AGREEMENT BETWEEN SELF AND OTHER RATINGS IN MULTI-RATER TOOLS: PERFORMANCE, ALTERNATIVE MEASURES, AND IMPORTANCE.

Myranda Grahek, MS

Dissertation Prepared for Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS
August 2008

APPROVED:

Linda Marshall, Major Professor and Chair of the Department of Psychology
Paul Lambert, Committee Member
Ed Watkins, Committee Member
Sandra L. Terrell, Dean of the Robert B. Toulouse School of Graduate Studies

Multi-rater tools also referred to as 360-degree feedback tools, are frequently used in addition to traditional supervisory appraisals due to sources (i.e., supervisor, peer, direct report) unique perspectives and opportunities to view different aspects of job performance. Research has found that the differences among sources are most prevalent between self and other ratings, and the direction of agreement is related to overall job performance. Research has typically focused on one form of agreement, the direction of an individual’s self-ratings compared to others’ ratings. The current study expanded on past research on rater agreement using a data set \( (n = 215) \) consisting of multi-rater data for professionals participating in a leadership development process. The study examined the ability to predict job performance with three different measures of self-other agreement (i.e., difference between overall mean scores (difference), mean absolute difference across items (difference), and mean correlation across items (similarity)). The study also examined how the relationships may differ across performance dimensions. The final purpose was to explore how the importance of the performance dimensions, as rated by the participant, may moderate the relationship between self-other agreement and job performance. Partial support for study’s hypotheses was found. The direction and difference measures of agreement on the overall multi-rater tool and performance dimensions accounted for a significant amount of the variance in job performance. The relationship between the similarity measure of agreement and job performance, and the moderating effect of importance were not supported in the current study.
Copyright 2008

by

Myranda Grahek
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>6</td>
</tr>
<tr>
<td>Assumptions of Multi-rater Tools and Development Process</td>
<td>8</td>
</tr>
<tr>
<td>Correlations among Rater Sources</td>
<td>9</td>
</tr>
<tr>
<td>Implications of Self-Other Agreement</td>
<td>11</td>
</tr>
<tr>
<td>Agreement and Performance</td>
<td>12</td>
</tr>
<tr>
<td>Theories Explaining Self-Other Agreement</td>
<td>14</td>
</tr>
<tr>
<td>Measurement of Agreement</td>
<td>17</td>
</tr>
<tr>
<td>Sources of Self-Other Agreement and Disagreement</td>
<td>22</td>
</tr>
<tr>
<td>Current Study</td>
<td>24</td>
</tr>
<tr>
<td>METHOD</td>
<td>29</td>
</tr>
<tr>
<td>Participants</td>
<td>29</td>
</tr>
<tr>
<td>Measures</td>
<td>29</td>
</tr>
<tr>
<td>Worthy Leadership Profile© for Executives (WLPe)</td>
<td>29</td>
</tr>
<tr>
<td>Measures of Agreement</td>
<td>30</td>
</tr>
<tr>
<td>Procedure</td>
<td>31</td>
</tr>
<tr>
<td>Data Analysis Plan</td>
<td>31</td>
</tr>
<tr>
<td>RESULTS</td>
<td>35</td>
</tr>
<tr>
<td>Hypothesis Testing</td>
<td>36</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>41</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>55</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Demographic Data (n, %) for Participants and Bosses</td>
<td>50</td>
</tr>
<tr>
<td>2.</td>
<td>Descriptive Data for Measures</td>
<td>51</td>
</tr>
<tr>
<td>3.</td>
<td>Measures of Agreement</td>
<td>51</td>
</tr>
<tr>
<td>4.</td>
<td>Correlations between Agreement on Performance Dimensions and Job Performance Rating</td>
<td>52</td>
</tr>
<tr>
<td>5.</td>
<td>Summary of Regression Analyses: Significant Associations of Measures of Agreement, Importance, and Interaction with Job Performance</td>
<td>52</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Histogram for direction measure of agreement.</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>Histogram for similarity measure of agreement.</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>Histogram for difference measure of agreement.</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>Histogram for overall job performance.</td>
<td>54</td>
</tr>
</tbody>
</table>
INTRODUCTION

Performance evaluation ratings provide the basis for numerous employment decisions, including retention, promotion, and salary increases. Traditionally, supervisors’ ratings have served as the primary source of performance information. As organizations shift from hierarchical structures to flatter organizations, supervisory ratings no longer provide sufficient information on individuals’ performance. As more emphasis is placed on individual accountability and there are fewer levels of management, the job requirements for managers change and expand (Murphy & Cleveland, 1995). Multi-rater feedback tools have been increasingly used as a substitute and/or complement to the traditional supervisor appraisal (Conway & Huffcut, 1996; Funderburg & Levy, 1997).

Multi-rater feedback tools, also referred to as 360-degree feedback or multisource tools, can increase the quantity and validity of the performance data for research, development, and decision-making purposes by utilizing a multitrait-multi-rater approach (Lawler, 1967). The multitrait-multi-rater approach refers to the use of ratings from several raters on several different traits or behaviors relevant to performance (Campbell & Fiske, 1959; Conway, 1998). Multi-rater tools obtain performance ratings from supervisors, peers, direct reports, and occasionally others (e.g., customers, vendors) outside of direct reporting relationships (Borman, 1997; Pollack & Pollack, 1996). The use of feedback from multiple sources provides perspectives from relevant viewpoints (London & Smither, 1995; Tornow, 1993). These tools have been used for performance appraisals (Bettenhausen & Fedor, 1997; Waldman, Atwater, & Antonioni, 1998), but are more predominantly used to provide developmental feedback (London & Smither, 1995; Pollack & Pollack, 2001; Tornow, 1993).
The use of multi-raters has continued to increase in organizations (Alimo-Metcalfe, 1998; Antonioni, 1991; Chruch & Bracken, 1997; Chruch & Waclawski, 2001). By 1996 a large majority of the Fortune 500, and 20 of the 32 most admired companies were using some form of multi-rater feedback (Yammarino & Atwater, 1997). These procedures are used by large companies and major consulting firms for individual feedback, coaching, leadership development, organizational development and change efforts, and large scale assessment and analysis (Alimo-Metcalfe, 1998; Church & Waclawski, 2001). Research in multi-rater feedback methodology is also a growing topic in human resource development, organization development, and industrial-organizational psychology (Church & Waclawski, 2001).

As multi-raters become more common tools for performance appraisal and development purposes, understanding the variance among rater sources becomes more important. This variance leads to difficulty in presenting congruent feedback, confounds relationships in empirical studies, and causes differences in test validities depending on the source when ratings are used as performance criteria (Conway, 1998; Podsakoff, MacKenzie, & Podsakoff, 2003). Researchers’ attempts to decrease the variance among sources through statistical techniques and construction of rating scales have had little success (Conway, 1998). When method variance, such as variance among sources, cannot be reduced, research has moved to understand why ratings differ across sources. In addition to understanding variance among rater sources, specific attention has focused on the relationship between self-other agreement and job performance.

The purpose of the current study was to expand upon research linking self-other agreement in multi-rater tools to job performance. Research has typically focused on one form of agreement, the direction of an individual’s self-ratings compared to others’ ratings. This method examines how different categories of individuals (i.e., under-raters, in-agreement, and over-
raters) differ in job performance. Less research has examined additional forms of rater agreement. The current study examined the ability to predict ratings of job performance with three different measures of self-other agreement (i.e., difference between overall mean scores (direction), mean correlation across items (similarity), and mean absolute difference across items (difference)). The study also examined how the relationships may differ across dimensions. The final purpose of the study was to explore how the importance of the dimensions, as rated by the individual, may moderate the relationship between self-other agreement and job performance.

Assumptions of Multi-rater Tools and Development Process

Multi-rater tools are used as feedback tools in place of traditional supervisory appraisals due to sources (i.e., boss, peer, direct reports) unique perspectives and opportunities to view different aspects of job performance (Borman, 2001). These differences in perspectives provide a more valid, reliable, and broader view of performance (Chruch & Braken, 1997; Holzback, 1978; Mabe & West, 1982; Podsakoff & Organ, 1986; Tornow, 1993) than a single perspective provided by a supervisor. A basic premise for multi-rater tools when used for developmental purposes is that individuals have a limited ability to evaluate themselves compared to objective criteria or others’ perceptions (Harris & Schaubroeck, 1988). Feedback from multiple relevant sources can assist individuals to see themselves as others do and identify areas where behavior change may be needed (Church & Braken, 1997; Tornow, 1993; Yammarino & Atwater, 1997). This increased knowledge of strengths, development opportunities, and others’ expectations can lead to improved leadership and management skills (London & Beatty, 1993) and support leadership development programs (Tornow, 1993).

The usefulness of multi-raters as a tool for individual development makes three
assumptions regarding the processing and use of feedback (Church & Waclawski, 2001). The first is behavioral consistency; individuals behave in consistent ways, allowing for measurement in a quantifiable manner such as ratings. Obtaining ratings from multiple sources increases the reliability and validity of this information (Landy & Farr, 1980). The second assumption is that raters are able and motivated to provide accurate ratings to the individual receiving feedback. The final assumption is that this feedback is useful to individuals in better understanding how they are viewed by others and increases individuals’ understanding of their strengths and development opportunities. The individual can then focus attention on his/her personal characteristics most related to the behaviors or performance dimensions and how to make improvements in those areas (Gibbons, 1983; London & Smither, 1990).

Atwater, Roush, and Fischthal (1995) found multi-rater feedback resulted in improved performance ratings from direct reports and higher self-other agreement in later appraisals. However, individual differences and other variables likely play a role in the degree to which multi-rater feedback improves performance and self-other agreement. For example, Parsons, Herold, and Turlington (1981) found individuals who valued external feedback had an increase in self-rating accuracy after receiving feedback.

Correlations among Rater Sources

One limitation of multi-rater tools is a low correlation typically found among sources (i.e. self, supervisor, direct report, peer). A meta-analytic study by Conway & Huffcut (1997) found both low mean interrater reliabilities (boss = .50; direct report = .30; peers = .37), and low mean correlations among sources. Self-ratings were most strongly correlated with supervisors ($r = .22$) and weakest with direct report ratings ($r = .14$). The mean correlation for self-peer ratings...
was .19. Similarly, an early meta-analysis by Mabe and West (1982) found low correlations between individuals’ self-ratings and objective measures of performance and ability ($r = .29$). In the 21 studies reviewed, self-ratings were higher than other-ratings in 15 of the studies.

The differences in the ratings among sources (i.e., bosses, peers, direct reports) may be due to their position in the organization relative to the target individual. Salam, Cox, and Sims (1997) found the aspects of performance that contribute to overall performance vary based upon source. The differences among sources may be due to unique perspectives and opportunities to observe performance, focusing on different facets of job performance, and/or weighting different dimensions of performance or importance of goals differently (Borman, 1974; Borman, 1997; Klimoski & London, 1974; Swann, 1984; Warr & Bourne, 1999).

Conway and Huffcutt (1997) compared the between source correlations to inter-rater reliabilities. With the exception of the relationship between supervisor and peer ratings, the very low correlations between sources suggest different sources maintain distinct views of performance. Similar results were found by Tsui and Ohlot (1988) who showed differences in performance ratings between sources were greater than differences within sources. In comparison, Mount, Judge, Scullen, Sytsma, & Hezlett (1998) found that the variance in multi-rater data was more attributable to the individual rater than the rater’s relationship to the participant (i.e., boss, peer, direct report). These findings suggest that ratings are affected by both the perspective/relationship of the rater, as well as individual characteristics or tendencies.

Research has found that the differences among sources are most prevalent between self and other ratings (Harris & Schaubroeck, 1988; Landy & Farr, 1980; Nilsen & Campbell, 1993). Individuals’ ratings of their own performance are typically more lenient than ratings by others (Thornton, 1980). Nilsen and Campbell (1993) found individuals tended to over-estimate their
performance by approximately .3 standard deviations. The tendency for individuals to provide ratings that are inflated and in low agreement with others’ ratings suggest that many individuals are not effective at evaluating themselves, (Harris & Schaubroeck, 1988) or may be motivated to positively skew their self-ratings. This self-leniency tendency has been found to be stable overtime (Nilsen & Campbell, 1993). Thornton (1968) found low correlations between self and other ratings even at the executive level. Additional research suggests the complexity of job requirements for higher level managers and executives than for lower level employees increases the discrepancy between self and other ratings (Sala, 2003).

Implications of Self-Other Agreement

Participants report multiple benefits of the individualized feedback from multi-rater tools including increased awareness of development needs and motivation to change and develop. Participants also reported they became more flexible with others, had increased awareness of the impact of their actions and decision styles on others, and perceived the feedback to be valid and useful for advancing their career development (McCarthy & Garavan, 1999). Farh, Werbel, and Bedeian (1988) found multi-rater tools provided individuals with a greater sense of participation and control in the performance appraisal process, as well as providing multiple perspectives of performance. The opportunity to provide self-ratings has also been related to satisfaction with the performance appraisal process (Miller, 2001).

The degree of congruence between self and other ratings has been shown to have meaningful implications for individuals and organizations. Individuals who tend to rate themselves higher than others receive lower leader satisfaction ratings from direct reports (VanVelsor et al., 1993), and are less promotable (Thornton, 1968). They tend to make less
effective job-relevant decisions, harbor resentment, and have higher absenteeism and turnover than individuals whose self-ratings are similar to others’ ratings, (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Bass & Yammarino, 1991; Fleenor, McCauley, & Brutus, 1996; Kozlowski, Kirsch, & Chao, 1998; Van Velsor, Taylor, & Leslie, 1993; Yammarino & Atwater, 1997). These individuals also demonstrate poor awareness of their own capabilities (Alimo-Metcalfe, 1998), and less behavioral change in response to feedback (Rynes, Gerhart, & Parks, 2005). Alternatively, there is some evidence that these individuals may be more likely to seek additional feedback from direct reports after receiving initial multi-rater feedback (Waldman & Atwater, 2001).

**Agreement and Performance**

Individuals who rate themselves in higher agreement with others have been found to be higher performers than low agreement individuals. More specifically than poor self-other agreement, over-raters are found to receive significantly lower overall performance ratings (Church, 1997). One of the most common methods of assessing the relationship of self-other agreement to overall performance classifies individuals into four groups based upon the relationship of self to other-ratings (over-estimators, in-agreement, under-estimators) and level of ratings (good, average, poor). Individuals who tend to rate themselves higher than others, or over-estimate their performance, are perceived to be less effective than individuals who accurately assess their performance. Individuals who under-estimate their performance, or provide self-ratings lower than others’ ratings are perceived to be more effective than in-agreement raters and raters who over-estimate their performance (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Fleenor, McCauley, & Brutus, 1996; Van Velsor, Taylor, & Leslie, 1993).
These findings suggest that there is an overall positive correlation between direction of agreement and performance, as self-ratings relative to others’ increase, performance increases.

Additional research examining direction of ratings and performance has found different results. Church (2000) and Mersman and Donaldson (2000) did not find differences in performance among categories of raters on a multi-rater tool. Fleenor, McCauley, & Brutus (1996) examined differences in performance among six categories of raters (over/good, over/poor, in-agreement/good, in-agreement/poor, under/good, under/poor). In this study no difference in performance was found between in/good, under/good and over/good individuals. Their findings suggest self-other agreement may only be related to performance for individuals with poorer performance.

Nowack (1997) compared individuals’ classification as an over-rater, in-agreement/good, in-agreement/poor and under-estimator to performance on an in-basket simulation in an assessment center. Self-ratings and supervisor ratings were significantly related to higher performance in the in-basket simulation. Unlike past research examining the link between self-other agreement and performance, in-agreement/good and over-estimators had better performance in the in-basket than in-agreement/poor and under-estimators. One possible explanation posited for their findings is that the relationship between agreement and performance may differ when agreement is measured between self-ratings and an objective measure of performance.

A single study was found that examined the relationship between agreement and performance using an alternative measure to the direction of agreement to the level method. Church (2000) measured agreement by calculating the sum of squared differences across items. This measure of agreement assesses the total difference or variance across items. In this study,
individuals whose self-ratings of performance were in higher agreement to other’s ratings received higher performance ratings than individuals lower in agreement, indicated by a negative relationship between the difference and performance.

Theories Explaining Self-Other Agreement

Past research has suggested multiple processes and theories to explain the low agreement among sources, especially between self and others’ ratings. As mentioned earlier, disagreement may be due to sources focusing on and weighing aspects of performance differently. Salam et al. (1997) found rating sources (i.e., boss and direct report) differed in the characteristics they viewed as related to performance. Hoojberg & Choi (2000) also found sources (i.e., bosses, peers, direct reports) differed in the characteristics they associated with effectiveness in the position. Tsui & Ohlot (1988) found that while sources agreed upon the characteristics and behaviors that were important for performance, sources weighted the importance differently. The higher agreement found within sources (e.g., direct report – direct report) than across sources (e.g., direct report-peer) is frequently attributed to the perspectives, focus on facets, and opportunity to observe an individual’s performance, which are similar within sources and different across (Borman, 1974; Harris & Schaubroeck, 1988; Klimoski & London, 1989). The particularly greater disagreement between self and other ratings may be due to individuals’ maintaining a broader view of their own performance. If individuals are motivated to provide ratings that are similar to others, they may need to balance perspectives of multiple sources in their self-ratings.

Other research has asserted that rater agreement is more due to individual differences and systematic rater biases than source (Klimoski & London, 1974; Mount, Judge, Scullen, Sytsma & Hezlett, 1998). Researchers suggest that rater biases are stable tendencies and may be a
predictable source of variance in performance ratings (Borman & Hallum, 1991; Kane et al., 1995; Thornton, 1980). The rater biases of the multiple individuals help to explain low correlations among sources (Borman & Hallum, 1991). For example, a tendency to be more lenient to the self than to others is one source of low self-other agreement. Different measures of agreement may relate to various forms of bias. While level of ratings may tap into leniency bias, difference and similarity may relate more to halo or central tendency error. Rater biases may describe the form of rater agreement but do not necessarily explain the underlying processes.

The agreement between self and other ratings in multi-rater tools is often labeled as a measure of self-awareness (Fletcher & Baldry, 2000). This research defines self-awareness as individuals’ understanding of their own strengths and weaknesses as defined by others (Fletcher & Baldry, 2000). Self-other agreement requires individuals to be able to assess and incorporate others evaluations of themselves into their own self-evaluations (Atwater & Yammarino, 1992). Self-awareness in this sense is a comparison to an external or objective standard set by others. As others’ ratings (e.g., boss, direct reports) are the external or objective standard for self-other agreement in multi-rater tools, an individual’s level of self-awareness and its relationship to other variables including performance may differ based upon the comparison source.

Although self-awareness is an individual’s understanding of his/her strengths and development opportunities, research suggests that it is not precisely equivalent to self-other agreement. Van Velsor, Taylor, and Leslie (1993) found individuals who under-estimated their performance were reported by direct reports to have the highest level of self-awareness, with over-estimators receiving the lowest self-awareness ratings. These findings may be due to defining self-awareness as a comparison to an external criteria (others’ ratings), an inappropriate measure of rater agreement (direction of overall ratings), or individuals who under-estimate their
performance may have high self-awareness but over emphasize the negative aspects. Examining rater agreement as a correlation or total difference across items may be needed to establish rater agreement as an appropriate measure of self-awareness.

Additional research has outlined how various cognitive or social processes underlie the rater bias and lack of agreement between rating sources. Theories are frequently used to hypothesize or explain self-other agreement. Some of the more frequently used include social comparison theory (Farh & Dobbins, 1989), egocentric bias (Harris & Schaubroeck, 1998), self-enhancement theory (Swann, 1984), and attribution theory (Fiske & Taylor, 1991).

Self-awareness as discussed earlier is an explanation for an individuals’ ability to accurately assess and incorporate other’s perceptions of their strengths and weaknesses into self-ratings of performance. Festinger’s (1954) social comparison theory provides a possible explanation for why individuals may be motivated to accurately assess their behaviors. Social comparison theory posits individuals seek social information to evaluate their opinions and behaviors. They will first seek to compare their performance to physical or objective measures of performance (e.g., sales per quarter). When objective measures are not available, as is typically the case for performance in the workplace, individuals will seek social comparison information (e.g., my performance is better than Pat and Chris), or use social information to add relative meaning to objective information (e.g., ratings of 3 and above, equate to average and above performance). Farh and Dobbins’ (1989) study found support for this hypothesis by showing individuals who were presented with comparative information provided more accurate self-ratings than individuals who had not received this information.

Harris and Schaubroeck (1988) outlined three ways egocentric bias may explain the self-other agreement. First, egocentric bias is shown in defensiveness, where an individual may be
motivated to inflate self-ratings of performance to protect his/her ego. A second explanation derives from self-enhancement theory. An individual may be motivated to downplay development opportunities and enhance strengths, by providing higher self-ratings of performance to protect or bolster self-esteem. The third approach suggested by Harris and Schaubroeck (1988) derives from attribution theory. Attribution theory proposes observers emphasize dispositional causes of behavior, whereas actors emphasize the situation. Individuals’ and others’ evaluations may be subject to this bias and combine to lead to differences in self-other ratings, and a tendency for the individual to rate themselves more highly. Harris and Schaubroeck’s (1988) hypotheses regarding egocentric bias may explain leniency in self-ratings and level of agreement.

The theories outlined here provide potential explanations for processes underlying self-other agreement in multi-rater tools. Systematic rater bias and unique perspectives due to relative level in the organization to the individual explain low correlations among sources. Self-raters’ leniency, egocentric, and self-evaluation bias may explain the tendency for individuals’ self-ratings to be higher than other ratings. Social comparison theory, self-consistency theory, and self-perception theory provide potential explanations for the processes underlying individuals’ responses to feedback. The various theories used to explain rater agreement highlight the multiple possible underlying processes and ways to conceptualize and measure agreement.

Measurement of Agreement

The largest variations in research regarding self-other agreement in multi-rater tools centers around the measurement of agreement. Rater agreement is the degree of congruence or fit between two sources (e.g., self-boss, peer-direct report) (Edwards, 1994). Initial research in rater
agreement primarily used correlations to explore the degree of relationship among sources’ (e.g. peer, direct report, other) overall mean ratings (Farh & Werbel, 1986; Holzbach, 1978; Mabe & West; Thornton, 1968). More recent research has been aimed at establishing a relationship between rater agreement and job performance. These studies have typically utilized a categorization method. Individuals are placed into categories based upon direction of agreement and job performance (i.e., over-rater, in-agreement/high, in-agreement/poor, under-rater). Analysis of variance (ANOVA) is then used to determine differences among groups in the dependent variable, such as job performance (Atwater et al., 1998; Fleenor, McCauley, & Brutus, 1996; Van Velsor, Taylor, & Leslie, 1993).

In comparison, research examining potential sources or moderators of self-other agreement have typically used correlation and difference scores across items, also referred to as profile similarity indexes (Edwards, 1993; Edwards, 1994; Fletcher & Baldry, 1999; Furnham & Stringfield, 1994; London & Wohlers, 1991; Mabe & West, 1982; Van Velsor et al., 1993; Warr & Bourne, 1999; Wohlers et al., 1993). Correlation measures assess the congruence in relative highs and lows or rank ordering of behaviors between sources (e.g., correlation across items). Difference measures assess the total variance or distance between sources’ ratings (e.g., mean absolute differences across items). While each of these are used to evaluate rater agreement, they differ in the underlying construct being assessed, appropriateness for various research questions, and face distinct criticisms resulting in limitations for interpretation (Edwards, 1993; Warr & Bourne, 1999).

A number of studies have used correlation across items to indicate similarity or agreement between sources’ ratings (Fletcher & Baldry, 2000; John & Robins, 1994; Harris & Schaubroeck, 1988; Levy, 1993; London & Wohlers, 1991; Sala, 2003; Warr, & Bourne, 1999).
Higher mean correlations indicate the individual’s self-ratings represent similar relative strengths and weakness as indicated by the other-ratings, while low mean correlations indicate the individual and others have more different views. These studies calculate a mean correlation across items between self and other ratings for each individual. The primary criticism of correlation measures is the inability to examine the relative distance between self and other ratings. The ability to assess direction and degree of relationship between sources, as well as the size of the differences between ratings, has led researchers to employ difference measures in examining self-other agreement.

Difference scores measure rater agreement by calculating a sum or mean difference across items. This form of measurement has slight variations. Edwards (1994) outlines four methods of calculating difference scores. The simplest method $D_1$ is the sum of differences between each item or dimension. An additional method calculates the sum of absolute differences, $|D|$. These first two methods assign equal weight to differences regardless of magnitude. The second two methods $D_2$, the sum of the squared differences, and $D$, the square root of $D_2$, assign greater weight to differences of larger magnitude. Difference measures as a whole are non-directional, in that they treat positive and negative differences the same (Edwards, 1993). A variety of studies have used some form of difference measures in examination of different aspects of multi-rater tools (Church & Waclawski, 1995; Farh & Dobbins, 1989; Farh & Werbel, 1986; Goffin & Anderson, 2007; Warr & Bourne, 1999; Wohlers, Hall, & London, 1993).

Difference scores are criticized for a number of reasons. The first is conceptual ambiguity, by combining items into a single score all items are given the same weight and source of variance is not identified (e.g., item one or item 7). The use of these methods also results in
lost information, similar differences are treated the same regardless of the level of individual scores (e.g. is 1 and 2 the same as 4 and 5). These criticisms are also made of correlation measures. Difference measures in particular are criticized for not accounting for pattern of agreement and being unreliable, systematically correlated with their components, and spuriously correlated with other variables (Edwards, 1994; Johns, 1981).

These criticisms have led to multiple methods and use of alternative methods of measuring agreement. Cheung (1998) outlined seven possible forms of rater agreement classified into two types, conceptual disagreement and psychometric disagreement. Correlation and difference score methods assume conceptual agreement, that performance is defined in the same manner for both sources, with the same items and weights, and equivalent reliability (Cheung, 1999). Measurement equivalence and confirmatory factor analysis are suggested methods of establishing conceptual agreement in factor form (i.e., what is important for performance) and factor loadings (i.e., how important) (Podsakoff et al., 2003; Diefendorff, Silverman, & Gregurag, 2005; Woehr, Sheehan, & Bennett, 2005). Cheung (1999) suggests that after establishing or assuming conceptual agreement psychometric agreement can be attributed to random measurement error, latent factor variability/range restriction, high correlation/halo effect, latent means/leniency, or comparable views of performance/correlation among sources. Confirmatory factor analysis and structural equation modeling are offered as alternative analyses for separating out the specific forms of agreement.

Another limitation of difference and correlation methods when used to identify moderators is the inability to determine the source of the relationship. A significant positive correlation found between self-esteem and low self-other agreement may be due to a positive relationship between self-esteem and self-ratings of performance, a negative relationship
between self-esteem and others’ ratings of performance, or an interaction between self and other ratings. A fourth method of measuring agreement, multiple regression techniques is proposed to circumvent this criticism (Cheung, 1999; Goffin & Anderson, 2007; Edward, 1994). This technique allows the identification of the source of disagreement (i.e., self, other, interaction), and opportunity to determine whether an independent variable is related to self-ratings, other-ratings, or both (Edwards, 1993). While the method is not as useful when agreement is the dependent variable, it provides a useful tool for assessing sources of disagreement.

Multiple regression, confirmatory factor analysis, and structural equation modeling procedures provide alternative methods of examining agreement. However, profile similarity indexes (i.e., correlation and difference scores) can be an appropriate measure of agreement. Edwards’ (1993) regression analysis methods are most appropriate when agreement is the criterion variable and less so when used as a predictor variable, as when using agreement to predict performance. When the measure is a reliable, multiple item scale difference scores and mean correlations are an acceptable method of examining individual differences on the same measure (Johns, 1981; Tisak & Smith, 1994). CFA and SEM approaches are effective in identifying conceptual agreement and type of agreement, but are complex, and require large sample sizes. They also do not provide simple indicators of self-other agreement that can be easily used in applied practice to assist individuals in understanding their own rating tendencies and perceptions of their performance.

The current study used three measures of agreement; direction (difference in overall mean scores), difference (mean absolute difference across items), and similarity (correlation across items). The three measures were used to assess how the relationship between agreement and job performance may differ. Each measure provides a simple index of agreement that could
be interpreted in applied practice and administration of multi-rater tools and represents a unique form of agreement, direction (i.e., leniency/severity in self-ratings), difference (i.e., variance across items), and similarity (i.e., similar rank ordering across items). Agreement was then examined at the dimension level to address the issue of conceptual agreement. Examining how agreement across dimensions may differ provides the ability to understand where disagreement most frequently occurs and what components contribute to self-other agreement (Edwards, 1993).

Sources of Self-Other Agreement and Disagreement

Meta-analyses using correlation methods have shown that sources often differ in their evaluations of individuals and recommend examining variables that can moderate the relationship between self and other ratings (Conway & Huffcutt, 1996; Harris & Schaubroeck, 1988; Mabe & West, 1982). Variables influencing self and other performance ratings are classified into common categories, including context, instrument, types of rater, demographic variables, and individual characteristics (Landy & Farr; 1980; Yammarino & Atwater, 1997). While rater agreement has examined potential sources of variability across categories, much of the research has focused on individual differences (i.e., demographic variables and individual characteristics).

Factors related to the context can include purpose of ratings, type of organization or industry, familiarity between the participant and rater, and opportunities to observe performance. Research in rater agreement has included type of organization (Brutus, Fleenor, & London, 1998; Wohlers, Hall, & London) and type of position (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; London & Wohlers, 1991, Sala, 2003). Research in this area has also
examined the variables in the second category, instrument. The instrument refers to the type and variations in the rating format used in the evaluation (e.g., graphic ratings, forced choice ratings, etc.), as well as the nature of the items (e.g., global versus specific). Instrument variables related to rater agreement in research included rating format and scale (Harris & Schaubroeck, 1988), rating instructions (Fox & Dinur, 1983), clarity of items (Wohlers & London, 1989), and participants’ knowledge of the appraisal process (Williams & Levy, 1992).

As discussed in previous sections the type of rater appears to be a significant source of variability in ratings. Ratings within sources (i.e., boss, peer, direct report) are more similar than across sources (Conway & Huffcutt, 1997; Mabe & West, 1982). Researchers assert that the poor agreement between self and other ratings is due to the relative level in organization to the individual. An increasing amount of research has begun to examine how individual differences may be related to the degree of agreement between sources. Demographic variables examined have included gender (Fletcher & Baldry, 1999; London & Wohlers, 1991; VanVelsor et al., 1993; Warr & Bourne, 1999; Wohlers et al., 1993), age (London & Wohlers, 1991; Warr & Bourne, 1999; Wohlers et al., 1993) and race (Furnham & Stringfield, 1994; Wohlers et al., 1993). Cognitive ability (Fletcher & Baldry, 2000; Mabe & West, 1982; Warr & Bourne, 1999) and self-esteem (Farh & Dobbins, 1989; Goffin & Anderson, 2007; John & Robins, 1994; Levy, 1993; Shrauger & Terbovic, 1976) are two of the individual characteristics most frequently examined in rater agreement research. Various personality measures have also been examined for their relationship to rater agreement; those examined have included the 16PF (Fletcher, 1999; Fletcher, Taylor, & Glanfield, 1996), MBTI (Church & Waclawski, 2001; Rousch & Atwater, 1997), Occupational Personality Questionnaire (Warr and Bourne, 1999), and Jackson Personality Inventory (Goffin & Anderson, 2007). Other constructs examined have included

A variable not extensively examined in rater agreement research is the importance of the performance dimension to overall effectiveness in the position. Many multi-rater tools are designed to include multiple aspects of performance in leadership positions and are thought to be relevant to many positions. While all performance dimensions may have some relevance to the individual’s effectiveness in the position they will likely not have equal importance. Rater agreement may differ based upon the importance of the dimension for the position, and the relationship between agreement and job performance may also be affected by the importance of the dimension. The current study expands upon past research by examining how importance of the dimension assessed by the participant may moderate the relationship between agreement and job performance.

Current Study

Research has shown self-other rater agreement is related to individuals’ job performance (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Brutus, Fleenor, & London, 1998; Church, 1997). Rater classification is frequently used to examine how under-raters, in-agreement, and over-raters differ in job performance (Atwater et al., 1998). Similarity measures, using the mean correlation across items, have also been used to examine the relationship between rater agreement and job performance (Church, 1997). The current study expands upon this research by examining the relationship of self-other agreement and job performance using three different measures of agreement, direction (difference between overall mean scores), similarity (mean correlation across items), and difference (mean absolute difference across items). The second
purpose of the study was to examine how the relationship varies across dimensions. Past research has focused on rater agreement on overall multi-rater tools versus agreement on specific dimensions. Examining agreement at the dimension level reduces some of the conceptual disagreement of examining across a broad measure by identifying which aspects of performance contribute to self-other agreement (Edwards, 1993). The final purpose of the study was to explore how the importance of the dimensions, as rated by the individual, may moderate the relationship between self-other agreement and job performance. This hypothesis is consistent with Tornow’s (1983) study which found rater agreement on dimensions related to people (e.g., interpersonal skills, relationships) had the strongest relationship to job performance.

Consistent with past research the current study hypothesized that agreement between individuals’ and their bosses’ ratings predict job performance (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Brutus, Fleenor, & London, 1998; Church, 1997). Individuals whose ratings are in high agreement (i.e., high similarity and low difference) with ratings from their boss will receive a higher overall performance rating from the boss than individuals in low agreement. It is also predicted that the direction of agreement is related to performance. Past research has utilized ANOVA designs to examine performance among categories of raters (over-estimators, in-agreement, and under-estimators). The current study utilizes a regression analysis to examine the relationship between job performance and agreement (i.e., direction, similarity, and difference). Individuals who indicate self-ratings which are lower than ratings provided by their bosses’ will receive the higher job performance ratings than individuals whose average overall mean scores are higher than their bosses’ overall mean scores.

Hypothesis 1: Agreement predicts job performance (similarity positively, direction and difference negatively).
Rater agreement was calculated for both the overall measure and for each dimension. The current study assesses if the relationship between rater agreement and job performance is consistent or differs across dimensions. It is hypothesized that rater agreement predicts job performance across dimensions. However, it was also expected that the strength of the relationship differs across dimensions.

Hypothesis 2a: Agreement on dimensions predicts job performance (direction, similarity, difference).

Hypothesis 2b: The relationship between agreement and job performance differs across the dimensions.

Past research on the relationship between self-other rater agreement and job performance have focused on direction of self-ratings compared to other-ratings (i.e., difference between overall mean scores). These studies have shown that individuals whose self-ratings are higher than others ratings receive lower job performance ratings (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Brutus, Fleenor, & London, 1998; Church, 1997). Similarity and difference measures are frequently used in studies examining potential sources or antecedents of agreement. The current study hypothesizes that similarity and difference measures of agreement will better predict job performance than the direction of agreement. While direction measures may appropriately assess an individual’s tendency to under or over-estimate his/her overall performance, this measure of agreement does not assess variance or similarity in self and other ratings across items. The current study uses the similarity and difference measures to assess how the relationship to job performance may differ. The underlying construct being measured with similarity and difference measures is distinct from the direction of rating. While direction measures leniency and severity, similarity more closely assesses an individual’s similar
perceptions of the relative high and lows across behaviors. The difference measure uses a mean of absolute differences across items, disregarding direction but accounting for total difference or variance between self and other ratings. It is expected that the similarity measure most closely assesses an individual’s ability and/or motivation to maintain self-perceptions that are similar to others and thus will have the strongest relationship to job performance. The difference measure is expected to have the second strongest relationship to job performance. This measure accounts for more information than the direction of agreement by accounting for variance in ratings, but provides less information than the similarity measure. Direction of agreement is expected to have the smallest relationship to job performance as it accounts for only the leniency/severity tendency. It is beyond the scope of this study to examine the possible underlying processes, varying relationships between job performance and the different measures of agreement will support the contention that multiple distinct processes may drive agreement.

Hypothesis 3: Similarity will better predict job performance than difference and direction. Difference will better predict job performance than direction.

The final purpose of the current study was to explore how the importance of the dimension may moderate the relationship between self-other agreement and job performance. The study’s second hypothesis predicts that the relationship between agreement and job performance will differ across dimensions. One possible explanation for the relationship between agreement and job performance is that individuals who attend to the perceptions and feedback of others are better able to make improvements in job performance. Individuals are more likely to attend to the dimensions that they view as most important for their position. It is hypothesized that relationship between agreement and job performance will be stronger for dimensions viewed as more important. Moderated regression analyses will be used to assess this hypothesis. It may
be that importance moderates the relationship for some dimensions while not affecting the relationship for other dimensions, or that the moderating effect may differ depending upon the measure of agreement (i.e., direction, similarity, difference).

Hypothesis 4: Importance will moderate the relationship between agreement and job performance, stronger relationship for important dimensions.
METHOD

Participants

The database used in the current study contained available multi-rater data for 215 participants who underwent a multi-rater process, and their bosses who also provided ratings in the process. Participants in the current study were primarily Caucasian, male mid-level managers or above, in their forties, holding a bachelor’s, with varying amounts of time in the position. Participants’ bosses were primarily Caucasian, male, executives, in their forties, holding a bachelor’s degree, with varying amounts of time in position. Frequency and percentage of participants and bosses in each demographic category are displayed in Table 1.

Measures

*Worthy Leadership Profile® for Executives (WLPe)*

The WLPe (Leadership Worth Following, LLC, 2006) is an online multi-rater tool administered as part of a leadership development program. Participants’ self-ratings and bosses’ ratings of their performance was included in the current study. Raters evaluate 120 behaviors on the extent to which the individual displays each characteristic on a 5-point Likert scale (1 = No Extent, 2 = Little Extent, 3 = Some Extent, 4 = Great Extent, 5 = Very Great Extent). These behaviors are organized into three dimensions and 12 sub-dimensions relevant to effective leadership. The first dimensions, Capacity, consists of Capacity to Reason and Make Good Decisions, Capacity to See and Realize the Future, Capacity to Communicate and Influence, Capacity to Know, and Capacity to Persevere and Adapt. The second dimensions Commitment include sub-dimensions Commitment to Excellence, Commitment to People and Relationships, Commitment to Learning and Personal Growth, and Commitment to Stakeholders. The final
dimension, Character, includes sub-dimensions Personal Integrity and Ethics, Organizational Integrity and Courage, and Humility, Gratitude and Forgiveness. The participant classifies each of the 12 dimensions as Important (1), Very Important (2), or Extremely Important (3) for the position. Participants and bosses are required to place four factors into each importance group. Importance ratings for the dimensions were calculated in this study by averaging the participant’s corresponding sub-dimension importance ratings. In addition raters indicated “the individual’s level of performance during the past year.” The primary boss’s response to this item was used as the measure of overall job performance in the current study.

Measures of Agreement

Three measures of rater agreement were used in the current study, direction (difference between overall mean scores), similarity (mean correlation across items), and difference (mean absolute difference across items). The first measure of agreement, direction, was calculated by subtracting the boss’s overall mean score from the self overall mean score and used to assess the direction of agreement. The second measure, similarity, calculated the mean correlation between self and boss ratings across all items. The similarity measure was used to examine similar pattern or relative highs and lows in self and other ratings. Difference scores were calculated by averaging the absolute differences between self and boss ratings across items, and used to assess overall variance between self and boss ratings. The same procedures were used for the dimensions by assessing agreement on the related items (Capacity = items 1 to 44; Commitment = 45 to 86; Character = 87 to 120).
Procedure

Data for the current study was provided by Leadership Worth Following, LLC (LWF). The data are collected continuously as a part of the firm’s research and continuous improvement efforts for their Worthy Leadership Profile for Executives™ (WLPe). The data consists of individuals who have participated in a leadership development assessment process with LWF. Part of this process entails completion of the WLPe, which is the focus of this study. Although the development program includes cognitive and personality inventories, interview and/or simulations (e.g., in-basket activity, role-plays), these data are not relevant to the current study. The participants’ organization identifies individuals to take part in the leadership development process. Participants have typically been identified as having ‘high potential’ for leadership in the organization.

Participants typically begin the multi-rater process six weeks prior to receiving feedback and participation in other assessment components. The participant receives an email with a link to the survey website, including instructions and deadlines for completing the process. The participant selects and enters names and email addresses for raters. The participants and respondents receive an email explaining the multi-rater process and a link to connect to an online survey. The WLPe feedback report is generated and provided to the participant by an individual trained in giving developmental feedback.

Data Analysis Plan

Stieger’s (1997) power analysis procedure was utilized to calculate the required sample size to achieve a desired level of power based upon an expected rho squared in the population, number of variables and alpha. The desired power (1-β) is the probability of finding a
relationship that actually exists or probability of rejecting false null hypothesis. Power in excess of .80 is considered a reasonable safeguard against Type II errors (Cohen & Cohen, 1975) and was set as the desired level. The regression equation for the second hypothesis includes four variables (i.e., agreement measures for the three performance dimensions and overall job performance). Past studies examining the relationship between rater agreement and job performance have found correlations ranging from .21 to .59 (Brutus et al., 1998; Church, 1997; Church 1997). Based upon this desired power, alpha (i.e., $\alpha = .05$, $1-\beta = .80$) and an average $\rho^2 = .16$, the required sample size for the current study was 64. The database for the current study had available data for 215 participants.

The results of the data analysis are presented and discussed in further detail in the results section of the current study. First, descriptives were calculated for the study variables. Demographic information for each rater group (self and boss) as provided in the multi-rater tool were calculated, including level and time in the position, education, age, and gender. Mean, range, standard deviation and Cronbach’s alpha for each rater group for the overall mean score and three performance dimensions and are reported.

To begin to understand the data, one-way ANOVAs were conducted for demographic variables (age, level, time in the position, education, race, and gender) to test for differences among groups in the rater agreement measures (i.e., direction, similarity, difference) were calculated. Previous research has found demographic variables may be related to rater agreement. Examining possible relationships within the study’s sample identifies factors that could affect or moderate the relationship between agreement and performance.

Prior to the regression analyses testing the hypotheses, examination of data was conducted to test for violations of assumptions relevant to multiple regression analysis. The
assumptions tested include linearity, normality, and homoscedasticity, in addition to an identification of possible outliers. Transformation of one or more variables, removal of extreme outliers and/or robust analysis were considered as possible procedures for reducing the effect of violated assumptions on Type II error rate, stability of regression coefficients and generalizability of findings. Robust analyses utilize statistical procedures, such as recoding outliers to limit the impact on regression coefficients.

Standard regression procedures were then used to assess the relationship between rater agreement (i.e., direction, similarity, difference) and job performance. Job performance as rated by the primary boss was used as the criterion variable and agreement on the overall mean score as the predictor variable. Three separate regression analyses were used to assess the ability to predict job performance with each of the agreement measures (i.e., direction, similarity, difference).

The second set of regression equations examined the second hypothesis, that agreement on dimensions would predict job performance. Agreement measures for the three dimensions were used as predictor variables with primary boss’s job performance rating as the criterion. Variations in the relationship between agreement and job performance across dimensions were examined by testing for significant differences between zero-order correlations using z-test proposed by Meng, Rosenthal, and Robin (1992).

The current study’s fourth hypothesis posits the agreement measures differ in the ability to predict job performance. It was hypothesized that similarity better predicts job performance than difference and direction measures, and difference would be a better predictor than direction. A larger total R² for the equation assessing one measure of agreement than the total R² from the equation assessing another measure regressed on job performance would indicate the former is a
better predictor of job performance. A z-test was used to test for a significant difference between correlations of predicted to observed scores of the two regression equations. The equation provided by Steiger (1980) was used in the current study.

The final purpose of the current study was to explore how the degree that the individual rates the dimension as important to the position may moderate the relationship between rater agreement and job performance. Hierarchical regression analyses were used to test this hypothesis. The main effects, measure of agreement and importance rating were entered in the first step, and an interaction term (i.e., rater agreement x importance rating) entered into the second step of a hierarchical regression equation. A significant r-squared change at the second step would indicate that the ability to predict job performance with rater agreement is moderated by the importance of the dimension to the overall position. A series of three regression analyses were run for each of the three agreement measures, resulting in nine analyses. It was hypothesized that importance would be a moderator for some dimensions, and not for others. It was also predicted that the moderating effect may differ across measures of agreement. Due to the increased likelihood of a Type I error by using nine separate regression analyses and lack of prior research examining importance as a potential moderator in the relationship between agreement and job performance this portion of the current study is exploratory.
RESULTS

The data set used in the current study included multi-rater data for 215 participants and their bosses. Data for 16 participants were removed due to the participant and/or boss not responding or marking not applicable more than ten percent of items. The remaining sample of 199 participants exceeds the required sample size determined by the power analysis.

The means, standard deviations, and internal consistencies (i.e., Cronbach’s alpha) for both participants and their bosses’ overall mean scores on the multi-rater and the performance dimensions (i.e., Capacity, Commitment, and Character) are shown in Table 2. Consistent with past studies participants’ self-ratings on the overall measure ($M = 3.98$, $SD = .34$) tended to be higher than bosses’ ratings ($M = 3.85$, $SD = .36$), $t(196) = 3.72$, $p < .001$ (Harris & Schaubroeck, 1988; Nilsen & Campbell, 1993). Boss’ rating on the single item assessing the participant’s job performance, and the participant’s rating of each performance dimension’s importance for his/her position are also displayed.

Descriptive data for three measures of agreement, direction (difference between self and boss overall mean scores), similarity (mean correlation between self and boss ratings across items), and difference (mean absolute difference between self and boss ratings) are displayed in Table 3. The correlation (i.e., similarity) between self and boss ratings ($r = .28$) was consistent with the mean correlation found in Conway & Huffcutt’s (1997) meta-analysis ($\bar{r} = .22$), $t(196) = -.23$, $ns$.

One-way analyses of variances were used to test for differences in measures of agreement among participants’ and leaders’ demographic data (i.e., age, education, level, time in position, and races), no significant differences were found. Independent samples t-tests were used to test for gender differences. There were no significant differences found between male and female
raters on the correlation (Male: $M = .27$, $SD = .14$; Female: $M = .30$, $SD = .13$) and difference
(Male: $M = .68$, $SD = .24$; Female: $M = .66$, $SD = .21$) measures of agreement. However, female
participants ($M = -.03$, $SD = .50$) were more likely than male participants ($M = .21$, $SD = .49$) to
rate themselves lower than their boss indicated by a significant difference in the direction
measure of agreement $t(197) = 3.18$, $p = .002$.

Hypothesis Testing

Prior to running regression analysis to test study’s hypotheses, data were examined for
violations of assumptions. Residual plots from the three regression analyses testing for the first
hypothesis were examined. Examination of predicted values against residuals suggested
normality and homoscedasticity in the regression analyses. Examination of the univariate data
was conducted to further assess the possible impact of violations of assumptions. Histograms of
measures of agreement and bosses’ overall job performance rating are displayed in Figures 1-4.
Examination of the figures suggests measures of agreement and bosses’ overall job performance
rating approximate normal distributions. Malanobis and Cook’s distances were also calculated
and plotted to test for outliers. Two cases were identified in all three regression analyses. Further
examination of multivariate data found both cases had overall job performance ratings of one.
Significantly lower than the mean performance rating ($M = 3.96$, $SD = .70$). These two cases
were removed from the regression analyses. Robust analyses including LS and MM were
conducted in S-Plus. In both instances r-squared dropped slightly .01 to .03. For ease of
discussion, SPSS results are presented here.

The study’s first hypothesis posited job performance would be related to agreement for
all three measures (i.e., direction, similarity, difference). As hypothesized, direction of agreement
accounted for 17 percent of the variance in job performance, \( R = .41, F(1,195) = 38.79, p < .001. \) These results indicate job performance is negatively related to direction of agreement, where positive direction indicates self-ratings are higher than boss-ratings and negative direction indicates self-ratings are lower than boss-ratings. The difference measure of agreement was also found to be related to performance as hypothesized, \( R = .20, F(1,195) = 7.44, p = .01. \) The difference measure of agreement accounted for 4 percent of the variance in bosses’ rating of the participants’ job performance. The mean absolute difference between participant and boss ratings was negatively related to job performance. In contrast, the hypothesized positive relationship between the similarity measure of agreement (i.e., mean correlation between self and boss ratings across items) and job performance was not supported, \( R = .01, F(1,195) = .01, ns. \)

The current study’s second hypothesis stated that the three measures of agreement (i.e., direction, similarity, and difference) would predict job performance for the three dimensions. Similar to the analyses for the first hypothesis standard regression analyses were used to test this hypothesis for each measure of agreement. Agreement measures were calculated for each dimension (i.e., Capacity, Commitment, and Character) and entered simultaneously into the respective regression analyses. Consistent with findings for the first hypothesis agreement on dimensions was related to job performance. The direction measure of agreement accounted for 17 percent of the variance in job performance, \( R = .41, F(3,193) = 13.17, p < .001. \) The difference measure of agreement accounted for four percent of the variance in job performance, \( R = .20, F(3,193) = 2.68, p = .05. \) Similarity in the dimension did not account for a significant amount of variance in job performance, \( R = .07, F(3,193) = .29, ns. \)

The second part of this hypothesis posited that the relationship between job performance and agreement would vary across dimensions. The zero-order correlations between measures of
agreement and job performance on the three dimensions are displayed in Table 4. A z-test suggested by Meng et al. (1992) was used to test for significant difference among the zero-order correlations. This hypothesis was not supported in the current study as z-values (.21 to .80) did not exceed the critical value of 1.96 for a two tail t-test.

The third hypothesis of the current study posited that the measures of agreement would differ in the ability to predict job performance. More specifically, similarity was expected to be most strongly related to performance followed by the difference measure, and finally direction of agreement. As shown in the preceding analyses this hypothesis was not fully supported by the current study, as similarity was not significantly related to job performance. The total r-squared for direction and difference measures indicate direction of agreement is likely a better predictor than the difference measure of agreement. To further test this difference a z-test provided by Steiger (1980) was used to test for a significant difference among the total r-squared from the regression analyses used to predict job performance with each measure of agreement. In the current sample, the direction measure of agreement was the strongest predictor of job performance. Direction of agreement resulted in a significantly higher total r-squared than correlation (Z* = 4.28, p < .05) and difference (Z* = 3.03, p < .05) measures of agreement. The difference between the total r-squared for the correlation and difference measures of agreement were not significantly different (Z* = 1.02, ns).

The final purpose of the current study was to explore if importance of the performance dimension would moderate the relationship between agreement and overall job performance. A series of nine hierarchical regression analyses were used to test this hypothesis, results are displayed in Tables 7 through 15. The main effects, measure of agreement for the performance dimension (e.g., Direction on Capacity) and the individual’s importance rating for the
performance dimension (e.g., Importance of Capacity) were entered into the first step. An interaction term of the measure of agreement and the importance rating for the dimension (e.g., Direction on Capacity x Importance of Capacity) was entered into the second step. A significant increase in total r-squared for the second step would indicate importance moderates the relationship between agreement and job performance. This hypothesis was not supported in the current study as entering the interaction term did not significantly increase the total r-square in any of the nine regression analyses.

The regression analysis examining the moderating effect of importance in the relationship between direction and importance for the capacity dimension found the main effects for direction ($\beta = -.40, p < .001$) and importance ($\beta = .14, p = .03$) were significant in step one account for eighteen percent of the variance in job performance, $R = .43, F(2,194) = 21.12, p < .001$.

Importance ($\beta = .16$) was the only variable to remain significant when the interaction was entered into the third step. Findings in the regression analyses for the commitment and character dimensions found similar results. The main effects of direction ($\beta = -.37, p < .001$) and importance ($\beta = .15, p = .03$) for the commitment dimension, accounted for sixteen percent of the variance in job performance, $R = .41, F(2,194) = 19.04, p < .001$. After entering the interaction in the second step, a negative beta weight for importance was the only significant contributor ($\beta = -.17, p = .02$). Only the main effect of direction ($\beta = -.38, p < .001$) was significant in the first step in the regression analysis for the character dimension, accounting for fifteen percent of the variance in job performance, $R = .39, F(2,194) = 16.93, p < .001$. Direction ($\beta = -.55, p = .05$) remained significant when the interaction was entered into the second step.

Consistent with the findings for previous hypotheses, the regression analyses for similarity, testing for the moderating effect of importance, did not account for a significant
amount of variance in job performance. However, significant beta-weights for the main effect of importance were found in the first step of the analyses for the capacity and commitment dimensions. The main effect for importance of the capacity dimension ($\beta = .14, p = .05$) was positively related to job performance. While, the main effect of importance of the commitment dimension was negatively related to job performance ($\beta = -.16, p = .03$). The beta-weights did not remain significant after the interaction term was entered into the second step. There were no significant associations with overall job performance found for the character dimension.

The regression analysis examining the moderating effect of importance for the difference measure of agreement on the capacity dimension found the main effects of difference ($\beta = -.16, p = .2$) and importance ($\beta = .15, p = .03$) were significant in step one, accounting for five percent of the variance in job performance, $R = .22, F(2,194) = 4.78, p < .01$. However, the beta weights did not remain significant when the interaction was entered into the second step. The regression analyses for commitment also found significant main effects, with both difference ($\beta = -.19, p < .01$) and importance ($\beta = -.16, p < .05$), both being negatively related to job performance and accounting for six percent of the variance in job performance, $R = .24, F(2,194) = 6.12, p = .003$. After entering the interaction term in the second step, the beta weights for main effects and the interaction were not significant. Only the main effect of difference ($\beta = -.20, p < .01$) was significant for the character dimension, accounting for four percent of the variance in job performance, $R = .20, F(2,194) = 4.07, p = .02$, neither the main effects nor interaction were significant in the second step.
DISCUSSION

Multi-raters are frequently used by organizations as a compliment to traditional performance feedback and development tools. Research has found low correlations between sources (i.e., self, boss, peer, direct report), and that differences among sources are most prevalent between self and other ratings (Harris & Schaubroeck, 1988; Landy & Farr, 1980; Nilsen & Campbell, 1993). Research has also found individuals’ ratings of their own performance are typically more lenient than ratings by others (Thornton, 1980). This tendency for individuals to provide self-ratings higher than others has been found to be related to multiple negative outcomes including lower job performance (Atwater et al., 1998; Fleenor et al., 1996; Van Velsor et al., 1993). The majority of these studies have primarily examined self other agreement as direction of agreement by classifying individuals into over, under, or in-agreement raters. Fewer studies have examined the link between job performance and other measures of agreement.

The purpose of the current study was to expand upon current research by examining the relationship between job performance and three different measures of self-other agreement, including direction (difference between self and boss overall mean scores), similarity (mean correlation between self and boss ratings), and difference (mean absolute difference between self and boss ratings across items). In addition, the study sought to explore how the relationship between agreement and job performance may differ across dimensions, and explore if importance of dimensions for the position may moderate this relationship. A sample of directors and executives from the retail industry, involved in leadership programs was used to test these hypotheses.

The direction measure of agreement was found to be the best predictor of overall job
performance in the current sample. The study’s findings in regards to direction of agreement were consistent with the findings of past research (Atwater et al., 1998; Fleenor et al., 1996; Van Velsor et al., 1993). Individuals who rated themselves higher than their boss, or over-estimated their performance, were perceived to be less effective than individuals who accurately assessed or under-estimated their performance.

The second measure of agreement examined in the current study was similarity or the mean correlation of self and boss ratings across items. The similarity measure of agreement was hypothesized to be positively related to the bosses’ rating of individuals’ overall job performance. It was expected that the similarity measure of agreement would to be a better predictor of performance than direction or difference measures. Examination of correlations and regression coefficients from the current sample did not support this hypothesis. Congruence in relative highs and lows or rank ordering of behaviors rated by the participant and their boss (i.e., correlation across items) was not found to be related to job performance.

The third measure of agreement examined in the current study, difference, was found to be related to bosses’ ratings of job performance. Consistent with Church’s (2000) findings variance between self and other ratings were related to job performance. In the current study, individuals with a low mean absolute difference between self and boss ratings received higher job performance ratings than individuals with greater difference between ratings.

The second hypothesis stated that agreement on dimensions would also predict overall job performance. Similar to the findings for the first hypothesis, the direction and difference measures were related to job performance. However, the hypothesis for the similarity measure was not supported. The second component of this hypothesis stated that the strength or size of the relationship would differ across dimensions. This hypothesis was not supported. The
differences between zero-order correlations for measures of agreement to job performance were not significantly different across dimensions. This finding differs from past research examining differences in the relationship between rater agreement and job performance, which found a stronger relationship for dimensions related to interpersonal skills (Tornow, 1983).

The third hypothesis predicted the similarity measure would be the best predictor of job performance followed by difference and finally direction. Past research examining the relationship between self-other agreement and job performance has focused on direction of agreement (Atwater et al., 1998; Brutus et al., 1998; Church, 1997). It was expected that measures of agreement which account for variance (i.e., difference) and correlation (i.e., similarity) between self and boss ratings across items would be more strongly related to job performance than measures assessing an individual’s tendency to under or over-estimate his/her overall performance (i.e., direction). In the current study measures did differ in their ability to predict job performance, but not as hypothesized. Direction was found to be a significantly better predictor than the difference and similarity measure. While the difference measure was found to significantly predict job performance, the difference between the total r-squared from the regression analyses for the difference and similarity measures of agreement were not significantly different.

The current study’s final hypothesis was aimed to explore how potential differences in the relationship between agreement and job performance across dimensions could be explained by the importance of the dimension for the position. It was hypothesized that as importance of the dimension increased the relationship between agreement and job performance would increase. Importance was not found to moderate the relationship between agreement and job performance in the current study.
The results of the regression analyses examining the moderation effect of importance on the relationship between agreement and job performance are summarized in Table 5. The moderating effect of importance on the relationship between agreement and job performance was not supported, indicated by non-significant changes in the $r^2$ following addition of the interaction term. The main effects of the direction and difference measures of agreement were significant in the first step across the three dimensions. However, the only significant beta-weight found after entering the interaction term was in the regression analysis examining the moderating effect of importance on the relationship between difference of agreement in the capacity dimension and job performance. It was not expected that importance would be directly related to job performance. However, a significant main effect for importance was found in the three regression analyses examining the moderating effect of importance on the relationship between job performance and agreement (i.e., direction, similarity, difference) in the capacity dimension. Both positive and negative beta-weights for importance were found in the regression analyses examining the moderating effect of importance on the relationship between job performance and agreement in the commitment dimension, while they were not significant for the character regression analyses.

The results of this study have implications for both research and practice. As the prevalence of multi-raters in organizations continues to grow, understanding the low correlations and notable differences among sources ratings becomes more important. The current study continues to expand on past research in multi-raters by examining the relationship of multiple measures of agreement to job performance and exploring importance as a potential moderator.

In the current study 17 percent of the variance in bosses’ ratings of the individuals’ job performance was accounted for by direction of agreement. This correlation is considered small.
against typical guidelines for interpreting correlation coefficients (Cohen, 1988). However, the impact of this variance is considerable when subjective performance ratings are used in empirical studies or for comparing individuals for pay, promotion, or other administrative decisions. The amount of variance accounted for by the difference measure (i.e., four percent) was significantly smaller than that accounted for by direction of agreement. However the impact of even this small amount of variance is worth considering in both research and practice.

The similarity measure of agreement was not found to account for a significant amount of variance in the current study. Additional research should examine its relationship to job performance in other samples, as well as a possible relationship to other important outcome variables. The similarity measure of agreement may not be related to job performance; however it is possible that an accurate understanding of relative strengths and weakness indicated by higher correlation between self and other ratings may be related to an increased ability to change and develop, perceptions of self-awareness, or satisfaction with leadership.

Results from the current study suggest direction of agreement is likely the most appropriate measure of agreement for examining the relationship to perceptions of job performance. Research should continue to use the direction measure, as well as a difference measure of agreement in examining the relationship between job performance and agreement. Particular attention should be given to further delineating measures and establishing if different measures of agreement are in actuality explained by different underlying processes.

Different underlying processes may partially explain the variations in the relationship to performance. Unique perspectives of performance are one possible explanation given to explain low correlations among sources’ ratings (Hoojberg & Choi, 2000; Tsui & Ohlot, 1988). Individuals weighting aspects of performance differently than others may not affect others
perceptions of job performance, as suggested by the current study’s non-significant findings for the moderating effect of importance. However, the egocentric biases often used to explain why self-ratings are higher than others’ ratings (Harris & Schaubroeck, 1988), may also be related to behaviors that result in individuals being viewed more or less favorably by others and ultimately impacting job performance ratings.

The relationship between the direction measure of agreement and performance may be partially be explained by the purpose of feedback. Individuals, particularly at higher levels of the organization as in the current sample, have likely had numerous opportunities to receive feedback. Individuals who have paid close attention to past feedback may be more likely to reflect this feedback in the performance ratings indicated by less lenient self-ratings. This attention to feedback may also result in continuous efforts to improve their performance and be reflected in higher performance ratings from others. Future studies could test this hypothesis the relationship between the direction measure of agreement and past opportunities to receive feedback.

Research also needs to further understand if the differences in job performance among individuals with varying levels of agreement are true differences in actual performance. The differences in studies examining the relationship between agreement and objective measures of performance (Nowack, 1997) may indicate that the relationship between agreement and job performance is limited to subjective ratings of job performance. This may suggest that direction of agreement may be descriptive of personality or situational variables that effect perceptions of overall job performance. If this is true, and the relationship is not descriptive of true differences, researchers should be careful in using subjective ratings of job performance as criterion in research studies. Organizations should also be cautious in using these ratings to make
administrative decisions.

The findings from the current and past studies regarding the relationship between direction of agreement and job performance are important for providing feedback to multi-rater participants. Past research suggests that individuals provided with feedback on their ratings were more likely to receive improved job performance ratings in future appraisals (London & Beatty, 1993). Providing feedback on individuals’ tendency to over or under-estimate their performance, in addition to research findings that direction of agreement is related to job performance, may provide increased motivation to more accurately assess their own performance and improve their chances for development and growth.

Importance of the dimension was not found to moderate the relationship between agreement and job performance as hypothesized. In the current study the mean importance rating for dimensions did not vary to a large extent (See Table 1). Further research should examine importance as a potential moderator where dimensions differ more in the importance for positions. Future research can also continue to examine other potential moderators or sources of agreement. Exploring potential sources and moderators will assist in developing a better understanding of the underlying sources for various measures of agreement.

One possible limitation of the current study is the available sample. The current sample included available data for 215 participants. While this sample met the requirements determined by the power analysis, larger and more diverse samples would increase the generalizability of the study’s findings. The sample includes primarily director-level and above participants in a single industry (i.e., retail). All of participants had been identified to participate in leadership development programs, and therefore likely as a whole to be above-average for their organizations. The overall mean score on the multi-rater tool ($M = 3.85, SD = .36$) and job
performance rating ($M = 3.96, SD = .70$) support this idea. It is suggested that future research continue to examine the relationship of agreement to job performance in larger, more diverse samples.

The performance criterion used in the current study was a single item rated by the participant’s boss at the end of the multi-rater tool. It is possible the study’s findings may be affected by common method variance. This could be inflating the relationship between agreement and performance, particularly for direction measure of agreement. For example, a boss with a tendency to provide overall lower ratings on the multi-rater tool would result in a larger direction of agreement score, due to common method variance the boss’ rating on the job performance criterion item is also likely to be affected by the rater bias, resulting in a lower overall job performance rating. This could result in an inflated relationship between agreement and performance. Future research could improve upon this study’s design by utilizing a different criterion measure (e.g., separate performance appraisal tool) or source (e.g., direct report).

The current study sought to examine three basic measures of agreement (i.e., direction, similarity, and difference). These measures were partially selected due to their use in past research. These measures were also selected because they could provide straightforward indicators that can be provided in feedback with participants to assist them in understanding their rating tendencies and the possible relationship to perceptions of job performance. More sophisticated measures of agreement are available including structural equation modeling, which could better identify the sources of agreement (e.g., leniency/severity in self or boss ratings). However, these methods would not provide a practical measure of agreement that could easily be used for development purposes.

As previously discussed, further research should be conducted to better understand the
differences in measures in agreement, the effects of potential sources and moderators, and relationships to multiple criterion items. The findings of past, current, and future research will provide information to efforts aimed at developing a better understanding of the underlying sources of agreement, as well as the relationship between agreement and important criterion variables relevant to leaders’ performance and development.
Table 1

Demographic Data (n, %) for Participants and Bosses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants</th>
<th></th>
<th>Bosses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>138</td>
<td>69.7</td>
<td>146</td>
<td>73.4</td>
</tr>
<tr>
<td>Female</td>
<td>60</td>
<td>30.3</td>
<td>52</td>
<td>26.1</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>3</td>
<td>1.5</td>
<td>7</td>
<td>3.1</td>
</tr>
<tr>
<td>Black/African American</td>
<td>10</td>
<td>5.1</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Caucasian</td>
<td>42</td>
<td>70.7</td>
<td>32</td>
<td>78.9</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>140</td>
<td>21.2</td>
<td>157</td>
<td>16.1</td>
</tr>
<tr>
<td>Others</td>
<td>3</td>
<td>1.5</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front-line Leaders</td>
<td>21</td>
<td>10.6</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Mid-level Leaders</td>
<td>79</td>
<td>39.6</td>
<td>6</td>
<td>3.0</td>
</tr>
<tr>
<td>Upper-level Leaders</td>
<td>73</td>
<td>36.7</td>
<td>78</td>
<td>39.2</td>
</tr>
<tr>
<td>Executive</td>
<td>26</td>
<td>13.1</td>
<td>110</td>
<td>55.3</td>
</tr>
<tr>
<td>Very Top Leaders</td>
<td>0</td>
<td>0.0</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 to 29</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>30 to 39</td>
<td>70</td>
<td>35.4</td>
<td>33</td>
<td>16.6</td>
</tr>
<tr>
<td>40 to 49</td>
<td>100</td>
<td>50.4</td>
<td>112</td>
<td>56.3</td>
</tr>
<tr>
<td>50 to 59</td>
<td>27</td>
<td>13.6</td>
<td>49</td>
<td>24.6</td>
</tr>
<tr>
<td>60 to 69</td>
<td>0</td>
<td>0.0</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-School/GED</td>
<td>15</td>
<td>7.5</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>Some College/Technical Training</td>
<td>42</td>
<td>21.1</td>
<td>25</td>
<td>12.6</td>
</tr>
<tr>
<td>Bachelor's Degree</td>
<td>74</td>
<td>37.2</td>
<td>72</td>
<td>36.2</td>
</tr>
<tr>
<td>Some Graduate Work</td>
<td>21</td>
<td>10.6</td>
<td>26</td>
<td>13.1</td>
</tr>
<tr>
<td>Master's Degree</td>
<td>43</td>
<td>21.6</td>
<td>67</td>
<td>33.7</td>
</tr>
<tr>
<td>Doctoral/ Professional Degree</td>
<td>4</td>
<td>2.0</td>
<td>6</td>
<td>3.0</td>
</tr>
<tr>
<td>Tenure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than One Year</td>
<td>39</td>
<td>19.6</td>
<td>20</td>
<td>10.1</td>
</tr>
<tr>
<td>1 to 2 Years</td>
<td>65</td>
<td>32.7</td>
<td>46</td>
<td>23.1</td>
</tr>
<tr>
<td>3 to 5 Years</td>
<td>60</td>
<td>30.2</td>
<td>59</td>
<td>29.6</td>
</tr>
<tr>
<td>6 to 10 Years</td>
<td>25</td>
<td>12.6</td>
<td>37</td>
<td>18.6</td>
</tr>
<tr>
<td>More than Ten Years</td>
<td>10</td>
<td>5.0</td>
<td>37</td>
<td>18.6</td>
</tr>
</tbody>
</table>
Table 2

*Descriptive Data for Measures*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Overall Mean Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Overall Mean Score</td>
<td>3.98</td>
<td>0.34</td>
<td>.97</td>
</tr>
<tr>
<td>Boss Overall Mean Score</td>
<td>3.85</td>
<td>0.36</td>
<td>.96</td>
</tr>
<tr>
<td>Boss Summary Job Performance Rating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Capacity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Overall Mean Score</td>
<td>3.89</td>
<td>0.36</td>
<td>.92</td>
</tr>
<tr>
<td>Boss Overall Mean Score</td>
<td>3.74</td>
<td>0.40</td>
<td>.92</td>
</tr>
<tr>
<td>Importance of Capacity</td>
<td>2.05</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td><strong>Commitment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Overall Mean Score</td>
<td>4.00</td>
<td>0.36</td>
<td>.94</td>
</tr>
<tr>
<td>Boss Overall Mean Score</td>
<td>3.87</td>
<td>0.36</td>
<td>.91</td>
</tr>
<tr>
<td>Importance of Capacity</td>
<td>1.98</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td><strong>Character</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Overall Mean Score</td>
<td>4.08</td>
<td>0.37</td>
<td>.92</td>
</tr>
<tr>
<td>Boss Overall Mean Score</td>
<td>3.95</td>
<td>0.40</td>
<td>.95</td>
</tr>
<tr>
<td>Importance of Character</td>
<td>1.94</td>
<td>0.44</td>
<td></td>
</tr>
</tbody>
</table>

Table 3

*Measures of Agreement*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Direction</th>
<th>Similarity</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Multi-rater Tool</td>
<td>0.13</td>
<td>0.50</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Capacity</strong></td>
<td>0.15</td>
<td>0.53</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Commitment</strong></td>
<td>0.13</td>
<td>0.51</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Character</strong></td>
<td>0.13</td>
<td>0.55</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
</tbody>
</table>
### Table 4

*Correlations between Agreement on Performance Dimensions and Job Performance Rating*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Direction</th>
<th>Similarity</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity and Job Performance</td>
<td>-0.40</td>
<td>0.04</td>
<td>-0.15</td>
</tr>
<tr>
<td>Commitment and Job Performance</td>
<td>-0.38</td>
<td>0.02</td>
<td>-0.19</td>
</tr>
<tr>
<td>Character and Job Performance</td>
<td>-0.39</td>
<td>0.06</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

### Table 5

*Summary of Regression Analyses: Significant Associations of Measures of Agreement, Importance, and Interaction with Job Performance*

<table>
<thead>
<tr>
<th>Direction</th>
<th>Similarity</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>β</td>
<td>R</td>
</tr>
<tr>
<td>Step One</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Agreement</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Importance</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Step Two</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Agreement</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Importance</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

- **Capacity**
  - Step One:
    - Agreement: -
    - Importance: +
  - Step Two:
    - Agreement: -
    - Importance: -

- **Commitment**
  - Step One:
    - Agreement: -
    - Importance: +
  - Step Two:
    - Agreement: -
    - Importance: -

- **Character**
  - Step One:
    - Agreement: -
    - Importance: -
  - Step Two:
    - Agreement: -
Figure 1. Histogram for direction measure of agreement.

Figure 2. Histogram for similarity measure of agreement.
Figure 3. Histogram for difference measure of agreement.

Figure 4. Histogram for overall job performance.
REFERENCES


