EXPLORING IDIOSYNCRATIC RATER EFFECTS IN TEAMS

Rusty T. Gillain

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

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
November 2017

ABSTRACT

EXPLORING IDIOSYNCRATIC RATER EFFECTS IN TEAMS

by Rusty T. Gillain

The present study examined the occurrence of rater idiosyncratic bias, bias specific to an individual rater, on peer performance ratings in teams. Specifically, ratee personality and biological sex, rater personality and biological sex, and the interaction of rater and ratee characteristics were assessed to determine any biasing effects on peer ratings of team member task and social role performance. Multilevel regression analyses indicated minor idiosyncratic effects from peers in assigning performance ratings. Of the effects found, most notable was that rater extraversion was positively related to peer ratings of task and social role performance and that female ratees tend to receive significantly lower scores in task performance roles than their male counterparts. Finally, evidence revealed that narrow facets of rater personality differentially predict ratings of role performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The use of performance appraisal, and specifically performance ratings, in making

personnel decisions has become common practice among organizations, and the use of

multisource feedback systems has been gaining popularity (Conway & Huffcutt, 1997). Multiple

sources may be utilized to gather performance ratings including supervisors, peers, and the

workers themselves (self-ratings) based on their observation and judgment of past behavior

(Conway & Huffcutt, 1997). Ratings are then used as the basis for personnel decisions regarding

promotion, compensation, training, and development (Cleveland, Murphy, & Williams, 1989;

Conway & Huffcutt, 1997).

Ideally, performance measurements are based on a combination of objective, personnel,

and judgmental indices. However, objective indices are difficult to observe, and therefore,

performance measurement is mostly dependent on judgmental indices (Landy & Farr, 1980).

Given this fact, the reliability and accuracy of performance ratings has been called into question

due to their largely judgmental nature (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988).

Subjective performance ratings have tended to result in several common biases including halo,

leniency or severity, and central tendency or restriction of range (Holzbach 1978, Saal, Downey,

& Lahey 1980).

Given that ratings are made by a number of different sources, the literature typically has

shown that rating source accounts for a significant portion of variance in 360-degree

performance ratings (Woehr, Sheehan, & Bennett, 2005). A number of different variables have

been investigated in order to understand why the reliability and correlations of performance

ratings among these different sources vary. Specifically, research has centered on

nonperformance factors such as rater characteristics, ratee demographic characteristics, and the interaction between rater and ratee characteristics in biasing performance ratings (Borman, 1997; Landy & Farr, 1980). Other research has tended to focus on the characteristics of the rating instrument itself and the context in which the rating occurs (Conway & Hufcutt, 1997; Harris & Schaubroeck, 1988). More recently researchers have examined idiosyncratic rater effects, biases specific to the rater, which may be to blame for the distortion of judgmental performance ratings (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen, Mount, & Goff, 2000).

While the literature on traditional multisource performance appraisal is abundant, relatively little research has been conducted on the reliability and validity of performance appraisals in a team context (May & Gueldenzoph, 2006). As a result of the increased utilization of teams in the workplace, many organizations use peer ratings to evaluate the performance of the individuals working in these teams (Devine, Clayton, & Phillips, 1999). Therefore, it is important to understand biases that may exist among peer performance appraisals in teams and how they may influence member evaluations.

Investigating peer performance ratings in a team context can unveil some critical similarities and differences to ratings in an individual context. This is because differences in the appraisal context can often influence the kinds of biases that are prevalent. For example, Farh, Cannella, and Bedeian (1991) found that peer ratings are more elevated and uniform when ratings are used for evaluative purposes rather than when used for developmental purposes. Peer ratings are affected by additional variables including rater competence (Barclay & Harland, 1995), rater self-esteem (Tziner, Murphy, & Cleveland, 2002), rater attitudes towards the rating process (Fedor, Bettenhausen, & Davis, 1999), and personality (Tziner, Murphy, & Cleveland, 2002).

According to Eby, Meade, Paris, and Douthitt (1999), team members rely on shared expectations for teamwork in task performing groups. Therefore, peers working in groups may be more familiar with one another and have a common perception of what is expected in terms of performance (Eby et al, 1999). Theoretically, then, team members may not be influenced by idiosyncratic sources of rater bias to the same extent as raters in a traditional appraisal context. Alternatively, it may be that team members are prone to certain types of idiosyncratic biases when rating but not others. For example, team members may not be influenced by surface-level (i.e. demographic) similarity among them because they have developed a sufficient level of familiarity with each other. On the other hand, idiosyncratic biases stemming from personality characteristics of the rater may exert greater influence on peer ratings in the team because members' expectations and perceptions of team behaviors can be colored by the general tendencies in behaviors and attitudes that are described by personality traits. While idiosyncratic biases appear to have a significant effect on performance ratings in an individual context (Mount et al, 1998; Scullen et al, 2000), there is a lack of literature regarding rater idiosyncrasies in groups. Therefore, it is important to investigate the effects of peer performance ratings in a team context.

The scarcity of research regarding the effect of nonperformance factors, including personal characteristics such as personality and demographics, on performance evaluations in a team context necessitates further research (May & Gueldenzoph, 2006). Do personality and demographic variables influence the quality of peer role performance ratings? Holding the content and the context of the rating instrument itself constant, examining personal characteristics of the rater and ratee and the interaction of these characteristics in a team context may help clarify whether nonperformance factors influence peer performance ratings in teams.

Close examination of this information may help to determine some of the specific sources of rater idiosyncrasies that are responsible for bias in peer performance ratings.

Performance Appraisal

According to Longenecker and Goff (1992), research estimates that around 90 percent of large organizations implement some type of performance appraisal system. Moreover, a reported $152 million was spent on multisource feedback rating systems in 1992 (Romano, 1994). Much of the research on performance appraisal in the past 75 years has focused on ways to improve performance appraisal and more importantly the relationship between performance ratings and actual job performance (Murphy, 2008).

Performance appraisal systems have a number of different uses for the organization. They may be used as a basis for employee motivation, salary decisions, employee development, human resource decisions, and for the purpose of setting goals and performance planning (Longenecker & Goff, 1992). According to Bernardin, Hagan, Kane, and Villanova (1998) three prescriptions help to characterize successful performance management systems. These include precision in defining and measuring performance, deriving content from internal and external customers, and taking into account and correcting for situational constraints in the performance management process (Bernardin et al, 1998). A successful performance appraisal system is one that accurately measures employee performance outcomes.

Most organizations utilize performance appraisal systems in which multiple sources are used to evaluate a target (Dalessio, 1998). Multisource feedback (MSF) is defined as "evaluations gathered about a target participant from two or more rating sources, including self, supervisors, peers, direct reports, internal customers, external customers, and vendors or

suppliers" (p. 278). The "rater" is the individual giving the rating and the "ratee" is the target of these ratings. 360-degree feedback is another type of multisource feedback that involves receiving feedback from above, below, and lateral positions in the organization. Upward feedback, receiving feedback from a subordinate, and team member feedback, receiving feedback from a peer, are also considered multisource feedback if evaluations include additional sources such as a self-rating (Dalessio, 1998).

According to Harris and Schaubroeck's (1988) meta-analysis, there was a moderate correlation between self-supervisor ratings (ρ = .35) and self-peer ratings (ρ = .36), and a relatively high correlation between peer and supervisor ratings (ρ= .62). This result lends mixed support for the use of performance ratings due to the variation in rating accuracy by different sources. Another meta-analysis conducted by Conway and Huffcutt (1997) examined the psychometric properties of multisource ratings in relation to job performance. The results showed varying reliability in performance ratings with subordinates being the lowest (.30), peers being second (.37), and supervisors with the highest reliability at (.50). Viswesvaran, Ones, and Schmidt (1996) found a similar result for supervisory ratings of job performance, with a mean interrater reliability of .52. Correlations between rating sources are relatively low among subordinate ratings, correlated (ρ = .14) with self-ratings, (ρ = .22) with peer, and (ρ = .22) with supervisor ratings, and for self-ratings, correlated (ρ= .19) with peer ratings and (ρ = .22) with supervisor ratings (Conway & Huffcutt, 1997). Supervisor-peer ratings correlations tend to be higher, around (ρ = .34), indicating that supervisors are more reliable raters than the other sources (Harris & Schaubroeck, 1988).

The judgmental nature of most performance appraisal systems opens them up to some common forms of rater biases (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Woehr

5

& Huffcutt, 1994). Such biases include halo, leniency or severity, and central tendency and restriction of range (Holzbach, 1978; Saal, Downey, & Lahey, 1980). Halo effect involves assigning a consistent rating based on some sort of global impression of the ratee, this in turn leads to a failure in discriminating differences in the certain aspects of the ratee's behavior. Leniency or severity involves assigning scores to raters that are either too low or too high in regards to the actual behavior of the ratee. Central tendency involves assigning performance ratings that are near the midpoint of the scale and not assigning scores too high or low in either direction. Restriction of range is often discussed in relation to central tendency and is defined by Saal et al (1980) as "the extent to which obtained ratings discriminate among different ratees in terms of their respective levels" (p. 417). While these errors are types of biases that raters make in the process of evaluating a target, Saal et al (1980) argued that it is also important to examine interrater reliability or agreement in the assessment of rating quality. This involves comparing the ratings from two independent raters to see the extent to which they overlap with one another. This helps us to establish convergent validity (Lawler, 1967). Decisions being made of the basis of performance ratings should take potential biases into account before coming to any definitive conclusions about the ratee.

More recently, idiosyncratic rater effects have been gaining attention as a source of bias in the performance appraisal literature. O'Neill, McLarnon, & Carswell (2015) explained that this tendency is a source of systematic variance that is unique to the particular rater involved and is not correlated across raters. According to Hoffman, Lance, Bynum, and Gentry (2010), an idiosyncratic rater effect is a "systematic effect that is common only to an individual rater" (p. 126). Studies have determined that this idiosyncratic rater effect accounts for a large amount, over half, of the variance in job performance ratings (Mount et al, 1998; Scullen et al, 2000).

Rater traits also attribute to the variance in performance ratings in addition to individual rater effects (Mount et al, 1998). Moreover, it is important to study the traits of the rater, and traits of the rater relative to the ratee as a way of investigating some of these idiosyncratic tendencies.

<div style="text-align:center">Performance Appraisal in Teams</div>

The personnel psychology literature is abundant with research on performance appraisal and the validity and reliability of their use has been well documented (May & Gueldenzoph, 2006). Different avenues have been investigated in order to understand why the reliability of performance ratings among different sources varies. A major reason includes differences in the characteristics of the rater and ratee, specifically personality and demographics (Borman, 1997; Landy & Farr, 1980). Other research has looked into the interaction, similarities or differences, of characteristics among the rater and ratee in biasing performance ratings (Antonioni & Park, 2001).

Teamwork and the structuring of work within teams has become commonplace in many modern organizations (Salas, Cooke, & Rosen, 2008). Kozlowski & Bell (2003) defined the team as a group composed of at least two or more people, working together, interdependently, to accomplish common goals. Work is divided among the group so that the knowledge, skills, and abilities (KSAs) of individual workers may complement one another and bolster the team's performance (Peeters, van Tuijl, Rutte, & Reymen, 2006).

Research on traditional multisource performance appraisal is widespread but few studies have examined the validity and reliability of peer performance ratings in a team context (May & Gueldenzoph, 2006). According to Kane & Lawler (1978), a peer rating is one in which team members use a specified rating scale to rate members based on their performance in the group. In

their review of 14 studies, they found discouraging median reliability and validity values for peer ratings, .45 and .35, respectively. According to Farh et al, (1991), peer ratings are the most useful for giving feedback but possess lower validity, reliability, and more biased measurements than peer nomination and peer rankings. Moreover, peer ratings used for evaluative purposes have low user acceptance and tend to suffer from halo, leniency, and restriction of range biases as well as low reliability and validity (Farh et al, 1991).

In order to examine peer ratings in a team context it is important to choose appropriate performance criteria. More specifically, it is important to designate criteria content that encompasses performance for each member of the team. Mumford, Campion, and Morgeson (2006) accomplished this task by developing a team role typology, which allows for the evaluation of individual role performance behaviors in a team context. Peer ratings can be used to assess performance of the individual behaviors conducted by the different members of a team. According to Mumford et al (2006), a useful way to conceptualize performance in a team context is to consider the individual role behaviors of each team member. A role can be defined as a cluster of goal-directed behaviors characteristic of a person in a specific situation (Stewart, Manz, & Sims, 1999). The successful completion of these roles is a determinant of team performance (Belbin, 1993). Each member of the team plays a separate but complementary role, which in turn contributes to the successful operation of the team as a whole (Mumford et al, 2006). Thus, team performance can be evaluated by assessing each individual's performance in a number of different roles.

Consensus on a comprehensive typology of team roles is a recent occurrence in the literature. For the better part of the 20th century team roles were loosely organized around two critical functions (Mumford et al, 2006). The two critical functions for successful team

performance were task execution and maintaining the social functions of the team (Benne &
Sheats, 1948; Bales, 1950). A comprehensive framework of team roles was not created until
Mumford et al (2006) reviewed the extant literature and developed a typology of team roles.
They further validated this typology through the use of a situational judgment test (SJT)
(Mumford, Van Iddekinge, Morgeson, & Campion, 2008).

Mumford et al (2006) considered roles both internal and external, which were necessary
for the successful performance of the team. This resulted in the creation of roles related to the
task, social, and boundary spanning functions of the team. Task roles are necessary for carrying
out the work of the team, and consist of five roles (Contractor, Creator, Contributor, Completer,
and Critic). An example of one task role is the Contractor role, which involves coordination and
organization of team members for task completion. Social roles involve maintenance of the
social environment to ensure the proper functioning of the team, and are divided into three roles,
(Communicator, Cooperator, and Calibrator). An example of social roles is the Communicator
role, which takes into consideration the feelings and opinions of other team members and
expresses personal sentiments. Finally, boundary-spanning roles are those behaviors individuals
perform outside of the team, and consist of two roles (Coordinator and Consul). For example, the
Consul role acts as a liaison between upper management and the team in order to procure the
resources necessary for successful performance (Mumford et al, 2006).

Although the identification of relevant member behaviors is an important development,
the utility of any framework of team member behaviors will be limited if it cannot be accurately
assessed. Past evidence has determined that idiosyncratic rating tendencies, which include
individual characteristics such as personality and demographics, are a potential source of bias in
performance ratings (Mount et al, 1998; Scullen et al, 2000). Given this evidence, there may be a

9

biasing effect of personal characteristics on peer performance ratings of individual task and social role performance in a team context. To determine the extent of this biasing effect, the personality and biological sex characteristics of raters, ratees, and characteristics of raters in relation to ratees should be examined.

<div align="center">Rater Effects</div>

A review by Landy & Farr (1980) identified several nonperformance factors attributable to performance rating biases, including personal characteristics that the rater possesses. The literature has tended to look at characteristics such as personality and demographics as a source of rater bias (Landy & Farr, 1980).

Rater Personality

Research has shown that idiosyncratic rating tendencies account for over half of the variance performance ratings (Mount et al, 1998; Scullen et al, 2000). These idiosyncratic tendencies are specific to the rater, and include such factors as personality. Therefore, factors such as individual personality traits may be responsible for the upward or downward bias of raters in assigning performance ratings of task and social roles in a team context. Researchers have examined the Five-Factor Model (FFM) of personality (openness, conscientiousness, agreeableness, extraversion, and emotional stability) as it relates to performance appraisal (Harari, Rudolph, & Laginess, 2015; Randall & Sharples, 2012; Tziner Murphy, & Cleveland, 2002; Sinha, Mesmer-Magnus, and Viswesvaran, 2012; Yun, Donahue, Dudley, & Mcfarland, 2005). Harari et al (2015) examined FFM personality traits as a source of systematic nonperformance variance among job performance ratings in their meta-analysis. They found that

rater personality traits (agreeableness, extraversion, and emotional stability) accounted for around 6-22% of the variance in performance ratings.

According to various researchers, personality traits characterize the ways in which people will react to different social demands (Mischel & Schoda, 1995; Roberts, 2009). According to *Trait Activation Theory,* personality traits help to direct individuals' behaviors in ways that are trait relevant (Tett & Guterman, 2000). Situational demands imposed by the team context may influence raters to behave in trait relevant ways, thus personality traits will affect the ratings given by individuals based upon the specific personality traits they possess. Additionally, this may influence idiosyncratic rater tendencies. For example, certain raters may be more likely to view specific personality traits, (e.g. conscientiousness), as more salient and related to task completion, attending to behaviors seen as more important or detrimental and ignoring less salient ones (Tett & Guterman, 2000). Moreover, raters may observe behaviors associated with personality traits incongruent to themselves, weighting them as less important for job completion. Therefore, it would follow that these incongruities in trait specific behaviors would result in a lower overall rating in task and social role performance.

According to the Five Factor Model (Costa & McCrae, 1992), people high in conscientiousness possess a high level of competence, achievement orientation, and organization. They tend to complete tasks in a dedicated and persistent manner. A study by Tziner, Murphy, & Cleveland (2002) specifically examined the moderating effects of conscientiousness on attitudes and beliefs about performance appraisal and rating behavior. Their findings indicated that raters high on conscientiousness discriminated among ratees more often, and were less likely to rate based upon their personal attitudes of performance appraisal than those low in conscientiousness. Additionally, Bernardin, Cooke, and Villanova (2000) found

11

that conscientiousness is negatively related to rating leniency (-.31). Therefore, it is assumed that raters high in the FFM personality trait of conscientiousness are less likely to assign inflated ratings.

> *Hypothesis 1a*: Rater conscientiousness will be negatively related to their ratings of peers on task and social role performance.

People high in emotional stability tend to be more level headed and consistent (Costa & McCrae, 1992), and less easily swayed by their emotions. Conversely, the low end of emotional stability, neuroticism, tends to encompass traits such as anxiety and depression (Costa & McCrae, 1992). Neurotic individuals tend to exhibit less stable behavior emotionally, which can be detrimental to rating objectivity and accuracy (Longenecker, Sims, & Gioia, 1987). This is relevant to peer performance ratings because it may be the case that raters exhibiting a low level of emotional stability exhibit less positive relationships with ratees, and as a result tend to rate them more negatively (Duarte, Goodson, & Klich, 1994). In support of this idea, Harari et al (2015) found evidence that raters exhibiting a high level of emotional stability (i.e. low in neuroticism) tend to rate peer performance more positively ($\rho=.12$) than those low in emotional stability (i.e. high in neuroticism). Therefore, in a team context, we expect that team members possessing a high level of emotional stability (i.e. low in neuroticism) will rate peers more positively than those exhibiting low emotional stability (i.e. high in neuroticism).

> *Hypothesis 2a:* Rater emotional stability will be positively related to their ratings of peers on task and social role performance

According to the Five Factor Model, people who are high in agreeableness tend towards a more trusting, altruistic, and cooperative attitude with others (Costa & McCrae, 1992). Harari et al (2015) studied the FFM personality factor of agreeableness in relation to performance ratings.

They found that agreeableness was significant and positively related to performance ratings, $p =$ .25 (Harari et al, 2005). Yun et al (2005) studied rater personality (agreeableness), rating format, and rating context and the effect that each of these factors exhibits on performance appraisal. They found that raters high on agreeableness were more likely to give elevated ratings when they expected to give face-to-face feedback. Raters low on agreeableness showed little difference in elevation of ratings across different formats (Yun et al, 2005). In contrast, Randall and Sharples (2012) found that raters high in agreeableness are more likely to give lenient ratings of poor performance regardless of the rating context, and Bernardin et al (2000) found that agreeableness positively related to rating level at (r=.33). Therefore, we expect that team members high in agreeableness will be more likely to give elevated ratings in both task and social roles.

Hypothesis 2b: Rater agreeableness will be positively related to their ratings of peers on task and social role performance.

People high in extraversion are usually more assertive, gregarious, and comfortable in group settings (Costa & McCrae, 1992). Sinha et al (2012) studied personality antecedents of self-other rating discrepancies. Consistent with the literature, their findings indicated that individuals high in extraversion were less likely to be self-reflective and more susceptible to social desirability in assigning self-ratings. When examining personality and peer ratings, Harari et al (2015) found a positive relationship ($\rho = .12$) between extraversion and peer performance ratings. Therefore, in a team context we expect that members scoring higher on personality trait extraversion will be more likely to assign elevated ratings than those scoring low in extraversion.

Hypothesis 2c: Rater extraversion will be positively related to their ratings of peers on task and social role performance

Rater Sex

Another possible source of rating variation can be attributed to the demographic characteristics specific to the rater (Landy & Farr, 1980). Several studies have examined demographic variables, the most prevalent characteristic being rater biological sex, in a performance appraisal context. Studies have looked the sex of the rater as a possible influence of bias in a general performance appraisal context (Mobley, 1982; Nevill, Stephenson, & Philbrick, 1983; Pulakos, White, Oppler, & Borman, 1989). Other research has examined rater sex in a peer-rating context, determining that sex does have a significant effect on performance ratings, albeit in an unexpected way (Ammons & Brooks, 2011).

The examination of rater gender effects on performance ratings has tended to yield small and non-significant results. Mobley (1982) conducted a field study, which looked at supervisor and employee gender effects on performance appraisals. He found that less than five percent of the variance in performance ratings was attributable to rater sex. Another study by Nevill et al (1983) found that female raters gave more severe scores, both high and low, to successful and unsuccessful job applicants than male counterparts, a finding contrary to the common stereotype. Finally, a study by Pulakos et al (1989) examined gender and performance ratings and found extremely small effect sizes.

Even though findings for rater gender effects on performance ratings were small, other studies have looked at rater gender in a group context (Ammons & Brooks, 2011). Ammons & Brooks (2011) collected self and peer assessments from students working on a group project in an undergraduate course in order to examine gender biases in performance ratings. Similar to the findings of Nevill et al (1983), in regards to specific work behaviors, females assign more severe scores when rating the extent to which a team member contributed to the group.

14

According to *Social Role Theory* (Eagly & Steffen, 1984), the observed differences between men and women, in regards to social behavior and personality, stem from the distribution of men and women into differing social roles (Eagly, 1997; Eagly, Wood, & Diekman, 2000). Historically women have performed more domestic work, they tend to hold more traditionally feminine occupations than men, and operate at lower levels of the organizational hierarchy. Conversely, men tend to hold more enterprising and male-dominated occupational roles (Eagly et al, 2000). As a result of the large amount of men holding high level, male-dominated positions women may be forced to act more competitively in the workforce. Moreover, according to *Conservation of Resources* (COR) theory, female workers may feel compelled to act more shrewdly, assigning lower performance ratings to protect their own position or status in the team.

According *Conservation of Resources* (COR) theory (Hobfoll, 1988, 1989, 1998), the loss of resources is the primary source of stress, while resource gain is of importance to preventing stress. Principle 2 of COR theory, resource investment, states that people invest resources in order to protect themselves from loss of resources, to recover from resource loss, and to gain resources. Similarly, corollary 1 of COR theory states that people with more initial resources are less vulnerable to loss and tend to gain resources more easily (Hobfoll, 2001). Given that women in the workforce have fewer resources to begin with, they may feel more vulnerable to loss of resources, i.e. position or status. This view aligns with the popular literature that women are more shrewd and severe than men in assigning ratings. They must take strides to ensure the appearance of power and status in the organization is maintained. In a team context, women may feel that they are outnumbered, that their voice is less important, and that their position within the group is dependent upon how accurately they perform in relation to other

team members. Therefore, consistent with the empirical literature on *Conservation of Resources* theory, we expect that females will assign more severe ratings than men in a team context.

Hypothesis 3: Female raters will provide generally lower ratings than males on peer evaluations of task and social role performance.

Ratee Effects

Another avenue of research has examined the personal characteristics of the ratee, including personality and biological sex, and the effects that these traits may have on performance appraisal (Landy & Farr, 1980). According to research by Tsui and Barry (1986), rater reports of interpersonal affect towards ratees can be linked to leniency, halo, and restriction of range rating errors. Taggar and Brown (2006) found that when team members gave positive or negative feedback to other members prior to performance ratings, these other members' ratings resulted in increased leniency or severity and restriction of range (Taggar & Brown, 2006). Evidence has shown that ratees receive different performance-rating scores based upon their biological sex (Ammons & Brooks, 2011; Hamner, Kim, Baird, & Bigoness, 1974; Hartman, Griffeth, Crino & Harris, 1991; Nevill et al, 1983; Schmitt & Hill, 1977). Additionally, there has been clear evidence that specific personality factors differentially influence performance in specific occupations (Barrick & Mount, 1991).

While it is important to consider personal characteristics of the ratee in assessing performance-rating bias, characteristics of the ratee such as personality are usually considered to be reflective of actual true performance scores, not bias. For this reason, we only examine ratee biological sex as a potential source of bias, as it is assumed that biological sex may be a source of idiosyncratic rating tendencies.

16

Ratee Sex

In examining performance appraisals, biological sex of the ratee should be a possible consideration as a factor contributing to idiosyncratic biases. According to Landy and Farr (1980) research has investigated the sex of the ratee as a possible source of bias in performance evaluations. Researchers have come to differing conclusions about the effect that biological sex has on performance rating bias. Some studies show that ratings tend to favor male ratees while others favor female ratees. A major consideration in regards to ratee sex and performance rating bias is *Social Role Theory* (Eagly & Steffen, 1984) and the stereotypical male or female nature of the job being rated (Landy & Farr, 1980).

According to *Social Role* Theory (Eagly & Steffen, 1984), women occupy more social and female-oriented occupational roles and men occupy dominant male-oriented occupational roles (Eagly et al, 2000). Thus, traditionally, when rating men in male-oriented jobs (e.g. managerial positions) and rating women in female-oriented jobs (e.g. secretarial positions), the ratings tend to favor men and women, to a lesser extent, respectively (Schmitt & Hill, 1977). However, several studies to date have found contrasting findings that women receive higher performance ratings in male oriented, nonprofessional non-managerial jobs (Hamner et al, 1974; Mobley, 1982). According to Ammons and Brooks (2011) women tend to receive more positive open-ended feedback than male students and are rated higher by both genders in peer evaluations. While these studies contrast the central tenant of *Social Role Theory*, there appears to be some underlying causes as to why women received higher performance ratings.

Several studies have looked at personal characteristics and the context surrounding the ratee. Nevill et al (1983) found that women in the branch bank manager role were viewed more positive when success was known, and were rated more positively when raters scored higher on

the Women as Managers Scale (WAMS) (Nevill et al, 1983). Additionally, a study by Hartman et al (1991) determined that female employees with more masculine personal characteristics were rated more highly regardless of the gender of the rater and the gender stereotypes regarding the job. Ratings depended more upon the ratee's personal characteristics than their gender (Hartman et al, 1991).

Thus, while some of the recent literature points to the idea that women receive higher ratings in certain circumstances, this tends to hold true only when personal characteristics and success of the particular female ratee are known (Hartman et al, 1991; Nevill et al, 1983). Therefore, consistent with *Social Role Theory* (Eagly & Steffen, 1984), we expect that women will receive more favorable performance ratings in social roles and men will receive more favorable ratings in task roles.

> *Hypothesis 4*: Female ratees will receive generally lower ratings than males on peer ratings of task role performance.

> *Hypothesis 5*: Female ratees will receive generally higher ratings than males on peer ratings of social role performance.

Ratee Personality

According to Barrick and Mount (1991) different FFM personality factors are related to performance in specific occupations. Given that ratee personality factors are considered antecedents of true performance in different jobs, and not necessarily an indicator of bias, we do not make any specific hypotheses regarding ratee personality and team role performance.

However, we expect that rater perceptions of specific ratee personality factors may be a potential source of idiosyncratic bias in performance ratings. In addition to the rater personality

main effects suggested previously, rater and ratee characteristics may interact in a way that will conjointly influence rater perceptions. More specifically, I suggest that similarity in rater and ratee characteristics, such as personality and biological sex, may cause an upward bias on peer performance ratings in a team context.

Interactions Between Rater and Ratee Characteristics

The relationship between rater and ratee in a team context is important when considering idiosyncratic bias. It may be the case that the perception of member behaviors depends conjointly on rater and ratee characteristics. Whereas ratee characteristics may influence their inclination to perform certain behaviors, the characteristics of the rater influences the salience and the value attached to that behavior when making an evaluation. As mentioned earlier, *Trait Activation Theory* helps to explain how individuals will respond in certain situations (Tett & Guterman, 2000). Some trait specific behaviors are more readily apparent in certain situations, and some behaviors are weighted more heavily in judging the success of performance outcomes based upon similarity.

Additionally, *Social Comparison Theory* helps explain why peers tend to focus on observing task relevant competencies and abilities (Mumford, 1983). Festinger (1954) developed *Social Comparison Theory*, which posits that individuals evaluate their abilities and opinions against others, usually peers, when no objective information exists. Regarding peer evaluation, individuals are likely to compare themselves to their peers in order to obtain information about their own performance based on the performance of those they are working closely with, either cooperatively or competitively. If an individual perceives another team member to possess deficient level of one trait, in comparison to their own level of that trait, necessary to perform a

role successfully, (e.g. lower conscientiousness than is needed for successful task role performance), this team member will be rated more severely. Individuals will focus on obtaining the implicit competencies and abilities necessary to perform a role successfully. They will also judge peers' performance based upon their possession of these task relevant competencies and abilities (Mumford, 1983). As a result, there may be bias in peer ratings when judging whether certain individuals possess the "necessary" or "required" traits (i.e. FFM personality traits) for the successful completion of the job.

Stemming from *Trait Activation Theory* (Tett & Guterman, 2000), *Social Comparison Theory* (Festinger, 1954), and the work of Mumford (1983), studies have examined the degree of similarity between rater and ratee personality traits as a source of rating bias (Antonioni & Park, 2001; Strauss, Barrick, & Connerly, 2001). Other studies have examined demographic variables such as biological sex and how the agreement or disagreement of these factors affects rating biases (Hartel, Douthitt, Hartel, & Douthitt, 1999; Pulakos & Wexley, 1983). While previous research has been beneficial, it is important to investigate the interactive effects occurring in teams. It could be that group members working closely together are less susceptible to interactive biases. On the other hand, team members may have more of an opportunity to observe and interpret specific behaviors, allowing more opportunity for rating bias.

Rater and Ratee Personality

According to Tsui et al (1986), performance ratings are affected by the interpersonal affect (e.g. positive affect) between the rater and the ratee. It may be the case that observation and perception of different personality traits differs from rater to rater. According to *Trait Activation Theory,* which may help to explain this effect, personality traits are consistent within

an individual, distinct across individuals, and dictate an identifiable pattern of behavior (Tett & Guterman, 2000). A situation in which there is a high amount of aggressive stimuli may trigger an individual to behave in a highly aggressive way. This may help to explain some of the idiosyncrasies specific to the rater in regards to his or her assessment of the ratee in specific situations. We can expect that an individual rater with a high level of a specific personality trait, e.g. conscientiousness, will be more likely to observe this trait in a team member. Thus, given that this trait is more salient we would expect the rater to assign more favorable ratings to those similar in personality, and less favorable ratings to those ratees dissimilar to one's self. Additionally, based on *Social Comparison Theory*, it is likely that individuals will judge their performance and the performance of others based on their personal characteristics such as FFM personality factors (Mumford, 1983). Therefore, similarity or disagreement of certain personality traits may have implications for bias in performance ratings.

Antonioni and Park (2001) examined at the effects of FFM personality similarity on peer ratings of contextual work behaviors in dyads. The authors found that when controlling for interpersonal affect, rater-ratee similarity in conscientiousness was significantly and positively related to peer ratings of three variables including consideration of others, interpersonal communication, and self-management responsibility. When interpersonal affect was not controlled for, peer ratings of consideration of others and interpersonal communication remained significant. Additionally similarity in agreeableness was significantly related to peer ratings of Self-Management Responsibility (Antonioni & Park, 2001). A similar study conducted by Strauss et al (2001) looked at actual differences and perceived differences in personality and their effect on performance ratings. The author's found that actual similarity in emotional stability and perceived similarity in conscientiousness, emotional stability, and extraversion,

were related to performance ratings. Furthermore, greater actual and perceived similarity was associated with higher peer and supervisor ratings of sales performance (Strauss et al, 2001).

As mentioned earlier, conscientiousness is related to performance in all types of occupations (Barrick & Mount, 1991). We expect that similarity in conscientiousness will be significant and positively related to peer task and social role performance ratings. Furthermore, we expect that raters possessing a higher level of conscientiousness will assign elevated ratings to peers that are also high in conscientiousness, and raters possessing a lower level of conscientiousness will assign elevated ratings to peers that are also low in conscientiousness.

> *Hypothesis 6***:** There will be a significant interaction among rater-ratee FFM trait conscientiousness such that greater similarity in this trait will be positively related to ratings of peers on task and social role performance.

Additionally, the FFM traits of extraversion, emotional stability, agreeableness, and openness had relationships with occupations possessing a social component (Barrick & Mount, 1991). We expect that these FFM traits will be significantly and positively related to peer ratings of social role performance. Furthermore, we expect that raters possessing a higher level of extraversion, emotional stability, agreeableness, and openness will assign elevated ratings to peers that are also high in these traits, and raters possessing a lower level of extraversion, emotional stability, agreeableness, and openness will assign elevated ratings to peers that are also low in these traits.

> *Hypothesis 7a***:** As similarity in rater-ratee FFM trait extraversion increases, greater similarity in this trait will be positively related to ratings of peers on social role performance.

Hypothesis 7b: As similarity in rater-ratee FFM trait emotional stability increases, greater similarity in this trait will be positively related to ratings of peers on social role performance.

Hypothesis 7c: As similarity in rater-ratee FFM trait agreeableness increases, greater similarity in this trait will be positively related to ratings of peers on social role performance.

Hypothesis 7d: As similarity in rater-ratee FFM trait openness increases, such greater similarity in this trait will be positively related to ratings of peers on social role performance.
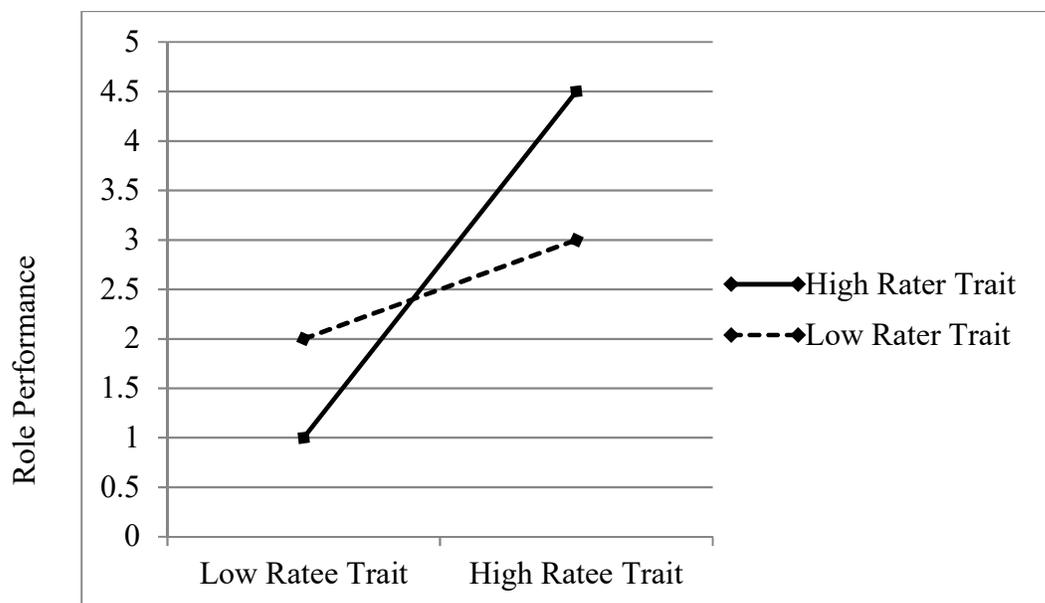
Figure 1. *Expected Relationship Between Rater and Ratee Personality Traits*

Rater and Ratee Sex

Another possible source of bias may be due to the *similarity* or *same group hypothesis* (Landy & Farr, 1980). The main idea of the *same group hypothesis* is that raters of one group will assign elevated performance ratings to members of the same group (e.g. men will assign

23

higher performance ratings to male ratees due to an inherent preference for individuals similar to themselves). Studies have examined whether the similarity or dissimilarity of rater-ratee gender have an effect on the quality of performance ratings both directly and indirectly (Hartel et al, 1999; Pulakos & Wexley, 1983).

Hartel et al (1999) examined participants' openness to dissimilarity using the Ideal Employee Inventory (IEI) as a potential predictor of gender discrimination in performance ratings. They found that participants who were less open to dissimilarity significantly and negatively rated members of a minority group, Native American women, when compared to ratings of Caucasian men. Pulakos & Wexley (1983) found evidence for the similarity hypothesis in a field setting. Their findings indicated that manager's perceived similarity of the subordinate interacted with both manager sex and subordinate sex, such that perceived similarity and sex led to higher peer performance ratings. We argue in favor of the *same group hypothesis*, in that members of the same gender will likely assign more favorable ratings in both task and social role performance.

*Hypothesis 8***:** There will be an interaction among rater-ratee gender such   that raters from the same gender as the ratee will rate peers more highly on task and social role performance.
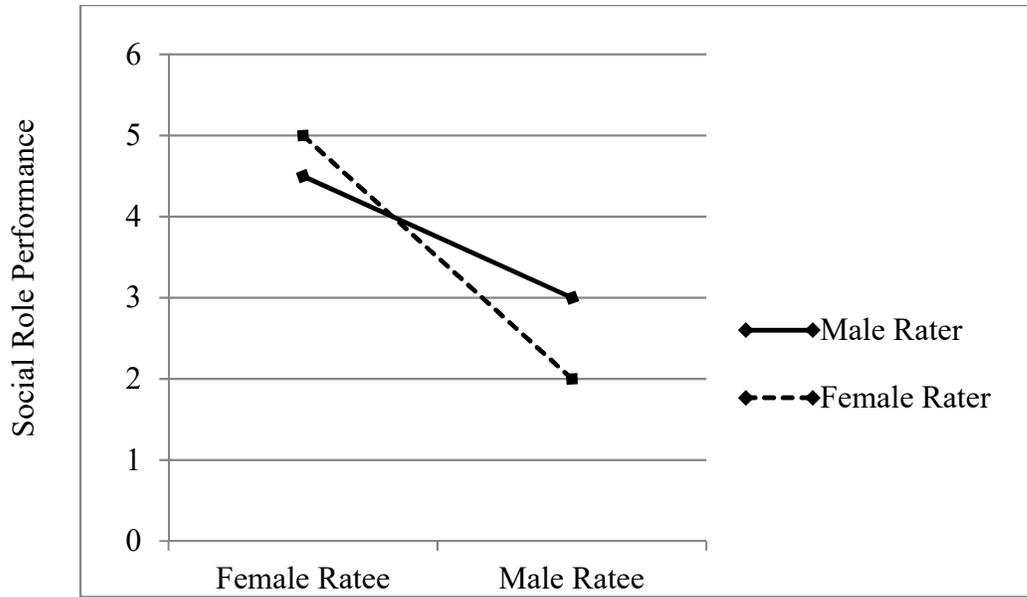
Figure 2. *Expected Relationship Between Rater and Ratee Sex and Social Role Performance*
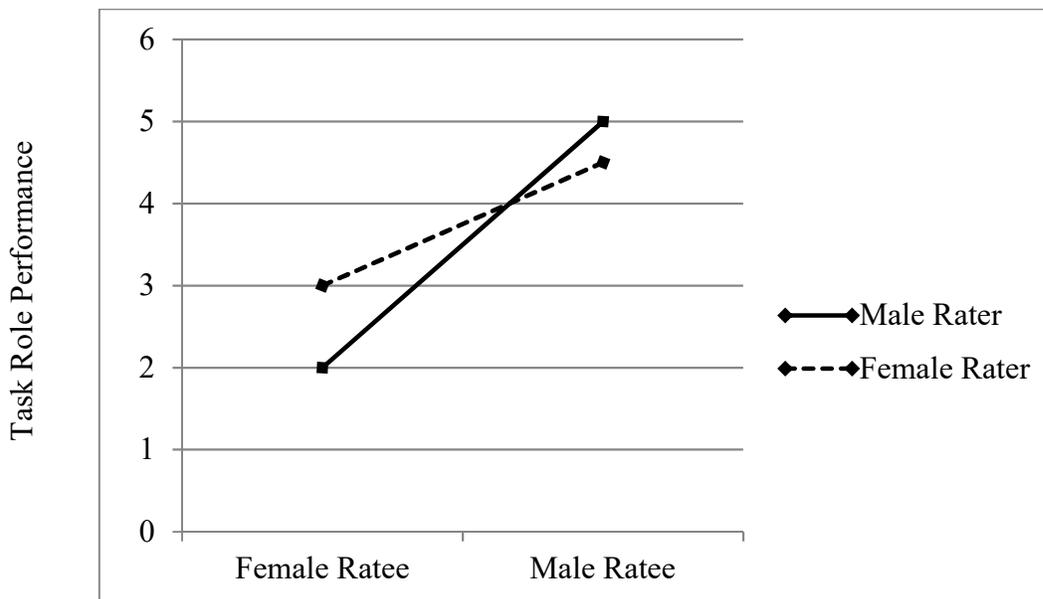
Figure 3. *Expected Relationship Between Rater and Ratee Sex and Task Role Performance*

Supplementary Analyses

The bandwidth-fidelity dilemma in personality assessment is an issue that has constantly

plagued personality researchers. A major question arises when it comes to deciding whether

narrow or broad personality traits possess a higher level of fidelity in predicting job performance. While conceding that it depends on the reason that personality assessment is being utilized, one camp believes that broad bandwidth personality traits used to measure performance outcomes possess greater predictive validity than narrow traits (Ones & Viswesvaran, 1996) Additionally, these researchers contend that narrow personality facets do not possess incremental validity over and above broad factors (Salgado, Moscoso, Sanchez, Alonso, Choragwicka, & Berges, 2015). The other camp of researchers tends to believe that the use of narrow personality traits does tend to possess predictive power over and above that of broad personality composites (Paunonen & Ashton, 2001).

While the usefulness of broad personality traits in predicting job performance cannot be overlooked, examining specific elements of broad traits may provide additional information that would be otherwise overlooked (Tett, Steele, & Beauregard, 2003). While both the predictors, FFM personality and biological sex, and criterion, task and social roles, in this study are both broad, it may be beneficial to examine some of the salient sub-facets of global FFM personality composites. For example, the global personality factor extraversion contains facets of friendliness and assertiveness (Goldberg, 1999). Thus, friendliness and assertiveness may differentially predict social and task role performance, respectively.

Research Question 1: Will examining facets of global personality composites differentially influence ratings of task and social role performance?

Present Study

The purpose of the following study is to discover some of the possible rater idiosyncrasies that may be responsible for bias of peer performance ratings in a team context.

Moreover, the goal is to assess rater and ratee personality factors and biological sex in a team context, comparing them to one another in order to examine the biasing effects they may have on performance ratings in teams. According to Murphy (2008), the correlation between performance and ratings of performance depends on the amount of nonperformance factors influencing ratings and the correlations between these nonperformance factors and job performance. The current study aims to identify some nonperformance factors that account for variance in peer performance ratings in teams. Specifically, we want to examine characteristics of personality and biological sex to determine if they attribute to upward or downward bias in performance ratings.

Unlike previous studies, this paper will assess peer ratings of team role performance in a group context to investigate possible rater idiosyncrasies that may occur in this type of setting. Mumford et al (2006) established a team role typology, which they validated through the use of a situational judgment test (Mumford et al, 2008). This typology can be broken down into broad categories of both task and social roles, which together contribute to the successful performance of the team. Task roles (Contractor, Creator, Contributor, Completer, and Critic) consist of behaviors that include organizing and coordinating other team members while social roles (Communicator, Cooperator, and Calibrator) include those behaviors contributing to a positive work climate.

This paper will take a multilevel modeling approach in order to examine Five Factor Model (FFM) personality characteristics, openness, conscientiousness, agreeableness, extraversion, and emotional stability (Costa & McCrae, 1992), and biological sex of raters and ratees to examine the main effects of these variables on task and social role performance in a team context. It will also examine the similarity and differences of these characteristics among raters and ratees in order to understand the upward or downward biasing effects these factors

may elicit on performance ratings. Ultimately, this information may help to shed information on

possible rater idiosyncrasies, specifically personality and sex, which may be responsible for

biasing peer performance ratings in a team context.

CHAPTER II

METHODS

Participants and Procedure

The data used in this study were archival in nature and derived from a thesis conducted

by (Quinn, 2014). The study participants consisted of undergraduate students attending Central

Michigan University (CMU) in the College of Business Administration and were recruited

through the school of Business Marketing. Students participated in a mandatory multidisciplinary

group project (CORE), which spanned the course of one semester during their junior year.

Participants worked in groups, ranging in size from four to six members, to create a business

plan for a participating company, which included topics regarding company finances, marketing,

accounting, and sales. Data were collected across two different time points, one at the beginning

of the academic semester collecting predictor variables such as personality and demographics

and one at the end of the semester collecting role performance ratings. Students who took part in

this study, both through the optional online and in person paper and pencil test, were given extra-

credit for their participation. The sample size for ratees (level-1) variables ranged from n= 245 to

n = 498, nested within n=90 raters (level 2) and n=47 teams (level-3).

Measures

Personality

GoldBerg's (1999) International Personality Item Pool (IPIP) facet scale was utilized to

measure personality. For theoretical reasons, lower-level facets were combined to create

corresponding FFM composites (e.g. facets of friendliness and assertiveness were combined to

create the global composite emotional stability). Participants were asked to rate the extent to which they 1 (strongly disagree) to 5 (strongly agree) with a set of statements on a 5-point Likert scale. The internal consistencies for each facet scale ranged from $\alpha = .73$ to $.87$.

Role Performance

Mumford et al's (2008) measure of team role performance was used to measure ratings of team task and social role performance. The scale consists of the critical role performance behaviors detailed in Mumford et al's (2006) team role typology. Members of every team were given a brief description of each of the ten roles and then asked to rate themselves and their peers on each one. Items tap into the extent to which each member and their peers, respectively, engage in each specific role. Ratings for the occurrence of these behaviors were recorded on a 5-point Likert scale ranging from 1(no extent) to 5 (a very great extent). Sample items on the TRT for the Calibrator role include the extent to which the team member: "Helps to settle conflicts between members of the team," "Suggests positive ways for the team members to interact, such as taking turns, showing respect, and being open to new ideas," and "Steps in to help resolve the difficulties, if there are negative feelings in the team." (Mumford et al, 2008). Self and other ratings of Contractor, Creator, Contributor, Completer, and Critic roles were combined to create task roles, and self and other ratings of Communicator, Cooperator, and Calibrator were combined to represent social roles. Boundary spanning roles, Consul and Coordinator, were excluded because they are roles tangential to the vital functioning of the group. Internal consistencies for the role performance rating scales ranged from $\alpha = .74$ to $.87$ and interrater reliabilities ranged from .81 to .88.

Data Analysis

According to LaHuis and Avis (2007), Multilevel Random Coefficient (MRC) models,

otherwise known as Hierarchical Linear Modeling (HLM), may be used to answer a number of

different research questions in regards to performance ratings. First, MRC can be used to parse

out how rater attributes influence performance ratings while controlling for ratee-level variables.

Next, it can be used to test ratee-by-rater interactions and the effect they have on performance

ratings. Finally, you can use MRC to control for team or organization level effects to see how

these characteristics influence rater effects.

Descriptive statistics, intercorrelations, and internal consistencies were assessed for all

study variables. Hypothesized rater effects, ratee effects, and ratee-by-rater interactions were

tested using MRC models. A three-level hierarchical model was specified with ratees (Level 1)

nested within raters (Level 2) and raters nested within teams (Level 3). If all ratees had

participated as raters in this research study, then raters and ratees would be best analyzed using a

cross-classified HLM model because these data points would be mutually nested within each

other. However, because not all ratees participated or provided ratings, there were more ratees

than raters for this study. As a result, we specified the aforementioned three-level model above.

Additionally, self-ratings of performance (where the rater and ratee are the same participant)

were excluded in data analysis to reduce the dependency between rater and ratees, as well as to

ensure that there was no self-serving bias confounding the results.

The first step was to assess variance in the intercept, the ICC for role performance, to

justify the use of a multilevel regression model (Hox, 2010). The next step was to assess rater

effects (Hypotheses 1-3), to examine whether rater personality variables and biological sex are

predictors of role performance. In this step we also assessed ratee level effects (Hypotheses 4 and

31

5), to see if rater personality and biological sex were significant predictors of role performance. Finally, in the third step rater predictor slopes were allowed to vary across ratees in order to assess interaction effects (Hypotheses 6-8). Specifically, rater personality traits were entered as level 2 predictors of the relationship between ratees' personality traits and the performance scores they received from the rater.

CHAPTER III

RESULTS

Data Cleaning

The data utilized in this study were derived from research conducted by Quinn (2014), and were assessed for quality at each time point during data collection by using reading check items. Reading check items asked respondents to indicate a specific response to ensure that they were carefully answering each item. Respondents who indicated an incorrect response option to any of the reading check items were removed from the sample. Only 4 participants from the original sample of 251 students were removed due to non-purposeful responding (Quinn, 2014). Given that data were collected at two different time points, the original sample decreased from n = 247 respondents at time one to n =175 respondents at time 2. Personality and demographic variables were collected at time 1 and role performance ratings were collected at time 2.

Descriptive Statistics and Intercorrelations

Descriptive statistics, intercorrelations, and internal consistencies of study variables were assessed prior to restructuring data for multilevel analyses. As mentioned previously, the sample consisted of 175 respondents at both time points. Reliabilities ranged from an acceptable .69 to .86 and correlations for study variables ranged from -.40 to .97.

Table 1. *Descriptive Statistics and Intercorrelations of FFM Personality*

Variable	M	SD	1	2	3	4	5	6	7
1. Openness	3.45	.59	(.69)						
2. Conscientiousness	4.00	.47	.17*	(.86)					
3. Extraversion	3.86	.59	.16*	.42**	(.82)				
4. Agreeableness	4.04	.54	.20**	.45**	.23**	(.78)			
5. Emotional Stability	2.26	.64	-.28**	-.25**	-.40**	-.35**	(.69)		
6. Task Role Ratings	.00	1.00	-.18	-.01	.19	-.04	-.14	(.98)	
7. SR Ratings	.00	1.00	-.16	-.13	.03	-.21*	-.10	.97**	(.98)

Note. n = 175. *p* < .05, **p* <. 01. SR= Social Role. Numbers in parentheses = reliability estimates.

Data Structuring

Next, data were restructured according to Raudenbush, Bryk, Cheong, Congdon, and Toit's (2011) suggestions to allow for multilevel analysis in HLM7 software. A three-level structure was specified in which ratees' biological sex and personality information and corresponding task and social role performance scores were situated at Level-1. Raters' biological sex and personality information and their corresponding task and social role performance scores were situated at Level-2 and aggregate team-level biological sex and personality information as well as task and social role performance scores were situated at Level-3. After running the data in HLM7, the number of ratees at Level-1 ranged from $n = 245$ to $n = 288$ and were nested in $n = 90$ raters at Level-2. Furthermore, team members were nested within $n = 47$ teams at the highest level of analysis, Level-3.

Moreover, prior to analyses, team members' performance ratings of each role were aggregated into an overall composite of task and social role performance ratings, respectively. Next, role performance scores as well as biological sex and personality predictor scores were linked to each rater. Each rater gave multiple ratings to the members in their team, therefore,

these ratings and corresponding biological sex and personality predictors were linked to each

ratee as indicated by the format in Table 2.

Table 2. *Data Format for Comparison of Rater and Ratee Characteristics*

Team	Rater	Ratee	Rater Sex	Ratee Sex	Rater Personality	Ratee Personality	Role Performance
1	1	1	1	1	3.33	3.33	-2.02
1	1	2	1	2	3.33	4.50	-2.60
1	1	3	1	2	3.33	…	1.33
1	1	4	1	1	3.33	…	2.67
1	1	5	1	1	3.33	…	0.83
1	1	6	1	2	3.33	…	-0.33
1	2	1	2	1	4.50	3.33	2.67
1	2	2	2	2	4.50	4.50	2.83
1	2	3	2	1	4.50	…	-2.55
1	2	4	2	2	4.50	…	2.47
1	2	5	2	2	4.50	…	3.88
1	2	6	2	1	4.50	…	4.50
…	…	…	…	…	…	…	…

Note. Not every rater assigned complete role performance ratings for each member of the team, therefore some instances of ratee role performance ratings were incomplete.

Data Aggregation

Prior to assessing hypothesized main effects and interactions, the residual files at each

respective level of data were examined to check assumptions including nesting, normality, and

heterogeneity of variances. After examining scatterplots and histograms it was concluded that the

data were normal and linearly related to one another.

Additionally, the intraclass correlation coefficients (ICC)s, which indicate how much

variance in the intercept is accounted for by group membership, for role performance were

calculated to assess nesting effects and to justify the use of HLM (Hox, 2010). The ICC(1) for

task role performance was .07 at the rater level and .59 at the team level. The ICC(1) for social

role performance was .02 at the rater level and .42 at the team level.

Multilevel Descriptive Statistics and Intercorrelations

Descriptive statistics and intercorrelations among ratee and rater level personality, biological sex, and role performance ratings were assessed and are listed in Table 3. These analyses were conducted in SPSS prior to analysis in HLM7 software. Study variables were standardized prior to analysis to allow for easier interpretation. Therefore, all study variables possessed a mean of zero and standard deviation of one except for rater and ratee biological sex. Biological sex was coded zero for males and one for females, such that negative coefficients indicate performance scores that favor males over females.

The sample size for ratee level variables ranged from $n = 245$ to $n = 409$. Correlations between ratee personality variables and peer ratings of task role performance ranged from $r = -.18$ to $r = .17$, the lowest being emotional stability and the highest being openness. Correlations between ratee variables and peer ratings of social role performance ranged from $r = -.18$ to $r = .10$, the lowest being emotional stability and the highest being openness. Finally, ratee sex was negatively correlated to peer ratings of task and social role performance at $r = -.15$ and $r = -.18$, respectively.

The correlations among rater level personality variables and peer ratings of task role performance ranged from $r = .10$ to $r = .24$, the lowest being emotional stability and the highest being extraversion. The correlations among rater level personality variables and peer ratings of social role performance ranged from $r = -.10$ to $r = .14$, the lowest being agreeableness and the highest being extraversion. Finally, rater sex was correlated with ratings of task and social role performance at, $r = -.01$ and $r = -.13$, respectively.

Table 3. *Descriptive Statistics and Intercorrelations of Study Variables*

	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Level 1: Ratee Variables															
1. Task Roles	407	—													
2. Social Roles	409	.90	—												
3. Ratee Sex	288	-.15	-.18	—											
4. Ratee Openness	262	.17	.10	-.07	—										
5. Ratee Conscientiousness	251	-.13	.05	.12	.02	—									
6. Ratee Extraversion	253	-.13	-.10	.14	-.07	.28	—								
7. Ratee Agreeableness	258	-.07	-.07	.24	.10	.40	.20	—							
8. Ratee Emotional Stability	245	-.18	-.16	.09	-.31	-.10	-.35	-.22	—						
Level 2: Rater Variables															
9. Rater Sex	90	-.01	-.13	.24	.07	.10	.13	.12	-.02	—					
10. Rater Openness	90	-.05	-.04	.08	.37	.08	.07	.11	-.04	-.03	—				
11. Rater Conscientiousness	90	.12	.03	.08	.03	.32	.11	.16	-.08	.13	.06	—			
12. Rater Extraversion	90	.24	.14	.09	.04	.11	.35	.03	-.09	.08	.06	.30	—		
13. Rater Agreeableness	90	-.02	-.10	.09	.10	.16	.01	.43	-.06	.27	.16	.42	.10	—	
14. Rater Emotional Stability	90	-.10	-.10	.04	-.05	-.07	-.08	-.07	.34	.04	-.32	-.19	-.34	-.23	—

Note. All ratee and rater level variables standardized (Mean = 0; SD = 1). All correlations analyzed in SPSS software without addressing the nested structure, therefore significance levels not listed.

Main Effects

Rater Personality

Main effects of rater FFM personality variables are listed in Table 4. Rater

conscientiousness was significant and positively related to ratings of task role performance, $\beta = .11; t(42) = 1.97, p = .03$, and was not significantly related to peer ratings of social role

performance, $\beta = .02; t(42) = .50, p = .62$. Therefore, Hypothesis 1 was not supported.

Emotional stability was not significantly related to ratings of peers in task role performance, $\beta = -.02; t(42) = -.55, p = .58$, or social role performance, $\beta = .01; t(42) = .21, p = .83$. Therefore, Hypothesis 2a was not supported. Additionally, rater agreeableness was not

positively related to ratings of peers in either task, $\beta = -.01; t(42) = -.18, p = .86$, or social

role performance, $\beta = -.08; t(42) = -1.77, p = .08$. Therefore, Hypothesis 2b was not

supported. Finally, in support of Hypothesis 2c, rater extraversion was significantly and

positively related to peer ratings of both task, $\beta = .20; t(42) = 4.49, p < .01$, and social role

performance $\beta = .09; t(42) = 2.23, p = .03$.

Table 4. *Main Effects of Rater Characteristics*

Variable	Coefficient	SE	d.f.	p-value
Predicting Task Role Performance				
Rater Sex	.02	.10	42	.72
Openness	-.02	.05	42	.66
Conscientiousness	.11*	.05	42	.03
Extraversion	.20**	.04	42	< .01
Agreeableness	-.01	.05	42	.86
Emotional Stability	-.02	.05	42	.58
Predicting Social Role Performance				
Rater Sex	-.02	.04	42	.47
Openness	-.06	.04	42	.14
Conscientiousness	.02	.04	42	.62
Extraversion	.09*	.04	42	.03
Agreeableness	-.08	.04	42	.08
Emotional Stability	.01	.04	42	.83

Note. $*p < .05$, $**p < .01$. Coefficient = Standardized Beta (β). *SE* = Standard Error, *d.f.* = degrees of freedom.

Rater Sex

Main effects for rater biological sex on peer ratings of task and social role performance are also listed in Table 4. In general, main effects of peer ratings on task role performance, $\beta = .02; t(42) = .36, p = .72$, and peer ratings on social role performance, $\beta = -.02; t(42) = -.74, p = .47$, were small and nonsignificant. Therefore, Hypothesis 3 was not supported.

Ratee Sex

Hypothesis 4 predicted that female ratees would receive generally lower ratings than males on peer ratings of task role performance which was significant and supported, $\beta = -.08; t(148) = -2.26, p = .03$. Surprisingly, females also received slightly lower ratings than males on peer ratings of social role performance, but this relationship was not statistically significant, $\beta = -.08; t(148) = -2.71, p = .13$. Therefore, Hypothesis 5 was not supported.

Rater and Ratee Personality

Hypotheses 6-7d predicted that similarity in specific FFM personality traits would upwardly bias peer ratings of task and social role performance. As shown in Table 5, there was significant interaction between rater and ratee conscientiousness such that greater similarity led to lower peer ratings of task role performance $\beta = -.03; t(40) = -2.78, p < .01$, however, this value was very small and revealed a pattern that was contrary to the hypothesized interaction. There was a nonsignificant interaction among rater and ratee conscientiousness and peer ratings of social role performance, shown in Table 6, $\beta = -.001; t(40) = -.14, p = .89$. Therefore, Hypothesis 6 was not supported. Additionally, as shown in Table 6, there were no other significant interaction effects among rater and ratee personality. Therefore, Hypotheses 7a-d were not supported.

Table 5. *Cross-Level Interaction Effects for Task Role Performance*

Variable	Coefficient	SE	d.f.	p-value
Ratee Sex*Rater Sex (γ_{110})	-.04	.03	40	.21
Intercept (γ_{000})	-.43*	.13	46	< .01
Ratee Sex (γ_{100})	-.07	.03	40	.05
Rater Sex (γ_{010})	.03	.06	40	.56
Ratee Conscientiousness*Rater Conscientiousness	-.03**	.01	40	< .01
Intercept (γ_{000})	-.45**	.13	46	< .01
Ratee Conscientiousness (γ_{100})	-.12*	.05	40	.03
Rater Conscientiousness (γ_{010})	.13*	.06	40	.04

Note. *p < .05, **p <. 01. Coefficient = Standardized Beta (*β*). *SE* = Standard Error, *d.f.* = degrees of freedom.
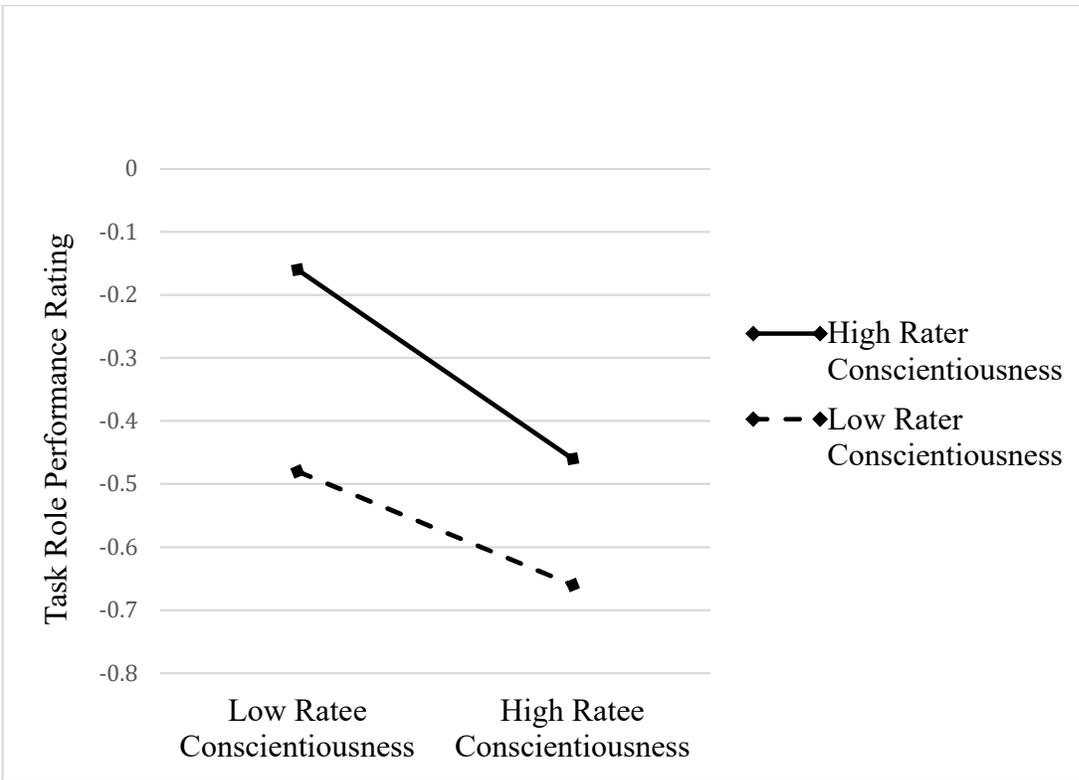
Figure 4. *Interaction Between Rater and Ratee Conscientiousness*
Note. Scale of x-axis ranges from 1 standard deviation below the mean to 1 standard deviation above the mean.

Rater and Ratee Sex

Hypothesis 8 suggested that similarity in rater and ratee sex would influence peer ratings of task and social performance in such a way that it upwardly biased ratings. As shown in Table 6, similarity in rater and ratee sex did not significantly impact peer task role performance ratings, $\beta = -.04; t(40) = -1.29, p = .21$, or peer ratings of social role performance, $\beta = -.04; t(40) = -1.48, p = .14$. Thus, Hypothesis 8 was not supported.

Table 6. *Cross Level Interaction Effects for Social Role Performance*

Variable	Coefficient	SE	d.f.	p-value
Ratee Sex*Rater Sex	**-.04**	**.03**	**40**	**.14**
Intercept (γ_{000})	-.44**	.12	46	< .01
Ratee Sex (γ_{100})	-.06	.03	40	.06
Rater Sex (γ_{010})	-.04	.05	40	.42
Ratee Openness*Rater Openness	**.01**	**.02**	**40**	**.61**
Intercept (γ_{000})	-.44**	.13	46	< .01
Ratee Openness (γ_{100})	.04	.03	40	.18
Rater Openness (γ_{010})	-.07	.05	40	.20
Ratee Conscientiousness*Rater Conscientiousness	**-.001**	**.01**	**40**	**.89**
Intercept (γ_{000})	-.46**	.12	46	< .01
Ratee Conscientiousness (γ_{100})	-.01	.03	40	.66
Rater Conscientiousness (γ_{010})	.03	.05	40	.55
Ratee Extraversion*Rater Extraversion	**-.02**	**.02**	**40**	**.32**
Intercept (γ_{000})	-.40**	.12	46	< .01
Ratee Extraversion (γ_{100})	-.01	.04	40	.71
Rater Extraversion (γ_{010})	.11*	.04	40	.02
Ratee Agreeableness*Rater Agreeableness	**-.02**	**.02**	**40**	**.38**
Intercept (γ_{000})	-.44**	.13	46	< .01
Ratee Agreeableness (γ_{100})	-.02	.06	40	.69
Rater Agreeableness (γ_{010})	-.05	.05	40	.35
Ratee ES*Rater ES	**-.03**	**.02**	**40**	**.17**
Intercept (γ_{000})	-.42**	.13	46	< .01
Ratee ES (γ_{100})	-.01	.04	40	.68
Rater ES (γ_{010})	.001	.05	40	.98

*Note. *p < .05, **p <. 01.* Coefficient = Standardized Beta (β). *SE* = Standard Error, *d.f.* = degrees of freedom.

Facet-Level Analyses

Additional analyses were conducted in order to assess whether narrow facets of personality exhibited a unique relationship with peer task and social role performance ratings. According to some researchers, the predictive power of narrow facets may possess incremental validity over and above that of broad personality factors (Paunonen & Ashton, 2001). In addition to FFM factors, it is important to consider both broad and narrow traits in determining whether personality exhibits a unique biasing effect on peer performance ratings in a team context.

It was expected that the global FFM facet conscientiousness would exhibit a negative relationship with peer ratings of task and social role performance. Additionally, it was expected that emotional stability, extraversion, and agreeableness would exhibit positive relationships with peer ratings of task and social role performance. In accordance with previous hypotheses, it is also expected that narrow personality facets of each trait will exhibit a similar relationship.

Additionally, it was hypothesized that similarity in certain FFM personality traits would inflate peer ratings in task or social role performance. Specifically, it was hypothesized that similarity in conscientiousness would result in more positive peer ratings of both task and social role performance. It was also hypothesized that similarity in emotional stability, extraversion, agreeableness, and openness would result in more positive peer ratings of social role performance. Thus, was expected that narrow personality facets will exhibit a similar effect on peer ratings of task and social role performance.

<div align="center">Facet-Level Descriptive Statistics and Intercorrelations</div>

Descriptive statistics, intercorrelations, and reliabilities of IPIP facets, (Goldberg, 1999) and peer performance ratings are listed in Table 7. Reliabilities for each facet ranged from $\alpha = .69$ to $\alpha = .88$. Narrow personality facet correlations ranged from $r = -.54$ to $r = .62$. The lowest correlation was between self-consciousness, which is a facet of emotional stability, and friendliness, a facet of extraversion. The largest correlation was between self-efficacy and achievement striving, both facets of the global FFM factor conscientiousness. The largest correlation between personality and peer role performance ratings were assertiveness and task role ratings at $r = .21$ and altruism and social role ratings at $r = -.25$, which was significant.

Reliabilities for peer task and social role performance ratings were both extremely high at $\alpha =$ .98.

Table 7. *Facet-Level Descriptive Statistics and Intercorrelations*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Emo	(.88)													
2. Intellect	.18*	(.69)												
3. SE	.12	.08	(.82)											
4. Order	.13	.01	.25**	(.81)										
5. Dutiful	.13	.06	.39**	.32**	(.86)									
6. AS	.18*	.10	.62**	.25**	.45**	(.83)								
7. Friend	.19**	.10	.36**	.10	.24**	.33**	(.79)							
8. Assert	.09	.05	.42**	.10	.20**	.42**	.36**	(.80)						
9. Altruism	.48**	.00	.26**	.20**	.22**	.41**	.36**	.14	(.76)					
10. Coop	.19**	-.03	.22**	.24**	.35**	.30**	.17*	.03	.34**	(.69)				
11. Anger	-.04	-.14	-.09	-.13	-.21**	-.14*	-.22*	.09	-.28**	-.49**	(.73)			
12. SC	-.13	.30**	-.10	-.10	.20**	-.12	-.54**	-.33**	-.06	.00	.08	(.82)		
13. Task	-.10	-.15	-.01	-.02	-.08	.09	.09	.21	-.09	.01	-.07	-.14	(.98)	
14. Social	-.15	-.09	-.09	-.11	-.15	-.03	-.02	.06	-.25*	-.12	.01	-.15		(.98

*Note. *p < .05, **p <.01. Emo = Emotionality. SE = Self-Efficacy. Order = Orderliness. Dutiful = Dutifulness. Friend = Friendliness. Assert = Assertiveness. Coop = Cooperation. Task = Task Roles. Social = Social Roles.* AS = Achievement Striving, SC = Self-Consciousness.

Next, the data were restructured so that rater and ratee facet-level personality were separated. Sample sizes and intercorrelations were then assessed between ratee facet-level personality and peer role performance ratings as well as rater facet-level personality and peer role performance ratings. Intercorrelations and sample sizes for ratee and rater personality variables are listed in Tables 8 and 9, respectively. All correlations were analyzed using SPSS software without addressing the nested structure. Therefore, significance levels are not discussed.

Sample size for ratees ranged from $n = 245$ to $n = 292$ for self-ratings of facet-level personality. The sample size for peer ratings of ratee personality were $n = 407$ and $n = 409$ for task and social role ratings, respectively. Narrow personality correlations among ratees ranged from $r = -.55$ and $r = .61$. The lowest correlation was between ratee anger, a negative facet of emotional stability, and ratee cooperation, a facet of agreeableness. The largest correlation was between self-efficacy and achievement striving, both facets of conscientiousness. Correlations between narrow personality facets and peer ratings of task role performance ranged from $r = -.19$ to $r = .19$, the lowest correlation being self-consciousness and the highest being assertiveness. Correlations between narrow personality facets and peer ratings of social role performance ranged from $r = -.09$ to $r = .26$, the lowest correlation being ratee emotionality and the highest being ratee dutifulness.

Table 8. *Ratee Facet-Level Intercorrelations*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Emotionality	--													
2. Intellect	.10	--												
3. Self-Efficacy	.00	.00	--											
4. Orderliness	.00	-.03	.20	--										
5. Dutifulness	.06	.09	.37	.42	--									
6. AS	.06	-.01	.61	.26	.49	--								
7. Friendliness	.14	.06	.32	.06	.13	.16	--							
8. Assertiveness	.04	-.09	.40	.01	.12	.31	.36	--						
9. Altruism	.40	-.05	.31	.12	.15	.38	.34	.07	--					
10. Cooperation	.09	-.09	.17	.21	.29	.26	.08	-.12	.27	--				
11. Anger	.04	-.10	-.10	-.21	-.19	-.16	-.30	.18	-.28	-.55	--			
12. SC	-.14	-.36	.11	-.08	-.15	.08	-.40	-.28	.07	.09	.13	--		
13. Task Roles	-.03	.11	-.03	.09	.19	.05	-.06	-.01	-.02	.08	-.11	-.19	--	
14. Social Roles	-.09	.10	-.03	.10	.26	.09	-.07	.02	-.05	.05	-.06	-.18	.90	--

Note. AS = Ratee Achievement Striving, SC = Ratee Self-Consciousness. All correlations analyzed in SPSS software without addressing the nested structure, therefore significance levels not listed.

47

Finally, the correlations between narrow rater personality facets and peer ratings of role performance were examined. Sample size for facet-level rater personality was $n = 90$ and the number of ratings for peer ratings of task and social role performance remained the same at $n = 407$ and $n = 409$, respectively. Correlations between rater personality facets ranged from $r = -.46$ to $r = .61$. The lowest correlation was between anger and cooperation and the largest was between self-efficacy and achievement striving. Correlations between narrow personality facets and peer ratings of task role performance ranged from $r = -.11$ to $r = .24$, the lowest correlation being rater self-consciousness and the highest being rater assertiveness. Correlations between narrow personality facets and peer ratings of social role performance ranged from $r = -.13$ to $r = .15$, the lowest correlation being rater self-consciousness and the highest being rater assertiveness.

Table 9. *Rater Facet-Level Descriptive Statistics and Intercorrelations*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Emotionality	--													
2. Intellect	.08	--												
3. Self-Efficacy	-.02	-.02	--											
4. Orderliness	-.03	.03	.27	--										
5. Dutifulness	.02	.04	.11	.14	--									
6. AS	.09	.04	.61	.32	.14	--								
7. Friendliness	.08	.05	.31	.07	.05	.18	--							
8. Assertiveness	.01	-.03	.40	.08	-.05	.31	.34	--						
9. Altruism	.43	.03	.24	.16	.06	.38	.29	.00	--					
10. Cooperation	.21	-.03	.18	.25	.11	.30	.01	-.02	.30	--				
11. Anger	-.03	-.14	-.12	-.21	-.08	-.22	-.28	.11	-.32	-.46	--			
12. SC	-.09	-.23	.08	-.04	-.06	.02	-.45	-.31	.08	.09	.22	--		
13. Task Roles	-.05	-.04	.13	.06	-.01	.18	.14	.24	-.03	-.01	-.05	-.11	--	
14. Social Roles	-.09	.02	.05	-.03	.01	.12	.07	.15	.11	-.07	-.03	-.13	.90	--

Note. AS = Rater Achievement Striving, SC = Rater Self-Consciousness. All correlations analyzed in SPSS software without addressing the nested structure, therefore significance levels not listed.

Rater Facet-Level Personality

The main effects of rater facet-level personality predicting peer ratings of task role performance are listed in Table 10. Global FFM results indicated that rater conscientiousness was significantly and positively related to peer ratings of task role performance. In accordance with this result, self-efficacy, $\beta = .15; t(42) = 2.38, p < .01$, and achievement striving, $\beta = .19; t(42) = 3.76, p < .01$, both positively and significantly predicted peer ratings of task role performance. In line with hypotheses and contrasting broad FFM findings, dutifulness negatively predicted peer ratings of task role performance, $\beta = -.10; t(42) = -.96, p = .34$, however this result was small and nonsignificant. Orderliness did not significantly predict peer ratings of task role performance, $\beta = .03; t(42) = .47, p = .64$.

Consistent with hypotheses and global FFM findings, friendliness, $\beta = .14; t(42) = 2.63, p < .05$, and assertiveness, $\beta = .20; t(42) = 4.06, p < .01$, facets of extraversion, were both significant and positively related to ratings of task role performance. For findings of agreeableness, neither altruism, $\beta = .06; t(42) = 1.07, p = .29$, nor cooperation, $\beta = -.05; t(42) = -.95, p = .35$, were significantly related to peer ratings of task role performance. Finally, neither facets of emotional stability, anger, $\beta = -.02; t(42) = -.38, p = .70$, and self consciousness, $\beta = -.03; t(42) = -.54, p = .60$, were significantly related to peer ratings of task role performance.

Table 10. *Rater Facet-Level Main Effects on Task Role Performance*

Variable	Coefficient	*SE*	*d.f.*	*p*-value
Conscientiousness				
Self-Efficacy	.15**	.10	42	< .01
Orderliness	.03	.06	42	.64
Dutifulness	-.10	.10	42	.34
Achievement Striving	.19**	.05	42	< .01
Extraversion				
Friendliness	.14*	.05	42	.01
Assertiveness	.20**	.05	42	< .01
Agreeableness				
Altruism	.06	.06	42	.29
Cooperation	-.05	.06	42	.35
Emotional Stability				
Anger	-.02	.06	42	.70
Self-Consciousness	-.03	.05	42	.60

Note. $*p < .05$, $**p < .01$. Coefficient = Standardized Beta (β). S*E* = Standard Error, *d.f.* = degrees of freedom.

Main effects of rater facet-level personality predicting peer ratings of social role performance are listed in Table 11. Consistent with global findings of conscientiousness, self-efficacy, $\beta = .07; t(42) = 1.74, p = .09$, and achievement striving, $\beta = .08; t(42) = 1.90, p = .07$, were both nonsignificant and positively related to peer ratings of social role performance. Additionally, while the finding was nonsignificant, dutifulness, $\beta = -.14; t(42) = -1.84, p = .07$, was negatively related to peer ratings of social role performance which was in line with hypotheses. Orderliness did not significantly predict peer ratings of social role performance, $\beta = -.03; t(42) = -.91, p = .37$.

Consistent with hypotheses and global findings of extraversion, assertiveness, $\beta = .10; t(42) = 2.48, p < .05$, was positively and significantly related to peer ratings of social role performance. Friendliness was not significantly related, $\beta = .03; t(42) = .68, p = .50$.

Consistent with previous findings of agreeableness, cooperation, $\beta = -.09; t(42) = -1.96, p =$

.06, was negatively related to peer ratings of social role performance although it was

nonsignificant. Altruism, $\beta = -.01; t(42) = .04, p = .74$, was not significantly related.

Finally, similar to global findings of emotional stability, neither anger, $\beta = .03; t(42) =$

$.65, p = .52$, nor self-consciousness, $\beta = .01; t(42) = .25, p = .81$, were significantly related

to peer ratings of social role performance.

Table 11. *Rater Facet-Level Main Effects on Social Role Performance.*

Variable	Coefficient	SE	d.f.	p-value
Conscientiousness				
Self-Efficacy	.07	.04	42	.09
Orderliness	-.04	.04	42	.37
Dutifulness	-.14	.08	42	.07
Achievement Striving	.08	.04	42	.07
Extraversion				
Friendliness	.03	.04	42	.50
Assertiveness	.10	.04	42	.02
Agreeableness				
Altruism	-.01	.04	42	.74
Cooperation	-.09	.04	42	.06
Emotional Stability				
Anger	.03	.05	42	.52
Self-Consciousness	.01	.04	42	.81

*Note. *$p < .05$, **$p < .01$. Coefficient = Standardized Beta (β). SE = Standard Error, d.f. =
degrees of freedom.

Facet-Level Interactions

Rater and Ratee Facet-Level Personality

Cross level interaction effects of rater facet-level personality and ratee facet-level

personality on peer ratings of task role performance are listed in Table 12. Overall, findings were

very small and nonsignificant. Contrasting previous findings of conscientiousness, cross level

effects of rater and ratee self-efficacy, $\beta = .05; t(40) = 1.20, p = .24$, dutifulness, $\beta =$

.001; $t(40) = .07, p = .95$, and achievement striving, $\beta = .02; t(40) = .32, p = .75$, were

positive. Orderliness, $\beta = -.07; t(40) = -1.70, p = .10$, was the only variable in which

personality similarity was negatively related to peer ratings of task role performance.
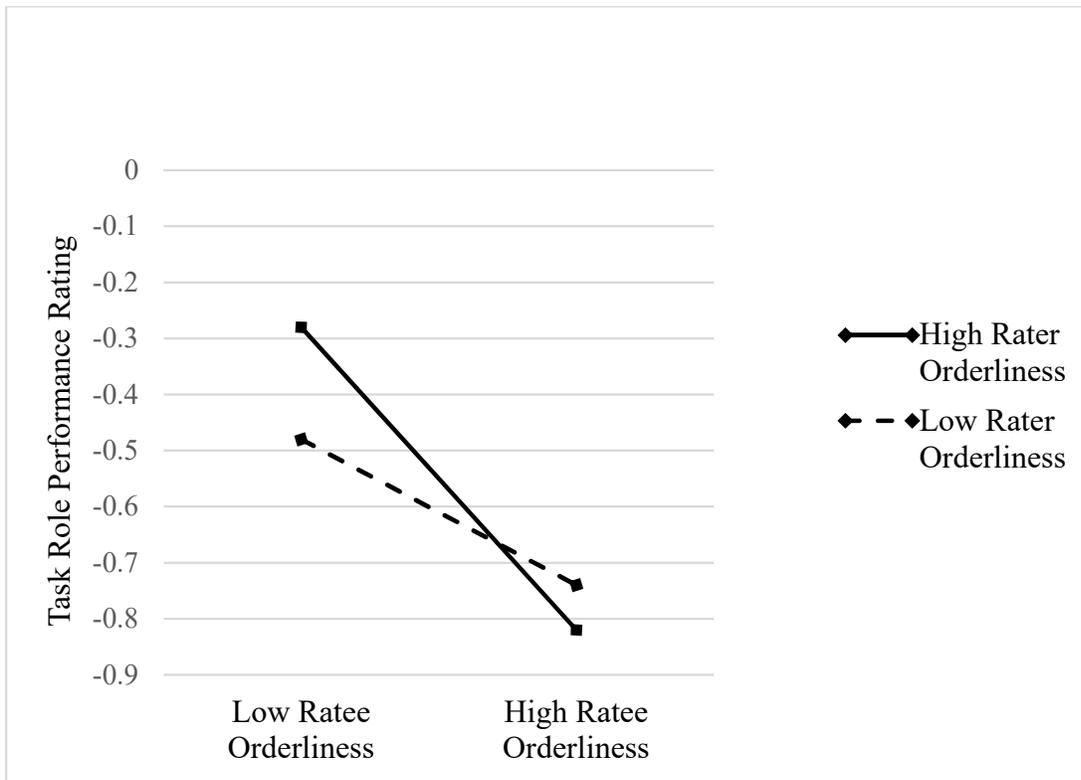
Figure 5. *Cross-Level Interaction Between Rater and Ratee Orderliness*
Note. Scale of x-axis ranges from 1 standard deviation below the mean to 1 standard
deviation above the mean.

Table 12. *Facet-Level Interaction Effects on Task Role Performance*

Variable	Coefficient	*SE*	*d.f.*	*p*-value
Conscientiousness				
Ratee Self-Efficacy*Rater Self-Efficacy	.06	.05	40	.24
Ratee Orderliness*Rater Orderliness	-.07	.05	40	.10
Ratee Dutifulness*Rater Dutifulness	.001	.02	40	.95
Ratee AS*Rater AS	.02	.07	40	.75

Note. *p < .05, **p <. 01. Coefficient = Standardized Beta (β). SE = Standard Error, d.f. =
degrees of freedom. AS = Achievement Striving.

Cross level interaction effects of rater facet-level personality and ratee-facet level personality on peer ratings of social role performance are listed in Table 13. Again, findings were very small and nonsignificant. Similarity in self-efficacy, $\beta = .03; t(40) = 1.50, p = .14$, and orderliness, $\beta = .02; t(40) = .66, p = .51$, was positively related to peer ratings of social role performance while similarity in dutifulness, $\beta = -.001; t(40) = -.156, p = .88$, and achievement striving, $\beta = -.03; t(40) = -.59, p = .56$, was nonsignificant and negatively related to peer ratings of social role performance. Similarity in friendliness, $\beta = -.02; t(40) = -.48, p = .63$, was nonsignificant and negatively related to peer ratings of social role performance while similarity in assertiveness, $\beta = .01; t(40) = .16, p = .87$. was nonsignificant and positively related to peer ratings of social role performance. Similarity in altruism, $\beta = .02; t(40) = .44, p = .66$, was positively related to peer ratings of social role performance while, cooperation, $\beta = -.01; t(40) = -.23, p = .82$, was negative. Similarity in rater and ratee anger, $\beta = .02; t(40) = .33, p = .74$, was positively related to peer ratings of social role performance while similarity self-consciousness, $\beta = -.001; t(40) = -.04, p = .97$, was negative. Finally, similarity in rater and ratee emotionality, $\beta = .002; t(40) = .03, p = .97$, and intellect, $\beta = .04; t(40) = .67, p = .51$, were both nonsignificant and positively related to peer ratings of social role performance.

Table 13. *Facet-Level Interaction Effects on Social Role Performance*

Variable	Coefficient	SE	d.f.	p-value
Conscientiousness				
Ratee Self-Efficacy*Rater Self-Efficacy	.03	.02	40	.14
Ratee Orderliness*Rater Orderliness	.02	.04	40	.51
Ratee Dutifulness*Rater Dutifulness	-.001	.01	40	.88
Ratee AS*Rater AS	-.03	.04	40	.56
Extraversion				
Ratee Friendliness*Rater Friendliness	-.02	.04	40	.63
Ratee Assertiveness*Rater Assertiveness	.01	.03	40	.87
Agreeableness				
Ratee Altruism*Rater Altruism	.02	.05	40	.66
Ratee Cooperation*Rater Cooperation	-.01	.04	40	.82
Emotional Stability				
Ratee Anger*Rater Anger	.02	.05	40	.74
Ratee SC*Rater SC	-.001	.03	40	.97
Openness				
Ratee Emotionality*Rater Emotionality	.002	.04	40	.97
Ratee Intellect*Rater Intellect	.04	.07	40	.51

Note. $*p < .05, **p < .01$. Coefficient = Standardized Beta (β). *SE* = Standard Error, *d.f.* = degrees of freedom. AS = Achievement Striving. SC = Self Consciousness

CHAPTER IV

DISCUSSION

The purpose of the following study was to identify some of the idiosyncratic effects which contribute to performance rating biases in teams. The literature on peer performance appraisal is abundant, however, there is a lack of information about how members make performance ratings in a team setting. (May & Gudeldenzoph, 2006). Moreover, on average idiosyncratic rater biases account for up to half of the variance in performance ratings from multiple different sources (Mount et al, 1998; Scullen et al, 2000). Due to the increasing number of teams being utilized in the workplace (Salas et al, 2008), it is important to understand which specific rater biases may exist in a team context and to what extent.

This project examined the biological sex and self-rated personality characteristics of business students working on a multidisciplinary group project. These factors were then compared to peer ratings of task and social role performance (Mumford et al, 2008). Specifically, the main effects of rater and ratee sex on peer ratings of role performance were analyzed to determine any biasing effects. Additionally, FFM personality composites and facets from the IPIP, (Goldberg, 1999), were utilized in order to compare rater personality to task and social role performance ratings. Finally, rater and ratee sex as well as rater and ratee personality scores were compared to one another to determine whether similarity in these respective factors inflated performance ratings. Analyses of personality were conducted at both the global and facet level to identify any unique effects narrow personality may have had over broad personality factors.

Analyses indicated that ICCs for rater sources of variance were small, (.07) for peer ratings of task role performance and (.02) for peer ratings of social role performance. Thus, the majority of the variance in peer ratings of role performance resided at the ratee and the team

56

levels. This indicates that peer ratings of team member role performance were relatively accurate, as the ratings were mostly explained by the personality of the ratee and team level variance. Only in the case of extraversion and conscientiousness were effects of rater personality found to influence ratings of role performance. However, in these instances effect sizes were extremely small. Additional findings lending to the accuracy of peer performance ratings is evidenced by the fact that rater characteristics did not appear to interact with ratee characteristics. In fact, the only circumstance that there appeared to be an effect of bias on team member performance ratings was when female ratees were being rated in task roles, albeit with a small effect size.

Overall, the hypotheses examined were largely unsupported. More specifically, only one hypothesis was fully supported, one hypothesis was partially supported, and the rest of the eight hypotheses were completely unsupported. Additional findings at the facet-level of personality tended to align with the global pattern of results except in a few instances. Study findings are listed in the following sections in addition to limitations and implications for future research.

Findings for Rater Personality

In terms of FFM personality, findings indicated mostly null results with the exception of rater extraversion. Specifically, rater extraversion was positively related to peer ratings of both task and social role performance (Hypothesis 2c). The hypothesis that conscientiousness (Hypothesis 1) would negatively predict peer role performance ratings and that agreeableness and emotional stability (Hypotheses 2 a & b) would positively predict role performance ratings were unsupported. Main effects of rater facet-level personality were also assessed in order to examine any unique effects beyond broad personality traits. Effect sizes were mostly small and

nonsignificant although findings were mixed such that specific facets of several global factors differentially predicted task and social role performance.

Research has consistently supported the idea that conscientiousness is related to job performance across all types of domains, both task and social (Barrick & Mount, 1991). Other studies have indicated that conscientiousness is negatively related to performance rating leniency (-.31) (Bernardin et al, 2000). It was expected that the shrewdness of highly conscientious students would lead them to assign less positive ratings of role performance. Findings indicated that conscientious individuals significantly and positively assign task role ratings ($\beta = .11$) and were not more likely to assign negative ratings to social roles ($\beta = .02$). Both findings contradicted hypotheses and the traditional literature on performance appraisal. An explanation for this result may be attributable to the fact that participants were working in project teams. Additionally, conscientiousness is comprised of facets of self-efficacy and achievement striving. An individual high in either one of these facets may assign more positive ratings to others due to a self-serving attributional bias. A self-serving bias is on in which individuals attribute positive outcomes to themselves (Heider, 1976). Given that individuals possessing these traits want to do well, this individual may also want the group to succeed and this results in higher performance ratings. Thus, a self-serving bias may explain why highly conscientious students assigned more positive ratings of role performance (Heider, 1976).

Similar to global findings, facets of conscientiousness, self-efficacy, orderliness, and achievement striving all positively predicted peer ratings of task role performance while both self-efficacy and achievement striving were significant. Contrary to global findings, dutifulness negatively predicted peer ratings of task role performance, although this finding was not significant. Dutiful individuals tend to keep promises and more importantly to tell the truth

58

(Goldberg, 1999). It would appear that this honesty aspect of dutifulness, which is missing from the other three facets of conscientiousness, may be responsible for the negative finding. More honest individuals may feel compelled to offer accurate performance ratings of peers independent of their consideration of team member success in the class. Additionally, individuals high in self-efficacy and achievement striving may feel compelled to succeed in a task regardless of honesty. Achievement striving may extend performance intentions from the individual to the entire group, resulting in a self-serving bias as mentioned in the previous paragraph (Heider, 1976).

Findings for conscientiousness facets and social role performance were all nonsignificant. Self-efficacy and achievement striving were positively related to peer ratings while orderliness and dutifulness were negative. It is worth noting that dutifulness had the largest effect size (β = -.14). Again, it may be the case that dutiful individuals offer more honest performance ratings regardless of the situation. While findings between facets and peer ratings of social role performance were nonsignificant, it was suspected that there may have been differences between facets. To further examine significant slope differences between facets of conscientiousness and peer ratings of social role performance, a dependent correlations test was utilized (Lee & Preacher, 2013; Steiger, 1980). Specifically, the dependent correlations test was conducted between dutifulness and the most positive facet of conscientiousness, achievement striving. Results of the one-tailed analyses indicated that slopes between the two facets were nonsignificant (z = -1.58, p = .06). Therefore, findings suggest that chance cannot be ruled out as the cause for the differences between these coefficients.

Findings indicated a positive main effect of rater extraversion on peer ratings of task (β = .20) and social role performance (β = .09), supporting Hypothesis 2c. This suggests that more

highly extraverted individuals assign more positive ratings across the board. According to Weinstein (1980), people possess an unrealistic optimism bias which sees them overestimating their own chances of positive life events and underestimating their chances of negative life events. Additionally, there appears to be a significant correlation between this optimism bias and extraversion at $r = .26$ (Darvill & Johnson, 1991). Given the correlation between these two factors, main effects of rater sociability may be attributable to an optimism bias in which highly extraverted individuals tend to give team members the benefit of the doubt when assigning performance ratings. This explanation would account for a leniency bias in extraverted individuals when assigning peer ratings of task and social role performance.

Similar to global findings, both facets of extraversion, friendliness and assertiveness, were positive and significantly related to peer ratings of task role performance. Extraversion facets were also positively related to peer ratings of social role performance however, in this instance only assertiveness was significant. The idea that highly extraverted individuals have an optimism bias (Weinstein, 1980), also helps to explain these findings as each facet possesses strong sociability components.

Finally, rater agreeableness and emotional stability were not significantly related to peer ratings of task or social role performance. Agreeableness and emotional stability are heavily comprised of social components such as cooperation and anger, respectively (Goldberg, 1999). Therefore, the null findings for peer ratings of task role performance are understandable. The nonsignificant findings for peer ratings of social role behavior are less readily apparent. Because these performance ratings were not used for student evaluations, even agreeable raters did not feel pressure to inflate their ratings. Alternatively, in high stakes situations, raters high in agreeableness may not wish to contribute to any negative consequences for the ratee. It is more

difficult to explain the weak findings regarding rater emotional stability, as negative emotions typically impact one's ratings. It may be that raters had ample opportunity to observe their peers' performance, leaving little room for internal emotional cues to influence their interpretation of member actions.

As for facet-level findings of agreeableness, Altruism was positively related to peer ratings of task role performance while cooperation was negative. In contrast, both altruism and cooperation were negatively related to peer ratings of social role performance. In both of these instances neither value was significant. It may be the case that altruistic individuals feel the need to help others by offering more positive ratings. Individuals high on cooperativeness tend to work well in groups (Goldberg, 1999). Therefore, cooperative individuals may be more sensitive to team members working poorly in the group, offering negative ratings to those they feel have been uncooperative.

Finally, findings for anger and self-consciousness were the same as they were for emotional stability. Anger and self-consciousness were both negatively related to peer ratings of task role performance and neither value was significant. Additionally, anger and self-consciousness were both nonsignificant and positively related to peer ratings of social role performance. This result reflected global findings of emotional stability and peer ratings of social role performance. Again, raters may have had plenty of time to assess performance, leaving little room for emotional cues to get in the way.

Findings for Biological Sex

It was expected that female raters would assign lower peer ratings of role performance in both task and social roles (Hypothesis 3). Findings indicated that females tend to assign slightly

higher ratings of task role performance and slightly lower ratings of social role performance. However, both of these findings were very small and neither value was significant. Therefore, Hypothesis 3 was unsupported.

Stemming from *Social Role Theory* (Eagly, 1984), and principles of *Conservation of Resources Theory* (Hobfoll, 2001), we reasoned that women might feel the need to invest more initial resources and protect existing resources in the organization. Thus, women would feel compelled to offer more negative ratings than men in both task and social roles. However, contrary to our findings, it appears that female raters offer ratings no different than men in terms of task and social roles. It may be the case that working in teams for a class project is not a strong enough situation to elicit this type of response. Given that women and men are placed on an equivalent playing field when working on a class project, they may not feel as much pressure to position themselves in a place of power such as women do in a large corporation.

Additionally, it was expected that female ratees would receive lower peer ratings of task role performance (Hypothesis 4) and higher peer ratings of social role performance (Hypothesis 5). Findings indicated that women do tend to receive slightly lower peer ratings of task role performance, supporting Hypothesis 4. This finding aligns well with *Social Role Theory* (Eagly, 1984), which states that men tend to inhabit more dominant, task-oriented roles than women. Given that women are more often associated with feminine roles, it follows that they tend to be rated less positively in these positions than men. A unique finding was that women do not receive significantly higher ratings of social role performance, contradicting Hypothesis 5. This finding is unique because it is in contrast to the central tenant of *Social Role Theory* (Eagly, 1984), that women tend to inhabit feminine roles, and this association leads to higher social role performance ratings. Perhaps females are held to a higher standard when evaluating their

communication and cooperation precisely because of these expectations. In support of this interpretation, research has found that citizenship behaviors are recognized more when performed by males than when performed by females (Allen & Rush, 2001). This finding is theorized to occur because females are expected to perform more citizenship behaviors, so it becomes noteworthy when males do perform such behaviors.

Overall, the role of biological sex in biasing peer ratings task and social role performance in a team context yielded null findings. The only result of significance was that women tend to be rated more negatively in task-oriented roles, which aligns well with the current literature on biological sex and performance ratings. Although it is a troubling finding, it seems to only occur to a small extent in a team context.

Rater and Ratee Personality and Biological Sex

Trait Activation Theory states that certain situations activate specific traits and identify a unique pattern of behavior (Tett & Guterman, 2000). Additionally, *Social Comparison Theory* posits that we are likely to evaluate ourselves against those more similar to ourselves, e.g. peers (Mumford, 1983). Using these popular theories and the popular research on performance appraisal as a basis (Ammons & Brooks, 2011; Antonioni & Park, 2001), it was expected that similarity in personality and biological sex would lead to higher ratings of task and social role performance.

Hypotheses 6-7d predicted that similarity in specific FFM personality traits would upwardly bias peer ratings of task and social role performance. Findings of cross-level interactions overwhelmingly indicated that this is not the case. Thus, none of these hypotheses were supported. The only significant cross-level interaction occurred for conscientiousness, such

that greater similarity in conscientiousness led to lower ratings of task role performance $\beta = -.03; t(40) = -2.78, p < .01$. Given the tiny effect size, this finding would usually be ignored, however, the standard error was .01 indicating that this result was highly consistent. As similarity in ratee conscientiousness increased, the interaction of rater and ratee similarity in personality decreased ratings of task role performance. Therefore, contrary to hypotheses, when ratee conscientiousness was low, raters high in conscientiousness tended to assign higher ratings of task role performance. When ratee conscientiousness was high, raters tended to assign lower ratings of task role performance. In light of this result, the interaction should still be interpreted with caution given the small effect size. Moreover, the reason for a significant finding is most likely attributable to the fact that interaction occurred with such a small margin of error.

Finally, Hypothesis 8 suggested that similarity in rater and ratee biological sex would upwardly bias performance ratings. Findings indicated no support for this hypothesis. Most likely, women assign low ratings regardless of gender due to having the same gender normed expectations as men. Given that women are occupying less male dominated jobs in most organizations and have fewer initial resources, they may also assign negative ratings to protect their position independent of gender being rated (Eagly, 1984; Hobfoll, 2001)

Rater and Ratee Facet-Level Personality Interactions

Finally, cross-level interaction effects were assessed for similarity in personality facets and the biasing effect this has on role performance. Similar to global findings, facet level findings were small and nonsignificant although some differential results were discovered for each FFM personality factor.

Findings for similarity in conscientiousness facets on peer ratings of task role performance were all positive except for orderliness. Given that orderliness has the largest effect size, this explains the negative interaction result for global conscientiousness. Additionally, findings for similarity in conscientiousness facets on peer ratings of social role performance were split down the middle. Self-efficacy and orderliness were both positive while dutifulness and achievement striving were both negative. It may be the case that team members high in self-efficacy and orderliness are more salient to others and thus similarity in these traits leads to higher performance ratings than does similarity in dutifulness and achievement striving, more abstract factors.

Facets of extraversion, agreeableness, emotional stability, and openness all had nonsignificant cross-level interaction effects on peer ratings of social role performance. These findings were in line with previous global findings of personality similarity. Thus, although we expected there to be a larger bias in performance ratings the more similar raters and ratees were, the opposite occurred.

According to *Trait Activation Theory,* certain situations activate specific traits and dictate how we will behave in that situation (Tett & Guterman, 2000). Working in an organization necessitates a level of competition and ingratiation that is not found in school. Competition for promotions and resources is abundant, even in project teams. In contrast, at the end of a college semester every individual gets a grade and students can part ways. Thus, working in a team on an undergraduate school project and providing evaluations simply for research purposes may not be a strong enough situation to activate specific traits.

Additionally, *Social Comparison Theory* states that individuals tend to compare themselves against those who are similar (Mumford, 1983). This may be true in an organization,

as the individual views himself as more valuable than individuals possessing different traits. Therefore, the individual assigns inflated ratings to those they see as "capable," or in this instance similar in ability. Working in a group for a school project has a finite end, unlike working in an organization. As there are no resources or promotions at stake, it is not of any benefit to rate specific traits as superior in an effort to outperform others.

Study Implications

Existing research on performance appraisal has tended to neglect performance ratings in a team context. Additionally, studies have not examined the specific characteristics that may lead to idiosyncratic rater biases in a team context. Overall findings suggest that peer ratings could prove to be a more reliable and accurate way of assessing performance in groups, though some training may be necessary to reduce biases associated with ratee sex.

Peers, especially those working closely together in a team context, possess nuanced information about their coworkers that others sources such as supervisors do not. Given that peers view behaviors closely and more often, sometimes daily in an organizational setting, they may be capable of making more accurate assessments of performance. Additionally, team members work in close conjunction with one another and tasks are usually interdependent. Peers may provide more accurate ratings of performance assessments as a way ensuring project milestones are completed and because there is an absence of self-serving bias that would be present with making self-ratings. Either way, practitioners could benefit from weighting confidential peer ratings more heavily than other sources or performance information.

Additionally, research may be well served to use peer sources for criterion measures. However, the results from this study might not generalize to organizational settings, where

political and/or social pressures could be stronger and thus distort peer ratings to a greater extent. For instance, Lievens, Conway, and Corte (2008) found that peers assign more weight to citizenship performance rather than to task performance when compared to supervisors, especially in team based organizational cultures. Therefore, raters may attend to these behaviors more readily than specific job-related behaviors in task and contextual domains. As suggested earlier, it may the case that working on an undergraduate group project is not a strong enough situation to elicit biases in performance ratings. Future research should examine the efficacy of peer ratings in an organizational context, especially in terms of task and social roles.

Study Limitations

According to Hox (2010), an appropriate heuristic for justifying the use of HLM is an Intraclass Correlation Coefficient (ICC) of .10. Other researchers such as Nezlek (2008) state that HLM should be used when the structure of the data is multilevel, regardless of how low the ICC is. The ICCs ideally tell us how much of the variance in a specific variable is accounted for by group membership (Nezlek, 2008). The criterion in this study had ICCs of .07 and .02 for peer ratings of task and social role performance, respectively. Moreover, 7% of the variance in peer ratings in task role performance and 2% of the variance in peer ratings of social role performance is attributable to group membership, in this case raters. The low ICCs indicate that raters did not generally have a standard set of biases across ratees. This also explains many small rater effects and nonsignificant findings. In introducing the study, participants were assured that ratings were confidential in order to emphasize the importance of honesty in assigning performance ratings (Quinn, 2014). Therefore, it may be the case that raters were much more

motivated to give honest ratings than in other studies. Regardless, findings in this study should be considered in light of the small ICCs.

An additional shortcoming in this study may involve the criterion utilized. Task and social roles are two large domains of role performance behaviors encompassing a combination of much smaller clusters of behaviors. While both task and social role performance showed good reliability and interrater agreement, it could be the case that using such a large criterion domain is too broad. In consideration of the bandwidth-fidelity dilemma in personality, it is always a good idea to match the breadth of criterion to the breadth of the personality variable utilized (Tett et al, 2003). Given that this study utilized the FFM of personality, it was decided that large criterion domains encompassing task and social behaviors was the best option.

While the current study helps to shed light on peer performance ratings in a team context, another potential shortcoming of this study is the utilization of a student sample. A student sample may elicit some surprising findings however, the magnitude of certain group dynamics may not be readily apparent. Additionally, working in a student team may not be equivalent to working in a project team in an organizational environment. In an organization the competition for resources and promotions is much stronger. Therefore, one would expect that ratings made in student environment might be less biased due to the absence of a high stakes environment. Sampling from project teams working within a large organization would go far to rectify this issue.

Conclusion

This study examined idiosyncratic rater biases in a team context. Specifically, it utilized personality and biological sex characteristics as predictors of task and social role performance

ratings of team members working on a group project. Additionally, similarity in these characteristics was considered as a potential source of bias in performance ratings. It was determined that peers working in teams assign performance ratings mostly based on merit, a finding which contrasted hypothesized relationships. This held true for all study variables except in the case of overly extraverted raters and for female ratees being rated on task role performance. Despite the lack of significant results, this finding is a generally positive result for organizational practice, as it implies that peers, especially team members, may assign more accurate ratings than other sources.

APPENDICES

APPENDIX A

Personality: International Personality Item Pool (Goldberg, 1999)

Indicate the degree to which the following statements generally describe you as a person.

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

Neuroticism	Anger	I get angry easily.	1	2	3	4	5
		I get irritated easily.	1	2	3	4	5
		I lose my temper.	1	2	3	4	5
		I am not easily annoyed.	1	2	3	4	5
	Self-Consciousness	I find it difficult to approach others.	1	2	3	4	5
		I am afraid to draw attention to myself.	1	2	3	4	5
		I am not bothered by difficult social situations.	1	2	3	4	5

Extraversion	Friendliness	I make friends easily.	1	2	3	4	5
		I feel comfortable around people.	1	2	3	4	5
		I avoid contact with others.	1	2	3	4	5
		I keep others at a distance.	1	2	3	4	5
	Assertiveness	I take charge.	1	2	3	4	5
		I try to lead others.	1	2	3	4	5
		I take control of things.	1	2	3	4	5
		I wait for others to lead the way.	1	2	3	4	5

Openness	Emotionality	I experience my emotions intensely.	1	2	3	4	5
		I feel others' emotions.	1	2	3	4	5
		I rarely notice my emotional reactions.	1	2	3	4	5
		I don't understand people who get emotional.	1	2	3	4	5
	Intellect	I love to read challenging material.	1	2	3	4	5
		I avoid philosophical discussions.	1	2	3	4	5
		I have difficulty understanding abstract ideas.	1	2	3	4	5
		I am not interested in theoretical discussions.	1	2	3	4	5

Agreeableness	Altruism	I love to help others.	1	2	3	4	5
		I am concerned about others.	1	2	3	4	5
		I am indifferent to the feelings of others.	1	2	3	4	5
		I take no time for others.	1	2	3	4	5
	Cooperation	I love a good fight.	1	2	3	4	5
		I yell at people.	1	2	3	4	5
		I insult people.	1	2	3	4	5
		I get back at others.	1	2	3	4	5

Conscientiousness	Self-Efficacy	I complete tasks successfully.	1	2	3	4	5
		I excel in what I do.	1	2	3	4	5
		I handle tasks smoothly.	1	2	3	4	5
		I know how to get things done.	1	2	3	4	5
	Orderliness	I like to tidy up.	1	2	3	4	5
		I often forget to put things back in their proper place.	1	2	3	4	5
		I leave a mess in my room.	1	2	3	4	5
		I leave my belongings around.	1	2	3	4	5
	Dutifulness	I keep my promises.	1	2	3	4	5
		I tell the truth.	1	2	3	4	5
		I break the rules.	1	2	3	4	5
		I break my promises.	1	2	3	4	5
	Achievement Striving	I do more than what's expected of me.	1	2	3	4	5
		I work hard.	1	2	3	4	5
		I put little time and effort into my work.	1	2	3	4	5
		I do just enough work to get by.	1	2	3	4	5

Role Performance (Mumford et al, 2008)

The following set of questions asks you to evaluate each team member (including yourself) on various behaviors pertaining to the team. Please refer to your assigned letters so that you know who you are evaluating. Indicate the extent to which each team member performed the behaviors listed below.

Not at All	A Little	Moderately	Considerably	Definitely
1	2	3	4	5

Task Roles

		1	2	3	4	5
Contractor	Organizes the team's work to get important work done on time.	1	2	3	4	5
	Coordinates the work done by others so that things are done in the right order.	1	2	3	4	5
	Helps the team focus on getting the job done efficiently.	1	2	3	4	5
Creator	Suggests creative ways to solve the team's problems.	1	2	3	4	5
	Helps the team take a fresh perspective on problems.	1	2	3	4	5
	Sees the "big picture" and has creative ideas for handling problems.	1	2	3	4	5
Contributor	Speaks out when he/she knows the most about the work to be done.	1	2	3	4	5
	Shares with the team any knowledge he/she has about the work to be done.	1	2	3	4	5
	Takes the lead in the team when he/she has a lot of experience in that area of work.	1	2	3	4	5
Completer	Takes personal responsibility for getting the work done.	1	2	3	4	5
	Finishes work for the team on time without being reminded.	1	2	3	4	5

	Follows through on commitments made to the team.	1	2	3	4	5
Critic	Speaks up if he/she has concerns with the work the team is doing.	1	2	3	4	5
	Makes sure the team talks about both positive and negative consequences of decisions.	1	2	3	4	5
	Shares honest opinions about how the team is working, even if the opinion is not favorable.	1	2	3	4	5

Social Roles

Cooperator	Supports the team and its goals after having given input, even if he/she would have personally set different goals.	1	2	3	4	5
	Admits when others have more experience in particular areas and trusts their judgment.	1	2	3	4	5
	Recognizes the expertise of others and allows them to take a leadership role in the team.	1	2	3	4	5
Communicator	Makes the work pleasant and comfortable by being happy and easy to work with.	1	2	3	4	5
	Communicates personal feelings and thoughts respectfully and without offending anyone.	1	2	3	4	5
	Listens carefully to the thoughts and feelings of others.	1	2	3	4	5
Calibrator	Helps settle conflicts between members of the team.	1	2	3	4	5
	Suggests positive ways for the team to interact such as taking turns, showing respect, and being open to new ideas.	1	2	3	4	5
	Steps in if there are negative feelings in the team to help resolve the difficulties.	1	2	3	4	5

APPENDIX C

Role Performance (MUMFORD ET AL., 2008): Abbreviated Wave II Measure.

What is the ID # assigned to your team? What is the letter assigned to you for this study (ex: A, B, C, etc.)? Please read through each team role behavior. Indicate the extent to which team member __ performs the given action. Use the roster to reference which team members correspond with each letter.

Member ___	Great Extent			Moderate Extent		No Extent	
Contractor Role: Organizes or coordinates the teams work to help the team maintain focus on getting it done efficiently.	7	6	5	4	3	2	1
Creator Role: Sees the big picture or suggests creative ways for solving problems and getting work done.	7	6	5	4	3	2	1
Contributor Role: Shares information or advice with the team. Takes the lead when he/she has a lot of experience in that area of work.	7	6	5	4	3	2	1
Completer Role: Takes personal responsibility for the team's work and finishes the work on time. Follows through on commitments made to the team.	7	6	5	4	3	2	1
Critic Role: Shares honest opinions about the team or makes sure the team talks about both positive and negative consequences of decisions.	7	6	5	4	3	2	1
Cooperator Role: Supports the team and other team members in their work even if he/she would have personally done it differently. Recognizes the expertise of others and allows them to take a leadership role in the team	7	6	5	4	3	2	1
Communicator Role: Communicates respectively and listens carefully to the thoughts and feelings of others. Makes the work atmosphere more comfortable because he/she is pleasant to work with	7	6	5	4	3	2	1
Calibrator Role: Helps the team get along together by helping to settle conflicts. Suggests positive ways for the team to interact, or steps in if there are negative feelings within the team.	7	6	5	4	3	2	1

REFERENCES

Allen, T. D., & Rush, M. C. (2001). The influence of ratee gender on ratings of organizational citizenship behavior. *Journal of Applied Social Psychology*, *31*(12), 2561-2587.

Ammons, J. L., & Brooks, C. M. (2011). An empirical study of gender issues in assessments using peer and self evaluations. *Academy of Educational Leadership Journal*, *15*, 49.

Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, *27*(4), 479-495.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups.* Cambridge. Addison-Wesley.

Barclay, J. H., & Harland, L. K. (1995). Peer performance appraisals: The impact of rater competence, rater location, and rating correctability on fairness perceptions. *Group & Organization Management*, *20*(1), 39-60.

Barrick, M. and Mount, M. (1991), ``The big five personality dimensions and job performance,'' *Personnel Psychology*, Vol. 44, pp. 1-26.

Belbin, R. M. (1993). *Team roles at work.* Oxford: Butterworth-Heinemann.

Benne, K. D., & Sheats, P. (1948). Functional roles of group members. *Journal of Social Issues,* 4(2), 41-49.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, *85*(2), 232.

Bernardin, H. J., Hagan, C. M., Kane, J. S., & Villanova, P. (1998). Effective performance management: A focus on precision, customers, and situational constraints. *Performance appraisal: State of the art in practice*, 3-48.

Borman, W. C. (1997). 360 ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, *7*(3), 299-315.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, 7 4 ,130-135.

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self- ratings. *Human Performance*, *10*(4), 331-360.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5.

77

Dalessio, A. T. (1998). Using multisource feedback for employee development and personnel decisions. *Performance appraisal: State of the art in practice*, 278-330.

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations prevalence, characteristics, and effectiveness. *Small Group Research*, *30*(6), 678-711.

Duarte, N. T., Goodson, J. R., & Klich, N. R. (1994). Effects of dyadic quality and duration on performance appraisal. *Academy of Management journal*, *37*(3), 499- 521.

Eagly, A. H. (1997). Sex differences in social behavior: Comparing social role theory and evolutionary psychology. *American Psychologist, 52*(12), 1380-1383.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology of Gender*, 123- 174.

Eby, L. T., Meade, A. W., Paris, A. G., & Douthitt, S. S. (1999). The development of an individual-level teamwork expectations measure and the application of a within-group agreement statistic to assess shared expectations for teamwork. *Organizational Research Methods*, *2*(4), 366-394.

Farh, J., Cannella, A. A.,&Bedeian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. *Group & Organization Studies*, 16, 367-386.

Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: employees' dual roles as raters and recipients. *Group & Organization Management*, *24*(1), 92-120.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117-140.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, *7*(1), 7-28.

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

Harari, M. B., Rudolph, C. W., & Laginess, A. J. (2015). Does rater personality matter? A meta-analysis of rater Big Five–performance rating relationships. *Journal of Occupational and Organizational Psychology*, *88*(2), 387-414.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*(1), 43-62.

Hartel, E. J., Douthitt, S. S., Hartel, G. & Douthitt, S. Y. (1999). Equally qualified but unequally perceived: Openness to perceived dissimilarity as a predictor of race and sex discrimination in performance judgments. *Human Resources Development Quarterly*, 10, 79-94.

Hartman, S. J., Griffeth, R. W., Crino, M. D., & Harris, O. J. (1991). Gender-based influences: The promotion recommendation. *Sex Roles*, *25*(5-6), 285-300.

Heider, F. (1976). A conversation with Fritz Heider. *New directions in attribution research*, *1*, 47-61.

Hobfoll, S.E. (1988). The ecology of stress. New York: Hemisphere Publishing Corporation.

Hobfoll, S.E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44, 513-524.

Hobfoll, S.E. (1998). Stress, culture, and community: The psychology and philosophy of stress. New York: Plenum.

Hobfoll, S. E. (2001). The influence of culture, community, and the nested-self in the stress process: advancing conservation of resources theory. *Applied Psychology*, *50*(3), 337-421.

Hoffman, B., Lance, C. E., Bynum, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, *63*(1), 119-151.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology*, 63, 579–588.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York, NY: Routledge.

Kane, J. S.,& Lawler,E.E. (1978). Methods of peer assessment. *Psychological Bulletin*, 85, 555-586.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), Handbook of psychology (Vol. 12): Industrial and Organizational Psychology (pp. 333-375). New York: Wiley.

Lahuis, D. M., & Avis, J. M. (2007). Using multilevel random coefficient modeling to investigate rater effects in performance ratings. *Organizational Research Methods*, *10*(1), 97-107.

Hoffman, B., Lance, C. E., Bynum, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, *63*(1), 119-151.

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

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, *51*(5p1), 369.

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lievens, F., Conway, J.M. and Corte, W. (2008), "The relative importance of task, citizenship and counterproductive performance to job performance ratings: do rater source and team-based culture matter?", *Journal of Occupational and Organizational Psychology*, Vol. 81 No. 1, pp. 11-27.

Longenecker, C. O., & Goff, S. J. (1992). Performance appraisal effectiveness: A matter of perspective. *SAM Advanced Management Journal*,*57*(2), 17.

Longenecker, C. O., Sims Jr, H. P., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *The Academy of Management Executive (1987-1989)*, 183-193.

May, G. L., & Gueldenzoph, L. E. (2006). The effect of social style on peer evaluation ratings in project teams. *The Journal of Business Communication (1973)*, *43*(1), 4-20.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246.

Mobley, W. H. (1982). Supervisor and Employee Race and Sex Effects on Performance Appraisals: A Field Study of Adverse Impact and Generalizability1. *Academy of Management Journal*, *25*(3), 598-606.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, *51*(3), 557-576.

Mumford, J. D. (1983). Social comparison theory and the evaluation of peer evaluations: A review and some applied implications. *Personnel Psychology*, 36, 867-881.

Mumford, T. V., Campion, M. A., & Morgeson, F. P. (2006). Situational judgment in work teams: A team role typology. *Situational judgment tests: Theory, Measurement, and Application*, 319-343.

Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2008). The team role test: development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, *93*(2), 250.

Murphy, K.R. (2008). Explaining theweak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology: Perspectives on Research and Practice*, 1, 148–160.

Nevill, D. D., Stephenson, B. B., & Philbrick, J. H. (1983). Gender effects on performance evaluation. *The Journal of Psychology*, *115*(2), 165-169.

O'Neill, T. A., McLarnon, M. J., & Carswell, J. J. (2015). Variance components of job performance ratings. *Human Performance*, *28*(1), 66-91.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 609-626.

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of personality and social psychology*, *81*(3), 524.

Peeters, M. A., Van Tuijl, H. F., Rutte, C. G., & Reymen, I. M. (2006). Personality and team performance: a meta-analysis. *European Journal of Personality*, *20*(5), 377-396.

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, 129-139.

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

Quinn, J. M. (2014). *Personality and team role performance: The mediating effect of role preferences*. (Unpublished master's thesis). Central Michigan University, Mt. Pleasant, MI.

Randall, R., & Sharples, D. (2012). The impact of rater agreeableness and rating context on the evaluation of poor performance. *Journal of Occupational and Organizational Psychology*, *85*(1), 42-59.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Du Toit, M. (2011). Hierarchical linear and nonlinear modeling (HLM7). *Lincolnwood, IL: Scientific Software International*.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, 43, 137–145.

Romano, C. (1994). Conquering the fear of feedback. *HR Focus*, *71*(3), 9-19

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*,*88*(2), 413.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 540-547.

Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2015). Validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology*, *24*(3), 325-349.

Schmitt, N., & Hill, T. E. (1977). Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. *Journal of Applied Psychology*, *62*(3), 261.

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, *85*(6), 956.

Sinha, N., Mesmer-Magnus, J., & Viswesvaran, C. (2012). Personality antecedents of self-other rating discrepancy. *Revista de Psicología del Trabajo y de las Organizaciones*, *28*(1).

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

Stewart, G. L., Manz, C. C., & Sims, H. P. (1999). *Team work and group dynamics.* New York: Wiley

Strauss, J. P., Barrick, M. R., & Connerly, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, 74, 637-657.

Taggar, S., & Brown, T. C. (2006). Interpersonal affect and peer rating bias in teams. *Small Group Research*, *37*(1), 86-111.

Tett, R. P.,& Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34, 397–423.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality–job performance relationship. *Journal of Organizational Behavior*, *24*(3), 335-356.

Tsui, A. S.,& Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal*, 29, 586-599.

Tziner, A., Murphy, K. R., & Cleveland, J. N. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and rating behavior? *International Journal of Selection and Assessment*, *10*(3), 218- 224.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology,* 81(5), 557- 574.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, *39*(5), 806.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*(3), 189-205.

Woehr, D. J., Sheehan, M. K., & Bennett Jr, W. (2005). Assessing measurement equivalence across rating sources: a multitrait-multirater approach. *Journal of Applied Psychology*, *90*(3), 592.

Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, *13*(2), 97-107.