University of Missouri, St. Louis
IRL @ UMSL

Dissertations   UMSL Graduate Works

11-2-2018

The Dreaded Performance Appraisal: Can the Process Ever be Comfortable?

Lauren LaBat
lrld49@mail.umsl.edu

Follow this and additional works at: https://irl.umsl.edu/dissertation

Part of the Industrial and Organizational Psychology Commons

Recommended Citation

The Dreaded Performance Appraisal: Can the Process Ever be Comfortable?

Lauren R. LaBat

M.A., Psychology, University of Missouri – St. Louis, 2015

B.A., Psychology, Rhodes College, 2013

A Dissertation Submitted to The Graduate School at the University of Missouri – St. Louis in partial fulfillment of the requirements for the degree Doctor of Philosophy in Psychology with an emphasis in Industrial and Organizational Psychology

December 2018

Advisory Committee

John Meriac, Ph.D.
Chairperson

Stephanie Merritt, Ph.D.

Jim Breaugh, Ph.D.

Nancy Singer, Ph.D.

Table of Contents

Abstract

Previous research has indicated that individuals dislike and resist the performance appraisal process. Fewer studies have examined reasoning for unintentional rating distortion that may result from a lack of training and clear understanding of how to effectively evaluate behaviors. Researchers have shown that the appraisal process is uncomfortable for raters, but empirical studies have yet to explore how to reduce this discomfort. Rater training research has revealed that trained raters have improved observational skills, a more precise vocabulary to describe behaviors, and improved rating accuracy. This research explored the relationship between performance appraisal discomfort and trait motivational factors (i.e., personality and self-efficacy) and rater behavior (i.e., leniency and accuracy), along with the impact of the experience of training. Additionally, these studies investigated whether performance appraisal discomfort could be reduced after experiencing rater training (specifically, Frame-of-Reference training). Results revealed that individual difference variables (personality and self-efficacy) were not consistently related to ratings of discomfort; however, rater training was found to be an effective mechanism for reducing discomfort with making performance ratings. Future research ideas and practical implications are discussed.

Keywords: performance appraisal, rater discomfort, rater training, personality

The Dreaded Performance Appraisal: Can the Process Ever be Comfortable?

A seemingly constant source of dissatisfaction, performance appraisal systems are often met with resistance from employees and employers. The process has been said to leave employees dejected and unfit for "productive" work for several weeks after making ratings (Deming, 1986), with research demonstrating that the system is unpleasant for both employees and supervisors (Longenecker, Sims, & Gioia, 1987). Why then, do organizations spend considerable time and money investing in a process that their employees despise? Several researchers have indicated that motivation and political factors contribute to lenient, biased, or inaccurate ratings (Cleveland & Murphy, 1992; Longenecker et al., 1987; Murphy, Cleveland, & Hanscom, 2018). Another explanation for inaccurate ratings that has received less attention could be that raters are not well-trained, leading to discomfort with the appraisal process. Researchers have called for the use of actual supervisor data to better understand performance appraisal discomfort (Bernardin & Villanova, 2005).

Using both field and lab investigations, these studies examined the impact of training on discomfort with performance ratings and explored the contributions of other performance appraisal factors and antecedents and outcomes of discomfort (i.e., individual difference characteristics). The objectives of this study included investigating the relationship between performance appraisal discomfort and motivational factors (i.e., personality and self-efficacy) and between performance appraisal discomfort and rating factors (i.e., leniency and accuracy). Additional objectives included determining whether performance appraisal discomfort could be reduced after training and delving deeper into

how training influences the relationships between discomfort with appraisals and

motivational factors.

Performance management processes include appraisals or reviews that are often

linked to compensation decisions, making them an important component of

organizations. Bretz, Milkovich, and Read (1989) reported that over 90% of Fortune 100

firms use performance appraisals for administrative purposes; however, even if

compensation was not a factor, performance ratings still have a place in businesses. For

example, the use of 360-degree performance evaluations has gained increasing popularity

(Nowack & Mashihi, 2012), even when used only for developmental purposes. Thus,

performance ratings continue to persist in organizations. In fact, not only are managers

making ratings of performance, but peers and subordinates are also being asked to weigh

in with their perspectives. Consequently, it is crucial that raters learn to view the

performance appraisal process as an opportunity to provide important and useful

behavioral feedback as opposed to a laborious chore they must endure. Rather than

exploring alternative reasons to overcome resistance to performance appraisals, previous

research has focused more on improving rating accuracy through changing rating formats

(Landy & Farr, 1980) or utilizing rater training (Roch, Woehr, Mishra, & Kieszczynska,

2012; Woehr & Huffcutt, 1994).

Earlier studies suggest that the main reason raters distort their ratings is because

they lack motivation (Fisher, 1989) or are influenced by social factors (i.e., trying to

please other employees, Bernardin & Villanova, 1986; Levy & Williams, 2004; Tziner,

Murphy, Cleveland, & Roberts-Thompson, 2001), when a more important issue may be

that raters are simply not *confident* in their performance appraisal abilities. Saffie-

Robertson and Brutus (2014) noted that the underlying cause of resistance to performance appraisals may be due to the unease evaluators feel when faced with assessing others' work performance. If performance appraisal research is to apply to practices in organizations, it is important not only to improve rating accuracy, but also to help raters feel confident in their ratings.

In order to accomplish this, an important motivational factor that warrants more investigation is performance appraisal discomfort, which concerns employees' feelings towards different components of the performance appraisal process (e.g., providing ratings, monitoring performance, justifying rating). Villanova, Bernardin, Dahmus, and Sims (1993) found that raters who reported higher levels of performance appraisal discomfort gave more lenient ratings than raters reporting lower levels of discomfort. Workers who experience discomfort with the appraisal process are expected to engage in more withdrawal and avoidance behaviors, such as turnover intentions or actual turnover (Villanova et al., 1993), which could be detrimental to performance rating and feedback processes.

Previous research has also indicated that all performance appraisal discussions are likely to produce some level of discomfort to the extent that appraisal is believed to be associated with important outcomes, such as compensation decisions (Smith, Harrington, & Houghton, 2000). While these and other studies have investigated predictors and outcomes of performance appraisal discomfort (Fried, Tiegs, & Bellamy, 1992; Smith et al., 2000; Villanova et al., 1993), fewer researchers have explored how to *reduce* rater discomfort. Rater training may be one way to address this important organizational issue, with an approach towards using it to reduce discomfort rather than increase accuracy.

The Performance Appraisal Process

Before proposing how rater training may help to reduce rater discomfort, it is important to understand more about the performance appraisal process, and what about it leads to discomfort with ratings. Performance appraisals have been defined as a "variety of activities through which many organizations seek to assess employees and develop their competence, enhance performance and distribute rewards" (Fletcher, 2001, p.473). Murphy and Cleveland (1995) maintain that the term "performance evaluation" has become synonymous with performance ratings, explaining that ratings are often the main format used for evaluations. The performance appraisal process has also been referred to as "one of the most emotionally charged activities in business life" (Pearce & Porter, 1986, p. 212). This perspective has largely been attributed to the relationship between performance evaluations and personnel decisions.

According to Bol (2011), rating inaccuracy is a common issue plaguing performance evaluations due to the subjective nature of the task and its use for administrative purposes as opposed to developmental. However, though employees are frustrated by performance appraisals, research indicates that managers are resistant to discarding performance appraisals completely, as they often believe them to be an essential component to human resource process (Meyer, 1991). As a result, in some organizational contexts, an inaccurate appraisal may be tolerated over no appraisal at all. Therefore, it is important to understand first why raters distort their ratings, and second, to consider novel methods for combatting this issue.

PERFORMANCE APPRAISAL DISCOMFORT

Performance Appraisals and Rating Distortion

Cleveland and Murphy (1992) indicated that it is often the motivation of a rater that determines the accuracy of an appraisal, and the distortion of ratings has been attributed to several different motives. Fisher (1989), for example, noted that even if raters are capable of rating accurately, there is no guarantee that they will choose to do so. This suggests that while raters may have the ability to make accurate ratings, they frequently lack the necessary motivation or drive to be accurate. Longenecker et al. (1987) noted that employees appear to dislike giving and receiving performance appraisals because though the process typically takes up a small piece of the year, it has a significant impact on individual career development and organizational performance.

Several researchers have shown that supervisors inflate ratings in order to avoid the aversive nature of appraisals (Latham, 1986; Longenecker et al., 1987), especially if they have to provide feedback face-to-face (Villanova et al., 1993). The prospect of future collaboration was found to predict higher performance appraisal ratings (Randall & Sharples, 2012). Moreover, supervisors may distort ratings for political reasons, such that though they may want to provide accurate appraisals, it may not be in their best interests to do so (Fisher, 1989; Poon, 2004). For example, Bernardin and Villanova (1986) found that superiors, administrators, and subordinates believed ratings were inflated to avoid confrontations with subordinates or to please employers. In another example, supervisors may inflate ratings in order to preserve their relationship with their subordinates. When the relationship between a manager and their employees was stronger, Bol (2011) found that managers made more lenient ratings.

In support of these political explanations for rating inaccuracy, there are also other motivational justifications. Expectancy theory of motivation (Mitchell, 1982) proposed that the ability to generate accurate ratings could depend on whether those ratings would result in desirable or undesirable outcomes. Similarly, there is evidence that attitudes (i.e., affective commitment, and use of appraisal), perceptions of fairness (Jawahar, 2007), and orientations toward performance appraisal (i.e., confidence in performance appraisal system) influence rating distortion, and as a result may influence comfort with performance appraisals (Tziner & Murphy, 1999; Tziner et al., 2001; Tziner, Murphy, & Cleveland, 2002).

The performance appraisal purpose also plays an important role (Jawahar & Williams, 1997). Individuals were found to be motivated to rate more leniently and less accurately when the appraisal was used for administrative purposes as opposed to for research, feedback, or employee development (Cleveland & Murphy, 1992; Jawahar & Williams, 1997). Jawahar and Williams (1997) referred to this as the performance appraisal purpose effect. In their study investigating the multiple uses of performance appraisal, Cleveland, Murphy, and Williams (1989) reported that the information collected during performance appraisals had the greatest impact on salary administration and performance feedback, lending support to the hypothesis that individuals rate more leniently when ratings are a factor in salary decisions. Raters may have different motivations depending on who they are rating. For example, making ratings for a supervisor or direct report that may influence a promotion or salary increase could hold more weight or make a larger impact than making ratings for a peer, as those are typically

used only for developmental purposes. It is clear having the power to influence another

employee's income may lead to discomfort, and subsequently lenient ratings.

Jawahar and Williams (1997) observed evidence of the performance appraisal

purpose effect in several scenarios. More lenient ratings were reported in field settings

with organizational raters as opposed to research settings with student raters when the

purpose was administrative. Furthermore, ratings were more lenient when appraisals were

made of a direct report or subordinate (downward appraisal) than of a manager or

supervisor (upward appraisal). Though ratings do not appear to be as lenient when a

performance appraisal is used for development, the process of making ratings can still be

difficult for raters who lack proper training. Considering these varied reasons for rating

leniently, Cleveland et al. (1989) explained how appraisals are often conflicting for raters

who must compare between-employee when making decisions about promotions or

salary, but then also compare within-employee to make decisions about development.

Jawahar and Williams (1997) proposed addressing the performance appraisal discomfort

effect by finding ways to decrease the discomfort raters and ratees experience with the

appraisal process. This research explored this recommendation and suggested that

training could be an effective means of reducing discomfort with performance appraisals.

Typically, inaccurate or distorted ratings result in elevated ratings that are skewed

in a positive direction, meaning employees receive higher ratings than they deserve

(Murphy & Cleveland, 1991). Specifically, rater leniency (also known as rating elevation

or inflation) is described as the tendency of raters to assign higher ratings than those that

would be justified or substantiated by actual performance (Bernardin, Cooke, &

Villanova, 2000; Bernardin, Tyler, & Villanova, 2009), and has been referred to as one of

the most prevalent and problematic issues related to performance appraisals (Pulakos & O'Leary, 2011).

Lenient ratings can lead to dissatisfaction with performance appraisals among managers, as it can be difficult to determine termination decisions, recognize and reward top performers, and portray the performance appraisal system as equitable (Bernardin et al., 2000). Unfortunately, research has shown rating elevation to be a relatively stable characteristic (Kane, Bernardin, Villanova, & Peyrefitte, 1995; Villanova et al., 1993), with Guilford (1954) suggesting that one underlying cause of this is a rater's personality. However, if rater leniency results from raters being uncomfortable with performance appraisals and therefore less confident in their ratings, it may be advantageous to explore ways to reduce discomfort rather than methods to increase rating accuracy.

Additional Performance Appraisal Factors

There are several factors that may influence the effectiveness of a performance appraisal system that would conceivably also impact comfort with appraisals. Smither (1998) indicated that effective performance appraisals focus on performance variables as opposed to personal traits. Gilliland and Langdon (1998) emphasized that employees must believe they can give meaningful input into the appraisal process in order for it to be considered effective. Fried et al. (1992) added that having time to observe an employee also strengthens the effectiveness of a performance appraisal. Time pressures can be burdensome for raters, impacting their ability to provide strong, or well-documented, evaluations (Bernardin & Villanova, 1986). Villanova et al. (1993) maintained the importance of considering how the nature of and frequency of supervisor feedback contribute to performance appraisal effectiveness. Moreover, appraisals should be task-

relevant, with managers being trained to observe pertinent skills (Coutts & Schneider, 2004). Appraisal systems bereft of these components may lead raters to feel uncomfortable with the evaluation system.

It is also worthwhile to consider the different components of a performance appraisal. The process includes several steps, such as observing behavior, evaluating behavior, and providing feedback. Although the Performance Appraisal Discomfort scale (PADS) developed by Villanova et al. (1993) proposes a factor structure that examines different components of performance appraisals (i.e., provision of negative feedback, solicitation of feedback, justifying/defending ratings, and encouraging performance monitoring), other studies have focused specifically on the feedback piece (Cox, Marler, Simmering, & Totten, 2011; Ilgen & Moore, 1987; Tumlin, 2004). Technological advances such as electronic rating forms and performance management software have facilitated ease of documenting qualitative and quantitative information (Brutus, 2010) from performance appraisals, making it easier to conduct evaluations; however, raters are often still uncomfortable with performance appraisals due to the psychological costs of having to share negative feedback (Cox et al., 2011; Rosen & Tesser, 1970).

With so many elements contributing to the performance appraisal process, more research should investigate ways to reduce performance appraisal discomfort. Furthermore, research is needed to examine the impact these elements have on employees with actual supervisory experience to get to the root cause of why managers and supervisors dislike performance appraisals, soliciting their feedback in order to improve their confidence and reliance on the process. Though performance appraisals are disliked, they continue to persist in organizations. Theoretically, reduced discomfort may improve

attitudes and beliefs about the performance appraisal system, which should influence

rating behavior and potentially rating accuracy. Furthermore, rating leniency should be

reduced if raters are trained to understand the impact of their ratings and to be confident

in their observations. To further explore and probe into performance appraisal discomfort

and how it may relate to rating accuracy, Villanova et al. (1993) developed a scale to

measure the construct.

Performance Appraisal Discomfort

Performance appraisal discomfort is an application of the theory of job

compatibility, which refers to the extent to which workers maintain preferences for job

characteristics that are consistent with the actual demands of the job (Villanova et al.,

1993). The Performance Appraisal Discomfort Scale (PADS) was developed by

Villanova et al. (1993) to capture and measure this construct. The job compatibility

framework suggests that workers whose preferences do not align with actual job

characteristics should report greater discomfort in performing job activities (Villanova et

al., 1993). Additionally, workers who report greater discomfort may engage in more

voluntary turnover behaviors at work (Villanova et al. 1993).

In their development of the PADS, Villanova et al. (1993) delineated four factors:

provision of negative feedback, solicitation of feedback, justifying/defending ratings, and

encouraging performance monitoring. Villanova et al. (1993) suggested that the factor

structure can be interpreted in light of demands placed on raters in appraisal situations.

Smith et al. (2000) investigated factors of the performance appraisal process that may

contribute as antecedents of performance appraisal discomfort. Their study results

revealed that the beliefs held by the rater significantly affected the extent to which raters

reported performance appraisal discomfort and suggested that discussions of performance problems would evoke greater discomfort than discussing acceptable performance if the rater believed appraisals influenced important work outcomes (Smith et al., 2000).

Since it is clear from previous research that there are factors (i.e., the belief that appraisals are important, Smith et al., 2000) that influence the level of performance appraisal discomfort, and that discomfort often leads to negative outcomes (i.e., rating distortion and withdrawal behaviors, Villanova et al. 1993), a question emerges as to which factors might contribute to reducing the level of discomfort supervisors feel when approaching performance ratings. Arguably, an important component that may play a role would be managerial experience.

In their study investigating leniency effects and performance appraisal discomfort, Villanova et al. (1993) found rater discomfort to be a relatively stable rater characteristic that was not subject to significant change as a result of moderate experience in appraisal situations, suggesting that limited experience in performance appraisal would not significantly influence responses on the PADS. Similar to the findings of Villanova et al. (1993), Smith et al. (2000) found that rater's age and managerial experience had no significant impact on performance appraisal discomfort. More experienced supervisors reported lower levels of performance appraisal discomfort than raters with less experience, but this finding was only marginally significant (Smith et al., 2000). Rather than the amount of time a rater has accrued in age or years of work experience, the type of experiences or skills acquired may be more important determinants of discomfort.

These findings contributed to the idea that performance appraisal discomfort does not change as a result of having more experience with conducting a performance

appraisal. If this is the case, and more experience does not significantly reduce discomfort with performance appraisals, what other factors may be able to increase comfort with appraising performance? Surprisingly, research had not investigated whether supervisors were uncomfortable making performance appraisals simply because they did not feel they had the skills or resources to provide accurate ratings supported by strong feedback. This research sought to demonstrate that rater discomfort is not stable, and could potentially be significantly reduced with appropriate rater training.

Rater Training

New managers are faced with the challenge of evaluating behavior without the experience more seasoned managers may have; however, even tenured or experienced managers, as indicated by previous research, may not have received the tools they needed to be successful at making effective and accurate ratings, resulting in rating distortion (Smith et al., 2000; Villanova et al., 1993). New managers move from focusing on their own tasks to coordinating the efforts of one or more individuals whose behavior they must evaluate (Hill, 2004). Being a manager not only includes stepping into a role of authority, but also requires developing interpersonal judgment in order to navigate and negotiate through politics and managing multiple relationships (Hill, 2004).

In an investigation into the performance appraisal process for police officers, Coutts and Schneider (2004) found that only 11.1% of participants who made ratings indicated substantial training for performance appraisals was provided while 67.3% indicated that very little training was provided. The vast majority of those sampled were receiving performance appraisals under the assumption that their supervisors were not provided with enough training to effectively evaluate their performance (Coutts &

Schneider, 2004). Previous research does not provide much insight into whether new manager training incorporates how to observe and evaluate others. While managers and supervisors may learn to *manage*, do they ever learn to effectively assess another person's performance?

Rater training research has mainly focused on the benefits of training for rating accuracy; however, rater training may have the potential to increase rater motivation to provide accurate ratings by helping raters to feel more confident in their appraisal abilities (Roch et al., 2012). There are several advantages of rater training, including improved observational skills, gaining new insights about behaviors, and having a more precise vocabulary with which to describe behaviors (Byham, 1971; Thornton & Rupp, 2005). One of the most widely used types of rater training, Frame-of-Reference Training (FORT), has been shown to most effectively improve rating accuracy compared to other types of rater training, namely Behavioral Observation Training (BOT) and Rater Error Training (RET) (Noonan & Sulsky, 2001; Stamoulis & Hauenstein, 1993; Woehr & Huffcutt, 1994).

Frame-of-Reference Training. FORT has received the most attention in empirical research and was first introduced by Bernardin and Buckley (1981) who suggested that establishing a common frame of reference would increase observational skills. FORT is designed to influence how a rater encodes, represents, organizes, and recalls information, and often incorporates other approaches of rater training, such as emphasizing performance dimensionality which is characteristic of Performance Dimension Training (Roch et al., 2012). This type of training improves rating accuracy by helping raters understand what behaviors are indicative of specific levels of

performance on specific dimensions and by establishing performance prototypes that allow raters to categorize ratee performance based on these prototypes presented during training (Roch et al., 2012; Sulsky & Day, 1992; Woehr & Huffcutt, 1994).

The focus of FORT is on the scaling of behaviors (i.e., determining effective or ineffective levels of performance), with the goal being that raters will learn to use common conceptualizations when providing ratings (Athey & McIntyre, 1987). The attention FORT places on the *evaluation* of behavior has been found to be more effective than the concern of BOT on the *observation* of behavior (Noonan & Sulsky, 2001; Woehr & Huffcutt, 1994). The emphasis placed on understanding rating scales and aligning observations with other raters should be beneficial for reducing performance appraisal discomfort.

Performance Appraisal Discomfort and Rater Training

If raters did not have to worry that their thoughts about behavior deviated too much from others in their organization and if they could be confident in their ability to observe and notice relevant behaviors, they would likely be more comfortable with the appraisal process and make less lenient, more accurate ratings. Typically, FORT involves providing the opportunity for practice and feedback, and feedback has been used to increase rater motivation to transfer skills learned in training to their job (Baldwin & Ford, 1988). If supervisors are taught which behaviors are indicative of certain levels of performance and are better able to discriminate between dimensions and levels of performance, it is hypothesized that as a result of FORT, they may experience less performance appraisal discomfort. Brutus and Donia (2010) examined the usefulness of a peer evaluation system and found that standardizing the experience of the peer evaluation

led to better performance and better observed performance the second time peer behavior was evaluated.

This finding suggested that uniting under a common guideline for evaluating behavior should allow individuals to learn to better observe performance, and theoretically, be more comfortable with evaluating behavior and make more accurate ratings. To further test this idea using FORT, which seeks to help raters align on how they view and observe behaviors, the opportunity to practice and receive feedback on practice ratings may help raters feel more comfortable with the appraisal process in addition to increasing their rating accuracy. Furthermore, training should help raters understand how to more effectively describe and notice behaviors, reducing discomfort and improving attitudes and beliefs about the performance appraisal system.

Additionally, research has indicated that evaluators often inflate subordinate ratings either to accomplish their own goals (Longenecker et al., 1987), because of organizational politics (Poon, 2004), or because of discomfort with the rating process (Villanova et al., 1993). Previous research has consistently found a positive relationship between discomfort and leniency (Bernardin & Villanova, 2005; Saffie-Robertson & Brutus 2014; Tziner & Murphy, 1999; Villanova et al., 1993). As a result, generating from previous research, it is expected that there will be a positive relationship between performance appraisal discomfort and leniency of ratings, such that those who report less performance appraisal discomfort after training should have less lenient ratings. Furthermore, because FORT and BOT have been found to improve rating accuracy (Woehr & Huffcutt, 1994) and because performance appraisal discomfort is associated

with more lenient ratings, it was expected that reduced discomfort would be related to greater rating accuracy.

If rater training was found to aid in reducing discomfort managers feel with conducting performance appraisals by clarifying what different levels of performance refer to for a given dimension, there could be strong implications for management training contents. Findings may lend support for the idea that all managers in supervisory roles should go through training to obtain the resources they need to feel more comfortable with conducting performance appraisals. If results indicate that more accurate scores are related to less performance appraisal discomfort, this study would indicate that not only does rater training assist in helping raters be more accurate, but it also helps them to become more confident in their rating abilities, reinforcing the importance of investing in training for employees. In addition to examining the impact training may have on performance appraisal discomfort, this study investigated how training might influence an expected existing relationship between individual difference characteristics and discomfort with performance appraisals.

Antecedents and Outcomes of Performance Appraisal Discomfort

Personality as an Antecedent. Saffie-Robertson and Brutus (2014) called for additional research on discomfort with performance appraisals by suggesting a more comprehensive model be tested that is inclusive of other variables that may influence discomfort. The proposed study examined how these variables might contribute when rater training was also introduced. One such variable is personality. According to Tziner et al. (2002), raters high in conscientiousness tended to better discriminate among ratees,

as they were more likely to make a good-faith effort to provide accurate evaluations than less conscientious individuals.

Relatedly, Bernardin, et al. (2000) found that ratings made by individuals who were low in conscientiousness or high in agreeableness made more elevated or lenient ratings. Randall and Sharples (2012) argued that conflict avoidance was a factor of motivation in ratings, reporting that high agreeableness predicted higher performance ratings. In another example, Bernardin et al. (2009) established that those low in conscientiousness or high in agreeableness made less accurate ratings and rated lower performers more leniently than other raters. Bernardin, Thomason, Buckley, and Kane (2015) concluded that there is a U-shaped relationship between personality and ratings, such that high and low levels of agreeableness were both related to less accurate ratings.

A meta-analysis conducted by Harari, Rudolph, and Laginess (2015) supported similar ideas regarding the influence of personality on performance ratings. Agreeableness, extraversion, conscientiousness, and emotional stability were all shown to have a positive relationship with performance ratings, while openness did not influence ratings (Harari et al., 2015). These effects were stronger in a field compared to a lab setting, and the effect of personality was stronger on ratings when they were collected for research purposes rather than administrative purposes and when accountability was low (Harari et al., 2015). Given these research findings, rater personality undoubtedly matters for performance ratings, but how might training and discomfort fit in? If personality factors influence the types of ratings made, would training reduce ratings of discomfort, regardless of a rater's personality? This study examined the relationship between

personality and discomfort to determine if there are differences if a participant has experienced training.

In their investigation of the relationships between academic motivation and personality among college students, Clark and Schroth (2010) found that those who were extrinsically motivated (i.e., those who engaged in behaviors to obtain an external reward or to avoid a punishment compared to those who were intrinsically motivated and engaged in behaviors for the enjoyment felt from learning or trying something new) tended to be more extroverted and conscientious, while those who lacked motivation tended to be disagreeable, suggesting that personality may influence motivation. Translated to the context of this study, because performance appraisal discomfort could be viewed as a motivational variable, it may be influenced by personality characteristics. Perhaps someone who is disagreeable and dislikes the performance appraisal process is less likely than agreeable individuals to become comfortable with ratings as a result of rater training simply because they are disagreeable.

In addition to research examining the relationship between personality characteristics and performance ratings, scholars have explored the association between personality dimensions and rating discomfort. Scholars have presented evidence that neuroticism and agreeableness are associated with conflict avoidance, contributing to discomfort with ratings (Sawyer, Hollis-Sawyer, & Pokryfke, 2002; Suls, Martin, & David, 1998). Specifically, positive relationships were identified between agreeableness, neuroticism, and rating discomfort (Sawyer et al., 2002).

Taken together, these findings indicated that those who tended to be less detail-oriented and aimed to please others tended to make ratings that were skewed in a positive

direction (Bernardin et al., 2000; Bernardin et al., 2009). Agreeable workers were often

found to be more empathetic towards others, as they were focused on social approval

(Costa & McCrae, 1992), suggesting that agreeable individuals could make more positive

ratings. It seems plausible that those who are more conscientious may focus more on the

individual components of performance evaluations, striving to do their best work at each

step. This level of focus may decrease their feelings of discomfort with the appraisal

process, while those who are more agreeable may experience more anxiousness

throughout the appraisal as they seek social approval, resulting in more discomfort with

performance appraisals. Earlier studies have yet to further explore these relationships in

an organizational setting along with additional individual difference measures, such as

performance appraisal discomfort. Additionally, scholars have not considered the effect

of training. The current study contributes to literature on individual differences and

performance appraisals by filling this void in research.

Self-Efficacy as an Outcome. In addition to personality, self-efficacy has been

explored as a potential contributing factor of performance appraisal discomfort

(Bernardin & Villanova, 2005; Tziner & Murphy, 1999), but could also be examined as

an outcome variable. Originating from social learning theory, self-efficacy is related to

outcome expectancies, and is the extent to which an individual is confident in their ability

to complete a task (Bandura, 1982). When considering self-efficacy towards the

performance appraisal process, specifically, scholars have investigated the confidence

raters feel in the ratings they make (Bernardin & Villanova, 2005). Raters may be more

likely to distort ratings if they are not confident in their ability to evaluate behavior

(Napier & Latham, 1986). Theoretically, raters who are high in self-efficacy should

collect better information about a ratee before making an evaluation and provide more

compelling justifications for their ratings (Bernardin & Villanova, 2005). Additionally,

raters may be more careful when observing performance and taking notes.

In one earlier study, the researchers examined self-efficacy training and rater

discomfort (Bernardin & Villanova, 2005). Self-efficacy training was directed at

increasing raters' confidence and ability to identify particular performance levels, as well

as to provide negative feedback to ratees (Bernardin & Villanova, 2005). This type of

training was found to significantly reduce rating elevation and performance appraisal

discomfort, providing support for the idea that training may help alleviate stresses

associated with conducting performance appraisals (Bernardin & Villanova, 2005).

Related to personality, positive relationships were identified between self-efficacy and

conscientiousness, suggesting that more conscientious raters are also more confident

raters (Tziner et al., 2002).

Given these findings, it is conceivable that self-efficacy would be influenced by

training. If training were found to reduce performance appraisal discomfort, raters should

theoretically report more confidence in their ratings compared to individuals who do not

experience training. Rater training may act as a moderator, impacting the strength of the

relationship between self-efficacy and performance appraisal discomfort. A few

hypotheses are offered regarding the influence of rater training and performance appraisal

discomfort.

The Current Study and Hypotheses

Overall, this research investigated whether rater discomfort could be reduced with

appropriate rater training. It is possible that performance appraisal discomfort stems from

a lack of confidence in appraisal skills, or results from the fact that appraisal skills were never learned. Surprisingly, few research studies have investigated whether supervisors are uncomfortable making ratings simply because they feel they lack the skills or necessary resources to provide accurate ratings. These studies explored how rater training could impact expected existing relationships between trait characteristics and performance appraisal discomfort.

The objectives of this research were to a) examine the relationship between performance appraisal discomfort and other trait motivational factors (i.e., personality and self-efficacy) and state motivational factors (i.e., reason for performance appraisal and type of performance appraisal), b) investigate the relationship between performance appraisal discomfort and rating factors (i.e., leniency and accuracy) and test whether there are differences between participants who experience training compared to those who do not, and c) explore whether performance appraisal discomfort is reduced after experiencing rater training. Two separate research studies addressed these research objectives. Study 1 was a field investigation of these variables relating to performance appraisal discomfort, and Study 2 was a lab investigation that sought to replicate findings of Study 1 with more control over the specific type of rater training and individual experiences. Both Study 1 and Study 2 were conducted concurrently.

Saffie-Robertson and Brutus (2014) indicated that due to the sensitive nature of the performance appraisal process, a limitation of performance appraisal research is that experimental studies are often conducted with university student samples. To address this limitation and examine the relationship between performance appraisal discomfort and personality and between performance appraisal discomfort and self-efficacy in the field,

Study 1 used a nonexperimental design to examine relationships between individual difference variables (i.e., personality and self-efficacy) and performance appraisal discomfort among individuals with management experience (i.e., a field study with managers who regularly rate performance). Study 2 explored similar relationships using a training manipulation and investigated how a specific type of training may influence rating behaviors (i.e., leniency and accuracy). Several hypotheses were offered with respect to personality, training, and performance appraisal discomfort.

Study 1: Effect of Training and Individual Differences on Rater Discomfort

Study 1 explored relationships between individual difference variables and performance appraisal discomfort using a nonexperimental design (see Appendix A for visual model). Participants with management or supervisory experience were recruited and asked to indicate whether they have experienced any type of rater training, which type (if identifiable, i.e., new manager training or performance evaluation training) they experienced, and how extensive the training was. Participants were also asked to respond to a questionnaire with personality, self-efficacy, and performance appraisal discomfort items, as well as other exploratory variables that may be related to discomfort (see Exploratory Variables). These variables included factors related to participants' attitudes and reactions towards the performance appraisal system. Results of Study 1 were expected to demonstrate how performance appraisal discomfort operates among individuals with management experience and to bolster external validity.

Study 1 investigated the impact of other types of interventions, such as leadership development training or new manager training courses on performance appraisal discomfort. Because Study 1 was conducted with a field sample, training was not

manipulated. Common types of rater training have been found to improve rater accuracy (Noonan & Sulsky, 2001; Woehr & Huffcutt, 1994), suggesting that individuals who have experienced rater training or performance evaluation training should be more comfortable making evaluations of behavior. Therefore, a negative relationship was expected between rater training and performance appraisal discomfort. Because rater training improves observational skills and gives raters a more precise vocabulary for describing and understanding behaviors (Byham, 1971; Thornton & Rupp, 2005), it was predicted that those who reported having gone through a type of performance management or evaluation training would show lower performance appraisal discomfort than those who reported having not experienced performance management or evaluation training.

In an academic scenario, Young (2004) found that classrooms are often updated with useful technology, but teachers are not provided with training to learn how to effectively utilize the updated systems. When students complain about this, professors feel pressured to try using the tools even when they are uncomfortable with them (Young, 2004). If this is true in an organizational context, supervisors may be trying to conduct effective performance appraisals when they are uncomfortable with the resources available to them due to a lack of training. This study examined whether supervisors who felt their organization provided them with the tools and resources needed be successful managers were more likely to have experienced training or mentorship in conducting performance appraisals. It was expected that those who began conducting appraisals with little or no assistance from their organization may feel that they were not afforded the

proper resources to learn what behaviors they should be looking for in their direct reports, and as a result, these individuals could be more uncomfortable with the appraisal process.

Hypothesis 1: There will be a negative relationship between the presence or absence of training and performance appraisal discomfort, such that participants who have experienced training will have lower ratings of performance appraisal discomfort compared to those who have not experienced training.

Hypothesis 2: There will be a negative relationship between the perceived quality of training and performance appraisal discomfort, such that participants who perceive a higher quality of the training they received will have lower ratings of performance appraisal discomfort compared to those who thought the training they experienced was of lower quality.

As discussed previously, research findings suggested that personality characteristics may contribute to rating behavior. Individuals who are agreeable tend to focus on social approval and conflict avoidance, and are therefore expected to experience more discomfort (Bernardin et al., 2009; Costa & McCrae, 1992). Based on this research, agreeableness was predicted to be positively related to discomfort due to the fact that these individuals may be motivated by or make their ratings based on potentially distressing and uncomfortable social or political factors within an organization (Bernardin et al., 2009). Conscientious raters should focus on avoiding mistakes, following instructions, and doing their best work (Costa & McCrae, 1992), suggesting that conscientious participants might experience less discomfort with evaluating behavior. Even if clear guidelines are not available, conscientious raters could be more

likely to explore and research on their own to ensure they complete their ratings

correctly.

Hypothesis 3: There will be a positive relationship between agreeableness and

performance appraisal discomfort, such that individuals with higher levels of

agreeableness will indicate higher ratings of performance appraisal discomfort.

Hypothesis 4: There will be a negative relationship between conscientiousness

and performance appraisal discomfort, such that those with higher levels of

conscientious will report lower ratings of performance appraisal discomfort.

Moreover, it was hypothesized that training would play a role in the relationship

between personality and performance appraisal discomfort. Study 1 sampled current

business professionals with managerial experience. Participants were asked if they have

ever attended a training specific to their company's performance appraisal rating process

and to provide details around this training (i.e., length of training, content, frequency).

Knowing that agreeable individuals may distort ratings for social approval while

conscientious individuals seek to avoid errors (Bernardin et al., 2009; Costa & McCrae,

1992; Sawyer et al., 2002), and because training should enable new raters to learn how to

navigate political relationships (Hill, 2004) as well as how to evaluate performance

accurately (Roch et al., 2012), it was expected that training experience would attenuate

the relationship between agreeableness and performance appraisal discomfort while

strengthening the relationship between conscientiousness and performance appraisal

discomfort.

Hypothesis 5: Training experience will moderate the relationship between

agreeableness and performance appraisal discomfort, such that the positive

relationship will be weaker if an individual has experienced training and stronger if an individual has not experienced training.

Hypothesis 6: Training experience will moderate the relationship between conscientiousness and performance appraisal discomfort, such that the negative relationship will be stronger if an individual has experienced training and weaker if an individual has not experienced training.

Self-efficacy was identified as a potential outcome of performance appraisal discomfort. As stated earlier, raters with higher levels of self-efficacy were expected to provide stronger justifications for their ratings as a result of making sure they can be confident and secure in their ratings (Bernardin & Villanova, 2005). Bernardin and Villanova (2005) created a scale to measure self-efficacy specifically in the performance appraisal context. In support of earlier work conducted by Bernardin and Villanova (2005), participants who experience lower levels of discomfort with evaluations and ratings were expected to feel more confident in their ratings. Previous research has not explored generally how performance management or evaluation training may impact rater self-efficacy or appraisal discomfort. As training should increase one's confidence in their ability to make accurate ratings by increasing the carefulness with which they approach performance ratings and observations (Roch et al., 2012) to reduce errors made (Woehr & Huffcutt, 1994), experiencing training should strengthen the negative relationship between rater self-efficacy and performance appraisal discomfort.

Hypothesis 7: There will be a negative relationship between rater self-efficacy and performance appraisal discomfort such that individuals with higher levels of rater self-efficacy will report lower ratings of performance appraisal discomfort.

Hypothesis 8: There will be a positive relationship between training experience and rater self-efficacy, such that individuals who have experienced rater training will report higher levels of self-efficacy.

Method

Participants. Participants were recruited using Qualtrics Panels. Participants were required to have at least one year of managerial or supervisory experience (i.e., a role in which the individual has/had direct reports and is/was responsible for evaluating employee performance) to take part in this study. All 250 participants passed the attention check items and no outliers were identified during data cleaning. The final sample was 64% female and 36% male, with the majority of the participants identifying as Caucasian (83%). Almost all of the participants reported that they were currently employed (90%) and the majority indicated that they worked 30 or more hours per week and were currently employed in a management or supervisory role (87%).

Procedure. Instructions and requirements for participation in this study were posted during recruitment for this study. Those who were eligible to participate could click on a survey link taking them to the informed consent form and questionnaire. The survey was expected to take no longer than one hour to complete. Once participants completed the online survey, they received credit/compensation for their participation per their agreement with Qualtrics Panels.

Measured Variables. The following variables were measured in Study 1. See Appendix D for a list of the items for each of these measures.

Rater Training Experience. Participants were asked whether they had experienced rater training or not (1 = "yes", 2 = "no"). For those who response "yes", a

number of exploratory questions were asked to further investigate the type of training

experienced.

   Perceived Quality of Training. Participants responded to one global measure of

perceived quality and a scale of items that were developed for this study to measure the

perceived quality of rater training. Generated items were evaluated by subject matter

experts (SMEs) who rated the items based on relevance to the experience of rater

training. Results were examined by looking at both the global measure of perceived

quality and the composite of the items developed to measure this construct ($\alpha = .94$). As

differences in findings between the two were negligible, the global measure was used for

reporting purposes.

   Agreeableness and Conscientiousness. Personality subscales were measured

using Bent-Martinez and John's (1998) English Big Five Inventory (BFI). Respondents

indicated their agreement to items on a five-point Likert scale (1 = "strongly disagree", 5

= "strongly agree"). The BFI ($\alpha = .76$) consists of five subscales, with Agreeableness ($\alpha = $

.79) and Conscientiousness ($\alpha = .83$) being used for this study.

   Rater Self-Efficacy. Rater Self-Efficacy was measured using the Performance

Appraisal Self-Efficacy Scale (PASES) developed by Bernardin and Villanova (2005).

Participants were asked to indicate the level of confidence they would have in being able

to successfully perform the behaviors described in 12 statements on a five-point Likert

scale (1 = with no confidence, 5 = with great confidence). Since its development, the

PASES has been used by Brutus, Fletcher, and Baldry (2009), who found that the four

subscales composed of three items each yielded acceptable reliability ($\alpha = .93$): process

features of the appraisal ($\alpha = .86$), rater subjectivity ($\alpha = .79$), appraisal discussion ($\alpha = $

.83), and suggesting performance improvement ($\alpha = .75$). Higher scores relate to higher levels of self-efficacy.

Rater Discomfort. Participants completed the Discomfort with Peer Evaluation Scale (Saffie-Robertson & Brutus, 2014). Items were amended for use with employees as opposed to students. Responses are meant to reflect the degree of discomfort felt by raters in performance appraisal situations. Participants rated their perceived discomfort level on a five-point Likert scale (1 = no discomfort, 5 = high discomfort). The discomfort scale ($\alpha = .92$) includes three subscales (collecting information, two items, $\alpha = 0.82$; rating, four items, $\alpha = 0.88$; and post-rating, four items, $\alpha = 0.78$), each reflecting different steps in a performance evaluation process (i.e., discomfort felt when collecting information used for evaluation purpose, discomfort felt when filling out rating form, and discomfort felt after the rating process is over).

Demographics. Participants were asked to respond to several demographic items asking them to report information about themselves (i.e., age, sex, race, and employment status, job title, tenure/years at organization, tenure/years in supervisory role, and number of hours worked).

Exploratory Variables. Participants were also asked to provide details around their supervisory role and organization. These items were used to learn additional information regarding the type of training experienced as well as context about their organization's performance appraisal process. Items included: opportunity to observe subordinate behavior, whether feedback is given face-to-face or electronically, whether the appraisal is used for development or administrative purposes, the frequency with which appraisals are conducted each year, the amount of time devoted to conducting an

individual appraisal, and how many people the participant conducts appraisals for each year. Participants were also asked to provide information about any training they have experienced related to management or learning to make evaluations. In addition, the following measures will be included for exploratory purposes.

Performance Appraisal Beliefs. Information on performance appraisal beliefs were collected in order to understand the impact of performance appraisals for each participant, as they are likely to work in different industries and organizations. Performance appraisal beliefs were measured with items used by Smith et al. (2000), who averaged five items from the Federal Employee Attitude Survey (DiMarco & Nigro, 1983). Responses were provided on a five-point scale with anchors of "strongly disagree" to "strongly agree", and a mid-point of "undecided". A high score on this scale would indicate strong beliefs that the organization's performance appraisal system affected organizational decisions. The coefficient alpha for this scale was .69 when used by Smith et al. (2000).

Performance Appraisal Views. In order to gain additional insight into participants' specific thoughts about individual performance appraisal systems, participants were asked to respond to relevant items from Coutts and Schneider's (2004) 18-item questionnaire measuring views about different aspects of the performance appraisal system. This measure was originally administered to employees rather than individuals who were supervisors or managers. Items were amended for use with this sample of employees with management or supervisory experience. This questionnaire covers the basis of appraisal systems, opportunity for input on an appraisal, the frequency and nature of performance feedback, perceptions of role in the performance appraisal

process to help clarify performance expectations and set goals, perceptions of accuracy of the appraisal, and extent of training received.

Results

The first few hypotheses examined the relationships between training, personality characteristics, and performance appraisal discomfort. Prior to examining these correlations, a preliminary analysis was conducted to examine whether there were mean differences on any of the key variables of interest between individuals who reported experiencing management or performance appraisal training ($n = 104$, 42%) and those who reported that they did not experience training ($n = 146$, 58%). No significant differences in reports of agreeableness, conscientiousness, self-efficacy, or discomfort with evaluation were found (See Appendix B for details; i.e., means, standard deviations, and correlations, etc.). Interestingly, regardless of training experience, the majority of the participants reported experiencing little or no discomfort with making performance evaluations (85%).

The first hypothesis examined the relationship between rater training and rater discomfort. Hypothesis 1 predicted that there would be a significant negative relationship between the presence or absence of training and rater discomfort, such that participants who experienced training would have significantly lower ratings of rater discomfort compared to those who had not experienced training. A bivariate correlation between training and rater discomfort revealed a nonsignificant positive relationship, $r(250) = 0.03$, $p = 0.697$. This finding aligned with the result that there were no significant differences found among ratings of rater discomfort between those who reported experiencing training ($M = 1.82$, $SD = 0.85$) and those who reported they had not

PERFORMANCE APPRAISAL DISCOMFORT

experienced training ($M = 1.87$, $SD = 0.89$), $t(248) = -0.39$, $p = 0.697$. Because no

relationship was found between rater discomfort and the experience of management or

performance appraisal training, this hypothesis was not supported.

The second hypothesis stated that there would be a negative relationship between

the perceived quality of training and rater discomfort, such that participants who perceive

a higher quality of training will have lower ratings of rater discomfort. Of the participants

who reported experiencing rater training, 60% reported that the training was "very" or

"extremely" useful. A bivariate correlation between training quality and rater discomfort

resulted in a nonsignificant negative relationship, $r(146) = -0.01$, $p = 0.953$.

Consequently, this hypothesis was not supported.

The third and fourth hypotheses investigated the relationships between personality

variables and rater discomfort. Hypothesis 3 stated that there would be a significant

positive relationship between agreeableness and rater discomfort, with more agreeable

individuals being expected to report higher ratings of discomfort. Hypothesis 4 postulated

that there would be a significant negative relationship between conscientiousness and

rater discomfort, such that individuals who were conscientious would report lower ratings

of discomfort. Bivariate correlations revealed significant negative correlations, both

between agreeableness and rater discomfort, $r(250) = -0.13$, $p = 0.044$, and between

conscientiousness and rater discomfort, $r(250) = -0.29$, $p < 0.001$. Interestingly,

participants who indicated higher agreeableness and conscientiousness also reported

lower feelings of discomfort with rating and evaluating performance, providing support

Hypothesis 4, but not for Hypothesis 3.

PERFORMANCE APPRAISAL DISCOMFORT

The fifth and sixth hypotheses explored the influence of training experience on the relationships between personality variables and rater discomfort. Hypotheses 5 and 6 suggested that training would moderate the relationship between personality characteristics and rater discomfort. Specifically, the presence of training was expected to attenuate the relationship between agreeableness and discomfort while strengthening the relationship between conscientiousness and discomfort. In order to examine these hypotheses, a multiple regression analysis was conducted. The relationship between personality and discomfort was expected to depend on whether a participant experienced training or not. A multiple regression analysis revealed a statistically significant relationship between agreeableness and rater discomfort ($r(250) = $ -0.13, $p = .022$), but not between rater discomfort and training experience ($r(250) = 0.03$, $p = .349$), nor between agreeableness and training experience ($r(250) = $ -0.08, $p = .096$). The interaction term between agreeableness and training experience was not significant, $\Delta R^2 = 0.013$, $\Delta F(1, 246) = 3.29$, $B = 0.11$, $p = 0.071$, suggesting that training experience did not influence the relationship between agreeableness and rater discomfort.

Similarly, a statistically significant relationship was found between conscientiousness and rater discomfort ($r(250) = $ -0.29, $p < .001$), but not between rater discomfort and training experience ($r(250) = 0.03$, $p = .349$), nor between conscientiousness and training experience ($r(250) = $ -0.03, $p = .051$). In the first step of the multiple regression, conscientiousness and training experience were included, and these two together did account for a significant amount of variance in rater discomfort, $R^2 = 0.089$, $F(2, 247) = 12.12$, $p < 0.001$). In testing for moderation, the interaction term between conscientiousness and training experience was added to the regression model.

The result was not significant, $\Delta R^2 = 0.00$, $\Delta F(1, 246) = 0.11$, $B = 0.02$, $p = 0.742$.

Examination of both interactions suggest that training experience does not influence the relationship between personality characteristics (agreeableness and conscientiousness) and rater discomfort, as neither hypothesis was supported.

The seventh and eighth hypotheses examined the relationships between rater self-efficacy and discomfort as well as between rater self-efficacy and training. Hypothesis 7 proposed that there would be a significant negative relationship between self-efficacy and rater discomfort. A bivariate correlation resulted in a significant negative correlation, $r(250) = -0.41$, $p < 0.001$, providing support for this hypothesis. Individuals who reported feeling increased confidence in their ability to make ratings during performance appraisals also indicated less discomfort with making ratings. This finding also suggests that those who may lack confidence in their ability to make ratings experience increased feelings of discomfort.

Hypothesis 8 stated that there would be a positive relationship between training experience and rater self-efficacy, such that those who have experienced rater training will indicate higher levels of rater self-efficacy. A bivariate correlation revealed a nonsignificant negative correlation between training experience and rater self-efficacy, $r(250) = -0.08$, $p = 0.220$. This finding aligns with the results that there were no significant differences found among ratings of rater self-efficacy between those who reported experiencing training ($M = 4.23$, $SD = 0.69$) and those who reported they had not experienced training ($M = 4.12$, $SD = 0.71$), $t(248) = 1.23$, $p = 0.220$. This hypothesis was not supported and could indicated that among those who participated in this study, training experience did not influence feelings of confidence with making ratings.

PERFORMANCE APPRAISAL DISCOMFORT

Exploratory Analyses

Exploratory analyses were conducted to gain additional insights into performance appraisals and rater training. Curvilinear relationships between personality variables and rater discomfort were explored, given that previous research has established that curvilinear relationships exist between certain personality traits and managerial performance (Bernardin et al., 2015). No specific hypotheses were proposed, but these non-linear relationships were tested using regression as supplemental analyses. Regression analyses revealed that there was not a significant curvilinear relationship between rater discomfort and agreeableness ($\beta = -0.187$, $F(2,247) = 2.04$, $p = .132$), but there was a significant curvilinear relationship between conscientiousness and rater discomfort ($\beta = -1.28$, $F(2,247) = 12.66$, $p < .001$).

Looking at other personality variables, Neuroticism was significantly correlated with rater discomfort ($r(250) = 0.15$, $p = 0.018$), while Extraversion was not significantly correlated with discomfort ($r(250) = -0.11$, $p = 0.088$). These findings suggest that individuals who reported increased neuroticism also reported higher rater discomfort. Based on the definitions of Extraversion and Neuroticism (Costa & McCrae, 1992), these findings are not surprising and are likely to have been the hypothesized direction. These results suggest further analysis into the relationship between personality and rater discomfort and could be especially interesting for selection or development assessments in organizations. It stands to reason that a combination of personality types could compile a "profile" that contributes to rater behavior rather than one personality factor alone.

Descriptive statistics were conducted to examine context behind participants' views and beliefs about performance appraisals, and also to better understand the

background and experiences of the sample. Of the participants who participated in Study 1, 90% indicated that performance appraisals could increase productivity when conducted correctly, and 80% reported that their organization considered performance appraisals to be an important part of a supervisor's duties. Interestingly, 69% of respondents agreed that the performance appraisal system in their organization helped to improve the performance of employees. When responding to questions about the purpose of the performance appraisal, 70% of participants reported that performance appraisals influence personnel actions while 60% reported the performance appraisal determine pay in their organization.

Respondents also reported other goals of their performance appraisal systems. A little over half (58%) of the sample reported that their performance appraisals are "always" or "frequently" based on personal traits (i.e., dependability, imitative, agreeableness) while 82% of the sample reported their performance appraisals are "always" or "frequently" based on work-related factors (i.e., quality of work outputs). About 70% of the sample indicated that they are regularly evaluated by a supervisor and 90% of the sample reported receiving face-to-face feedback after the performance appraisal "always" or "sometimes." Over half of the sample (80%) have a meeting to discuss their performance appraisal with their supervisor, and most (90%) respondents indicated that the performance appraisal helps them clarify performance expectations to a "moderate" or "large" extent. Regarding training, 53% of respondents reported that very little training is provided concerning the effective use of the performance appraisal system, while 27% reported that substantial training is provided and 20% reported that no training is provided.

These measures were interesting and showed that the majority of the sample has had similar experiences with performance appraisals. It would be interesting to examine response to these items among a sample of individuals who may be less comfortable with making ratings to determine whether beliefs about the performance appraisal system in a given organization relate to feelings of discomfort with making ratings or evaluations. Because most of this sample indicated that they were not uncomfortable with making ratings, these data did not yield any particularly groundbreaking outcomes. One finding of note would be that people indicated several different experiences with training, from going through a process annually, to only going through once becoming a new manager. As only half of the sample indicated that they experienced training, an inference could potentially be drawn that these managers learned how to complete performance appraisals "on-the-job" and through experience supervising others. Future research should explore this avenue.

Study 1 Discussion

The first aim of this study was to examine the relationship between performance appraisal discomfort and other trait motivational factors (i.e., personality and self-efficacy) and performance appraisal factors. In addition, this study sought to investigate the role of training on these relationships. The relationship between training experience and rater discomfort was not significant, which could be due to the finding that the majority of those who participated in Study 1 indicated that they were not uncomfortable with evaluating performance, regardless of their experience with rater training. It would be interesting to examine these findings among a sample of new managers with less supervisory experience than the current sample. Additionally, a nonsignificant

relationship was found between perceived quality of training and rater discomfort. When asked to describe the perceived quality of training overall, among those who had received training in their careers, the majority indicated that it was "very" or "extremely" useful (60%). Only 14% indicated that training was "not useful at all" or "slightly useful," providing some support for the importance of rater training in organizations.

Results regarding personality factors showed that both agreeableness and conscientiousness were significantly and negatively correlated with rater discomfort, suggesting that those who reported higher agreeableness and conscientiousness also reported lower rater discomfort. Though the original hypothesis was that less agreeable individuals would experience less discomfort, perhaps these findings call for deeper analysis into relationships between personality variables and rater discomfort. It may be plausible that agreeableness and conscientiousness are working together here, such that being higher on both would lead people to not only desire to make the best rating, but also to work hard to learn and ensure they are being fair raters. Exploratory analyses revealed that there was not a significant curvilinear relationship between rater discomfort and agreeableness, but there was a significant curvilinear relationship between conscientiousness and rater discomfort. This finding could indicate that more research in this area could be fruitful in terms of which personality assessments organizations might consider when selecting employees for management or supervisory roles.

This study also examined the influence of training experience on the relationship between personality and rater discomfort. Training was expected to moderate these relationships, weakening the positive relationship between agreeableness and rater discomfort and strengthening the negative relationship between conscientiousness and

rater discomfort; however, neither outcome was supported by a significant finding. Again, this could be due to the fact that the majority of the sample indicated that they were not uncomfortable evaluating performance. While it was important to start investigating these relationships with a sample of individuals who had experienced training and who had experience supervising others, future research should certainly address these findings with a sample of newer managers. It stands to reason that individuals who are less comfortable with making ratings would value rater training, and these hypothesized outcomes could yield different results and may further bolster the argument for training among employees who are responsible for rating performance.

The final aim of this study was to examine whether self-efficacy would be negatively correlated with rater discomfort, and that this relationship would be strengthened by the experience of training. A significant negative correlation was found between self-efficacy and rater discomfort, suggesting that those who felt less confident about making ratings or with the performance appraisal process experienced increased rater discomfort, though training was not found to influence this relationship. Because participants in this study did not experience the same training and could have experienced training at any point in time of their career, it might be interesting to explore this relationship more among a sample of people who undergo the same training at the same time. Both of these findings do provide additional insight into what organizations should consider when promoting or selecting supervisors. It may be beneficial to look for individuals who have more self-efficacy, or to determine whether self-efficacy is increased by having supervisors attend performance evaluation trainings.

The results of Study 1 contributed to the understanding of how the experience of training impacts performance appraisal discomfort. These findings could serve hiring and selection research regarding the use of personality assessments. Organizations could consider examining self-efficacy and conscientiousness when choosing managers; however, it would also be important to take one's actual ability into account. This study is limited in that data were collected at one point in time and from only one source. Additional sources could have rated the accuracy of an individuals' evaluation. Another limitation of Study 1 is that it was not a true experiment and lacked the ability to test the impact of training in a controlled manner, as managers in various organizations could have a broad interpretation of what "training" is. In order to account for this limitation, Study 2 was developed as a controlled laboratory experiment in which a specific form of training was manipulated to determine how it influences performance appraisal discomfort.

Study 2: The Effects of Training and Individual Difference Factors on Rater Behaviors

Previous research has demonstrated that raters who report increased performance appraisal discomfort make more lenient ratings (Villanova et al., 1993). Additionally, studies have investigated the relationship between personality characteristics and rater leniency (Bernardin et al., 2000; Jawahar, 2001; Kane et al., 1995) as well as the relationship between personality characteristic and rater accuracy (Bernardin et al., 2009). What is missing from research is the role training may play in these relationships as well as how training could influence rater discomfort.

Study 2 investigated whether a specific type of rater training (specifically, FORT) would result in reduced performance appraisal discomfort, as well as how FORT might impact how individual difference variables (personality and self-efficacy) influence rater accuracy and rater leniency. Study 2 was an experimental design (within and between-subjects) and conducted in a research laboratory. Students were randomly assigned to one of two conditions: rater training (i.e., FORT) or a control condition (i.e., no training). Accuracy indices were calculated to determine whether there was a relationship between rating accuracy and performance appraisal discomfort. Results of Study 2 were expected to reveal whether performance appraisal discomfort could be reduced by experiencing FOR training as well as whether raters who are more comfortable with the appraisal process may make more accurate ratings.

Cronbach's (1955) accuracy indices were utilized to calculate and examine rater leniency and rater accuracy. These indices include four components: (1) elevation measures how accurate a rater is over all rated dimensions and ratees and can be used to determine how lenient an individual person is rating; (2) differential elevation measures how a rater is able to discriminate among ratees averaging across all rated dimensions and is particularly important for administrative decisions in which a rater must effectively compare individuals; (3) stereotype accuracy measures the ability of a rater to discriminate among the rated dimensions, averaging across ratees, and is important for training purposes as the rater must be able to accurately distinguish between the different dimensions; and (4) differential accuracy measures how well a rater is able to discriminate among ratees within each performance dimension, and is important for

giving feedback because a rater must be able to distinguish between dimensions for a particular ratee. A lower accuracy score indicates increased accuracy.

As previous research findings suggest that individuals who are more agreeable tend to focus on social approval and conflict avoidance, agreeable participants are expected to make more lenient ratings and less accurate ratings (Bernardin et al., 2009; Costa & McCrae, 1992). People who distort their ratings for social approval or to avoid confrontation may feel uncomfortable with their ratings and were expected to make less accurate ratings on all Cronbach's (1955) accuracy indices. Individuals with higher levels of agreeableness were not expected to discriminate between ratees and dimensions as well due to their desire to maintain harmony and gain social approval. It was expected that individuals with higher levels of agreeableness may rate everyone similarly at more elevated levels/more leniently while individuals with lower levels of agreeableness may rate more accurately and less leniently because they may be less focused on being a polite rater and more focused on being a correct rater. Study 2 sought to replicate earlier findings that agreeableness is positively related to elevated ratings or rater leniency and negatively related to accuracy (Bernardin et al., 2009).

Hypothesis 9: There will be a positive relationship between agreeableness and rater elevation, such that individuals who indicate higher levels of agreeableness will make more lenient ratings.

Hypothesis 10: There will be a negative relationship between agreeableness and rater accuracy, such that individuals who indicate higher levels of agreeableness will make less accurate ratings.

(a) There will be a negative relationship between agreeableness and differential elevation. Individuals with higher levels of agreeableness will not discriminate as well between ratees across all rated dimensions.

(b) There will be a negative relationship between agreeableness and stereotype accuracy. Individuals with higher levels of agreeableness will not discriminate as well among rated dimensions.

(c) There will be a negative relationship between agreeableness and differential accuracy. Individuals with higher levels of agreeableness will not discriminate among ratees within individual dimensions as well as individuals with lower levels of agreeableness.

Regarding conscientiousness, people who are considered to be conscientious tend to be more detail-oriented and strive to do their best and most accurate work when making ratings (Costa & McCrae, 1992). These raters may be more likely to ensure that they avoid errors and have evidence to support their claims and assertions. Additionally, another characteristic of conscientious raters includes being more focused on doing one's personal best than appeasing others (Costa & McCrae, 1992). As a result, in order to replicate previous findings (Bernardin et al. 2009), it was hypothesized that conscientious individuals would make more accurate and less lenient ratings.

Hypothesis 11: There will be a negative relationship between conscientiousness and rater elevation, such that those who are more conscientious will make less lenient ratings.

Hypothesis 12: There will be a positive relationship between conscientiousness and rater accuracy, such that those who are more conscientious will make more accurate ratings.

(a) There will be a positive relationship between conscientiousness and differential elevation. Individuals with higher levels of conscientiousness will discriminate well between ratees across all rated dimensions.

(b) There will be a positive relationship between conscientiousness and stereotype accuracy. Individuals with higher levels of conscientiousness will discriminate well among rated dimensions.

(c) There will be a positive relationship between conscientiousness and differential accuracy. Individuals with higher levels of conscientiousness will discriminate well among ratees within individual dimensions.

Previous research has demonstrated that raters exposed to self-efficacy training produced less elevated ratings after training and reported lower levels of performance appraisal discomfort (Bernardin & Villanova, 2005). As a result, though this study does not investigate self-efficacy training, raters with higher levels of rater self-efficacy were expected to be able to provide justifications for their ratings to demonstrate their confidence and self-reliance (Bernardin & Villanova, 2005). In alignment with earlier research (Bernardin & Villanova, 2005), participants who are self-assured and confident in their ratings should experience less discomfort with evaluations and ratings and were predicted to make less lenient and more accurate ratings.

Hypothesis 13: There will be a negative relationship between rater self-efficacy and rater elevation, such that individuals with higher levels of rater self-efficacy will make less lenient ratings.

Hypothesis 14: There will be a positive relationship between rater self-efficacy and rater accuracy, such that individuals with higher levels of rater self-efficacy will make more accurate ratings.

    (a) There will be a positive relationship between rater self-efficacy and differential elevation. Individuals with higher levels of self-efficacy will discriminate well between ratees across all rated dimensions.

    (b) There will be a positive relationship between rater self-efficacy and stereotype accuracy. Individuals with higher levels of self-efficacy will discriminate well among rated dimensions.

    (c) There will be a positive relationship between rater self-efficacy and differential accuracy. Individuals with higher levels of self-efficacy will discriminate well among ratees within individual dimensions.

Knowing that providing the opportunity to practice and working to unite all raters under a common language for evaluating behavior has been most effective for increasing rating accuracy (Noonan & Sulsky, 2001; Roch et al., 2012), FORT was used as the training manipulation in Study 2. Participants were randomly assigned to a control condition (no training) and FORT condition and were asked to rate performance appraisal discomfort after training. The training content included an overview of performance evaluation, FORT details, and the opportunity to practice making ratings and receiving feedback. In a within-subjects design, participants who experience FORT were predicted

to report reduced performance appraisal discomfort (measured before and after training). In a between-subjects design, participants in the FORT condition were predicted to report lower discomfort than participants who did not experience training.

Hypothesis 15: Participants in the FORT condition will report lower ratings of performance appraisal discomfort after experiencing FORT compared to ratings made before training.

Hypothesis 16: Compared to the control group, participants who receive FORT will report lower ratings of performance appraisal discomfort than participants who do not have any training at all.

Finally, Study 2 explored relationships between performance appraisal discomfort and rater leniency and accuracy. After learning appropriate evaluation skills during FORT, participants should be prepared and able to make less lenient and more accurate ratings and experience less performance appraisal discomfort. Lenient ratings should be positively related to discomfort, while accurate ratings should be negatively related to discomfort. Also, those in the FORT condition should make less lenient and more accurate ratings compared to participants in the control condition who do not experience training.

Hypothesis 17: Among those in the FORT condition, there will be a positive relationship between performance appraisal discomfort and evaluation leniency after training.

Hypothesis 18: Among those in the FORT condition, there will be a negative relationship between performance appraisal discomfort and rating accuracy after training.

(a) There will be a negative relationship between performance appraisal discomfort and differential elevation. Individuals with lower levels of performance appraisal discomfort will discriminate well between ratees across all rated dimensions.

(b) There will be a negative relationship between performance appraisal discomfort and stereotype accuracy. Individuals with lower levels of performance appraisal discomfort will discriminate well among rated dimensions.

(c) There will be a negative relationship between performance appraisal discomfort and differential accuracy. Individuals with lower levels of performance appraisal discomfort will discriminate well among ratees within individual dimensions.

Hypothesis 19: The rater leniency (elevation) scores of participants on the FORT condition will be lower than the rater leniency (elevation) scores of those in the control condition.

Hypothesis 20: The rater accuracy scores of participants it the FORT condition will be higher than the rater accuracy scores of those in the control condition.

(a) Differential elevation scores of participants in the FORT condition will be lower (indicating greater accuracy) than the differential elevation scores of those in the control condition. Those in the FORT condition will discriminate better between ratees across all rated dimensions than those in the control condition.

(b) Stereotype accuracy scores of participants in the FORT condition will

be lower (indicating greater accuracy) than those in those in the

control condition. Individuals in the FORT condition will discriminate

better among rated dimensions than those in the control condition.

(c) Differential accuracy scores of participants in the FORT condition will

be lower (indicating greater accuracy) than those in the control

condition. Individuals in the FORT condition will discriminate better

among ratees within individual dimensions than those in the control

condition.

Method

Participants. Undergraduate students at a Midwestern university aged eighteen

and above were recruited to participate in this study. A total sample of 190 students

participated in this study, with 137 individuals completing both Part 1 (short online

survey) and Part 2 (lab session). Assumptions of normality and homogeneity of variance

were tested and attention checks were examined prior to data analysis. Twenty-three

(16.8%) participants who failed attention check items were removed from the dataset and

no alarming outliers were identified. After data cleaning, 114 participants were included

for analysis, with 56 students (49%) in the Training condition and 58 students (51%) in

the Control condition. The final sample was 56% male and 44% female between the ages

of 18 and 64 years old, with 63% identifying as Caucasian, 18% identifying as African

American, 6% identifying as Asian, and 1% identifying as Hispanic/Latin-American;

11% were of other ethnicities. The majority of participants indicated that they were

employed (74%) and most reported they did not have prior management or performance evaluation experience (68%).

Design. Participants were randomly assigned to two conditions: FORT and a control condition (no training). Lab research assistants served as trainers for FORT. They attended practice sessions and were provided with scripts to facilitate a consistent experience for participants. Each lab session was allotted no more than 90 minutes. Athey and McIntyre (1987) found that group size did not have an effect on the retention of training information, and as a result, session size was not specifically controlled (apart from random assignment). Several participants could join a given training session, and these were held in computer labs. All participants viewed three vignettes (i.e., good, average, and poor performance for different rated dimensions) as practice and three vignettes (also with good, average, and poor performance for different rated dimensions) as target/test ratings. In total, participants viewed six videos. A job talk presentation was rated for practice, and role play between a subordinate and a manager was evaluated as the target/test rating.

Procedure. Overall, the protocol followed steps developed by Stamoulis and Haustein (1993). Participants signed-up for a lab time-slot to participate in this study. Before signing up for a lab time, participants were asked to complete a brief survey responding to personality, self-efficacy, and performance appraisal discomfort items. After completing this survey, participants chose a lab session at a time that was convenient for them. Once they arrived to the lab, they filled out an informed consent form and were randomly assigned to either the FORT condition or the control condition. Next, participants in the control condition watched six videos and made ratings – three

for practice, and three for test trials. Participants were told to imagine they were making ratings that would influence whether someone was going to be hired or receive a promotion. They were asked to envision that their ratings would impact the future of those they would observe in order to simulate an actual performance appraisal situation.

Participants in the FORT condition participated in a brief training in which they learned specific content about "effective" and "ineffective" behaviors to help them improve their performance ratings. These individuals also watched the three practice trials with a trained lab assistant and received feedback. Verbal feedback was provided to those in the training condition. The groups discussed how participant ratings compared to true score ratings developed by subject matter experts and trainers answered any questions. After training, the participants watched the final three videos and completed the test trials, then filled out a performance rating form on their own. Once all participants completed the test ratings, they took a final survey. The survey was a second measure of performance appraisal discomfort along with relevant demographic variables. After completing this questionnaire, lab assistants debriefed with the participants and awarded them credit for participating. This entire study (pre-work survey and lab session) lasted no longer than 90 minutes in total.

Materials. The following stimulus and training materials were used for Study 2.

Stimulus Materials. An existing set of laboratory videos that were developed by graduate students were used for practice and target/test ratings. The videos were based on scenarios scripted by experienced assessment center researchers and practitioners and filmed with actors. Subject matter experts made "true score" ratings. In Noonan and Sulsky's (2001) study, two ratees were used for training practice and three ratees were

used to assess rating accuracy. Using this method as a model, a stimulus set consisted of

six videos; three videos were used for practice and three videos were used for target/test

ratings. These videos depicted an assessment center appraisal context and job talk

presentation context. Participants were responsible for rating the supervisor's behavior in

the appraisal context and the job candidate in the job talk presentation context.

Participants used three dimensions to assess performance (see Meriac, Hoffman, &

Woehr, 2014). Critical incidents exemplifying performance at alternative levels (poor,

average, excellent) were exhibited in each video that was approximately five to eight

minutes in length.

   Frame-of-Reference Training. The procedures for FORT followed those

developed by Noonan and Sulsky (2001) and Pulakos (1984; 1986). Participants who

were randomly selected to receive FORT reviewed information about performance

appraisals and the importance of ratings (see Appendix E for a sample FORT handout

with training details). They examined videotaped incidents and received evaluative and

behavioral cues corresponding to each rating scale item. Participants received feedback

from trained laboratory research assistants and were able to discuss their practice ratings

so that they would have a strong understanding of appropriate behaviors indicative of

each performance level. Lab assistants briefly explained the strengths and weaknesses of

participants' rationales and answered any questions as they discussed examples of

effective and ineffective behaviors. Participants reviewed three videos for practice before

watching three videos on their own as a "test trial" where they did not receive feedback.

   Control Training. Participants who were randomly assigned to the control

condition viewed all of the same videos as those in the FORT condition; however, they

did not receive feedback or discuss the performance appraisal process. These individuals watched the three practice videos and then made test ratings for the three test videos on their own with no discussion detailing specific behaviors that they should observe for each performance level of the different dimensions.

Measured Variables. The following variables were measured in Study 2. See Appendix D for a list of items for each measure.

Agreeableness and Conscientiousness. Personality subscales were measured using Bent-Martinez and John's (1998) English Big Five Inventory (BFI), as used in Study 1. Respondents will indicate their agreement to 44 items on a five-point Likert scale (1 = "strongly disagree", 5 = "strongly agree"). Two personality characteristics from the BFI (α = .73) were measured in this study: Agreeableness (α = .74) and Conscientiousness (α = .78).

Rater Self-Efficacy. Self-Efficacy was measured using the Performance Appraisal Self-Efficacy Scale (PASES) developed by Bernardin and Villanova (2005), as used in Study 1. Participants were asked to indicate the level of confidence they would have in being able to successfully perform the behaviors described in 12 statements on a five-point Likert scale (1 = with no confidence, 5 = with great confidence). The PASES (α = .92) was created by Bernardin and Villanova (2005) and has since been used by Brutus, Fletcher, and Bailey (2008). The four subscales composed of three items each yielded acceptable reliability: process features of the appraisal (α = .85), rater subjectivity (α = .79), appraisal discussion (α = .83), and suggesting performance improvement (α = .80). Higher scores relate to higher levels of self-efficacy.

Rater Discomfort. Participants completed the Discomfort with Peer Evaluation Scale (Saffie-Robertson & Brutus, 2014), as used in Study 1 but tailored for a student population. Responses are meant to reflect the degree of discomfort felt by raters in performance appraisal situations. Participants rated their perceived discomfort level on a five-point Likert scale (1 = no discomfort, 5 = high discomfort). The discomfort scale (α = .93) included three subscales (collecting information, two items, $\alpha = 0.83$; rating, four items, $\alpha = 0.84$; and post-rating, four items, $\alpha = 0.87$), each reflecting different steps a performance evaluation process (i.e., discomfort felt when collecting information used for evaluation purpose, discomfort felt when filling out rating form, and discomfort felt after the rating process is over).

Measured Dimensions. New and specific to Study 2, three dimensions of performance were rated to examine the effect of training on performance appraisal discomfort. Meriac et al. (2014) reviewed the structure of assessment dimensions. Their results supported a three-factor model of dimensions for use with making ratings of behavior: Administrative Skills, Drive, and Relational Skills. In order to allow for behaviors that fall into each of these categories to be more clearly delineated by student raters, three narrowed subscale dimensions based on the dimensions in a handbook for managers (Gebelein et al., 2010) were used in this study. These dimensions included: Builds Realistic Plans (e.g., considering feasible expectations, determining resources needed to be successful, stating clear objectives), Shows Drive and Initiative (e.g., showing motivation, tenacity, and energy, initiating, putting in extra effort, focusing on results), and Establishes Relationships (e.g., showing sensitivity and an awareness of others' feelings, creating and maintaining positive partnerships with others). These

dimensions were rated on a five-point Likert scale (1 = very ineffective, 5 = very

effective). See [Appendix E] for a Sample Rating Form with definitions and behavioral

examples of each dimension.

Rating Scale and Comparison Scores. Specific to Study 2, a five-point Likert

rating scale (1 = very ineffective, 5 = very effective) was used to measure job

performance seen in the videos. Comparison scores were derived from subject matter

experts on all performance dimensions and were used as "true scores" and to calculate

Cronbach's (1955) accuracy measures. A group of three subject matter experts made

ratings alone and then discussed any discrepancies between ratings. The subject matter

experts were consultants who assess performance and calibrate ratings together regularly.

The final consensus rating was used as the "true score".

Overall Performance. In Study 2, participants were asked to assess the overall

performance of the target person in the video on a seven-point Likert scale (1 = strongly

disagree, 5 = strongly agree). Saffie-Robertson and Brutus (2014) utilized two items to

rate satisfaction with the target person's performance and an overall rating of the target

person's performance.

Rater Leniency. To evaluate performance in Study 2, rater leniency was

measured as it has been in previous studies (Saffie-Robertson & Brutus, 2014; Villanova

et al., 1993). The difference between the rater's score and the true score across all ratees

and dimensions was used to measure leniency. This is also the first component of

Cronbach's (1955) accuracy indices, namely elevation.

Rater Accuracy. Rating accuracy was computed by comparing participants'

ratings to the comparison scores that were derived from expert raters in Study 2.

Specifically, Cronbach's (1955) accuracy component scores, which are an index that

represents the squared difference between subject ratings and true scores averaged across

the number of ratees and number of dimensions were assessed. Sulsky and Balzer (1988)

provided formulas for these computations. The component scores are derived using the

logic of analysis of variance (ANOVA), with each component expressing a different

portion of the distance between rater ratings and true scores (Sulsky & Balzer, 1988).

Using ANOVA terminology, the components include:

(a) elevation (E), the differential grand mean and average rating over all

ratees and items, which is calculated by computing the squared difference

between the rater's score and the true score to compare the two for across all

ratees and dimensions (i.e., (overall rater mean rating – overall true score mean

rating)^2);

(b) differential elevation (DE), the differential main effect of ratees, which

is calculated by computing the sum of the squared differences between two

deviation scores to determine the average rating assigned by a rater to each ratee

across all performance dimensions, as if ranking the best performer to the worst

performer (i.e., 1/#ratees*$\sum$[(rater mean for ratee – overall rater mean rating) –

(true score mean for ratee – overall true score mean rating)]^2;

(c) stereotype accuracy (SA), the differential main effect of dimensions

and refers to accuracy in discriminating among performance dimensions,

averaging over ratees to identify which performance for a ratee (in this case, out

of 3 dimensions) was best, which is calculated by computing the sum of the

difference in *dimension* scores between the rater and true score (i.e.,

1/#dimensions*$\sum$[(rater mean for dimension – overall rater mean rating) – (true

score mean for dimension – overall true score mean rating)]^2); and finally

(d) differential accuracy (DA), the differential ratee by dimension

interaction and shows a rater's ability to distinguish among ratees within

dimension and recognize patterns of performance, which is calculated by

computing the difference between a rater's score and the true score taking both

the dimension and ratee into account (i.e., 1/(#ratees*#dimensions) *$\sum$[(rater

score for each ratee on each dimension – rater mean for ratee – rater mean for

dimensions + overall rater mean rating) – (true score for each ratee on each

dimension – true score mean for ratee – true score mean for dimensions + overall

true score rating)]^2) (see Sinclar, 2000; Sulsky & Balzer, 1988).

Elevation was used to calculate rater leniency (as indicated above) while the other

three components were used to calculate rater accuracy. Smaller values of the resulting

squared component scores denote greater accuracy.

Training Reactions. Items evaluating reactions to training were used to gauge

participants' response to the training procedures, length, content, and lab rater trainer (α =

0.86). These items allowed for the identification of any effects due to the rater trainer.

Only participants in the FORT condition were asked to answer questions about their

reaction to training in Study 2.

Demographics. Participants were asked to respond to a variety of demographic

questions, including information about their age, sex, race, employment status, and

management or performance rating experience.

Results

Hypotheses 9-12 explored the relationship between personality variables and rater behaviors. Hypothesis 9 stated that there would be a significant positive relationship between agreeableness and rater leniency, such that those who indicate higher levels of agreeableness would make more lenient ratings. A bivariate correlation revealed a nonsignificant negative relationship between agreeableness and elevation, $r(108) = -0.07$, $p = 0.444$, which did not provide support for this hypothesis.

Hypothesis 10 stated that there would be a significant negative relationship between agreeableness and rater accuracy, such that individuals who indicated higher levels of agreeableness would make less accurate ratings. Specifically, it was expected that there would be a negative relationship between agreeableness and differential accuracy, stereotype accuracy, and differential accuracy. It is important to note that when using Cronbach's accuracy component scores, smaller values denote greater accuracy. Consequently, a positive correlation would support this hypothesis. A bivariate correlation showed nonsignificant negative correlations between agreeableness and differential elevation, $r(108) = -0.05$, $p = 0.619$, between agreeableness and stereotype accuracy $r(108) = -0.04$, $p = 0.719$, and between agreeableness and differential accuracy, $r(108) = -0.15$, $p = 0.114$. Therefore, Hypothesis 10 was not supported.

Hypothesis 11 postulated that there would be a significant negative relationship between conscientiousness and rater elevation. Those who are more conscientiousness were expected to make less lenient ratings. Lenient ratings were indicated by higher elevation scores. A bivariate correlation revealed a nonsignificant positive relationship

between conscientiousness and elevation, $r(108) = 0.07$, $p = 0.474$, which did not support the hypothesized outcome.

Conversely, Hypothesis 12 proposed that there would be a significant positive relationship between conscientiousness and rater accuracy. Those who indicated higher levels of conscientiousness were expected to make more accurate ratings. Because smaller values denote greater accuracy, a negative correlation would support this hypothesis. A bivariate correlation showed nonsignificant positive correlations between conscientiousness and differential elevation, $r(108) = 0.02$, $p = 0.847$, and between conscientiousness and stereotype accuracy, $r(108) = 0.05$, $p = 0.601$. A bivariate correlation showed a nonsignificant positive correlation between conscientiousness and differential accuracy, $r(108) = -0.12$, $p = 0.150$. Hypothesis 12 was not supported.

Hypotheses 13 and 14 investigated the relationship between rater self-efficacy and rater behaviors. Hypothesis 13 stated that individuals with higher levels of rater self-efficacy would make less lenient ratings, indicating a significant negative relationship between rater self-efficacy and rater leniency. A bivariate correlation revealed a nonsignificant negative correlation between self-efficacy and elevation or leniency, $r(108) = -0.01$, $p = 0.912$. Therefore, Hypothesis 13 was not supported.

Hypothesis 14 suggested that individuals with higher levels of rater self-efficacy would make more accurate ratings, hypothesizing that there would be a significant positive relationship between self-efficacy and rater accuracy. Because smaller values denote greater accuracy, negative correlations would support this hypothesis. A bivariate correlation showed nonsignificant positive correlations between self-efficacy and differential elevation, $r(108) = 0.14$, $p = 0.138$, and between self-efficacy and differential

accuracy, $r(108) = 0.02$, $p = 0.881$. A bivariate correlation showed a nonsignificant

negative correlation between self-efficacy and stereotype accuracy, $r(108) = -0.16$, $p =$

0.093. Hypothesis 14 was not supported by significant results.

Hypothesis 15 and 16 examined the influence of a training manipulation on

performance appraisal discomfort. Participants were asked to indicate the quality of the

rater training they experienced and also rated their rater trainer. Most of the participants

indicated that the FOR training was "very" or "extremely" useful (63%) and the majority

agreed the training was directly relevant to the task of rating performance (73%).

Participants reported that they learned to distinguish between high and low performers

and between "effective" and "ineffective" performance (82%). The majority also

indicated that the trainer/facilitator was knowledgeable about the training content (86%)

and 80% of participants reported that they would recommend the training to others.

Overall, participants indicated that training purpose was clear and rater trainers

effectively presented the information.

Hypothesis 15 stated that participants who experienced FORT would report

significantly lower ratings of performance appraisal discomfort after the training

compared to ratings made before the training. A repeated-measures paired-samples t-test

was conducted to compare ratings of performance appraisal discomfort before and after

training. Results revealed a significant difference in ratings of performance appraisal

discomfort among the 54 participants in the FORT condition before ($M = 2.39$, $SD =$

1.02) and after training ($M = 2.00$, $SD = 0.93$), $t(53) = 2.81$, $p = 0.007$. Specifically,

participants indicated experiencing less performance appraisal discomfort after receiving

training than they did before the training session. This finding supported the hypothesis and suggests that training does have an impact on performance appraisal discomfort.

Hypothesis 16 postulated that participants who experienced FORT would have significantly lower ratings of rater discomfort compared to participants in a control group who were not exposed to any type of training. An independent samples t-test revealed nonsignificant differences in ratings of rater discomfort between those in the FORT group ($M = 2.00$, $SD = 0.93$) and those in the Control group ($M = 2.05$, $SD = 0.94$), $t(112) = -0.29$, $p = 0.769$. Though the finding was not significant, reports of rater discomfort were lower among those in the FORT group.

Hypotheses 17-20 investigated the influence of training on performance appraisal discomfort and rater behaviors. Hypothesis 17 stated that there would be a significant positive relationship between performance appraisal discomfort and evaluation leniency after training among those in the FORT condition. A bivariate correlation revealed a nonsignificant negative correlation between rater discomfort and elevation or leniency, $r(56) = -0.05$, $p = 0.730$. This finding does not support the hypothesis.

Hypothesis 18 stated that there would be a significant negative relationship between rater discomfort and rating accuracy after training among those in the FORT condition. Because smaller values denote greater accuracy, positive correlations would support this hypothesis. A bivariate correlation revealed nonsignificant positive correlations between performance appraisal discomfort and differential elevation, $r(56) = 0.16$, $p = 0.242$, and between rater discomfort and differential accuracy, $r(56) = 0.13$, $p = 0.353$. A bivariate correlation showed a nonsignificant negative correlation between rater

discomfort and stereotype accuracy, $r(56) = -0.16$, $p = 0.226$. Unfortunately, none of the findings were significant, and Hypothesis 18 was not supported.

Hypothesis 19 asserted that there would be a significant difference in leniency scores between the control and FORT conditions. Specifically, those in the control condition will have significantly higher leniency scores than those in the FORT condition. An independent samples t-test revealed significant differences in leniency scores between those in the FORT group ($M = 0.16$, $SD = 0.27$) and those in the Control group ($M = 0.38$, $SD = 0.68$), $t(112) = -2.26$, $p = 0.025$. Those in the Control group had higher elevation scores, indicating more lenient ratings than those in the FORT condition and providing support for Hypothesis 19.

Hypothesis 20 indicated that there would be a significant difference in accuracy scores between participants in the control and FORT conditions. Specifically, those in the control condition were expected to have significantly lower accuracy scores than those in the FORT condition. An independent measures t-test resulted in significant differences in differential elevation scores between those in the FORT group ($M = 0.99$, $SD = 0.86$) and those in the Control group ($M = 1.40$ $SD = 1.17$), $t(112) = -2.09$, $p = 0.039$. Additionally, there were significant differences in differential accuracy scores between those in the FORT group ($M = 0.21$, $SD = 0.11$) and those in the Control group ($M = 0.27$, $SD = 0.14$), $t(112) = -2.44$, $p = 0.016$. There were nonsignificant differences in stereotype accuracy scores between those in the FORT group ($M = 0.12$, $SD = 0.11$) and those in the Control group ($M = 0.09$, $SD = 0.10$), $t(112) = 1.37$, $p = 0.175$. A multivariate analysis of variance was also conducted to examine all accuracy indices together. The MANOVA resulted in a nonsignificant difference in accuracy indices based on the

training group (F (3,110) = 2.56, p = .059; Wilks Λ = 0.935, partial η^2 = 0.065).

Participants in the FORT condition were more accurate than those in the Control

condition on differential elevation and differential accuracy, indicating that they were

significantly more accurate when judging between ratees and identifying which ratee had

the highest score on a given dimension. There was not a significant difference between

the two groups on stereotype accuracy.

Study 2 Discussion

The objective of Study 2 was to explore the relationship between performance

appraisal discomfort and rating factors (i.e., leniency and accuracy). Individuals who

reported lower levels of agreeableness, but higher levels of conscientiousness and self-

efficacy were expected to make less lenient and more accurate ratings. Results revealed

that personality factors were not significantly related to rater accuracy or leniency.

Additionally, self-efficacy was not significantly related to leniency or accuracy. Since

these individual difference factors were not found to influence rater outcomes, rater

behavior could be a factor that all different types of managers could learn to improve

upon. These findings could suggest that any personality type could be a successful rater.

In terms of selection processes for managers, organizations may not necessarily need to

select based on personality indicators. Additional research into other nuances of

personality could be advantageous.

The final objective of Study 2 was to investigate the influence of FORT on rater

discomfort. Results showed that participants reported experiencing reduced discomfort

after receiving FORT than they did prior to the training session. Additionally, though the

finding was not significant, the FORT group reported lower ratings of rater discomfort

than those in the Control group. This research lends some support for using FORT for new managers or supervisors to help them learn behaviors they should be observing for different levels of performance as well as develop a vocabulary to discuss these behaviors during performance evaluations.

Regarding rater behavior among those on the FORT condition, rater discomfort was not significantly related to lenient or accurate ratings. When comparing the Control and FORT groups, those in the FORT condition made significantly less lenient ratings. Differences in rater accuracy between the Control and FORT conditions were also found, and these results were consistent with previous rater training research showing that training was generally effective at improving rater accuracy and decreasing leniency (Noonan & Sulsky, 2001).

Though rater discomfort did not appear to be related to accuracy or leniency, there were differences between the FORT and Control conditions suggesting that training does play a role. As ratings could become less formal, perhaps accuracy and leniency are not as important as the opportunity to be trained in how to interact with your subordinate. Making sure that employees are trained should afford individuals the tools they need to observe and discuss behavior productively with their direct reports should allow managers and supervisors to feel more comfortable with the evaluations they make, as they will have the support and evidence they need to back up their claims.

In terms of limitations for Study 2, the three dimensions that were rated, Builds Realistic Plans, Shows Drive and Initiative, and Establishes Relationships, may have been too broad and harder for student participants to learn how to distinguish between during the brief training session. An open text box was provided under each dimension

for participants to take notes while watching the videos to use when making their final ratings. Analysis of the notes showed that some behaviors could have been classified under multiple dimension.

Furthermore, the rated dimensions were derived and theorized from assessment center, performance appraisal, and leadership literature (Meriac et al., 2014). More specific dimensions based on a job analysis for the two contexts in the videos may have assisted students with making more robust assessments of behaviors. While the framework was intended to be general and each dimension was presented with examples and definitions, it is possible that even more specific and narrowly defined dimensions would have been preferable for novice raters. Future research should investigate rated dimensions to further contribute to performance appraisal research. Such research could provide insights for training development purposes and also for actual performance appraisal or evaluation scenarios when competencies and/or rated dimensions need to be specific and job-related.

Another limitation and future research opportunity is related to the experience of FORT. In the future, it would be interesting to have a condition where the "expert" ratings are offered as an "average" rating, and compare how participants rate based on this information. Additionally, explaining the anchors used for rating purposes more clearly might also assist with bolstering the impact of training. For those in the FORT condition, participants did review the rating scale and identified "effective" and "ineffective" behaviors for each rated dimension; however, students may not have clearly understood the difference between each behavioral anchor. For example, is a "5" rating reserved for Rockstar performers and should only be given out sparingly? These types of

discussions may firm up understanding regarding the best rating to make in an appraisal or evaluation context.

General Discussion

In summary, there were a few interesting findings regarding similarities and inconsistencies between the two studies when looking at rater discomfort. While both agreeableness and conscientiousness were related to discomfort among the sample of managers, neither were related to rater discomfort among the sample of students. Self-efficacy was found to be related to rater discomfort in both studies. Surprisingly, participants in both studies indicated "little" or "no" discomfort with making ratings, reducing the likelihood of finding significant relationships with rater discomfort.

A couple of reasons for some of these findings were identified. First, it seems plausible that the context and purpose of the appraisal may have played a role. For example, if participants had to justify their ratings for an important administrative decision, perhaps ratings of rater discomfort may have varied more extensively. Secondly, participants in both studies were anonymous. Perhaps if some of the items around rater discomfort were framed more closely around the face-to-face component of providing feedback, there may have been different reactions. In Study 2, there were no specific consequences for the ratings made, which may be why student raters felt no concerns with making ratings. They tend to feel comfortable appraising the performance of their professors and end-of-semester evaluations, and similarly, may have simply felt comfortable with the anonymity of the rating process in this research context.

Taken together, all these findings have important implications for the workplace, for both selection and development purposes. Considering the relationships between self-

efficacy and rater discomfort, organizations might consider further investigation into self-efficacy and rater ability when considering who to promote or select for a management position. Additionally, although rater training did not appear to be significantly related to personality variables, results of both studies suggested that training could help raters be more confident in their ratings, lending justification and endorsement for the idea that employees in supervisory roles should attend training to obtain the resources they need to conduct effective performance appraisals. This study fills a void in research by contributing to what we know about rater training and rater discomfort; however, there are limitations to consider.

Limitations and Future Research

Because no previous research has investigated how to reduce performance appraisal discomfort, this study is largely exploratory in nature; however, overall, the results related to discomfort where somewhat disappointing. It is likely that the samples used for this research influenced results, or it is possible the measure of discomfort could have been improved. The subscales of rater discomfort might be more closely delineated and perhaps a clearer explanation of what rater discomfort looks like to make the experience of discomfort feel real could have made a difference. Research has shown that there are several components to the performance appraisal process (i.e., data collection, making ratings, and giving performance feedback) that could be investigated in more detail.

As mentioned previously, the performance appraisal process includes more than making ratings about performance. Results from Study 1 showed that most participants were not uncomfortable with making ratings, but many indicated that they engage in

face-to-face feedback discussions. Future researchers should investigate the relationship among feedback, training, and performance appraisal discomfort. Macan, Mehner, Havill, Roberts, Heft, and Meriac (2011) found that managers who were trained as assessors made more behaviorally specific comments on performance evaluations than managers who were not trained. Other research has shown that training aids individuals in developing a more precise vocabulary for describing behaviors (Byham, 1971; Thornton & Rupp, 2005), implying that it may be important to explore components of feedback as they relate to performance appraisal discomfort. Future research should explore the relationship between performance appraisal discomfort and performance feedback quality, and the potential mediating effect of rater training.

An additional consideration is related to the samples of both studies. When conducting online and lab studies, it can be difficult to get the participant invested in the outcomes of interest. One particular challenge for Study 2 was getting students to be dedicated to the task at hand. Accountability is a key area that could confound these results. The notion of accountability is very important in performance appraisals (Bernardin et al., 2015; Curtis, Harvey, & Ravden, 2005; Mero, Guidice, & Brownlee, 2007). People who believe they are actually going to impact another person's career are apt to make different ratings than people who are in a simulated lab environment and making ratings, or people who are reflecting back on their careers and experiences. While the videos used in Study 2 were vetted by subject matter experts, no one was required to actually have a conversation with a person afterwards after making ratings. In the absence of this accountability, perhaps the students did not take the training exercise as seriously as they could have. Having to look a subordinate in the eye and being held accountable to

justify ratings is important for performance appraisals and is a component that should be included in future research studies that explore rater behavior and training impact.

Another limitations and potential avenue for future research is that neither study included physiological measures of discomfort. In order to improve the way rater discomfort is understood, it would be beneficial to have multiple measures of discomfort. It may be quite easy for participants to make a quick rating of discomfort whereas physiological measures could contribute new information regarding stress response in performance evaluation contexts. Further related to this would be to have additional sources of information about the ratings made instead of only participant self-report.

Among all of these variables for future research considerations, a model of the stages of the performance appraisal system demonstrating the influence of personality, discomfort, and training on feedback and rating accuracy should be constructed and tested. Such a model could be integral for combining performance appraisal literature and could also be practical and useful for organizations.

Practical Implications and Conclusions

The results of this research yield a few implications. Organizations could consider measuring performance appraisal discomfort as a training needs assessment for employees in management positions with supervisory performance appraisal duties. This study did not address the "amount" of training specifically, and because the majority of the managerial sample indicated being comfortable with making performance ratings, this subject was not addressed. Since performance appraisal outcomes are linked to important organizational outcomes, such as compensation decisions, it is important that organizations ensure that their managers are providing accurate evaluations of their

subordinates. Additionally, investigating performance appraisal discomfort could be important for reducing turnover and absenteeism among individuals in supervisory roles.

There are also practical implications for other facets of appraisal, specifically 360-degree feedback. Gillipse, Rose, and Robinson (2006) found that comments provided by supervisors and subordinates were clearer than comments from peers in 360-degree feedback. This finding may suggest that anyone who is evaluating behaviors should be trained first. For example, though training may be most important for supervisors, it could also be useful for other employees who would be asked to evaluate someone's behavior at work, such as peers or direct reports. Perhaps supervisors should have access to multiple trainings so that they can continually refresh their skills, while other employees could be exposed to information related to FORT during employee onboarding so that they are familiar with how to observe the behaviors going on around them and provide constructive feedback when appropriate. Research into 360-degree feedback and performance appraisal discomfort may be important to evaluate as more organizations are utilizing 360-degree evaluations. Training peers, direct reports, and supervisors to be comfortable with ratings could improve the effectiveness and use of 360-degree appraisals.

It may also be worthwhile to consider the types of relationships that exist between managers and their subordinates. Examining case studies by investigating organizations where ratings have a strong impact could yield interesting findings. There can be both generational differences and managers who were once peers but were promoted into a supervisory role where they now must manage their friends. These types of experiences could benefit from training, allowing the new manager to understand how to manage

generational differences or even how to effectively manage and evaluate a subordinate who was once a peer. Aspects of FORT could be useful in both situations. Training could be an easy solution to help organizations avoid accusations of discrimination. Making sure that managers have the resources they need to provide evidence for their performance evaluations should reduce the likelihood that employees will feel they are treated unfairly. This could be an important practical implication of training research in the future.

Furthermore, there may be implications for the stipulations placed on managers when conducting performance appraisals. For example, some organizations may force managers to rank order their employees, potentially resulting in poor performers receiving an unfair rating simply because management requires a distribution of ratings. Training could improve managers' ability to truly understand their employees' performance and to be better able to defend the ratings they give. Rater training may also provide support in opposition of forced distribution performance appraisals, showing that given the tools, managers can offer strong feedback and identify developmental needs for their people that are more meaningful than rank ordering.

Knowing whether a manager is comfortable with performance appraisal or not should help organizations to identify which managers may need more training in order to develop comfort with the skill of appraising performance. The different dimensions of performance appraisal discomfort could be used to identify and address development needs. It is important that researchers consider that raters may not be distorting ratings only for political reasons. New managers in supervisory roles may lack confidence in

evaluating behavior and providing them with training could ease their fears about

conducting evaluations of performance.

References

Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy:
Levels-of-processing theory and social facilitation theory perspectives. *Journal of
Applied Psychology*, *72*(4), 567-572.
http://dx.doi.org/10.1037/0021-9010.72.4.567

Baldwin, T. T., & Ford, J. (1988). Transfer of training: A review and directions for future
research. *Personnel Psychology*, *41*(1), 63-105.
http://dx.doi.org/10.1111/j.1744-6570.1988.tb00632.x

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American
Psychologist*, *37*(2), 122-147. http://dx.doi.org/10.1037/0003-066X.37.2.122

Benet-Martinez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic
groups: Multitrait multimethod analyses of the big five in Spanish and English.
Journal of Personality and Social Psychology, 75, 729-750.
http://dx.doi.org/10.1037/0022-3514.75.3.729

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of
Management: The Academy of Management Review, 6*(2), 205-212.
http://dx.doi.org/10.2307/257876

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and
agreeableness as predictors of rating leniency. *Journal of Applied
Psychology*, *85*(2), 232-236. http://dx.doi.org/10.1037/0021-9010.85.2.232

Bernardin, H.J., Thomason, S., Buckley, M.R., and Kane, J. S. (2015). Rater rating-level
bias and accuracy in performance appraisals: The impact of rater personality,

performance management competence, and rater accountability. *Human Resource Management, 55*, 321–340. http://dx.doi.org/10.1002/hrm.21678

Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment*, *17*(3), 300-310. http://dx.doi.org/10.1111/j.1468-2389.2009.00472.x

Bernardin, H.J., & Villanova, P. (1986). Performance appraisal. In E.A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp.43-62). Lexington, MA: Lexington Books.

Bernardin, J.H., & Villanova, P. (2005). Research streams in rater self-efficacy. *Group & Organization Management*, 30, 1, 61–88. http://dx.doi.org/10.1177/1059601104267675

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, *86*(5), 1549-1575. http://dx.doi.org/10.2308/accr-10099

Bretz, R. D., Milkovich, G. T., & Read, W. (1989). Comparing the performance appraisal practices in large firms with the directions in research literature: Learning more and more about less and less. *Center for Advanced Human Resources Studies Working Paper Series #89-17*, 414. Ithaca, NY: Cornell University, School of Industrial and Labor Relations. http://digitalcommons.ilr.cornell.edu/cahrswp/414

Brutus, S. (2010). Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal. *Human Resource Management Review*, *20*(2), 144-157. http://dx.doi.org/10.1016/j.hrmr.2009.06.003

Brutus, S., & Donia, M. B. (2010). Improving the effectiveness of students in groups with

a centralized peer evaluation system. *Academy of Management Learning &

Education*, *9*(4), 652-662. http://dx.doi.org/10.5465/AMLE.2010.56659882

Brutus, S., Fletcher, C., & Baldry, C. (2009). The influence of independent self-construal

on rater self-efficacy in performance appraisal. *The International Journal of

Human Resource Management*, 20(9), 1999–2011.

http://dx.doi.org/10.1080/09585190903142431

Byham, W. C. (1971). The assessment center as an aid in management development.

Training and Development Journal, *25(12),* 10-22.

Clark, M. H., & Schroth, C. A. (2010). Examining relationships between academic

motivation and personality among college students. *Learning and Individual

Differences*, *20*(1), 19-24. http://dx.doi.org/10.1016/j.lindif.2009.10.002

Cleveland, J. N., & Murphy, K. R. (1992). Analyzing performance appraisal as goal-

directed behavior. *Research in Personnel and Human Resources

Management*, *10*(2), 121-185.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance

appraisal: Prevalence and correlates. *Journal of Applied Psychology*, *74*(1), 130-

135. http://dx.doi.org/10.1037/0021-9010.74.1.130

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory and the NEO

five-factor inventory*. Odessa, FL: Psychological Assessment Resources.

Coutts, L. M., & Schneider, F. W. (2004). Police officer performance appraisal systems:

How good are they?. *Policing: An International Journal of Police Strategies &

Management*, *27*(1), 67-81. http://dx.doi.org/10.1108/13639510410519921

Cox, S. S., Marler, L. E., Simmering, M. J., & Totten, J. W. (2011). Giving feedback:

    Development of scales for the mum effect, discomfort giving feedback, and

    feedback medium preference. *Performance Improvement Quarterly*, *23*(4), 49-69.

    http://dx.doi.org/10.1002/piq.20094

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and

    "assumed similarity." *Psychological Bulletin, 52(3),* 177-193.

    http://dx.doi.org/10.1037/h0044919

Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in

    performance appraisals: Appraisal purpose and rater accountability. *Group &*

    Organization Management, *30*(1), 42-60.

    http://dx.doi.org/10.1177/1059601104267666

DeMarco, J. J., & Nigro, L. G. (1983). Using employee attitudes and perceptions to

    monitor supervisory implementation of CSRA performance appraisal

    systems. *Public Personnel Management*, *12*(1), 43-51.

    http://dx.doi.org/10.1177/009102608301200105

Deming, W. E. (1986). *Out of the crisis*. Massachusetts Institute of Technology, Center

    for Advanced Engineering Study. Cambridge, MA.

Fisher, C. D. (1989). Current and recurrent challenges in HRM. *Journal of*

    Management, *15*(2), 157-180. http://dx.doi.org/10.1177/014920638901500203

Fletcher, C. (2001). Performance appraisal and management: The developing research

    agenda. *Journal of Occupational and Organizational Psychology*, 74, 473–487.

    http://dx.doi.org/10.1348/096317901167488

Fried, Y., Tiegs, R. B., & Bellamy, A. R. (1992). Personal and interpersonal predictors of

    supervisors' avoidance of evaluating subordinates. *Journal of Applied*

    Psychology, *77*(4), 462-468. http://dx.doi.org/10.1037/0021-9010.77.4.462

Gebelein, S. H., Nelson-Neuhaus, K. J., Skube, C. J., Lee, D. G., Stevens, L. A.,

    Hellervik, L. W., & Davis, B. L. (2010). *Successful Manager's Handbook* (8th

    ed.). Roswell, GA: PreVisor.

Gillespie, T., Rose, D. S., & Robinson, G. N. 2006. May. *Narrative comments in 360*

    degree feedback: Who says what? Paper presented at the 21st Annual Conference

    of the Society for Industrial and Organizational Psychology. Dallas, TX.

Gilliland, S.W and Langdon, J.C. (1998). "Creating performance management systems

    that promote perceptions of fairness", in Smither, J.W. (Ed.), *Performance*

    appraisal: State of the art in practice, Jossey-Bass, San Francisco, CA.

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

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.

    http://dx.doi.org/10.1111/joop.12086

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable

    mediation, moderation, and conditional process modeling [White paper].

    Retrieved from http://www.afhayes.com/ public/process2012.pdf

Hill, L. A. (2004). New manager development for the 21st century. *The Academy of*

    Management Executive, *18*(3), 121-126.

    http://dx.doi.org/10.5465/ame.2004.14776182

Ilgen, D. R., & Moore, C. F. (1987). Types and choices of performance feedback. *Journal of Applied Psychology*, *72*(3), 401-406. http://dx.doi.org/10.1037/0021-9010.72.3.401

Jawahar, I. M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology*, *86*(5), 875-883. http://dx.doi.org/10.1037/0021-9010.86.5.875

Jawahar, I. M. (2007). The influence of perceptions of fairness on performance appraisal reactions. *Journal of Labor Research*, *28*(4), 735-754. http://dx.doi.org/10.1007/s12122-007-9014-1

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*(4), 905-925. http://dx.doi.org/10.1111/j.1744-6570.1997.tb01487.x

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal, 38(4),* 1036-1051. http://dx.doi.org/10.5465/256619

Landy, F.J., & Farr, J.L. (1980). Performance rating. *Psychological Bulletin, 87(1),* 72-107. http://dx.doi.org/10.1037/0033-2909.87.1.72

Latham, G. P. (1986). Job performance and appraisal. *International Review of Industrial and Organizational Psychology*, *1*, 117-155.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*(6), 881-905. http://dx.doi.org/10.1016/j.jm.2004.06.005

Longenecker, C.O., Sims, H.P., & Gioia, D.A. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1,* 183-193. http://dx.doi.org/10.5465/ame.1987.4275731

Macan, T., Mehner, K., Havill, L., Roberts, L., Heft, L., & Meriac, J. P. (2011). Two for the price of one: Assessment center training to focus on behaviors can transfer to performance appraisals. *Human Performance*, *24*(5), 443-457. http://dx.doi.org/10.1080/08959285.2011.614664

Meriac, J. P., Hoffman, B. J., & Woehr, D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management*, *40*(5), 1269-1296. http://dx.doi.org/10.1177/0149206314522299

Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management, 33*(2), 223-252. http://dx.doi.org/10.1177/0149206306297633

Meyer, H.H. (1991). A solution to the performance appraisal feedback enigma. *Academy of Management Executive, 5*(1), 68-76. http://dx.doi.org/10.5465/ame.1991.4274724

Mitchell, T. R. (1982). Expectancy-value models in organizational psychology. In N. T. Feather (Ed.), *Expectations and actions: Expectancy-value models in psychology,* 293-312. Hillsdale, NJ: Erlbaum.

Murphy, K. R., & Cleveland, J. N. (1991). *Human resource management series. Performance appraisal: An organizational perspective.* Needham Heights, MA, US: Allyn & Bacon.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social,
    organizational, and goal-based perspectives*. Newbury Park, CA: Sage.

Murphy, K. R., Cleveland, J. N., & Hanscom, M. E. (2018). Performance appraisal and
    management. Newbury Park, CA: Sage Publications.

Napier, N. K., & Latham, G. P. (1986). Outcome expectancies of people who conduct
    performance appraisals. *Personnel Psychology*, *39*(4), 827-837.
    http://dx.doi.org/10.1111/j.1744-6570.1986.tb00597.x

Noonan, L.E., & Sulsky, L.M. (2001). Impact of frame-of-reference and behavioral
    observation training on alternative training effectiveness criteria in a Canadian
    military sample. *Human Performance, 14(1),* 3-26.
    http://dx.doi.org/10.1207/S15327043HUP1401_02

Nowack, K. M., & Mashihi, S. (2012). Evidence-based answers to 15 questions about
    leveraging 360-degree feedback. *Consulting Psychology Journal: Practice and
    Research*, *64*(3), 157-182. http://dx.doi.org/10.1037/a0030011

Pearce, J.L., and Porter, L.W. (1986). Employees responses to formal performance
    appraisal feedback. *Journal of Applied Psychology*, *71*(2), 211–218.

Poon, M.L. (2004). Effects of performance appraisal politics on job satisfaction and
    turnover intention. *Personnel Review, 33(3),* 322-334.
    http://dx.doi.org/10.1108/00483480410528850

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and
    accuracy training. *Journal of Applied Psychology, 69*(4)*,* 581–588.
    http://dx.doi.org/10.1037/0021-9010.69.4.581

Pulakos, E. D. (1986). The development of training programs to increase accuracy with

different rating tasks. *Organizational Behavior and Human Decision Processes,
38(1),* 78–91. http://dx.doi.org/10.1016/0749-5978(86)90027-0

Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management broken?

Industrial and Organizational Psychology, *4(2)*, 146–164.

http://dx.doi.org/10.1111/j.1754-9434.2011.01315.x

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.

http://dx.doi.org/10.1348/2044-8325.002002

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training

revisited: An updated meta-analytic review of frame-of-reference training.

Journal of Occupational and Organizational Psychology, *85*(2), 370-395.

http://dx.doi.org/10.1111/j.2044-8325.2011.02045.x

Rosen, S., & Tesser, A. (1970). On reluctance to communicate undesirable information:

The mum effect. *Sociometry, 33*(3), 253–263. http://dx.doi.org/10.2307/2786156

Saffie-Robertson, M. C., & Brutus, S. (2014). The impact of interdependence on

performance evaluations: The mediating role of discomfort with performance

appraisal. *The International Journal of Human Resource Management*, *25*(3),

459-473. http://dx.doi.org/10.1080/09585192.2013.792864

Sawyer, T. P., Hollis-Sawyer, L. A., & Pokryfke, A. (2002). Personality and social-

evaluative anxieties associated with rating discomfort in anticipated positive and

negative feedback conditions. *Social Behavior and Personality: An International Journal*, *30*(4), 359-371. http://dx.doi.org/10.2224/sbp.2002.30.4.359

Sinclair, A. L. (2000). *Differentiating rater accuracy training programs* (Doctoral dissertation, Virginia Polytechnic Institute and State University).

Smith, W. J., Harrington, K. V., & Houghton, J. D. (2000). Predictors of performance appraisal discomfort: A preliminary examination. *Public Personnel Management*, *29*(1), 21-32. http://dx.doi.org/10.1177/009102600002900102

Smither, J.W. (1998). Lessons learned: Research implications for performance appraisal and management practice. In Smither, J.W. (Ed.), *Performance Appraisal.* San Francisco: Jossey-Bass, 537-547.

Stamoulis, D.T., & Hauenstein, N.M.A. (1993). Rater training and rating accuracy: Training for dimensional accuracy versus training for ratee differentiation. *Journal of Applied Psychology, 78(6),* 694-1003. http://dx.doi.org/10.1037/0021-9010.78.6.994

Suls, J., Martin, R., & David, J.P. (1998). Person-environment fit and its limits: Agreeableness, neuroticism, and emotional reactivity to interpersonal conflict. *Personality & Social Psychology Bulletin, 24(1)*, 88-98. http://dx.doi.org/10.1177/0146167298241007

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

Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of*

Applied Psychology, *77*(4), 501-510. http://dx.doi.org/10.1037/0021-
9010.77.4.501

Thornton III, G. C., & Rupp, D. E. (2005). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. New York: Psychology Press.

Tumlin, G. R. (2004). *Communicating negative feedback in performance appraisal interviews: An experimental study* (Order No. 3145870). Available from ABI/INFORM Collection; ProQuest Dissertations & Theses A&I. (305128086).

Tziner, A., & Murphy, K. R. (1999). Additional evidence of attitudinal influences in performance appraisal. *Journal of Business and Psychology*, *13*(3), 407-419. http://dx.doi.org/10.1023/A:1022982501606

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. http://dx.doi.org/10.1111/1468-2389.00211

Tziner, A., Murphy, K. R., Cleveland, J. N., & Roberts-Thompson, G. P. (2001). Relationships between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, *9*(3), 226-239. http://dx.doi.org/10.1111/1468-2389.00176

Villanova, P., Bernardin, H. J., Dahmus, S. A., & Sims, R. L. (1993). Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement*, *53*(3), 789-799. http://dx.doi.org/10.1177/0013164493053003023

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. http://dx.doi.org/10.1111/j.2044-8325.1994.tb00562.x

Yi, M. Y., & Davis, F. D. (2003). Developing and validating an observational learning

model of computer software training and skill acquisition. *Information Systems

Research*, *14*(2), 146-169. http://dx.doi.org/10.1287/isre.14.2.146.16016

Young, J. (2004). When good technology means bad teaching: Giving professors gadgets

without training can do more harm than good in the classroom, students say. *The

Chronicle of Higher Education*, *51*(12), A31-A37.

Appendix A: Hypothesized Model and Final Model
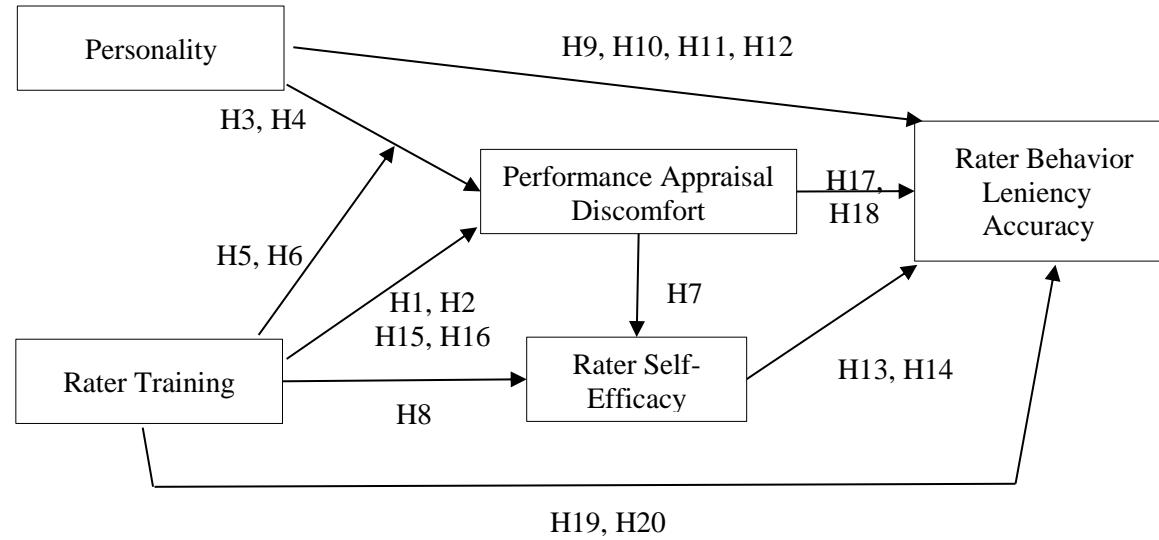
Figure 1. Hypothesized Model
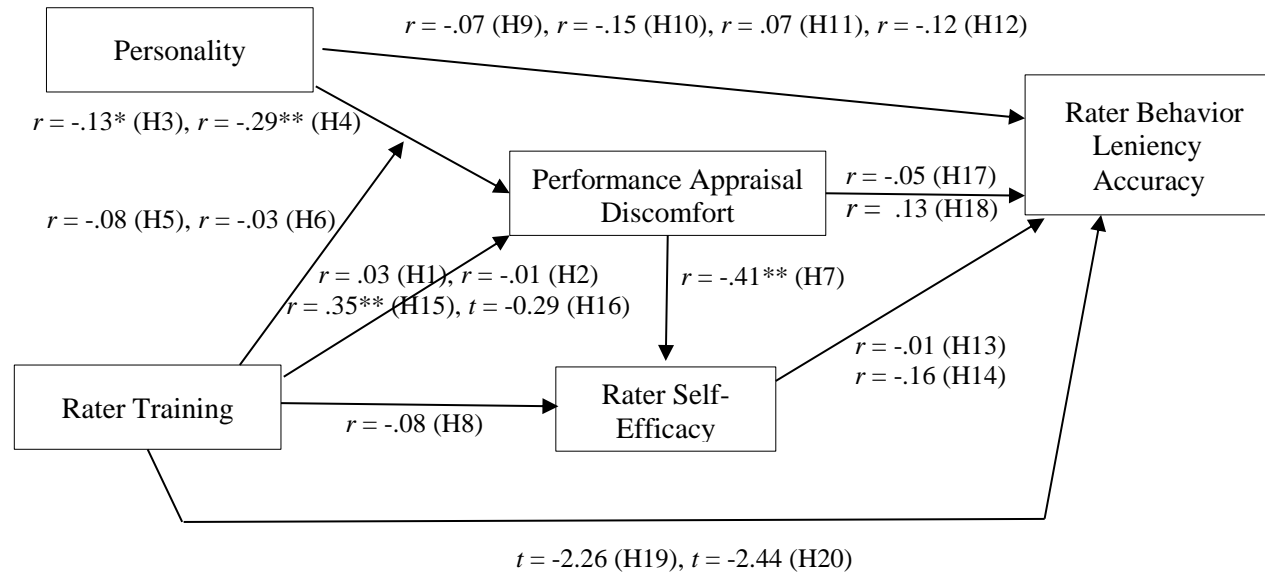
Figure 2. Final Model

Note:

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

Accuracy indices for rater behaviors are represented by Differential Accuracy. This indicator is reported in the figure above, as it encompasses accuracy judgments across both dimensions and raters. See Appendices for additional accuracy indices.

Appendix B: Study 1 Results Tables

Table 1

Descriptive Statistics and Correlations Among Study 1 Variables

Variable	N	M	SD	1	2	3	4	5	6	7	8	9
Training Experience	250	1.42	0.49	1								
Agreeableness	250	4.04	0.70	-0.083	1							
Conscientiousness	250	4.25	0.69	0.103	0.408**	1						
Self-Efficacy	250	4.18	0.70	-0.078	0.302**	0.452**	1					
Rater Discomfort	250	1.84	0.87	0.025	-0.127*	-0.294**	-0.414**	1				
Rater Discomfort (Collecting Info)	250	1.71	0.91	-0.006	-0.128*	-0.232**	-0.321**	0.811**	1			
Rater Discomfort (Rating)	250	1.82	0.86	-0.014	-0.136*	-0.234**	-0.395**	0.906**	0.796**	1		
Rater Discomfort (Post-Rating)	250	1.97	0.85	0.044	-0.136*	-0.249**	-0.394**	0.866**	0.648**	0.770**	1	
Perceived Quality of Training	146	3.63	1.01	a	0.231**	0.248**	0.179*	-0.005	0.018	-0.010	-0.046	1

Note: N = Sample Size, M = Mean, SD = Standard Deviation, **p<.01, *p<.05.

a. Cannot be computed because at least one of the variables is constant.

Table 2

Descriptive Statistics by Training Experience (Study 1)

	Training Experience (N = 146)		No Training Experience (N = 104)	
	M	SD	M	SD
Agreeableness	4.09	0.70	3.97	0.70
Conscientiousness	4.19	0.70	4.34	0.68
Self-Efficacy	4.23	0.69	4.12	0.72
Rater Discomfort	1.82	0.85	1.87	0.89

Note. N = Sample Size, M = Mean, SD = Standard Deviation.

Table 3

Regression Results for Impact of Training on the Relationship between Personality and Rater Discomfort (Study 1)

	B	t	p
Agreeableness	-0.126	-1.99	0.047*
Training Experience (Model 1)	0.018	0.28	0.781
Agreeableness X Training Experience (Model 2)	0.144	1.81	0.071
Conscientiousness	-0.299	-4.88	0.000**
Training Experience (Model 1)	0.055	0.89	0.370
Conscientiousness X Training Experience (Model 2)	0.020	0.33	0.742

Note. N = 250, **$p < .01$, *$p < .05$.

Table 4

Study 1 Gender Demographics

Gender	Frequency	Percent
Male	89	35.6
Female	161	64.4
Total	250	100

Table 5

Study 1 Race/Ethnicity Demographics

Race	Frequency	Percent
African American/Black	14	5.6
Asian	10	4.0
Caucasian/White	207	82.8
Native American	3	1.2
Hispanic/Latin-American	11	4.4
Other	5	2.0
Total	250	100

Table 6

Study 1 Age of Sample

Age	Frequency	Percent
18-24	6	2.4
25-34	50	20.0
35-44	60	24.0
45-54	57	22.8
55-64	61	24.4
65+	16	6.4
Total	250	100

Table 7

Study 1 Employment Status

Status	Frequency	Percent
Employed	224	90.3
Not Employed	24	9.7
Total	248	100

Table 8

Study 1 Hours Worked Per Week

Hours	Frequency	Percent
1-20	15	6.0
21-30	17	6.8
31-40	76	30.4
40+	142	56.8
Total	250	100

Table 9

Study 1 Management/Appraisal Training

Exp.	Frequency	Percent
Yes	146	58.4
No	104	41.6
Total	250	100

Table 10

Study 1 Prior Management Experience

Exp.	Frequency	Percent
Yes	123	49.6
No	125	50.4
Total	248	100

Table 11

Study 1 Experience in Supervisory Role

Exp.	Frequency	Percent
Yes	221	88.4
No	29	11.6
Total	250	100

Table 12

Study 1 Years in Supervisory Role

Years	Frequency	Percent
1-2 years	20	9.1
3-5 years	64	29.1
6-10 years	59	26.8
10+ years	77	35.0
Total	220	100

Appendix C: Study 2 Results Tables

Table 13

Correlations Among Study 2 Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Agreeableness (Part 1)	1														
Conscientiousness (Part 1)	0.417**	1													
Self-Efficacy (Part 1)	0.176	0.331**	1												
Rater Discomfort (Part 1)	-0.185	-0.183	-0.500**	1											
Rater Discomfort (Collect Info, Part 1)	-0.268**	-0.101	-0.453**	0.751**	1										
Rater Discomfort (Rating, Part 1)	-0.086	-0.126	-0.398**	0.866**	0.693**	1									
Rater Discomfort (Post-Rating, Part 1)	-0.174	-0.158	-0.426**	0.877**	0.633**	0.705**	1								
Elevation (Part 2)	-0.074	0.070	-0.011	0.004	0.033	-0.006	-0.025	1							
Differential Elevation (Part 2)	-0.048	0.015	0.144	-0.101	-0.115	-0.103	-0.051	0.231*	1						
Stereotype Accuracy (Part 2)	-0.035	0.051	-0.163	0.047	0.050	-0.003	-0.023	0.324**	-0.101	1					
Differential Accuracy (Part 2)	-0.153	-0.139	0.015	-0.029	0.036	-0.026	0.014	0.084	0.595**	-0.140	1				
Rater Discomfort (Part 2)	0.085	-0.107	-0.355**	0.351**	0.253**	0.362**	0.412**	-0.003	0.131	-0.030	0.081	1			
Rater Discomfort (Collect Info, Part 2)	-0.024	-0.141	-0.339**	0.234*	0.290**	0.304**	0.261**	-0.030	-0.015	-0.035	0.023	0.759**	1		
Rater Discomfort (Rating, Part 2)	0.057	-0.095	-0.364**	0.355**	0.253**	0.384**	0.397**	0.052	0.145	0.026	0.104	0.856**	0.651**	1	
Rater Discomfort (Post-Rating, Part 2)	0.158	-0.078	-0.354**	0.380**	0.250**	0.355**	0.426**	0.009	0.018	-0.039	0.048	0.862**	0.599**	0.695**	1

Note: Part 1 N = 108, Part 2 N = 114, **p<.01, *p<.05. Part 1 variables were measured as a baseline prior to a training session where Part 2 variables were measured.

Table 14

Correlations Among Study 2 Variables - Frame-of-Reference Condition Only

Variable	1	2	3	4	5
Elevation (Part 2)	1				
Differential Elevation (Part 2)	-0.065	1			
Stereotype Accuracy (Part 2)	0.142	-0.120	1		
Differential Accuracy (Part 2)	-0.130	0.636**	-0.010	1	
Rater Discomfort (Part 2)	-0.047	0.159	-0.164	0.126	1

Note: $N = 56$, **$p < .01$, *$p < .05$.

PERFORMANCE APPRAISAL DISCOMFORT

Table 15
Descriptive Statistics for Study 2 Variables

Variable	N	M	SD	Min.	Max.
Agreeableness (Part 1)	108	3.84	0.63	2	5
Conscientiousness (Part 1)	108	3.76	0.71	2	5
Self-Efficacy (Part 1)	108	3.76	0.67	2	5
Rater Discomfort (Part 1)	108	2.30	0.96	1	5
Rater Discomfort (Collect Info, Part 1)	108	2.28	1.01	1	5
Rater Discomfort (Rating, Part 1)	108	2.38	0.95	1	5
Rater Discomfort (Post-Rating, Part 1)	108	2.41	1.03	1	5
Elevation (Part 2)	114	0.27	0.53	0	3.57
Differential Elevation (Part 2)	114	1.20	1.05	0.02	4.77
Stereotype Accuracy (Part 2)	114	0.10	0.10	0	0.47
Differential Accuracy (Part 2)	114	0.24	0.13	0.02	0.58
Rater Discomfort (Part 2)	114	2.03	0.94	1	5
Rater Discomfort (Collect Info, Part 2)	114	2.05	0.96	1	5
Rater Discomfort (Rating, Part 2)	114	2.14	0.97	1	5
Rater Discomfort (Post-Rating, Part 2)	114	2.25	1.13	1	5

Note. N = Sample Size, M = Mean, SD = Standard Deviation, *Min.* = Minimum, *Max.* = Maximum. Sample size differed between Part 1 and 2 of Study 2.

Table 16
Descriptive Statistics by Training Group

	FORT Condition (N = 56)		Control Condition (N = 58)	
	M	SD	M	SD
Elevation	0.16	0.27	0.38	0.68
Differential Elevation	0.99	0.86	1.40	1.17
Stereotype Accuracy	0.12	0.11	0.09	0.10
Differential Accuracy	0.21	0.11	0.27	0.14
Rater Discomfort	2.00	0.93	2.05	0.94

Note. N = Sample Size, M = Mean, SD = Standard Deviation. Rater Discomfort ranged from 1 (no discomfort) to 5 (high discomfort). For all rater accuracy variables, smaller values denote greater accuracy. FORT = Frame-of-Reference Training.

Table 17

Correlations Among Study 2 Rated Dimensions

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Practice 1 - Builds Realistic Plans	1																	
Practice 1 - Shows Drive & Initiative	0.709**	1																
Practice 1 - Establishes Relationships	0.617**	0.744**	1															
Practice 2 - Builds Realistic Plans	0.055	-0.030	0.014	1														
Practice 2 - Shows Drive & Initiative	-0.260**	-0.224*	-0.171	0.353**	1													
Practice 2 - Establishes Relationships	-0.104	-0.063	0.019	0.355**	0.668**	1												
Practice 3 - Builds Realistic Plans	-0.205*	-0.137	-0.060	0.143	0.510**	0.485**	1											
Practice 3 - Shows Drive & Initiative	-0.266**	-0.109	-0.163	-0.001	0.392**	0.249*	0.582**	1										
Practice 3 - Establishes Relationships	-0.209*	-0.182	-0.221*	0.174	0.224*	0.039	0.417**	0.556**	1									
Test 1 - Builds Realistic Plans	0.041	-0.046	-0.051	0.022	-0.030	-0.101	0.084	-0.056	-0.053	1								
Test 1 - Shows Drive & Initiative	-0.006	-0.070	-0.146	-0.096	0.087	-0.003	0.068	-0.021	-0.055	0.644**	1							
Test 1 - Establishes Relationships	-0.011	-0.001	-0.078	-0.134	0.249**	0.089	0.381**	0.412**	0.268**	0.347**	0.427**	1						
Test 2 - Builds Realistic Plans	-0.181	-0.070	-0.012	-0.088	0.163	0.213*	0.232*	0.195	0.092	-0.089	-0.040	0.058	1					
Test 2 - Shows Drive & Initiative	-0.205*	-0.087	-0.064	-0.025	0.189*	0.130	0.172	0.152	0.262*	-0.073	-0.004	0.044	0.699**	1				
Test 2 - Establishes Relationships	0.001	-0.020	-0.053	0.144	0.132	0.041	0.024	-0.121	-0.021	0.159	0.140	0.069	0.394**	0.451**	1			
Test 3 - Builds Realistic Plans	0.237*	0.207*	0.132	0.092	-0.166	-0.043	-0.251*	-0.325**	-0.171	0.162	0.166	-0.146	0.016	0.060	0.087	1		
Test 3 - Shows Drive & Initiative	0.209*	0.225*	0.218*	0.117	-0.103	0.042	-0.195	-0.288**	-0.224*	0.138	0.103	-0.188*	0.052	0.056	0.042	0.832**	1	
Test 3 - Establishes Relationships	0.151	0.182	0.155	0.128	-0.147	-0.025	-0.239*	-0.399**	-0.268**	0.180	0.130	-0.294**	0.040	0.090	0.064	0.792**	0.828**	1

Note: Part 2 N = 114, **p<.01, *p<.05.

Running head: PERFORMANCE APPRAISAL DISCOMFORT

Table 18

Descriptive Statistics by Training Group (Part 2 Rated Dimensions)

	SME Ratings	FORT Condition (N = 56)		Control Condition (N = 58)	
	M	M	SD	M	SD
Practice 1 - Builds Realistic Plans	5.00	4.58	0.74	3.98	1.13
Practice 1 - Shows Drive & Initiative	4.00	4.38	0.80	3.98	1.19
Practice 1 - Establishes Relationships	4.00	4.16	0.94	3.62	1.17
Practice 1 - Overall Satisfaction		4.44	0.79	4.03	1.06
Practice 1 - Overall Performance		4.20	0.99	3.90	1.05
Practice 2 - Builds Realistic Plans	4.00	3.77	0.97	3.59	1.01
Practice 2 - Shows Drive & Initiative	1.00	1.95	0.86	2.53	1.30
Practice 2 - Establishes Relationships	1.00	2.07	0.93	2.62	1.20
Practice 2 - Overall Satisfaction		2.38	1.08	2.69	1.26
Practice 2 - Overall Performance		2.09	0.95	2.38	1.18
Practice 3 - Builds Realistic Plans	1.00	1.36	0.72	1.81	0.96
Practice 3 - Shows Drive & Initiative	2.00	1.25	0.44	1.78	0.92
Practice 3 - Establishes Relationships	3.00	1.64	0.90	2.09	0.94
Practice 3 - Overall Satisfaction		1.50	0.85	1.71	0.94
Practice 3 - Overall Performance		1.36	0.80	1.64	0.95
Test 1 - Builds Realistic Plans	4.00	2.59	1.19	2.34	1.42
Test 1 - Shows Drive & Initiative	4.00	2.48	1.18	2.29	1.18
Test 1 - Establishes Relationships	1.00	1.25	0.51	1.47	0.96
Test 1 - Overall Satisfaction		1.63	0.84	1.74	1.04
Test 1 - Overall Performance		1.54	0.74	1.66	0.91
Test 2 - Builds Realistic Plans	2.00	2.80	1.23	3.17	1.26
Test 2 - Shows Drive & Initiative	1.00	3.14	1.21	3.21	1.27
Test 2 - Establishes Relationships	4.00	4.27	0.84	4.05	0.98
Test 2 - Overall Satisfaction		3.39	1.15	3.34	1.10
Test 2 - Overall Performance		3.09	1.08	3.24	1.19
Test 3 - Builds Realistic Plans	5.00	4.84	0.50	4.62	0.95
Test 3 - Shows Drive & Initiative	5.00	4.79	0.46	4.52	0.96
Test 3 - Establishes Relationships	5.00	4.77	0.47	4.41	1.03
Test 3 - Overall Satisfaction		4.68	0.72	4.62	0.83
Test 3 - Overall Performance		4.68	0.69	4.59	0.86

Note. *N* = Sample Size, *M* = Mean, *SD* = Standard Deviation. Dimension Ratings ranged from 1 (very ineffective) to 5 (very effective). Video Satisfaction and Overall Performance ratings ranged from 1 (strongly disagree) to 5 (strongly agree).

Table 19

Paired Samples Test - Rater Discomfort Before and After Frame-of-Reference Training

	df	t	p
Rater Discomfort Pre/Post	53	2.81	0.007*

Note. *N* = 54, *p* < .05

Table 20

Independent Samples Test - Rater Discomfort between FORT and Control Groups

	df	t	p
Rater Discomfort	112	-0.294	0.769

Note. *N* = 114, *p* < .05

Table 21

Independent Samples Test - Accuracy Variables between FORT and Control Groups

	df	t	p
Elevation	112	-2.26	0.025*
Differential Elevation	112	-2.09	0.039*
Stereotype Accuracy	112	1.37	0.175
Differential Accuracy	112	-2.44	0.016*

Note. FORT *N* = 56, Control *N* = 58, *p* < .05.

Running head: PERFORMANCE APPRAISAL DISCOMFORT

Table 22

Study 2 Training Condition

Condition	Frequency	Percent
FORT	56	49.1
Control	58	50.9
Total	114	100

Table 23

Study 2 Gender Demographics

Gender	Frequency	Percent
Male	64	56.1
Female	50	43.9
Total	114	100

Table 24

Study 2 Race/Ethnicity Demographics

Race	Frequency	Percent
African American/Black	21	18.4
Asian	7	6.1
Caucasian/White	72	63.2
Hispanic/Latin-American	1	0.9
Other	13	11.4
Total	114	100

Table 25

Study 2 Age of Sample

Age	Frequency	Percent
18-24	88	77.2
25-34	22	19.3
35-44	2	1.8
45-54	1	0.9
55-64	1	0.9
Total	114	100

Table 26

Study 2 Employment Status

Status	Frequency	Percent
Employed	84	74.3
Not Employed	29	25.7
Missing	1	0.0
Total	114	100

Table 27

Study 2 Prior Management Experience

Exp.	Frequency	Percent
Yes	37	32.5
No	77	67.5
Total	114	100

Table 28

Study 2 Experience in Supervisory Role

Exp.	Frequency	Percent
Yes	39	34.2
No	75	65.8
Total	114	100

Table 29

Study 2 Conducted Performance Appraisals

Exp.	Frequency	Percent
Yes	34	29.8
No	80	70.2
Total	114	100

Appendix D: Measures
Study 1 Measures
Big Five Inventory (BFI)
Benet-Martinez & John (1998)
Please circle a number for each statement to indicate the extent to which you agree or disagree with that statement.

Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1--------------------2------------------------------3----------------------------------4--------------------5				

I see myself as *someone who…*

1. is talkative	1-----2-----3-----4-----5
2. tends to find fault with others	1-----2-----3-----4-----5
3. does a thorough job	1-----2-----3-----4-----5
4. is depressed, blue	1-----2-----3-----4-----5
5. is original, comes up with new ideas	1-----2-----3-----4-----5
6. is reserved	1-----2-----3-----4-----5
7. is helpful and unselfish with others	1-----2-----3-----4-----5
8. can be somewhat careless	1-----2-----3-----4-----5
9. is relaxed, handles stress well	1-----2-----3-----4-----5
10. is curious about many different things	1-----2-----3-----4-----5
11. is full of energy	1-----2-----3-----4-----5
12. starts quarrels with others	1-----2-----3-----4-----5
13. is a reliable worker	1-----2-----3-----4-----5
14. can be tense	1-----2-----3-----4-----5
15. is ingenious, a deep thinker	1-----2-----3-----4-----5
16. generates a lot of enthusiasm	1-----2-----3-----4-----5
17. has a forgiving nature	1-----2-----3-----4-----5
18. tends to be disorganized	1-----2-----3-----4-----5
19. worries a lot	1-----2-----3-----4-----5
20. has an active imagination	1-----2-----3-----4-----5
21. tends to be quiet	1-----2-----3-----4-----5
22. is generally trusting	1-----2-----3-----4-----5
23. tends to be lazy	1-----2-----3-----4-----5
24. is emotionally stable, not easily upset	1-----2-----3-----4-----5
25. is inventive	1-----2-----3-----4-----5
26. has an assertive personality	1-----2-----3-----4-----5
27. can be cold and aloof	1-----2-----3-----4-----5
28. perseveres until the task is finished	1-----2-----3-----4-----5
29. can be moody	1-----2-----3-----4-----5
30. values artistic, aesthetic experiences	1-----2-----3-----4-----5
31. is sometimes shy, inhibited	1-----2-----3-----4-----5
32. is considerate and kind to almost everyone	1-----2-----3-----4-----5
33. does things efficiently	1-----2-----3-----4-----5
34. remains calm in tense situations	1-----2-----3-----4-----5
35. prefers work that is routine	1-----2-----3-----4-----5
36. is outgoing, sociable	1-----2-----3-----4-----5
37. is sometimes rude to others	1-----2-----3-----4-----5
38. makes plans and follows through with them	1-----2-----3-----4-----5
39. get nervous easily	1-----2-----3-----4-----5
40. likes to reflect, play with ideas	1-----2-----3-----4-----5
41. has few artistic interests	1-----2-----3-----4-----5
42. likes to cooperate with others	1-----2-----3-----4-----5
43. is easily distracted	1-----2-----3-----4-----5
44. is sophisticated in art, music, literature	1-----2-----3-----4-----5

Performance Appraisal Self-Efficacy Scale for Raters (PASES)
(Bernardin & Villanova, 2005)

Instructions: Using the scale below, please indicate what level of confidence you have in being able to successfully perform the behavior as it is described in each statement.

With no confidence	With little confidence	With some confidence	With a good level of confidence	With great confidence
1	2	3	4	5

Subscale: Process Features of the Appraisal
1. Establishing accurate and fair standards for judging the job performance of others
2. Using my observations of others' job performance in assigning ratings
3. Collecting observations and records of employee performance to produce accurate ratings

Subscale: Rater Subjectivity
4. Setting aside any personal biases to arrive at accurate employee ratings
5. Assigning ratings that are accurate but that may disagree with others' expectations
6. Evaluating employee performance independent of personal like or dislike for the employee

Subscale: Appraisal Discussion
7. Explaining to employees how it is that I arrived at a performance rating when they believe higher ratings were deserved
8. Justifying poor ratings to employees who believe poor ratings are undeserved
9. Discussing my reasons for assigning specific ratings to employees suspicious of my motives

Subscale: Suggesting Performance Improvements
10. Providing suggestions for improving job performance to more senior or more experienced employees
11. Setting aside employees' personal life accounts for poor performance
12. Suggesting ways to improve job performance to employees resistant to change their ways of doing the work

PERFORMANCE APPRAISAL DISCOMFORT

DPE (Discomfort with Peer Evaluation) Scale
Saffie-Robertson & Brutus (2014)

When evaluating the performance, how comfortable do you feel…?

No Discomfort		Undecided		High Discomfort
1	2	3	4	5

Subscale: Collecting Information
1. Collecting Information of your subordinate's/peer's performance to assign accurate ratings**
2. Using and trusting your observations to assign ratings**

Subscale: Rating
3. Evaluating subordinate's/peer's performance independent of your personal like or dislike for that person**
4. Assigning ratings that are accurate but which you know may disagree with your subordinate's/peer's expectations**
5. Distributing points among your subordinates/peers according to their performance
6. Providing written feedback or comments regarding subordinate's/peer's performance*

Subscale: Post-rating
7. Talking to a subordinate/peer about the evaluation you gave him/her*
8. Telling a subordinate/peer how his/her performance can improve if he/she asks for your advice*
9. In future courses, being in the same work group/[collaborating on a future work team] with a subordinate/peer whose performance you evaluated below average
10. Developing a friendship or social relationship with a subordinate/peer whose performance you evaluated as below average

*Item modified from PADS (Performance Appraisal Discomfort Scale)
**Item modified from PASES (Performance Appraisal Self-Efficacy Scale)
Items with no other indication correspond to those specifically created for the DPE scale

Perceived Quality of Training

Using the scale below, please indicate how useful you believe the performance appraisal training you experienced to be.

Not Useful at All	Slightly Useful	Neutral	Somewhat Useful	Very Useful
1	2	3	4	5

Study 1 Demographics

1. What is your age?
2. What is your gender?
3. Which race/ethnicity do you identify as?
4. Do you have prior management or performance evaluation experience?
   a. Yes (If yes, please explain _____)
   b. No
5. What is your job title?
6. How many years have you been with your current organization?
7. How many years have you been in a supervisory/management role?
8. Approximately how many hours do you work each week?

Study 1 Exploratory Variables

Performance Appraisal Beliefs
(Smith, Harrington, & Houghton, 2000)
Items averaged from Federal Employee Attitude Survey (DeMarco & Nigro, 1983)

Instructions: Using the scale below, please indicate your level of agreement with each statement.

Strongly Disagree		Undecided		Strongly Agree
1	2	3	4	5

1. Performance appraisals when conducted correctly can increase productivity.
2. My organization considers performance appraisal an important part of supervisor's duties.
3. The performance appraisal system in my organization helps improve the performance of its employees.
4. The quality of one's performance determines one's pay in my organization.
5. Performance appraisals influence personnel actions taken in my organization.

Performance Appraisal Views
(Coutts & Schneider, 2004)
Note: Amendments for this study are underlined

I: Bases of Performance Appraisal

Never	Sometimes	Frequently	Always
1	2	3	4

1. Extent to which appraisal is based on personal traits (dependability, initiative, agreeableness)
2. Extent to which appraisal is based on work-related results (quality of work outputs)
3. Extent to which appraisal is based on work-related behaviors

II: Opportunity for Performance Appraisal Input
4. Person who completes ~~your~~ the performance appraisal
   a. Current Supervisor only
   b. Current Supervisor with input from other supervisors
   c. Current supervisor with input from his/her (their) supervisor
   d. Other

5. Extent to which ~~you~~ subordinates are given the opportunity to provide input into the preparation of ~~you're~~ their appraisal
   a. No opportunity
   b. Some opportunity
   c. A great deal of opportunity
6. Extent to which ~~your supervisor~~ you discusses the content of your <u>direct reports'</u> appraisal with ~~you~~ them before completing it
   a. Never
   b. Sometimes
   c. Frequently
   d. Always

III: Frequency and Nature of Supervisor Feedback

7. Extent to which ~~your supervisor~~ you provide~~s~~ informal performance feedback throughout the year
   a. Seldom
   b. Sometimes
   c. On a regular basis
8. Frequency of formal performance appraisals
   a. Once every two years
   b. Once a year
   c. Twice a year
   d. Three times a year
   e. Others
9. Extent to which ~~your supervisor~~ you provide~~s~~ face-to-face feedback following completion of ~~your~~ a performance appraisal
   a. Seldom
   b. Sometimes
   c. Always
10. Method of performance appraisal feedback ~~from your supervisor~~
    a. Lengthy ~~interview~~ meeting in which ~~I am asked to~~ you ask to discuss ~~my~~ your subordinate's feelings and perceptions about ~~my~~ performance to discuss future work goals and objectives
    b. Brief ~~interview~~ meeting in which ~~I am~~ you ask~~ed~~ if ~~I have~~ there are any questions about ~~my~~ the performance appraisal
    c. After ~~receiving my~~ sending a written appraisal~~, I am asked~~ you ask to set up a meeting with ~~my supervisor~~ your subordinate if ~~I want to discuss it~~ they want to discuss
    d. Other

IV: Perceived Impact and Benefits of Performance Appraisal

Very little	Moderate Extent	Large Extent
1	2	3

11. Extent to which ~~your~~ the appraisal helps you to clarify performance expectations and standards <u>with your subordinate</u>
12. Extent to which ~~your~~ the appraisal helps you to establish performance goals <u>with your subordinate</u>
13. Extent to which <u>your subordinate's</u> job performance has improved as a result of the performance appraisals
14. Extent to which you appraisal helps you to identify career development objectives <u>for your subordinate</u>

V: Perceptions of own Appraisal and Performance
15. Statement that best describes your perception of your own performance appraisals during the past few years
    a. My appraisals have been accurate evaluations of my true level of performance
    b. My appraisals have been reasonable accurate but incomplete evaluations of my true level of performance
    c. My appraisals have reflected an overestimate (positive bias) of my true level of performance
    d. My appraisals have reflected an underestimate (negative bias) of my true level of performance
16. Statement that best describes your overall feelings about the performance appraisal system in your department
    a. The appraisal system provides me with useful information so that I can continue to improve my job performance/ <u>The appraisal system allows me to provide my subordinates with useful information so that I can continue to help them improve their job performance</u>
    b. The appraisal system frequently confuses me regarding performance expectations and standards/ <u>The appraisal system frequently confuses me regarding performance expectations and standards for my subordinates</u>
    c. The appraisal system generally demoralizes me and reduces my job motivation/ <u>The appraisal system generally demoralizes my subordinates and reduces their job motivation</u>
    d. The appraisal system has little, if any, impact on my subsequent job performance / <u>The appraisal system has little, if any impact, on my subordinates' subsequent job performance</u>
17. ~~Rating of your own job performance in comparison to the performance of your peers~~
    a. ~~Top 10-20 of members~~
    b. ~~Top 25-50 of members~~
    c. ~~Top 60-90 of members~~

VI: Training
18. In your opinion, to what extent do supervisors in your department receive training concerning the effective use of the performance appraisal system?
    a. No training is provided
    b. Very little training is provided

      c. Substantial training is provided

Additional Exploratory Items

1. Who is sent to training at your organization? (For example, new managers, all employees, higher level managers?)
2. Is training voluntary?
3. How many direct reports do you supervise/manage?
4. How many performance appraisals do you conduct each year?
5. Do you have the opportunity to observe your direct reports' behaviors?
6. How often do you conduct performance appraisals?
7. Approximately how much time do you devote to performance appraisals each year?
8. Do you feel pressured for time when conducting performance appraisals?
9. Do you give feedback on your direct reports' behaviors face-to-face or through a virtual medium (i.e., email)?
10. Are performance appraisals in your organization used for developmental or administrative purposes?
11. Have you ever received management or performance appraisal training?
    a. If yes, please explain your experience, listing the name of the training if you remember it.
    b. If yes, please explain the amount of training you received (i.e., one training, one training each year, etc.).
    c. If yes, please list when during your tenure at your organization that you received this training (i.e., after becoming a new manager or before?).

Performance Appraisal Discomfort Scale
(Villanova et al., 1993, wording revised by Smith et al., 2000)

No Discomfort		Undecided		High Discomfort
1	2	3	4	5

1. Telling an employee that he or she must stop coming to work late
2. Telling an employee that his or her work is only satisfactory, when you know he or she expects above satisfactory rating (required for merit pay increases)
3. Talking to an employee about his or her performance on the job
4. Conducting a formal performance appraisal interview with an ineffective employee
5. Asking an employee if he or she has any comments about your ratings of his or her performance
6. Telling an employee who has problems in dealing with other employees that he or she should do something about it (take a course, read a book, seek counseling, etc.)
7. Telling a male subordinate that his performance must improve
8. Responding to an employee who is upset over your rating of his or her performance
9. Conducting a formal appraisal interview with an effective employee
10. Letting an employee give his or her point of view regarding a problem with performance
11. Giving a satisfactory rating to an employee who has done a satisfactory (but not exceptional) job
12. Letting a subordinate talk during an appraisal interview
13. An employee's challenging you to justify your evaluation in the middle of an appraisal interview
14. An employee's accusing you of playing favorites in the rating of your staff
15. Recommending that an employee be discharged

16. Telling an employee that his or her performance can be improved
17. Warning an ineffective employee that unless performance improves he or she will be discharged
18. Telling a female employee that her performance must improve
19. Encouraging an employee to evaluate his or her own performance
20. Telling an employee that you will not tolerate his or her taking extended breaks

Study 2 Additional Scales

Training Reactions
(Based on Noonan & Sulsky, 2001)

Poor   Excellent

1   2   3   4   5

1. I found the training procedure to be
2. The content of training as
3. The presentation of information by the rater trainer was
4. The clarity of information presented by the rater trainer was
5. I found the length of training to be

Overall Performance
(Saffie-Robertson & Brutus, 2014)

Strongly Disagree   Strongly Agree

1   2   3   4   5

1. Overall, I am satisfied with the performance of ~~my team-mate in the group project~~ the supervisor/job candidate in the video
2. The performance of ~~this team-mate~~ the supervisor/job candidate has been, in general, excellent

Study 2 Demographics

1. What is your age?
2. What is your gender?
3. Which race/ethnicity do you identify as?
4. Do you have prior management or performance evaluation experience?
   a. Yes (If yes, please explain how much and describe your experience_____)
   b. No

Appendix E: Study 2 Materials
Sample Rating Form

Builds Realistic Plans

Shows Drive and Initiative

Establishes Relationships

Please choose one rating:

	Very Ineffective	Somewhat Ineffective	Neither Effective nor Ineffective	Somewhat Effective	Very Effective
Builds Realistic Plans	○	○	○	○	○
Shows Drive and Initiative	○	○	○	○	○
Establishes Relationships	○	○	○	○	○

PERFORMANCE APPRAISAL DISCOMFORT

Sample Frame-of-Reference Training Handout

Frame-of-Reference Handout
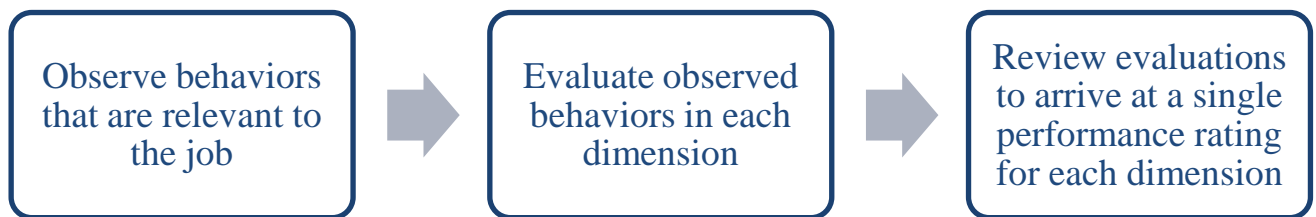
The Purpose of Rater Training
- The purpose of rater training is to assist in preparing you to accurately and effectively evaluate the performance of job candidates and subordinates
- This training will help you to understand general assessment and factors to consider when making judgments about behavior

The Performance Rating Process
First, it is important to gather information about the job the person you will be rating holds or is applying for by conducting a job analysis. Learn about the tasks the person must complete, and what skills and abilities are required to complete the work effectively. Ratings should be made based on the individual's ability to perform the job tasks. Performance ratings can be made for several purposes, including employee selection and employee development.
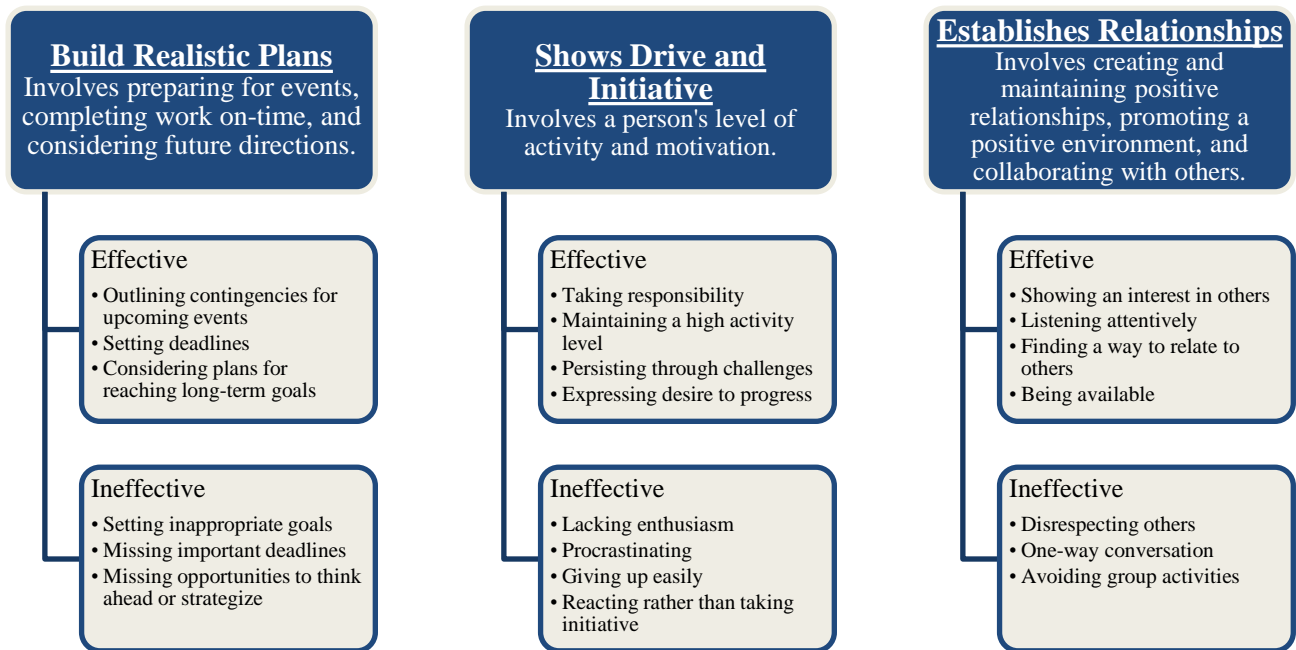
For this rater training, you will watch videos of a manager and a job applicant. The manager will be speaking to their subordinate about an issue their subordinate is having. You will be rating the manager's behaviors and interactions with their subordinate. In the second scenario, you will rate an individual who is applying for a job in the Office of Admissions. You will receive a job description and resume of the candidate, which will provide you with the relevant job duties that person should have when evaluating them for the role.

Judgments should follow 3 distinct steps:

Observe behaviors that are relevant to the job	→	Evaluate observed behaviors in each dimension	→	Review evaluations to arrive at a single performance rating for each dimension

Frame-of-Reference Training
- The goal of FOR is to reach a shared idea of what high and low performance on each rated dimension should look like
- All raters should leave with an aligned understanding regarding what "effective" and "ineffective" behaviors look like for a given dimension
- We will review ratings made by subject matter experts and discuss how they arrived at their ratings, then we will compare these ratings to the ones you make during practice
- Refer to the example dimensions below

Dimensions to be Rated:

Build Realistic Plans
Involves preparing for events, completing work on-time, and considering future directions.

Effective
- Outlining contingencies for upcoming events
- Setting deadlines
- Considering plans for reaching long-term goals

Ineffective
- Setting inappropriate goals
- Missing important deadlines
- Missing opportunities to think ahead or strategize

Shows Drive and Initiative
Involves a person's level of activity and motivation.

Effective
- Taking responsibility
- Maintaining a high activity level
- Persisting through challenges
- Expressing desire to progress

Ineffective
- Lacking enthusiasm
- Procrastinating
- Giving up easily
- Reacting rather than taking initiative

Establishes Relationships
Involves creating and maintaining positive relationships, promoting a positive environment, and collaborating with others.

Effetive
- Showing an interest in others
- Listening attentively
- Finding a way to relate to others
- Being available

Ineffective
- Disrespecting others
- One-way conversation
- Avoiding group activities

Let's Practice!

We will watch three videos together and make ratings of the manager's performance. As you watch, consider the following questions:

- What would a strong or weak performer actually do?
- What behavior indicate strong behavior? What behaviors indicate weak behaviors?

Sample Job Description

Sample Job Description: Associate Director of Admissions

Position Overview: This position reports to the Director of Admissions assisting with all aspects of the admissions and recruitment process.

In addition, s/he is responsible for the internal department management of admissions print publications, advertising, and marketing efforts in coordination with the director of marketing.

Specific Responsibilities include but are not limited to:

- Meet enrollment goals for assigned target recruitment area
- Travel on behalf of the University for recruitment and other college related events
- Engage in outreach actions to promote the University's visibility and general recruitment efforts
- Assist in the development, maintenance and execution of the department's communication plan to prospective and admitted students
- Be responsible for the consistent, timely processing of all admissions applications and communications
- Evaluate and execute admission decisions on applicant files
- Supervise and manage the support staff of the department
- Other duties as assigned

Qualifications and Experience:

- A bachelor's degree is required
- A master's degree is preferred with at least 4 years of admissions or higher education experience
- Previous supervisory experience managing professional and student staff is preferred
- A proven track record as an Admissions recruiter is preferred
- Ability to effectively communicate in person and in writing and be able to speak publicly
- Should possess high energy, attention to detail, and the ability to organize effectively
- A valid driver's license and the ability to travel and work nights and weekends as necessary

Retrieved from: http://www.nacacnet.org/career-center/Tools/JobDescriptions/Pages/AssociateDirectorofAdmissions.aspx

Amended based on sampling several Associate Director of Admissions job descriptions found using LinkedIn.

PERFORMANCE APPRAISAL DISCOMFORT

Sample Resume

TAYLOR FORTH

3891 Tyler Drive ■ Essex Junction, VT 99993 ■ Cell: (900) 999-9000 ■ Email:tfadmissions@gmail.com

OBJECTIVE

To work as an Associate Admissions Director position bringing knowledge of managing strategic plans for admissions and community outreach in order to bring in new students for quality education purposes.

SUMMARY OF QUALIFICATIONS

• Six years of experience working in admissions for Vermont State University
• Highly skilled in establishing and forecasting educational plans and outcomes
• Working experience of generating marketing plans to support applicant
• In depth knowledge of managing budgets and other financial information in support of the major program

CORE STRENGTHS

• Complete knowledge of education systems of the USA
• Excellent marketing acumen
• Exceptional communication skills
• Well versed in creating and managing budgets
• Strong interpersonal and multitasking skills

PROFESSIONAL EXPERIENCE

Vermont State University, Essex Junction, VT   Jan 2011 – Present

Admissions Counselor

• Create and implement marketing and student admissions plans
• Communicate admission criteria as and when needed
• Manage paperwork and admission documentation
• Manage budgets and promote the university to the community
• Assist new students in providing information and helping with the registration procedures
• Determine scholarships for worthy students

Vermont State University, Essex Junction, VT   May 2009 – Jan 2011

Admissions Assistant

• Provided information to students about the admissions process
• Assisted in filling out forms and registration procedures
• Provided information to students regarding courses and curriculum
• Assisted the admissions director with community outreach and marketing activities

EDUCATION

Bachelors of Arts in Education – Vermont State University – 2007
Masters of Arts in Higher Education – Vermont State University – 2009

Retrieved from: http://coverlettersandresume.com/director/admissions-director-resume-sample/
Amended slightly