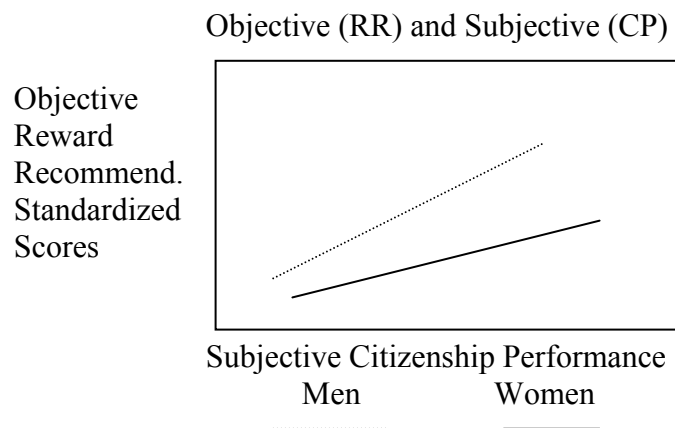


Next, the effects of gender when reward recommendations are measured using an objective scale and citizenship performance is measured using a subjective scale, will be discussed. Gender is expected to moderate this relationship, however, the difference

between men and women is not expected to be as great as when objective measures are used for both reward recommendations and citizenship performance. Figure 6 displays the expected results for Hypothesis 4b.

*Hypothesis 4b:* When an objective scale is used to measure reward recommendations and a subjective scale is used to measure citizenship performance, gender is predicted to moderate the relationship between reward recommendations and citizenship performance. The relationship between citizenship performance and reward recommendations is predicted to be greater for men than for women.

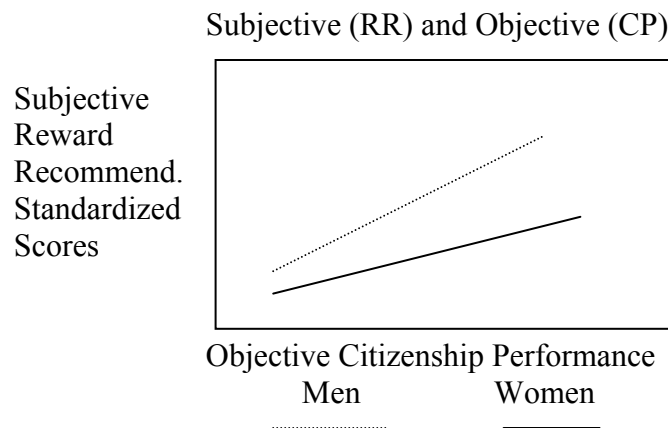
Figure 6. Predicted Results for Hypothesis 4b.



Next will be a discussion of the expectations when using a subjective scale of reward recommendations. First, when reward recommendations is measured with a subjective scale and citizenship is measured using an objective scale, gender is predicted to moderate the relationship between reward recommendations and citizenship performance. Men are expected to have a stronger relationship between reward recommendations and citizenship performance than are women, but the difference is not expected to be as great as when the two are measured with objective scales. Figure 7 displays the expected results for Hypothesis 5.

*Hypothesis 5:* When citizenship performance is measured using an objective scale and reward recommendations are measured using a subjective scale, the relationship between the two is predicted to be moderated by gender. There is predicted to be a stronger relationship for men than for women.

Figure 7. Predicted Results for Hypothesis 5.



Gender is not expected to moderate the relationship between reward recommendations and citizenship performance when they are measured with subjective scales. However, since these results are unknown, this question will be an exploratory one.

*Exploratory question 2:* Does gender moderate the relationship between citizenship performance and reward recommendations?

### *Citizenship Performance Dimensions*

Of the three dimensions proposed by Coleman and Borman (2000), it seems likely that the personal support dimension may be the strongest link to gender stereotypes. The citizenship behaviors included in the personal support dimension appear to have the strongest similarities to gender stereotypes that describe women as unselfish and supportive. If personal support is more representative of gender stereotypes than are initiating structure or organizational support, it will likely provide the largest rating differences between male and female citizenship behaviors.

Lovell et al. (1999) provided support for the hypothesis that gender has more impact on the personal support dimension than the other two dimensions. The researchers found that of the three dimensions (altruism, sportsmanship, and mediation) identified in the study, the altruism dimension was the only one that demonstrated a significant difference between ratings of men and women.

Because of the lack of research on the differential influence of dimensions, this issue will be investigated in an exploratory manner.

Exploratory question 3: Is there a larger difference in ratings of citizenship performance between men and women when the personal support dimension is measured rather than initiating structure and organizational support?

## Chapter 2

### Method

#### *Participants*

Two hundred and seventy-two University of South Florida students received one extra credit point for participating in this study. Of those 272, 254 (93%) answered the manipulation check correctly. The following demographics are based on the sample of 254 who passed the manipulation check.

Most of the respondents were females (N = 210, 83%) between the ages of 17 and 23 years (N = 194, 76%). The majority were white (N = 150, 59%), psychology majors (N = 138, 55%), and were juniors or higher (N = 165, 65%). Almost all (N = 244, 96%) respondents had work experience and 43% (N = 110) had worked as supervisors.

#### *Design*

A 2x2 between subjects factorial design was applied. The independent variables were ratee gender and citizenship performance scale type (objective or subjective). Participants were randomly assigned to one of the following conditions, male target rated on a subjective Citizenship Performance (CP) scale (N= 64), male target rated on an objective CP scale (N = 63), female target rated on a subjective CP scale (N = 63), or female target rated on an objective CP scale (N = 64).



Table 1. Distribution of Participants into 4 Conditions.

Gender	Scale Type	
	Objective	Subjective
Female	N = 64 13 Men 51 Women	N = 63 9 Men 54 Women
Male	N = 63 9 Men 54 Women	N = 64 13 Men 51 Women

### *Procedure*

Participants completed the materials alone or in groups no larger than five. Participants began by signing an informed consent form (Appendix A). Next, participants were asked to read a cover story (Appendix B). The story explained that we have created a professor evaluation and reward system and that students are needed to test the new items. They were given a performance log (Appendix C) of either a male or a female professor. It was explained in the instructions and verbally by the experimenter that these statements were collected last semester from students and other members of the psychology department. Then, participants were asked to evaluate the professor's performance and recommend rewards. The gender manipulation was evident throughout the descriptions by the frequent use of gender specific pronouns. The development and piloting of these materials will be described in the next section.

After reading the performance log, participants filled out an evaluation of citizenship performance. Citizenship performance scale type was a between-subjects variable, with half the participants filling out an objective scale and half filling out a subjective scale.

The citizenship performance evaluation was followed by either an overall performance evaluation or reward recommendations. The order of presentation of overall performance and reward recommendations was counterbalanced with half the subjects receiving overall performance first and half receiving reward recommendations first. The remaining scale (reward recommendations or performance evaluation) was presented as the final measure. The order of the subjective and objective scales, for reward recommendations and overall performance, was also counterbalanced. Statistical analyses were done to test for order effects and are discussed in the results section.

Finally, participants were asked to provide demographic, employment, and education information about themselves. Following the demographic information was a one-item manipulation check asking participants to indicate the gender of the professor they evaluated (Appendix G). After the participants had completed the materials, they were given written information that debriefed them about the true intent of the study (Appendix H).

### *Materials*

In order to identify a job that students perceive to be occupied by an equivalent number of men and women, a pilot study was conducted. It was important to use a job not preconceived to be dominated by one sex and therefore gender stereotyped. Forty-two University of South Florida undergraduate psychology students took part in this pilot study. Participants were provided with a list of 14 occupations and were asked to indicate “What percentage of men and women do you think are employed in these occupations?” Subjects filled in the percentage of men they believed occupied each of the 14 jobs and the corresponding percentage of women they felt occupied the 14 jobs. A

t-test was used to test whether the perceived percentages of men and women were significantly different. Nine out of the 14 jobs were found to be significantly different, with either men or women thought to occupy a larger percentage of the positions in that field. Table 2 displays the findings including the means, standard deviations, and t-values for all 14 comparisons. Of the five jobs that were not significantly different, three of them were professors including a sociology professor ( $t(42) = -.513, \underline{n.s.}$ ), a psychology professor ( $t(42) = .16, \underline{n.s.}$ ), and a communications professor ( $t(42) = -1.07, \underline{n.s.}$ ). A psychology professor was chosen because psychology students were the planned study participants and would therefore be familiar with this job.

Table 2. Perceived Number of Males and Females in 14 Occupations

Occupation	Men		Women		T-Test	
	M	SD	M	SD	T	p
English Prof	78.60	16.38	21.40	16.38	11.45**	p < .000
Video Store Clerk	55.30	10.63	44.70	10.63	3.27**	p = .002
Psychology Prof	50.42	17.56	49.58	17.56	.156	p = .877
Fast Food Worker	51.81	7.93	48.19	7.93	1.50	p = .141
Communications Prof	47.72	14.00	52.28	14.00	-1.07	p = .292
Nurse	28.60	23.77	71.40	23.77	-5.90**	p < .000
Real Estate Agent	42.49	14.02	57.52	14.02	-3.51**	p = .001
Sociology Prof	48.74	16.04	51.26	16.04	-.513	p = .610
Manager	62.05	12.18	37.95	12.18	6.49**	p < .000
English Professor	38.61	14.45	61.40	14.45	-5.17**	p < .000
Veterinarian	55.14	13.14	44.86	13.14	2.57*	p = .014
Bartender	59.70	16.00	40.30	16.00	3.99**	p < .000
Social Worker	27.47	12.84	72.53	12.84	-11.51**	p < .000
Bus Driver	51.80	19.38	48.21	19.38	.606	p = .548
N = 43, *significant at .05, **significant at .01						

Participants were provided with the name of the professor, either Michelle Smith or Michael Smith, and informed that he/she is a psychology professor. Then they received the performance log with 19 statements about the professor (Appendix B). The list included 11 task performance and 9 citizenship performance comments. The task performance statements were obtained from Sauser, Evans, and Champion's (1979) critical incidents of professors. Sauser et al. used undergraduate students to develop a list of 251 critical incidents (samples of good or bad performance) of college professors teaching performance. All incidents were coded on their perceived level of effective

teacher performance. The items chosen for this study were appropriate for the job of a professor and were average in effectiveness (4 to 7 on an 11 point scale).

Citizenship performance statements were based on Borman, Buck, Hanson, Motowidlo, Stark, and Drasgow (2001). Borman et al. developed 124 citizenship performance statements that corresponded to one of Coleman and Borman's (2000) three dimensions (personal support, organizational support, and conscientious initiative). These statements were evaluated by two sets of raters, 37 officers in the Air Force and 26 employees of Personnel Decisions Research Institute's (PDRI). Both groups sorted the statements into the three dimensions and provided effectiveness ratings for each statement. The effectiveness ratings were a way to identify the effective and ineffective citizenship behaviors. The dimensions were correctly sorted into the three dimensions with 90 and 96 percent agreement for the two groups, respectively. Effectiveness scores were given on a 4-point scale. Raters agreed within .5 points with 80 and 96 percent agreement for the two groups. Seventeen items were taken from the 124 behavioral statements created by Borman et al. All 17 items had an effectiveness rating of three. Six items were taken from both the personal support and organizational support dimensions, and five items from the conscientious initiative dimension.

A pilot study was conducted of the 35 (17 citizenship and 18 task performance) performance statements in order to verify that they could consistently be identified as task and citizenship performance. Ten graduate students (5 females and 5 males) in the Industrial/Organizational Ph.D. program at the University of South Florida participated. Students were provided with definitions of task and citizenship performance and were asked to sort the behavioral statements into either task or citizenship performance. Half

of the students (3 females and 2 males) received female performance logs and half of the students (3 males and 2 females) received male performance logs, in order to balance rater gender effects. The graduate students were able to sort most of the items successfully. The final list for this study included 10 task and 9 citizenship performance statements. Of the ten task performance statements chosen, six of these statements were sorted with 100% accuracy and four with 90% accuracy. Of the nine citizenship performance statements chosen (three from each dimension), seven items were sorted with 100% accuracy and two items (1 organization support dimension and 1 conscientious initiative dimension) were sorted with 80% accuracy.

The final profile developed is that of an average task performer, but a slightly above average citizenship performer. Prior to data collection, another pilot study was conducted in order to test the credibility of the cover story and the flow of the materials. The results of this last pilot study will be discussed in the results section.

### *Measures*

*Citizenship Performance.* Nine subjective citizenship performance items were developed for the current study (Appendix D). These items were created from Coleman and Borman's (2000) taxonomy of three dimensions, described earlier. Each dimension emphasized three factors in their definition. The personal support dimension included helping, cooperating, and courtesy. The organizational support dimension included representing, loyalty, and compliance. Finally, the conscientious initiative dimension included persistence, initiative, and self-development. These factors were used to create items for the citizenship performance scale. The intent was to develop a composite score of subjective citizenship performance and an average dimension score for the three

dimensions. The resulting coefficient alpha for the overall scale was .77. The scale was factor analyzed and the three dimension scales all produced coefficient alphas of .65.

The factor analysis is discussed further in the results section.

For the objective measure of citizenship performance, participants were asked to give the professor a letter grade, a percentage score from 0 to 100, and to indicate what they would rank the professor if they were comparing him/her to 9 other professors.

Biernat (1995) claims that, “there is general agreement in our culture that a letter grade should be based on meeting some unwavering criteria for performance and that an A means an A, regardless of who has received it.” (p. 93). Along the same lines, a percentage is expected to have the same meaning for all ratees. These measures are common and familiar, especially to a college student population, and is expected to have universal meaning. Finally, because participants were asked to rank the professor, they directly compared men and women. The internal consistency of the objective scale was  $\alpha = .83$ .

*Overall Performance.* The following five items were used as a subjective measure of overall performance (Appendix E): “This professor makes an important contribution to the university,” “This professor is extremely valuable to the University of South Florida,” “This professor would be extremely costly for the University of South Florida to replace,” “This professor is a vital part of the University of South Florida,” and “This professor is indispensable to the University of South Florida.” These five items were developed by Allen and Rush (2001) to measure overall performance of professors. Participants responded on a 5-point scale ranging from strongly disagree to strongly

agree. Responses from Allen and Rush's study resulted in a high internal reliability ( $\alpha = .92$ ). The present study resulted in an internal reliability of .86.

The same items used as objective measures of citizenship performance, letter grade, a percentage score, and ranking, were used as objective measures of overall performance. The objective overall performance scale resulted in a coefficient alpha of .86.

*Reward Recommendations.* Allen and Rush (1998) developed a five-item measure of reward recommendations, which demonstrated high internal reliability ( $\alpha = .90$ ). Three out of the five items were chosen for the subjective measure of reward recommendations (Appendix F) because of their applicability to the job of a professor. Participants were asked to indicate the extent that they would "recommend the professor for each of the following rewards": a promotion to a more prestigious teaching position, a teaching award, or a salary bonus. Participants provided their answers on a 5-point scale from "would definitely not recommend" to "would recommend without reservation." The coefficient alpha obtained for the present study was .82.

For the objective measure of reward recommendations, three items were used. Participants were asked to indicate which salary bonus they would recommend with 6 choices ranging from \$0 to \$500. Next, participants were asked if 10 professors were competing for a more prestigious teaching position and a teaching award where they would rank this professor. Participants were given 10 choices ranging from first (most deserving) to tenth (least deserving). The resulting internal consistency for this scale was  $\alpha = .80$ .



## Chapter 3

### Results

#### *Pilot Study*

The final pilot study was conducted to determine the credibility of the materials and the time needed to complete the 8 to 9 paged questionnaire. Twelve undergraduate psychology students at the University of South Florida were asked to complete the questionnaires. After the questionnaire was completed, participants were asked three questions about the materials. First, they were asked if there was anything in the questionnaire that they did not understand. All participants responded that the materials were clear. Next, the respondents were asked about the believability of the questionnaire and all the participants agreed that the materials were realistic.

Finally, the students were asked what their reference point was in making the ratings on the subjective and then on the objective scales. The common response was that they were comparing the professor in the materials to the best professor they have had and that they did not change their reference point from the subjective to the objective scale. The pilot study provided support for the materials credibility and helped determine the length of time (10 to 15 minutes) needed to complete the questionnaire.

#### *Manipulation Check*

The manipulation check consisted of one question, “What was the gender of the professor that you just evaluated?” Participants were instructed not to flip back to answer this question and were given the choice to circle either a male or a female professor. Of the 272 participants, 93% (254) correctly identified the gender of the professor. The 18

participants who failed to correctly identify the gender of the professor were not used in further analyses.

### *Order Effects*

The first step in analyzing the data was converting them into z-scores for a standard unit of comparison across different scales (subjective and objective). There were two types of scales that could cause order effects in this study. The first was the presentation order of reward recommendations and overall performance. T-tests were run to determine if there was a significant difference on any of the four dependent measures between those who had reward recommendations first and those who had overall performance first. There were no significant differences found for order of performance and rewards on subjective performance ratings ( $t(252) = .95$ , ns), subjective reward recommendations ( $t(252) = 1.47$ , ns), or objective reward recommendations ( $t(252) = 1.68$ , ns). There was a significant difference found on objective performance ratings ( $t(252) = 2.45$ ,  $p < .025$ ). When participants rated the professor on overall performance before reward recommendations, they rated objective performance significantly higher ( $M = .13$ ,  $SD = .84$ ) than when they rated the professor on reward recommendations first ( $M = -.13$ ,  $SD = .91$ ).

The significant order effect could potentially affect Hypothesis 2a and 2b, which both have objective overall performance as the dependent variable. Hypothesis 2a involves participants who evaluated the citizenship performance of the professor with an objective measure and Hypothesis 2b involves participants who evaluated the citizenship performance of the professor with a subjective measure. In other words, the independent variable for Hypothesis 2a is objective citizenship performance ratings and the

independent variable for Hypothesis 2b is subjective citizenship performance ratings. A t-test was computed to check for order effects on the dependent measure (objective overall performance). There were no significant differences found that would affect Hypothesis 2a (objective CP;  $t(125) = .51, ns$ ). However, when a t-test was computed to check for order effects that would affect Hypothesis 2b (subjective CP), it was significant ( $t(125) = 2.90, p < .01$ ). Therefore, the order of outcome variables for participants who filled out a subjective evaluation of citizenship performance impacted their ratings on the objective measure of overall performance. Participants who rated the professor's overall performance first gave the professor higher ratings on objective overall performance ( $M = .22, SD = .71$ ) than did participants who recommended rewards before evaluating performance ( $M = -.22, SD = .97$ ). As a consequence, when Hypothesis 2b was evaluated, order of overall performance and reward recommendations was entered first (step 1) to remove the variance attributable to the order effect.

T-tests were also run to test for order effects of subjective and objective scales. There were no significant differences found for order of subjective and objective scales on objective overall performance ratings ( $t(252) = .62, ns$ ), subjective overall performance ( $t(252) = .32, ns$ ), objective reward recommendations ( $t(252) = -1.28, ns$ ) or subjective reward recommendations ( $t(252) = -1.72, ns$ ).

#### *Rater Gender Effects*

One possible influence on raters' stereotypes is the gender of the rater. Although rater gender effects on performance evaluations are generally considered to be minimal (Pulakos, White, Oppler, & Borman, 1989), there is still the possibility that rater gender could influence the results. Analysis of variance was used to test for gender differences

on ratings of citizenship performance, overall performance, and reward recommendations by testing for an interaction between rater gender and the ratee gender. There were no significant interactions found for ratings on citizenship performance ( $F(1,250) = .16, \underline{ns}$ ), objective overall performance ( $F(1,250) = .08, \underline{ns}$ ), subjective overall performance ( $F(1,250) = .11, \underline{ns}$ ), objective reward recommendations ( $F(1,250) = .04, \underline{ns}$ ) or subjective reward recommendations ( $F(1,250) = .13, \underline{ns}$ ).

T-tests were also run to test for main effects of rater gender on ratings of citizenship performance, overall performance, and reward recommendations. There were no significant differences found between the ratings provided by male and female participants on citizenship performance ( $t(252) = -.26, \underline{ns}$ ), overall performance ( $t(252) = 1.47, \underline{ns}$ ), or reward recommendations ( $t(252) = -1.27, \underline{ns}$ ).

#### *Descriptive statistics*

Tables 3 and 4 display the means and standard deviations for the outcome variables separated by scale type and professor gender. In Table 3, the values are given in z-scores and in Table 4 the values are given in raw scores. The means are provided for each item on the objective scales because the items were measured with different scales and the objective scale means cannot be computed without converting the values first to z-scores.

Table 3. Descriptive Statistics By Gender and Scale Type Using Z-Scores.

Measures	Gender of Professor			
	Female		Male	
	M	SD	M	SD
	Subjective Scales			
Citizenship Performance	.0006	.59	-.0006	.60
Overall Performance	.010	.89	-.010	.71
Reward Recommendations	.051	.88	-.051	.84
	Objective Scales			
Citizenship Performance	-.049	.93	.050	.80
Overall Performance	.0029	.84	-.0029	.92
Reward Recommendations	-.0024	.84	.0024	.85
	Total			
Citizenship Performance	-.025	.78	.025	.71
Overall Performance	.0063	.79	-.0063	.72
Reward Recommendations	.024	.81	-.024	.79

Note: Values are in z-scores.

Table 4. Descriptive Statistics By Gender and Scale Type Using Raw Data.

Measures	Gender of Professor			
	Female		Male	
	M	SD	M	SD
	Subjective Scales			
Citizenship Performance	4.14	.42	4.14	.42
Overall Performance	3.61	.75	3.59	.59
Reward Recommendations	3.54	.85	3.44	.81
	Objective Scales			
Citizenship Performance Letter Grade (A+ to F, 1 to 13)	3.31	1.65	3.41	1.66
Citizenship Performance Percentage score	86.25	8.82	88.29	6.22
Citizenship Performance Ranking (1 to 10, 1 = the best)	3.56	1.63	3.41	1.59
Overall Performance Letter Grade(A+ to F, 1 to 13)	4.11	1.85	4.20	1.97
Overall Performance Percentage score	84.17	8.65	84.60	10.53
Overall Performance Ranking(1 to 10, 1=the best)	4.13	1.88	4.17	1.96
Reward Recomm Ranking teacher award (1 to 10, 1 = the best)	4.37	1.89	4.21	2.03
Reward Recomm Salary bonus	261.42	118.22	255.18	120.49
Reward Recomm. Ranking promotion(1 to 10, 1=the best)	4.30	1.94	4.33	2.04

Table 5 contains the zero order correlations between objective citizenship performance and the dependent measures, and Table 6 contains the zero order correlations between subjective citizenship performance and the dependent measures.

Table 5. Zero-Order Correlation Coefficients Among Variables with Objective Citizenship Performance.

Variable	1	2	3	4	5	6	7
1. Prof Gender	-						
2. Rater Gender	-.08	-					
3. Objective Citizenship Performance	.06	-.03	-				
4. Objective Overall Performance	-.02	.08	.49**	-			
5. Subjective Overall Performance	-.04	.00	.44**	.59**	-		
6. Objective Reward Recomm.	-.05	.05	.43**	.70**	.66**	-	
7. Subjective Reward Recomm.	-.11	.06	.41**	.65**	.66**	.73**	-
Mean			0	0	3.58	0	3.47
SD			.87	.89	.64	.84	.82

Note: Gender was dummy coded (females = 0, males = 1).

Means for the objective scales are in z-scores and the means for the subjective scales are based on the raw values. \* =  $p < .05$ , \*\* =  $p < .01$ .

Table 6. Zero-Order Correlation Coefficients Among Variables with Subjective Citizenship Performance.

Variable	1	2	3	4	5	6	7
1. Prof Gender	-						
2. Rater Gender	.08	-					
3. Subjective Citizenship Performance	.00	-.18*	-				
4. Objective Overall Performance	.01	.09	.44**	-			
5. Subjective Overall Performance	.01	-.13	.55**	.61**	-		
6. Objective Reward Recomm.	.05	.11	.40**	.81**	.53**	-	
7. Subjective Reward Recomm.	-.01	.07	.43**	.76**	.64**	.76**	-
Mean			4.14	0	3.62	0	3.51
SD			.42	.87	.68	.85	.84

Note: Gender was dummy coded (females = 0, males = 1).

Means for the objective scales are in z-scores and the means for the subjective scales are based on the raw values. \* =  $p < .05$ , \*\* =  $p < .01$ .

### *Hypothesis testing*

For each hypothesis, the first step in analyzing the data was converting them into z-scores. Hypothesis 1 was tested with a 2 x 2 analysis of variance. The dependent variable was citizenship performance scores. The two independent between-subject variables were gender and scale type. This first hypothesis predicted that there would be an interaction between scale type and gender on ratings of citizenship performance. More specifically, it was predicted that females would receive higher objective ratings of citizenship performance than males and that no differences would be found between males and females on the subjective scales. Tables 3 and 4, displayed earlier, have the means for both types of citizenship performance by gender. The interaction between gender and scale type was not significant ( $F(1, 250) = .29, \underline{ns}$ ). Additionally there were no main effects for scale type ( $F(1, 250) = .00, \underline{ns}$ ) or gender ( $F(1, 250) = .275, \underline{ns}$ ). Therefore, no support was found for Hypothesis 1.

Hierarchical multiple regression was used to test the remaining hypotheses. At step one, citizenship performance scores were entered. At step two, gender was entered into the equation. At the final step, the interaction term was entered. For significant interactions, Aiken and West's (1991) procedure for plotting and testing slopes was used. Table 7 displays the correlations between objective citizenship performance, overall performance, and reward recommendations, separated by gender. Table 7 can be used to help interpret the findings for Hypotheses 2a, 2b, 4a, and 4b. Table 8 displays the correlations between subjective citizenship performance, overall performance, and reward recommendations, separated by gender. This table will assist in interpretation of Hypotheses 3 and 5 and exploratory questions 1 and 2.

Table 7. Zero-Order Correlation Coefficients Among Variables with Objective Citizenship Performance, Split by Professor Gender.

	Variable	1	2	3	4	5	Means (SD)
Female Prof	1. Objective Citizenship Performance	-					-.05 (.93)
	2. Objective Overall Performance	.51**	-				.02 (.84)
	3. Subjective Overall Performance	.46**	.63**	-			3.60 (.71)
	4. Objective Reward Recommendations	.49**	.75**	.70**	-		.04 (.80)
	5. Subjective Reward Recommendations	.49**	.62**	.64**	.71**	-	3.56 (.87)
Male Prof	1. Objective Citizenship Performance	-					.05 (.80)
	2. Objective Overall Performance	.49**	-				-.02 (.95)
	3. Subjective Overall Performance	.43**	.55**	-			3.55 (.56)
	4. Objective Reward Recommendations	.38**	.67**	.63**	-		-.04 (.88)
	5. Subjective Reward Recommendations	.34*	.69**	.70**	.75**	-	3.38 (.77)

Note: Means for the objective scales are in z-scores and the means for the subjective scales are based on the raw values. \* =  $p < .05$ , \*\* =  $p < .01$ .

Table 8. Zero-Order Correlation Coefficients Among Variables with Subjective Citizenship Performance, Split by Professor Gender.

	Variable	1	2	3	4	5	Mean (SD)
Female Prof	1. Subjective Citizenship Performance	-					4.14 (.42)
	2. Objective Overall Performance	.50**	-				-.01 (.85)
	3. Subjective Overall Performance	.60**	.70**	-			3.61 (.75)
	4. Objective Reward Recommendations	.39**	.84**	.56**	-		-.04 (.88)
	5. Subjective Reward Recommendations	.45**	.80**	.70**	.80**	-	3.51 (.83)
Male Prof	1. Subjective Citizenship Performance	-					4.14 (.42)
	2. Objective Overall Performance	.39**	-				.01 (.90)
	3. Subjective Overall Performance	.49**	.53**	-			3.63 (.62)
	4. Objective Reward Recommendations	.40**	.78**	.49**	-		.04 (.83)
	5. Subjective Reward Recommendations	.41*	.72**	.59**	.71**	-	3.50 (.85)

Note: Means for the objective scales are in z-scores and the means for the subjective scales are based on the raw values. \* =  $p < .05$ , \*\* =  $p < .01$ .



Hypothesis 2a stated that when overall performance and citizenship performance were measured with an objective scale, the strength of the relationship between citizenship and overall performance would be stronger for men than for women. The regression results for Hypothesis 2a are displayed in Table 9. The interaction between citizenship performance and gender was not significant ( $\beta = .07$ ,  $t(125) = .71$ , ns) providing no support for Hypothesis 2a. The only variable that contributed a significant amount of variance in explaining objective overall performance scores was objective citizenship performance (step 3;  $\beta = .45$ ,  $t(125) = 4.34$ ,  $p < .05$ ). Citizenship performance accounted for 24% of the variance in objective overall performance at step 1. The main effect of gender was not significant (step 3;  $\beta = -.05$ ,  $t(125) = -.66$ , ns).

Table 9. Hierarchical Regression Analysis with Objective Overall Performance as the Dependent Variable.

	Standardized regression weights			
	Objective Overall Performance			
	Step 1	Step 2	Step 3	
Hypothesis 2a				
Objective Citizenship Performance	.49**	.50**	.45**	
Gender		-.05	-.05	
Interaction			.07	
R <sup>2</sup> at each step	.24	.25	.25	
R <sup>2</sup> change		.003	.003	
F	39.95**	.44	.50	
Hypothesis 2b	Step 1	Step 2	Step 3	Step 4
Presentation order of reward recomm. and overall performance	-.25**	-.29**	-.29**	-.29**
Subjective Citizenship Performance		.46**	.46**	.52**
Gender			.01	.01
Interaction				-.07
R <sup>2</sup> at each step	.06	.27	.26	.26
R <sup>2</sup> change		.21	.00	.002
F	8.40**	36.59**	.02	.42

Note: Gender and Presentation order were dummy coded (female/overall performance first = 0 and male/reward recommendations first = 1). \* =  $p < .05$ , \*\* =  $p < .01$ .

Hypothesis 2b proposed that when citizenship performance was measured with an objective scale and overall performance was measured with a subjective scale, men

would have a stronger relationship between their citizenship performance and overall performance than would women. Table 9 shows the regression results for Hypothesis 2b. Because presentation order of overall performance and reward recommendations had a significant impact on objective overall performance ratings when a subjective citizenship performance scale was used, the variable, order of the dependent measure, was entered first. Presentation order was dummy coded with participants receiving performance evaluation first coded “0” and participants receiving reward recommendation first coded “1.” Presentation order was significant (step 4;  $\beta = -.29$ ,  $t(125) = -3.78$ ,  $p < .05$ ) and accounted for 6% ( $F(1, 123) = 8.40$ ,  $p < .05$ ) of the variance in objective overall performance at step 1.

The interaction between gender and citizenship performance was not significant ( $\beta = -.07$ ,  $t(125) = -.64$ , ns). Subjective citizenship performance was significant (step 4;  $\beta = .52$ ,  $t(125) = 4.66$ ,  $p < .05$ ) and it accounted for 21% ( $F(1, 123) = 36.59$ ,  $p < .05$ ) of the variance in objective overall performance ratings, beyond the order effect. The main effect of gender (step 4;  $\beta = .01$ ,  $t(125) = .15$ , ns) was not significant. Therefore, no support was found for Hypothesis 2b.

Hypothesis 3 predicted that when citizenship performance was measured using an objective scale and overall performance was measured using a subjective scale, the relationship between citizenship performance and overall performance would be greater for males than for females. Table 10 displays the results for Hypothesis 3. The interaction term was not significant ( $\beta = -.05$ ,  $t(125) = -.48$ , ns). The only variable that was a significant predictor of subjective overall performance was objective citizenship performance (step 3;  $\beta = .44$ ,  $t(125) = 4.52$ ,  $p < .05$ ) and it accounted 20% of the variance

in subjective overall performance (step 1;  $F(1, 123) = 30.62, p < .05$ ). The gender term was not significant (step 3;  $\beta = -.06, t(125) = -.79, ns$ ). Consequently, the results did not support Hypothesis 3.

Table 10. Hierarchical Regression Analysis with Subjective Overall Performance as the Dependent Variable.

	Standardized regression weights		
	Subjective Overall Performance		
	Step 1	Step 2	Step 3
Hypothesis 3			
Objective Citizenship Performance	.44**	.44**	.48**
Gender		-.06	-.06
Interaction			-.05
R <sup>2</sup> at each step	.20	.20	.20
R <sup>2</sup> change		.004	.002
F	30.62**	.63	.23

Note: Gender was dummy coded (females = 0, males = 1). \* =  $p < .05$ , \*\* =  $p < .01$ ,  $\Psi = .15$ .

The remaining hypotheses concern the interaction of gender and citizenship performance on reward recommendations. Hypothesis 4a predicted that when both citizenship performance and reward recommendations were measured using an objective scale, the correlation between citizenship performance and reward recommendations would be higher for men than for women. Table 11 displays the regression results for Hypothesis 4a. The beta weights were not significant for the interaction term ( $\beta = -.007, t(125) = -.07, ns$ ) or for gender (step 3;  $\beta = -.07, t(125) = -.88, ns$ ). Citizenship performance was significant at all stages (step 3;  $\beta = .44, p < .05$ ) and accounted for 18% (step 1;  $F(1, 125) = 27.77, p < .05$ ) of the variance in objective reward recommendations. Therefore, there was no support found for Hypothesis 4a.

Table 11. Hierarchical Regression Analysis with Objective Reward Recommendations as the Dependent Variable.

	Standardized regression weights		
	Objective Reward Recommendations		
	Step 1	Step 2	Step 3
Hypothesis 4a			
Objective Citizenship Performance	.43**	.43**	.44**
Gender		-.07	-.07
Interaction			-.007
R <sup>2</sup> at each step	.18	.19	.19
R <sup>2</sup> change		.005	.00
F	27.77**	.78	.004
Hypothesis 4b			
Subjective Citizenship Performance	.40**	.40**	.41**
Gender		.05	.05
Interaction			-.02
R <sup>2</sup> at each step	.16	.16	.16
R <sup>2</sup> change		.003	.00
F	23.28**	.37	.02

Note: Gender was dummy coded (females = 0, males = 1). \* =  $p < .05$ , \*\* =  $p < .01$ .

Hypothesis 4b stated that when citizenship performance was measured with a subjective scale and rewards were measured with an objective scale, a stronger relationship between reward recommendations and citizenship performance would be expected for men than for women. As seen in Table 11, the beta coefficient did not support a significant interaction ( $\beta = -.02$ ,  $t(125) = -.15$ , ns). The gender main effect was similarly not significant (step 3;  $\beta = .05$ ,  $t(125) = .61$ , ns). Subjective citizenship performance did predict objective reward recommendations (step 3;  $\beta = .41$ ,  $t(125) = 3.45$ ,  $p < .05$ ) and accounted for 16% of the variance in objective reward recommendations.

The final hypothesis concerns subjective reward recommendations. Hypothesis 5 predicted that when objective citizenship performance and subjective reward recommendations were used, there would be a stronger relationship for men between citizenship performance and reward recommendations than there would be for women.

Table 12 displays the regression results for Hypothesis 5. The beta weight did not support a significant interaction between citizenship performance and gender ( $\beta = -.09$ ,  $t(125) = -.85$ , ns) nor was the effect of gender significant (step 3;  $\beta = -.13$ ,  $t(125) = -1.66$ ,  $p = .10$ ). With the addition of gender at step two, the model explained an additional 2% ( $F(1, 125) = 2.78$ ,  $p = .10$ ) of the variance in subjective reward recommendations, which was marginally significant at an alpha level of .10 (Females:  $M = .09$ ,  $SD = .91$ , Males =  $M = -.10$ ,  $SD = .80$ ). Subjective citizenship performance did predict objective measures of reward recommendations (step 3;  $\beta = .48$ ,  $t(125) = 4.49$ ,  $p < .05$ ) and accounted for 17% (step 1;  $F(1, 125) = 25.66$ ,  $p < .05$ ) of the variance in objective reward recommendations. However, no support was found for Hypothesis 5.

Table 12. Hierarchical Regression Analysis with Subjective Reward Recommendations as the Dependent Variable.

	Standardized regression weights		
	Subjective Reward Recommendations		
	Step 1	Step 2	Step 3
Hypothesis 5			
Objective Citizenship Performance	.41**	.42**	.48**
Gender		-.14 $\Psi$	-.13 $\Psi$
Interaction			-.09
R <sup>2</sup> at each step	.17	.19	.19
R <sup>2</sup> change		.02	.005
F	25.66**	2.78 $\Psi$	.72

Note: Gender was dummy coded (females = 0, males = 1). \* =  $p < .05$ , \*\* =  $p < .01$ ,  $\Psi = p = .10$ .

### *Exploratory Questions*

The first two exploratory questions were analyzed using hierarchical multiple regression. The first exploratory question concerned the relationship between subjective measures of both citizenship and overall performance and how the relationship differed for men and women. The results are displayed in Table 13. The beta weight for the interaction was not significant ( $\beta = -.17$ ,  $t(125) = -1.57$ ,  $p = .15$ ) and neither was the beta weight for gender (step 3;  $\beta = .01$ ,  $t(125) = .15$ , ns). The incremental validity from the

addition of the interaction term was 1%, which was not a significant addition ( $F(1, 123) = 2.47, p = .15$ ). Citizenship performance was a significant predictor (step 3;  $\beta = .67, t(125) = 6.22, p < .05$ ) and it accounted for 30% (step 1;  $F(1, 125) = 53.19, p < .05$ ) of the variance in subjective overall performance.

Table 13. Hierarchical Regression Analysis with Subjective Measures of Both Overall Performance and Citizenship Performance.

	Standardized regression weights		
	Subjective Overall Performance		
	Step 1	Step 2	Step 3
Exploratory question 1			
Subjective Citizenship Performance	.55**	.55**	.67**
Gender		.01	.01
Interaction			-.17 $\Psi$
R <sup>2</sup> at each step	.30	.30	.31
R <sup>2</sup> change		.00	.01
F	53.19**	.02	2.47 $\Psi$

Note: Gender was dummy coded (females = 0, males = 1). \* =  $p < .05$ , \*\* =  $p < .01$ ,  $\Psi = p = .10$ .

The second exploratory question investigated differences for men and women in the relationship between subjective measures of reward recommendations and citizenship performance. Table 11 displays the results. Both the interaction term and the gender variable were not significant ( $\beta = -.03, t(125) = -.25, ns$  &  $\beta = -.01, t(125) = -.10, ns$ , respectively). Subjective citizenship performance was a significant predictor of subjective reward recommendations (step 3;  $\beta = .45, t(125) = 3.88, p < .05$ ) and accounted for 19% of the variance ( $F(1, 125) = 28.53, p < .05$ ).

Table 14. Hierarchical Regression Analysis with Subjective Measures of Both Reward Recommendations and Citizenship Performance.

	Standardized regression weights		
	Subjective Overall Performance		
	Step 1	Step 2	Step 3
Exploratory Question 2			
Subjective Citizenship Performance	.43**	.43**	.45**
Gender		-.01	-.01
Interaction			-.03
R <sup>2</sup> at each step	.19	.19	.19
R <sup>2</sup> change		.00	.00
F	28.53**	.01	.06

Note: Gender was dummy coded (females = 0, males = 1). \* =  $p < .05$ , \*\* =  $p < .01$ .

The third exploratory question investigated if men and women are rated differently on the three types of citizenship performance. If differences were found on the three types of citizenship performance, then the magnitude of those differences would be contrasted with three t-tests. Although there were no significant differences found between men and women on the composite, subjective measure of citizenship performance ( $t(125) = .01$ , ns; Males:  $M = .00$ ,  $SD = .59$ ; Females:  $M = .00$ ,  $SD = .60$ ), is it possible that differences could be found on one of the dimensions.

A principle factor analysis was done to verify that the citizenship performance dimensions (personal support, organizational support, and conscientious initiative) do, in fact, load on three separate factors. The factor analysis was followed by a promax oblique rotation with a forced three-factor solution. The eigenvalues before rotation were 3.23, 1.22, 1.04 and accounted for 61% (36%, 14% and 12%, respectively) percent of the variance. The scree plot (see Figure 8 for details) appears to support a one to four factor solution, with the line leveling out after the fifth eigenvalue. After the rotation, the eigenvalues became 2.10, 2.02, and 1.88.

Figure 8. Scree Plot of Eigenvalues for Factor Analysis of Citizenship Performance Dimensions.

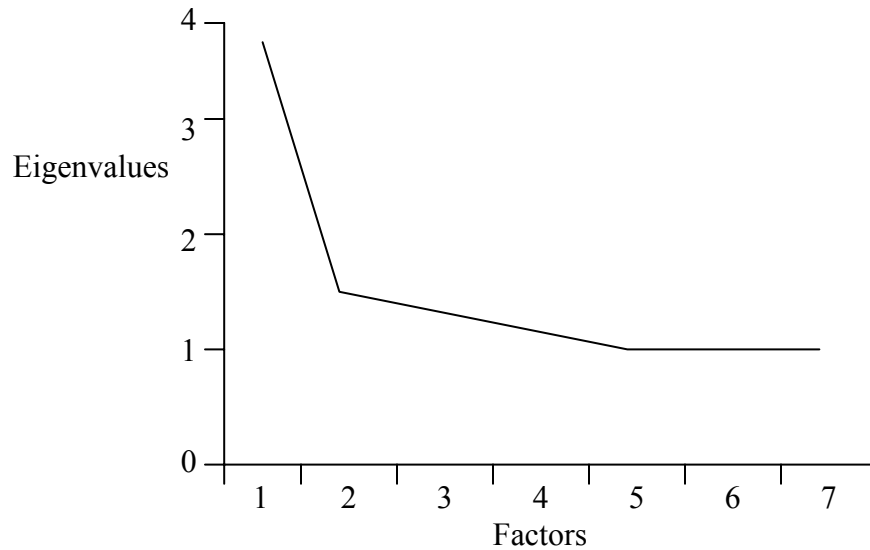


Table 15 displays the pattern matrix of the factor loadings after the promax rotation. With a sample size of approximately 140, the criterion for significant factor loadings was set at .43 (Stevens, 2002). All three organizational support items loaded on factor 1. All three conscientious initiative items loaded on factor 2. Only two of the three personal support items loaded on factor 3. The third personal support item, “Dr. Smith is cooperative when working with colleagues and students” did not sufficiently load on the personal support factor (.27). Therefore, this item was dropped leaving only two items in the personal support factor. The first personal support item, “Dr. Smith is helpful to co-workers and students,” had a high loading on both conscientious initiative (.42) and personal support (.50). Because the personal support loading was larger, met the loading criterion, and was in accordance with past research (Coleman & Borman, 2000), this item was included in the personal support factor. The final scales all yielded coefficient alphas of .65.



Table 15. Promax Oblique Factor Rotation Pattern Matrix.

Items	Factor 1	Factor 2	Factor 3
5. Represents school well to outsiders (OS2)	.65	.02	.00
6. Complies with school rules and procedures (OS3)	.62	-.08	-.06
4. Loyalty to school (OS1)	.59	.08	.10
3. Cooperation with co-workers and students (PS3)	.37	.15	.27
7. Persistence and extra effort given when needed (CI1)	.07	.76	-.18
9. Dedicated to self development (CI3)	-.07	.56	.04
8. Takes initiative to get things done (CI2)	.07	.46	.03
2. Courteous with co-workers and students (PS2)	.05	-.16	.92
1. Helpful with co-workers and students (PS1)	-.15	.42	.50
Eigenvalues after rotation	2.10	2.02	1.88

Table 16 displays the mean factor scores by gender. Differences between citizenship ratings for men and women were evaluated with t-tests. There were no significant differences found between men and women on any of the citizenship performance dimensions (ps:  $t(125) = -.66$ , ns; os:  $t(125) = .62$ , ns; ci:  $t(125) = -.24$ , ns). The effect sizes were calculated and all three effect sizes were small (ps:  $d = .12$ , os:  $d = .11$ , and ci:  $d = .04$ ) according to Cohen's effect size criteria ( $d < .2 =$  small effect; Stevens, 2002).

Table 16. Mean Factor Scores by Gender.

Gender	Factor 1 Organizational Support		Factor 2 Conscientious Initiative		Factor 3 Personal Support	
	M	SD	M	SD	M	SD
Female	4.21	.56	4.03	.60	4.20	.47
Male	4.15	.52	4.06	.61	4.26	.52
Coefficient Alpha	.65		.65		.65	
Eigenvalues	2.10		2.02		1.88	

### *Post Hoc Analysis*

Following the main analyses, outliers were identified and evaluated on the hypotheses and exploratory questions. Outliers were defined as studentized residuals greater than two. Following the identification of the outliers, an influence analysis was conducted, including Cook's distance, leverage, and DF betas. DF Betas with scores over .18 ( $2/\sqrt{N} = 2/\sqrt{127}$ ) were considered to have significant impact on the intercept or

slopes. Leverage scores near  $.047$  ( $2(k+1)/N = 2(2+1)/127$ ) and cook's distance values above  $.2$  were considered to have impact on the variables. The outlier identification and influence criteria were based on recommendations by Pedhazur (1997).

There were no changes in significance as a result of the removal of outliers, but exploratory question one had marginal significance as a result of removing outliers. Exploratory question one was interested in how the relationship between subjective citizenship performance and subjective overall performance was different for men than it was for women. Before the outliers were deleted, the incremental variance attributed to the interaction term was 1% ( $F(1, 123) = 2.47, p = .15$ ). Several outliers were identified and three outliers were deleted based on the influence analysis detailed in Table 17. After the removal of these outliers, the incremental  $R^2$  for the interaction term increased to 2% ( $F(1, 120) = 3.56, p = .06$ ) and was marginally significant at the  $.10$  alpha level.

Table 17. Outliers and their Influence on the Variables.

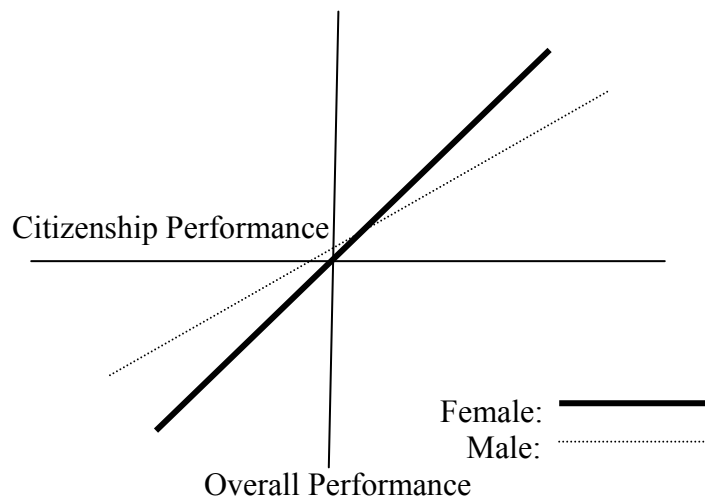
ID Number	Studentized Residual	Cook's D	Leverage	DF Betas			
				Intercept	Subj. CP	Gender	Interaction
Exploratory Question 1							
64	-2.13	.05	.04	-.28	-.37	.20	.27
105	-3.35	.05	.01	-.45	.14	.32	-.06
231	-2.78	.03	.01	-.36	.03	.26	-.07

The regression equation now predicts 37% of the variance in subjective reward recommendations. The beta weight for the interaction term is  $-.20$  and is marginally significant ( $t(122) = -1.89, p = .06$ ). The correlation between subjective citizenship and overall performance was stronger for women ( $r = .69, p < .05$ ) than it was for men ( $r = .49, p < .05$ ).

To explore this finding further, the procedures described by Aiken and West (1991) for plotting and interpreting interactions was employed. Figure 9 shows the

results from plotting the interaction. A t-test was performed to determine if the slopes for women and men were significantly different from zero. The female slope was significantly different from zero ( $b = .96$ ;  $t(118) = 6.98$ ,  $p < .001$ ), but the male slope was only marginally significant ( $b = .60$ ;  $t(118) = 1.98$ ,  $p < .10$ ). Therefore, the relationship between citizenship performance and overall performance, when they were measured with subjective scales, is significant for females but not for males.

Figure 9. Interaction plotted according to Aiken and West's procedure.



## Chapter 4

### Discussion

Several lines of recent research suggest that employees who perform citizenship behaviors receive higher performance evaluations and more rewards when compared to employees who do not perform citizenship behaviors (e.g. Allen & Rush, 1998; Van Scotter et al., 2000). Further research was needed to determine whether all types of people receive equal consequences for their citizenship performance or whether this phenomenon holds true for only a specific portion of the general population. The present study was designed to determine whether performing citizenship behaviors would have different consequences for women when compared to men. More specifically, the present study investigated whether men and women are equally evaluated and rewarded for their citizenship performance.

As a first step, the differences in ratings of citizenship performance for men and women were studied. Next, differences in the relationship between citizenship performance and overall performance for men and for women were tested. Along similar lines, differences in the relationship between citizenship performance and reward recommendations were investigated. Finally, two different scales were used to measure citizenship performance, overall performance, and reward recommendations to determine whether scale type was a moderator of these relationships as proposed by the shifting standards model (Nelson et al., 1990). Unfortunately, the differential relationships hypothesized for men and women were not supported by this study.

### *Gender and Citizenship Performance*

Eagly and Crowley (1986) found that women are expected to assist others in their goals and nurture them through their personal problems. These stereotype expectations of women resemble the expectations of an organizational citizen. Further, Allen and Rush (2001) found that there are greater expectations for women to display citizenship performance than for men. Building on this stereotype, the shifting standards model (Nelson et al., 1990) would predict that when raters use a subjective scale to rate citizenship performance, they will rate men relative to men and women relative to women. The subjective scale would consequently hide the stereotypes held by the raters. However, when an objective scale is used, raters' stereotypes are expected to be more evident. In the present study, it was hypothesized that when citizenship performance was measured with an objective scale, women would be rated higher than would men on citizenship performance. Conversely, when citizenship performance was measured with a subjective scale, no differences were expected between men and women. This hypothesis was not supported.

Several explanations are offered to explain these null results. The first consideration is the shifting standards model, which has been widely supported through numerous studies (Biernat & Manis, 1994; Biernat et al., 1991; Nelson et al., 1990). Because of the growing support of this model, it seems unlikely that the null results are the outcome of a faulty model, but rather a problem with the present studies application of that model. Three other possible explanations include that there is no gender stereotype of citizenship performance, the gender stereotype was not triggered by the materials, and

the objective scale did not force raters to compare men and women. In what follows, each of these possibilities will be considered.

The explanation that there is no gender stereotype of citizenship performance is contrary to the findings by Allen and Rush (2001). Allen and Rush (2001) found that there were more expectations for women to perform citizenship performance than for men. These results were found for male stereotyped jobs and for gender neutral jobs. Therefore, perhaps the gender stereotypes of citizenship performance are dependent on the type of job. The present study tested the gender stereotypes of psychology professors and found that students did not expect one gender to be more prevalent than the other for that profession. However, there may be other aspects of the job that can affect the gender stereotype of citizenship performance.

One example might be jobs in which citizenship performance overlaps with task performance. In this case, the job might not possess the gender stereotypes of citizenship performance because these behaviors are equally expected of both genders. Especially from students' perspective, citizenship performance might be an expectation of all teachers and therefore would not be considered extra-role. This idea will be discussed further in the limitations section.

The second explanation is that the study did not trigger the gender citizenship performance stereotype. First, the definition of citizenship performance might not have been clear enough for participants. In order for the stereotype to be triggered, participants needed to understand the meaning of citizenship performance. The participants might not have attended to the definition given or it might not have been clear. One example of participants not attending to the definition was observed during

the pilot study. After filling out the materials, participants were asked if there was anything about the questionnaire they did not understand. One person commented that a couple of the measures were repetitious. She identified the citizenship performance and overall performance measures as repetitious and it was evident that she had disregarded the directions. One suggestion for future research is to orally explain the meaning of citizenship performance and provide more examples of it.

Another reason why the stereotype might not have been triggered is that the professor's performance was given in a vignette. In these circumstances, the gender of the professor might not have been a salient characteristic. In a field setting, the gender of the employee is typically an unmistakable characteristic that is noticed about someone. However, a lab setting might not trigger the same stereotype as a field setting. One suggestion for future research is to provide a picture with the vignette. Chen and Heilman (2001) used vignettes to describe citizenship performance and overall performance. These researchers provided a picture with performance descriptions and they found support for their hypothesis that women and men are not equally evaluated or rewarded for their citizenship performance.

The final explanation for the null results suggests there could be problems with the objective measure of citizenship performance. A number of participants expressed confusion with the objective item that asks them how this professor would compare against nine other USF professors. Part of Nelson et al.'s (1990) definition of an objective scale is that it is familiar. It is supposed to be a type of measurement that is common to raters. Therefore, it is possible that this item was not acting as an objective measure that unmask the stereotype. To test this possibility further, the results were

rerun with the objective ranking item excluded. There was no change in significance after this item was deleted. Therefore, problems with the ranking objective measure seem like an unlikely explanation for the unsupported hypothesis.

Along the same lines, there is no evidence that raters using the objective scales were rating men and women against the same standard, as implied by the objective measure. When raters were using letter grades and percentages they could have been rating men against men and women against women. Biernat and Manis (1994) used letter grades in their study of stereotypes of verbal ability. These researchers found support for the shifting standards model, with women receiving a significantly higher letter grade for verbal ability (objective measure) than men. However, when a subjective measure was used, there was no significant difference found between men and women. This finding provides support for the use of letter grades as an objective measure and reduces the likelihood that the explanation for these results is a problem with the objective measures. However, future research should consider having participants rate several employees so that the objective measure can be ranking them and therefore forcing raters to compare men and women directly.

### *Overall Performance*

The present study found that both measures (objective and subjective) of citizenship performance explained a significant amount of variance in both measures (objective and subjective) of overall performance. This result is consistent with results found in previous research (e.g. Allen & Rush, 1998), indicating that citizenship performance is related to ratings of overall performance.



Several studies have investigated if the correlation between citizenship performance and overall performance is different for men and women (Allen & Rush, 2001; Chen & Heilman, 2001; Lovell et al., 1999), with mixed results. The present study was designed to extend these findings by testing whether the scale type used to measure performance affects the relationship between citizenship and overall performance for men and women. The hypotheses predicted that the relationship between citizenship and overall performance would be larger for men than for women and the relationship would vary between objective and subjective scale types. These hypotheses were not supported. Further, no significant differences were found between men and women on their relationship between citizenship and overall performance.

The explanations provided for the null results from the previous section are applicable here. Another possible explanation is that the two scales, subjective and objective, were presented as a within subjects variable with one following the other, rather than between. Participants may not have differentiated between the scales. This explanation was tested with factor analysis. First, the correlation between objective overall performance and subjective overall performance was .60 and the reliability coefficient when the two scales were combined was .89. A principal factor analysis extracted two factors and with a promax oblique rotation, the subjective items loaded on factor one and the objective items loaded on factor two. Therefore, it appears as if the two scales were distinct and the explanation that these two scales were not discriminated between is not likely.

### *Reward Recommendations*

Several studies have found a significant relationship between citizenship performance and organizational rewards (e.g., Van Scotter et al., 2000). The present study is consistent with these findings with both measures of citizenship performance relating to both measures of reward recommendations. There have been mixed results in studies that tested whether citizenship performance relates to reward recommendations differently for men than for women (Allen, 2000; Allen & Rush, 2001; Chen & Heilman, 2001). For the current study, these results were further investigated by testing whether the type of measurement used affects the relationship between citizenship performance and reward recommendations differently for men than for women. Based on the shifting standards model (Nelson et al., 1990), it was predicted that the relationship between citizenship performance and reward recommendations, for the different scales, would be stronger for men than for women. These hypotheses were not supported. No significant differences were found in the relationship between citizenship performance and reward recommendations for men and for women.

The explanations provided for the earlier hypotheses, there is no gender stereotype, the objective scales were not performing as defined, scale type should be treated as a between subjects variable, and the stereotype was not triggered by the materials, are applicable for the reward recommendations hypotheses as well. The proposed explanation that the null results were a consequence of scale type being a within subjects variable is explored further with reward recommendations. The correlation between objective and subjective reward recommendations was .74, and the reliability coefficient for the scale as a whole was .88. A factor analysis was performed and only

one component was extracted. Therefore, it appears as if the two types of reward recommendation scales were not clearly distinguished by participants. Future studies should use scale type as a between subjects factor or conduct a field study where current salary and position are previously established objective measures and subjective measures can be filled out by participants, therefore reducing possible carryover effects.

#### *Citizenship Performance Dimensions*

There are three dimensions of citizenship performance used in this study. As opposed to the organizational support and conscientious initiative dimensions, the personal support dimension most closely resembles the female stereotype of women as nurturers. Therefore, the final research question investigated whether men and women received different ratings on the citizenship performance dimensions and whether the effect sizes were different across dimensions. The factor analysis verified that the scale truly did capture three distinct dimensions and that the factors correspond, for the most part, with the expected items. However, no significant differences were found between men and women on any of the dimensions.

Two previously mentioned explanations are relevant for these null findings. The first is that there are no gender stereotypes of citizenship performance. The second explanation is that the gender stereotype was not triggered by the materials.

#### *Post Hoc*

Following the hypothesis testing, outliers were identified that may be influencing the results. Exploratory question one investigated how the correlation between subjective measures of citizenship and overall performance was different for men and women. With the removal of three outliers, there was a marginally significant result ( $p = .06$ ). The

correlation between citizenship performance and overall performance was stronger for women ( $r = .69$ ) than for men ( $r = .49$ ). After plotting the interactions, the slope was significant for women and not for men. Further, women received a higher correlation than did men on all but one of the hypotheses and exploratory questions. This finding was contrary to expectations. More research is needed to determine the cause of these findings. One explanation could be the job used in this study. This idea will be discussed further in the following section.

### *General Limitations*

One limitation of this study was the participant pool. Participants were taken from a student population and therefore, might not have been invested in this project since they were doing it for extra credit. If the participants were not invested, they would have been less likely to remember the performance statements, read the directions and definitions carefully, or take their time filling out the materials. If participants are not focused on the task, the stereotype might not be triggered and, consequently, no differences would be expected.

Another possible limitation with the study is that the task statements were given from the perspective of students, and the citizenship performance statements were given from the perspective of co-workers or supervisors. This is a problem because the participants were students and might have focused more on the task performance than on the citizenship performance, since students wrote the task statements. Another possibility is that the performance statements reminded the participants of a particular teacher, making it hard to distinguish between the professor triggered and the performance of the professor in the vignette. As evidence, several students asked whom the professor was

they had evaluated because they thought they recognized the professor. In one case, the participant said she was thinking of a female professor at USF when she was making the ratings even though she knew the professor in the vignette was a man. Therefore, this participant got the manipulation check correct because she knew the vignette was about a male professor, but she was thinking about a particular female professor while filling out the materials.

Students are likely to have role schemas (expectations about how a person should act in a particular role; Baron, Byrne, & Johnson, 1998) for the performance of a professor. Statements that fit in that schema are more likely to be paid attention to, encoded, and recalled (Baron, Byrne, & Johnson, 1998). Therefore, the citizenship performance statements may not have been attended to as much as the task statements because the task statements met the students' schema of teacher performance.

A third limitation concerns the job chosen for this study. The stereotype that women are expected to be "concerned with others" and possess "communal qualities, such as being friendly" might be expectations of a teacher and not extra-role behaviors. Sauser et al. (1979) identified five dimensions of teacher performance from their list of college professor behavioral incidents. Some examples of effective performance from this dimension are "this professor offers help at night" and "this professor made appointments at her students' convenience to discuss problems with classwork." Citizenship performance might be considered a requirement of a teacher and consequently, expectations for citizenship performance might be equal for men and women.

The hypotheses proposed for the present study might have had different results in an occupation where citizenship performance is more clearly extra role. Consistent with this idea, Allen and Rush (2001) used the job of a teacher and found no significant differences in ratings of overall performance or reward recommendations between men and women. Further, Lovell et al. (1998) studied college resident advisors in their investigation of citizenship and overall performance ratings for men and women. Task expectations of a resident advisor are similar to the expectations of an organizational citizen. Resident advisors are expected to help the residents when needed and act as a counselor when the residents need someone to talk to. Further, Lovell et al. (1998) claimed that one possible reason for their findings was that resident advisors may have been chosen for the job based on their abilities to perform citizenship performance. Therefore, citizenship performance for the resident advisor job might also have equal expectations for both men and women.

Chen and Heilman (2001) provided an example of a job that did exhibit gender differences in evaluating and rewarding citizenship performance. The name of the job was not given, but the description of the procedures made it clear that it was an office job. The OCB provided was either staying late to help a co-worker with an important copy job or not staying late to help the co-worker. This sample of OCB is more clearly extra-role. In a more traditional office setting than a university, citizenship performance might be considered extra-role.

Another limitation of the present study is using a multiple regression interaction as the main level of analysis. A great deal of power is needed to detect an interaction in multiple regression (Aiken & West, 1991). The important considerations for power are

sample size, effect size, and alpha level. Effect sizes were calculated with a formula provided by Aiken and West (1991;  $f^2 = (r^2_{Y.MI} - r^2_{Y.M}) / (1 - r^2_{Y.MI})$  MI = interaction and main effects, M = main effects} and they ranged from .01 to .0001. The largest effect was .01, which is considered small by Aiken and West, (1991). Therefore, no further power analysis was done on this sample. It is unlikely that lack of power is the reason for the null results.

Another limitation is that there was not an equal number of men ( $N = 44$ ) and women ( $N = 210$ ) in the sample. Although this unequal number is stated as a limitation, there was no evidence of a difference between male and female raters. However, one consideration is that differences might be found if the stereotypes held by the raters are studied rather than their gender. Dobbins, Cardy, and Truxillo (1988) found that raters who held traditional stereotypes about women's roles in society rated women less accurately than raters who held nontraditional stereotypes of women. There was no effect found for rater gender. Future research might want to test the stereotypes held by raters, rather than rater gender, as a possible moderator of citizenship performance ratings.

#### *Implications and Future Research*

The null results of the current study left two general explanations. The first possibility is that the design or method of the study was flawed. This possible explanation was discussed previously. The second is that raters do evaluate and reward men and women the same for their citizenship performance. Therefore, the relationship between citizenship performance and overall performance and citizenship performance and reward recommendations is not different for men than it is for women. These

conclusions would be supported by findings from Allen and Rush (2001) and partially by Lovell et al. (1998). If this conclusion is accurate and there are no significant differences in performance ratings for men and women (Landy & Far, 1988), then it is likely that the men and women perform the same amount of citizenship behaviors in the work place. Future research might expand this research further by measuring baserates of male and female citizenship behaviors. Perhaps women perform one type of citizenship behaviors more and men perform another type of citizenship behaviors more.

It is possible that the present results may have been different if not for the above mentioned limitations. Future research should consider adding to the knowledge in this area with a field study. Based on the limitations mentioned, several suggestions are made for future research. First, scale type should be a between-subjects variable. Second, the type of job chosen should be one in which citizenship performance is not an expectation of the job. Third, the difficulty in choosing an appropriate objective measure should be realized when putting the design together. Finally, make sure that the citizenship performance definition is clear and that the rater understands its meaning. With these limitations met in a field study, a more clear understanding can be reached about whether the null results are a consequence of the design of the study or the hypotheses.

Two variables were identified as possible moderators of the gender stereotype of citizenship performance. The first possible moderator is the stereotype of job. As mentioned earlier, Allen and Rush (2001) found that the citizenship performance stereotype was moderated by the stereotype of the job. Future research could perform the present study with a gender-neutral occupation, a male stereotyped occupation and a female stereotyped occupation. The second possible moderator suggests that gender



stereotypes held by the rater may have an impact the gender stereotypes of citizenship performance (Dobbins, Cardy, & Truxillo, 1988). Future research could consider rater stereotypes in their analyses.

In conclusion, the concept of citizenship performance has important considerations for individuals because it influences their performance evaluations and rewards. The current study was unable to find differences between males and females, which might lead to the conclusion that there are no differences between ratings for males and females on citizenship performance. Future research is needed to test these hypotheses in a field setting creating a realistic setting for gender citizenship performance stereotypes to materialize.

## References

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Allen, T. (2000, April). *Rewarding good citizens: The relationship between citizenship behavior, gender and organizational rewards*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Allen, T. D., & Rush, M. C. (2001). The influence of rater gender on ratings of organizational citizenship behavior. *Journal of Applied Social Psychology, 31*, 2561-2587.
- Allen, T. D. & Rush, M. C. (1998). The effects of organizational citizenship behavior on performance judgments: A field study and a laboratory experiment. *Journal of Applied Psychology, 83*, (2), 247-260.
- Atwater, L. E. & Van Fleet, D. D. (1997). Another ceiling? Can males compete for traditionally female jobs? *Journal of Management, 23*, (5), 603-626.
- Auster, E. R. (1999). Task characteristics as a bridge between macro- and microlevel research on salary inequality between men and women. *Journal of Management Review, 14*, (2), 173-193.
- Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.
- Baron, Byrne, & Johnson (1998). *Exploring Social Psychology* (4<sup>th</sup> ed.). Needham Heights, MA: Allyn and Bacon.

Basket, G. D. (1973). Interview decisions as determined by competency and attitude similarity. *Journal of Applied Psychology*, 57, (3), 343-345.

Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In Y. T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 87-114). Washington, DC: American Psychological Association.

Biernat, M., Crandall, C. S., Young, L. V., Kobrynowicz, D., & Halpin, S. M. (1998). All that you can be: Stereotyping of self and others in a military context. *Journal of Personality and Social Psychology*, 75, (2), 301-317.

Beirnat, M. & Manis, M. (1994). Shifting standards and stereotype-based judgements. *Journal of Personality and Social Psychology*, 66, (1), 5-20.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgement. *Journal of Personality and Social Psychology*, 60, (4), 485-499.

Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., and Drasgow, F. (2001). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology*, 86, (5), 965-973.

Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10 (2), p. 99-109.

Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, 9, (1-2), 52-69.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee talk performance and interpersonal factors on supervisor and peer ratings. *Journal of Applied Psychology, 80* (1), 168-177.

Brief, A. P. & Motowidlo, S. J. (1986). Prosocial organizational behaviors. *Academy of Management Review, 11*, (4), 710-725.

Byrne, D. & Griffitt, W. (1966). Similarity versus liking: A clarification. *Psychometric Science, 6*, (6), 295-296.

Cardy, R. L. & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati, OH: South-Western Publishing.

Chen, J. J. & Heilman, M. E. (2001, April). Need some help? Gender-specific rewards for organizational citizenship behaviors. Paper presented at the meeting of the Society for the Industrial and Organizational Psychology, San Diego, CA.

Coleman, V. I. & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review, 10*, 25-44.

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.

Eagley, A. H. & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*(3), 283-308.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, 111 (1), p. 3-22.

Field, A. (2000). *Discover statistics, using SPSS for windows*. Thousand Oaks, CA: Sage.

Golightly, C., Huffman, D. M., & Byrne, D. (1972). Liking and loathing. *Journal of Applied Psychology*, 56, 521-523.

Graen, G. B., Novak, M. A., Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Decision Processes*, 30, (1), 109-131.

Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. (1974). Race and sex determinants of ratings by potential employers in a simulated work-sampling task. *Journal of Applied Psychology*, 59, (6), 705-711.

Heilman, Madeline, E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality*, 10, (6), 3-26.

Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, 74 (6), 935-942.

Heilman, M. E., Martel, R. F., & Simon, M. C. (1988). The vagaries of sex bias: Conditions regulating the undervaluation, equivalation, and overvaluation of female job applicants. *Organizational Behavior and Human Decision Processes*, 41, 98-110.

Hui, C., Lam, S., & Law, K. (2000). Instrumental values of organizational citizenship behavior for promotion: A field quasi-experiment. *Journal of Applied Psychology, 85*, (5), 822-828.

Judge, T. A., Cable, D. M., Boudreau, J. W., & Bretz, Jr., R. D. (1995). An empirical investigation of the predictors of executive success. *Personnel Psychology, 48*, 485-519.

Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral Science, 9*, 131-46.

Kiker, D. S. & Motowidlo, S. J. (1999) Main and interaction effects of task and contextual performance on supervisory reward decisions. *Journal of Applied Psychology, 84* (4), 602-609.

Landy, F. J. & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, (1), 72-107.

Lewis, K. & Taylor, M. S. (2001, April) Reciprocity from the organization's side: Manager reactions to employee psychological contract breach. Paper presented at the meeting of Society for Industrial Organizational Psychologist, San Diego.

Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study on the early development of leader-member exchanges. *Journal of Applied Psychology, 78*, (4), 662-674.

Lovell, S. E., Kahn, A. S., Anton, J., Davidson, A., Dowling, E., Post, D., & Mason, C. (1999). Does gender affect the link between organizational citizenship behavior and performance evaluation? *Sex Roles, 41* (5-6), p. 469-478.

Lyness, K. S. & Thompson, D. E. (1997). Above the glass ceiling? A comparison of matched samples of female and male executives. *Journal of Applied Psychology, 82*, (3), 359-375.

MacKenzie, S. B., Podsakoff, P. M., & Ahearne, M. (1998). Some possible antecedents and consequences of in-role and extra-role salesperson performance. *Journal of Marketing, 62*, 87-98.

MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salesperson' performance. *Organizational Behavior and Human Decision Processes, 50* (1), 123-150.

Manis, M., Biernat, M., & Nelson, T. (1990). Comparison and expectancy processes in human judgment. *Journal of Personality and Social Psychology, 61*, (2), 203-211.

Martel, R. F. & DeSmet, A. L. (2000, April). Gender stereotyping in the managerial ranks: A Bayesian approach to measuring beliefs about the leadership abilities of male and female managers. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Massengill, D. & di Marco, N. (1979). Sex-role stereotypes and requisite management characteristics: A current replication. *Sex Roles, 5*, (5), 561-570.

Miles, R. E. (1964). Attitudes toward management theory as a factor in managers' relationships with their superiors. *Academy of Management Journal, 7*, 303-313.

Morrison, E. W. (1994). Role definitions and organizational citizenship behavior: The importance of the employee's perspective. *Academy of Management Journal*, 37, (6), 1543-1567.

Motowildo, S. J. & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. Al. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557-576.

Nelson, T. E., Biernat, M. R., & Manis, M. (1990). Everyday base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology*, 59, (4), 664-675.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, 10, 85-97.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.

Organ, D. W. (1977). A reappraisal and reinterpretation of the satisfaction causes performance hypothesis. *Academy of Management Review*, 2, 46-53.

Organ, D. & Moorman, R. (1993). Fairness and organizational citizenship behavior: What are the connections? *Social Justice Research*, 6, p. 5-18.



Organ, D. W. & Paine, B. J. (1999). A new kind of performance for industrial and organizational psychology: Recent contributions to the study of organizational citizenship behavior. *International Review of Industrial and Organizational Psychology*, 14, p. 337-368.

Organ, D. W. & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, 48, (4), 775-802.

Orr, J. M., Sackett, P. R., & Mercer, M. (1989). The role of prescribed and nonprescribed behaviors in estimating the dollar value of performance. *Journal of Applied Psychology*, 74 (1), 34-40.

Pedhazor, E. J. (1997). *Multiple regression in behavioral research* (3<sup>rd</sup> ed.). Orlando, FL: Harcourt Brace College.

Podsakoff, P. M., Ahearne, M. & MacKenzie, S. M. (1997). Organizational Citizenship Behavior and the Quantity and Quality of Work Group Performance. *Journal of Applied Psychology*, 82(2), 262-270.

Podsakoff, P. M. & Mackenzie, S. B. (1997). Impact of organizational citizenship behavior on organizational performance: A review and suggestions for future research. *Human Performance*, 10(2), 133-151.

Podsakoff, P. M. & MacKenzie, S. B. (1994). Organizational citizenship behavior and sales unit effectiveness. *Journal of Marketing Research*, 31, 351-363.

Podsakoff, P. M, MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26, (3), 513-563.

Powell, G. N. & Butterfield, D. A. (1989). The “good manager:” Did androgyny fare better in the 1980s? *Group and Organization Studies*, 14, (2), 216-233.

Pulakos, E. D. & Wexley, K. N. (1983). The relationship among perceptual similarity, sex, and performance ratings in manager-subordinate dyads. *Academy of Management Journal*, 26, (1), 129-139.

Pulakos, E. D., White, L. A., Oppler, S. H. & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74 (5), p. 770-780.

Robbins, T. L. and Denisi, A. S. (1993). Moderators of sex bias in the performance appraisal process: A cognitive analysis. *Journal of Management*, 19, (1), 113-126.

Rosen, B. & Jerdee, T. H. (1973). The influence of sex-role stereotypes on evaluations of male and female supervisory behavior. *Journal of Applied Psychology*, 57, 44-48.

Roethlisberger, F. J. & Dickson, W. J. (1939). *Management and Worker*. Cambridge, MA: Harvard University Press.

Sausser, W. I., Evans, K. L., & Champion, C. H. (1979, March). *Two hundred and fifty scaled incidents of college classroom behavior*. Paper presented at the Southeastern Psychological Association annual conference, New Orleans.

Schwab, D. P. & Grams, R. (1985). Sex-related errors in job evaluation: A “real-world” test. *Journal of Applied Psychology*, 70, (3), 533-539.

Shein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, 57, (2), 95-100.

Schein, V. E., Mueller, R., & Jacobson, C. (1989). The relationship between sex role stereotypes and requisite management characteristics among college students. *Sex Roles, 20*, 103-110.

Senger, J. (1971). Managers' perceptions of subordinates' competence as a function of personal value orientations. *Academy of Management Journal, 14*, 415-423.

Skarlicki, D. P. & Latham, G. P (1996). Increasing citizenship behavior within a labor union: A test of organizational justice theory. *Journal of Applied Psychology, 81*, 161-169.

Smith, A., Organ, D., & Near, J. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology, 68*, p. 653-663.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4<sup>th</sup> ed.). New Jersey: Lawrence Erlbaum Associates.

Stroh, L. K., Brett, J. M., & Reilly, A. H. (1992). All the right stuff: A comparison of female and male managers' career progression. *Journal of Applied Psychology, 77*, (3), 251-260.

Sutton, M. (1998). *Does the type of behavior and gender make a difference?* Unpublished master's thesis, University of South Florida, Florida.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan mckay versus john mckay: Do gender stereotypes bias evaluations? *Psychological Bulletin, 105*(3), 409-429.

Turban, D. B. & Jones, A. P. (1988). Supervisor-subordinate similarity: Types, effects, and mechanisms. *Journal of Applied Psychology, 73*, (2), 228-234.

Van Scotter, J. R. (2000). Relationships of task performance and contextual performance with turnover, job satisfaction, and affective commitment. *Human Resource Management Review, 10*, (1), 79-85.

Van Scotter, J. R., Motowidlo, S. J., & Cross, T. C. (2000). Effects of task performance on contextual performance on systematic rewards. *Journal of Applied Psychology, 85*, (4), 526-535.

Wakabayashi, M., Graen, G., Graen, M., & Graen, M. (1988). Japanese management progress: Mobility into middle management. *Journal of Applied Psychology, 73*, (2), 217-227.

Wayne, S. J. & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. *Journal of Applied Psychology, 75*(5), 487-499.

Wayne, S. J. & Green, S. A. (1993). The effects of leader-member exchange on employee citizenship and impression management behavior. *Human Relations, 46*, (12), 1431-1440.

Wexley, K. N., Alexander, R. A., Greenawalt, J. P., & Couch, M. A. (1980). Attitudinal congruence and similarity as related to interpersonal evaluations in manager-subordinate dyads. *Academy of Management Journal, 23*, (2), 320-330.

Zalesny, M. D. & Highhouse, S. (1992). Accuracy in performance evaluations. *Organizational Behavior and Human Decision Processes, 51*, 22-50.

## Appendices

## Appendix A

### Consent Form

The purpose of this research study is to test a new system for evaluating professor performance in the psychology department at USF. You will be helping to pilot the new evaluation materials. You are being asked to participate because as a student, your input is important to the pilot process. You will be placed

After reading a list of performance statements, you will be asked to evaluate the professor's performance and recommend rewards for the professor. Following the evaluations, you will be asked a few demographic and educational questions. The entire process will take approximately 20 minutes to complete. There are no risks for your participation.

You will receive no compensation for your participation. Your privacy and research records will be kept confidential to the extent of the law. Authorized research personnel, employees of the Department of Health and Human Services and the USF Institutional Review Board may inspect the records from this research project.

The results of this study may be published. However, the data obtained from you will be combined with data from other people in the publication. The published results will not include your name or any other information that would in any way personally identify you. Numbers will be used to identify your survey and only the investigators will have access to the questionnaires, which will be kept in a locked office on campus.

Your decision to participate in this research study is completely voluntary. You are free to participate in this research study or to withdraw at any time. If you choose not to participate, or if you withdraw, there will be no penalty or loss of benefits that you are entitled to receive.

If you have any questions about this research study, contact Lisa Wilkinson at: [lvwilkin@helios.acomp.usf.edu](mailto:lvwilkin@helios.acomp.usf.edu) or 974-5034.

If you have questions about your rights as a person who is taking part in a research study, you may contact a member of the Division of Research Compliance of the University of South Florida at 813-974-5638.

Appendix A: (Continued)

**Your Consent—By signing this form I agree that:**

- I have fully read or have had read and explained to me this informed consent form describing a research project.
- I have had the opportunity to question one of the persons in charge of this research and have received satisfactory answers.
- I understand that I am being asked to participate in research. I understand the risks and benefits, and I freely give my consent to participate in the research project outlined in this form, under the conditions indicated in it.
- I have been given a signed copy of this informed consent form, which is mine to keep.

\_\_\_\_\_  
Signature of Participant

\_\_\_\_\_  
Printed Name of Participant

\_\_\_\_\_  
Date

**Investigator Statement**

I have carefully explained to the subject the nature of the above protocol. I hereby certify that to the best of my knowledge the subject signing this consent form understands the nature, demands, risks and benefits involved in participating in this study.

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Printed Name of Investigator

\_\_\_\_\_  
Date

**Institutional Approval of Study and Informed Consent**

This research project/study and informed consent form were reviewed and approved by the University of South Florida Institutional Review Board for the protection of human subjects. This approval is valid until the date provided below. The board may be contacted at (813) 974-5638.

**Approval Consent Form Expiration Date:**

**Revision Date:** \_\_\_\_\_

## Appendix B

### Cover Story

#### **Professor Evaluations**

We are working on the development of a new system for evaluating and rewarding professors in the psychology department at USF. We are currently collecting pilot data that will be used to help make decisions about the implementation of the new system. As students, your input is important to the pilot process.

On the next several pages you will see comments regarding the performance of "Dr. Smith" (for confidentiality purposes, the real name of the professor is protected). This information was collected from students and other members of the psychology department during the last semester. Following the performance statements, you will be asked to review these comments and then evaluate various aspects of the professor's performance and provide your opinion on rewards that could be offered to this professor.

A fair and objective evaluation and reward process for professors is key to the university system. Please read and respond to the materials carefully. Thanks in advance for your assistance with this project.



## Appendix C

### Performance Log

Psychology Professor

Michael Smith or Michelle Smith

These comments are taken from evaluations provided by students and colleagues of Dr. Smith during Fall 2001 semester. The following comments are in no particular order. Please read them carefully because you will be asked to evaluate Dr. Smith's performance.

- He gave details about the material in class but never elaborated beyond them.
- His test questions are usually reasonable, but are sometimes tricky.
- He expresses his own personal satisfaction in being a faculty member at USF when asked by outsiders.
- He can generally be persuaded to sacrifice own personal interests for the good of the psychology department.
- He looks for additional productive work to do when his own normally scheduled duties are completed.
- He accepts invitations to attend teaching and research enhancement courses offered by the university.
- When students or other faculty ask for help, he can usually be counted on to suggest solutions to their problems.
- He assigned ten pages of reading before each class period.
- He leaves promptly after giving his lecture.
- He required a term paper, oral presentation, and weekly tests.
- He continuously referred back to his notes while attempting to lecture.
- When others ask for his help because they are overloaded, he can usually be counted on to take on some additional tasks.
- When students or faculty ask for help, he can usually be counted on to listen to their personal problems and provide emotional support.

### Appendix C: (Continued)

- He requires a lot of memorization for his class.
- He generally completes work on time, unless deadlines are very short.
- His tests usually cover 3 or 4 chapters of the book.
- He might offer suggestions for changes to university procedures to make them more efficient.
- He marks off for poor class attendance.
- He takes advantage of available opportunities to develop own research and teaching skills when such opportunities present themselves.

## Appendix D

### Citizenship Performance Scale

We are interested in evaluating professors at the University of South Florida on their citizenship performance. Please read the following definition of citizenship performance carefully and circle the number that corresponds to the answer you think is appropriate for Dr. Smith.

Citizenship Performance: behaviors that go beyond an employee's normal job duties. These behaviors involve doing extra tasks or making an extra effort that supports the organization and assists co-workers in performing their duties.

#### Subjective Scale

**1. Dr. Smith is helpful to co-workers and students:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**2. Dr. Smith is courteous with co-workers and students:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**3. Dr. Smith is cooperative when working with colleagues and students:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**4. Dr. Smith is loyal to USF:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**5. Dr. Smith represents the USF well to outsiders:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

Appendix D: (Continued)

**6. Dr. Smith complies with University rules and procedures:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**7. Dr. Smith is persistent and gives extra effort when needed:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**8. Dr. Smith takes initiative to get things done when it is needed:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**9. Dr. Smith is dedicated to self-development in order to improve his own performance:**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

Appendix D: (Continued)

**Objective Scale of Citizenship Performance**

**1. Compare Dr. Smith's performance against the average citizenship performance of all professors and circle the letter grade you feel that Dr. Smith deserves for his citizenship performance:**

A      A-      B+      B      B-      C+      C      C-      D+      D      D-      F

**2. . Compare Dr. Smith's performance against the average citizenship performance of all professors and place an x in the box that corresponds to the percentage you would give to Dr. Smith for his citizenship performance, with one hundred percent equaling the best citizenship performance that a professor can have.**

0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	

**3. If you were to review the citizenship performance of nine other USF psychology professors in addition to this one, how would you guess this professor would compare?**

1	2	3	4	5	6	7	8	9	10
Professor									Professor
Would									Would
Rank First									Rank Tenth
(The Best)									(The Worst)

Appendix E

Overall Performance Scale

Use the information from the performance log of Dr. Smith to evaluate his performance. Please circle the number that corresponds to your answer.

**1. This professor makes an important contribution to the University of South Florida.**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**2. This professor is extremely valuable to the University of South Florida.**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**3. This professor would be extremely costly for the University of South Florida to replace.**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**4. This professor is a vital part of the University of South Florida.**

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

**5. This professor is indispensable to the University of South Florida.**

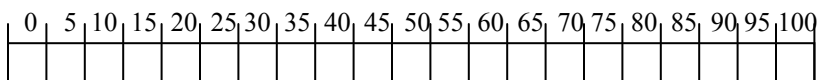
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

Appendix E: (Continued)

**6. Circle the letter grade you feel that Dr. Smith deserves for his overall performance:**

A      A-      B+      B      B-      C+      C      C-      D+      D      D-      F

**7. Compare Dr. Smith's performance against the average performance of all professors and place an x in the box that corresponds to the percentage you would give to Dr. Smith for his performance, with one hundred percent equaling the best performance that a professor can have.**



**8. If you were to rank the overall performance of nine other professors in addition to this one, how would you guess this professor would compare?**

1	2	3	4	5	6	7	8	9	10
Professor									Professor
Would									Would
Rank First									Rank Tenth
(The Best)									(The Worst)

Appendix F

Reward Recommendations Scale

**Directions: Please indicate the extent that you would recommend the professor for each of the following rewards. Circle the number that corresponds to your choice.**

**1. Would you recommend this professor for a teaching award?**

Would definitely not recommend	Would probably not recommend	Neutral	Would recommend with some minor reservations	Would recommend without reservation
1	2	3	4	5

**2. Would you recommend this professor for a promotion into a more prestigious teaching position?**

Would definitely not recommend	Would probably not recommend	Neutral	Would recommend with some minor reservations	Would recommend without reservation
1	2	3	4	5

**3. Would you recommend this professor for a salary bonus?**

Would definitely not recommend	Would probably not recommend	Neutral	Would recommend with some minor reservations	Would recommend without reservation
1	2	3	4	5

**4. If 10 professors were competing for a teaching award, where would you rank this professor?**

1	2	3	4	5	6	7	8	9	10
Most Deserving Of the Award (Best)									Least Deserving Of the Award (Worst)



Appendix F: (Continued)

**5. Compare Dr. Smith's performance against the average performance of all professors, and circle the amount you would recommend for this professor to receive as a salary bonus:**

0            100            200            300            400            500

**6. If 10 professors were competing for a more prestigious teaching position, where would you rank this professor?**

1	2	3	4	5	6	7	8	9	10
Most									Least
Deserving									Deserving
Of the									Of the
Position									Position
(Best)									(Worst)

## Appendix G

### Demographic Information

Please provide some information about yourself:

1. Age: \_\_\_\_\_
2. Gender: **M** \_\_\_\_\_ **F** \_\_\_\_\_
3. Race or ethnicity:  
African-American \_\_\_\_\_  
White, non-Hispanic \_\_\_\_\_  
Hispanic/Latina \_\_\_\_\_  
Asian/Pacific Islander \_\_\_\_\_  
Native American/Alaskan \_\_\_\_\_
4. What is your college grade level?  
Freshman \_\_\_\_\_  
Sophomore \_\_\_\_\_  
Junior \_\_\_\_\_  
Senior \_\_\_\_\_  
Graduate Student \_\_\_\_\_
5. Please list your major: \_\_\_\_\_
6. Do you have any work experience? Yes No
7. Do you have any experience as a supervisor? Yes No

**When answering the following question, do not flip back.**

8. What was the gender of the professor that you just evaluated? Male Female

**Thank you for your time, we appreciate your participation!**

## Appendix H

### Participant Debriefing

The purpose of this study was to investigate how raters differentiate rateses when evaluating performance and when recommending rewards. Specifically, we are interested in how citizenship performance scores effect performance ratings and reward recommendations. However, in past studies there have been mixed results with studies finding that people are differentially evaluated for their citizenship performance (eg. Lovell, et. al., 1999) and studies that did not find a significant difference in how people are evaluated for their citizenship performance (Organ & Ryan, 1995).

On account of these mixed findings the shifting standards model was applied. This social psychology theory predicts that stereotyped differences will be more apparent when objective rather than subjective scales are used (Biernat, Manis, & Nelson, 1991). Therefore, someone who is stereotyped to receive higher ratings on citizenship performance will receive higher ratings on objective scales, but not subjective scales. Further, this study will investigate how citizenship performance expectations will affect their ratings on overall performance evaluations and reward recommendations.

**On account of the nature of this study, please do not reveal the purpose of this study to other USF students because they may be future participants.**

Please view the following references for more information:

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgement. Journal of Personality and Social Psychology, 60, (4), 485-499.

Lovell, S. E., Kahn, A. S., Anton, J., Davidson, A., Dowling, E., Post, D., & Mason, C. (1999). Does gender affect the link between organizational citizenship behavior and performance evaluation? Sex Roles, 41 (5-6), p. 469-478.

Organ, D. W. & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. Personnel Psychology, 48, (4), 775-802.