

WHO IS THE BEST JUDGE OF PERSONALITY: INVESTIGATING THE ROLE OF
RELATIONSHIP DEPTH AND OBSERVATIONAL BREADTH ON THE ACCURACY OF THIRD-
PARTY RATINGS

by

MITCHELL J. TINDALL
B.S. Psychology, University of Central Florida, 2006
M.S., IO Psychology, University of Central Florida, 2009

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: Kimberly Smith-Jentsch

© 2015 Mitchell Tindall

ABSTRACT

To date, the vast majority of research regarding personality in IO Psychology has relied on self-report assessments. Despite support for the utility of third-party assessments, IO Psychologists have only just begun extensive research in this area. Connelly and Ones (2010) conducted a meta-analysis that demonstrated that accuracy of third-party ratings improved as intimacy between the judge and the target grew. This remained true with the exception of predicting behavioral criteria, where non-intimates maintained superior predictability (Connelly & Ones, 2010). This was later contradicted by a recent investigation that found the best predictive validity for third-party assessments when they are taken from personal acquaintances as opposed to work colleagues (Connelly & Hulsheger, 2012). The current study is intended to investigate how the *depth* of the relationship and *breadth* of behavioral observations differentially moderate the relationship between third-party personality assessments and accuracy criteria (i.e., self-other overlap, discriminant validity and behavior). Results indicate that both depth and breadth impact accuracy criteria and they do so differentially based on trait visibility and evaluativeness. These findings will be discussed along with practical implications and limitations of the following research.

ACKNOWLEDGMENTS

This dissertation was written with the oversight of Dr. Kimberly Smith-Jentsch and with the support of the University of Central Florida.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES	vii
LIST OF TABLES.....	viii
INTRODUCTION	1
The Development and Validation of the Big Five Personality Dimensions	4
Research and Practice of the Five Factor Model (FFM) in IO Psychology	7
Variance in Self-report Personality Assessment	8
Variance in Third-Party Personality Assessment.....	12
The Accuracy Paradigm in Third-party personality Assessment.....	16
Moderators of Accuracy in Third-Party Personality Assessment	20
Hypotheses.....	27
METHODS	38
Participants.....	38
Procedures.....	38
Measures	38
RESULTS	42
Descriptive Statistics and Intercorrelations.....	42
Regression Results for Self-Other Overlap.....	45
Regression Results for Discriminant Validity.....	47
Moderated Regression Results for Predictive Validity	48
Moderated Regression Results Incremental Validity.....	51
DISCUSSION	56
Implications for Self-Other Overlap	56
Implications for Discriminant Validity	56
Implications for Criterion-Related Validity	57
Implications for Incremental Validity.....	58
Limitations and Directions for Future Research	60

Conclusions..... 62
APPENDIX: UCF IRB APPROVAL LETTER 64
REFERENCES 67

LIST OF FIGURES

Figure 1. Interaction between depth and third-party rated openness predicting GPA.	50
Figure 2. Interaction between breadth and third-party rated agreeableness predicting GPA.....	50
Figure 3. Interaction between breadth and third-party rated extraversion for predicting GPA.	53
Figure 4. Interaction between breadth and third-party rated agreeableness for the incremental prediction of GPA.	53
Figure 5. Interaction between breadth and third-party rated extraversion for the incremental prediction of GPA.	54

LIST OF TABLES

Table 1. Descriptive statistics and intercorrelations of study variables.	44
Table 2. Regression Results for the Predictability of Breadth and Depth on Self-Other Overlap and Discriminant Validity.....	48
Table 3. Moderated Regression Results for the Predictive Validity (GPA) of Third-party Ratings of Big Five Personality Traits and Their Interactive Effects with Breadth and Depth.	51
Table 4. Moderated regression results for the incremental validity (GPA) of third-party reports beyond self-reports of Big Five personality traits and their interactive effects with breadth and depth.	55

INTRODUCTION

Whether we realize it or not we make trait based judgments about those with whom we interact and these judgments inform the decisions we make about other people. As an illustration, when asked about a new colleague on the job, we might say he seems to be outgoing, disorganized but extremely nice. Employees often describe their bosses to one another in trait based terms (e.g., he's a pushover, she is very authoritative, and she is very approachable). These judgments likely influence the way in which we interact with these people. Likewise, some of our current more formal methods of personnel assessment are inherently trait based (i.e., letters of recommendation and reference checks). For example, people responding to references or writing letters of recommendation might describe candidates as punctual, well organized, adaptable, easy to work with, creative, easy going, etc. While trait based opinions of others are used more often than conventional wisdom would indicate, attempts to solicit this information using validated personality scales (i.e., using existing validated measures of personality) are rare in personnel selection settings.

Surprisingly, one of the most hotly debated topics in IO psychology over the past decade has been whether or not the continued use of self-report personality assessments are appropriate in selection given their notoriously low predictive validities (Guion & Gottier, 1965; Hough, 1992; Morgeson et al., 2007; Ones, Dilchert, Viswesvaran & Judge. 2007). Proceedings from a recent symposium at the Society for Industrial and Organizational Psychologists (SIOP) added to the debate of using personality assessments in selection contexts (Tett, Christiansen, Swanavelder, Meyer & Bartram, 2012; A sticky topic: Using personality tests in selection settings). Those advocating for the use of personality measures offered suggestions for improving prediction and reasons why low validities are not concerning. Specifically, it was argued that while personality measures may not predict performance, they may inadvertently predict other constructs as important as performance in industry (e.g., the ability to identify desired professional conduct or behavior, turnover, Organizational Citizenship Behaviors (OCBs), absenteeism). Suggestions for improving the validity of self-report personality measures included using 'force choice scales',

providing ‘warnings that lying can be detected’, and changing the context in which personality is measured from a maximum performance environment to a typical performance environment. Curiously, omitted from the debate was the use of third-party sources for the assessment of a target person’s personality. This approach to the assessment of personality has demonstrated incremental validity in the prediction of job performance beyond self-report assessment (Mount, Barrick & Strauss 1994; Connelly & Ones, 2010). While this approach to assessment has seen an abundance of research attention in clinical psychology (Achenbach, McConaughty & Howell, 1987; Duhig, Renk, Epstein & Phares, 2000; Renk & Phares, 2004), personality and individual differences psychology (Paunonen & O’Neill, 2010; Beer & Watson, 2008; Vazire, 2005; Vazire, 2010,) and social psychology (Bernieri, Zuckerman, Koestner & Rosenthal, 1994; Letzring, 2005; Letzring, 2010; Selfhout, Denissen, Branje & Meeus, 2009;), little has been done in IO Psychology to better understand the potential benefits of utilizing “other reports” of a target’s personality in the recruitment, placement and selection of personnel. In fact, in a recent handbook chapter of IO Psychology Oswald & Hough (2011) called for more research to be done on the use of third-party personality assessment in industry.

Though it is apparent that more work needs to be done in this area, there is evidence that third-party personality assessments hold considerable promise for improving the predictive validity of personality in industry. Mount and colleagues (1994) found that personality as assessed by an informant predicted job performance “at least” as well as the self-report assessment. In fact, they found that self-reports failed to add incrementally to the predictive validity of third-party assessments. If this is true, consultants may think twice before solely requesting the self-assessment of job candidates and instead include assessments from at least one outside source. Connelly & Ones (2010) confirmed the findings of Mount and colleagues (1994) in their meta-analysis and discussed several theoretical justifications for it:

- Third-party personality assessments are not contaminated with self-report bias

- Third-party assessors possess more behavioral observation in relevant contexts (e.g., school, work)
- Third-party assessments are more representative of reputation than internally held motives

In addition to these findings, Connelly and Hulsheger (2010) found that friends' ratings of a target's personality from outside the work context predicted job performance better than did self-reports. This is an indication that the context of observation may not be the sole driver of the effect as was thought by Connelly and Ones (2010). Thus, there is uncertainty as to who would be the best source of personality information in selection contexts. As previously stated, many of the findings from Mount and colleagues (1994) were corroborated by Connelly and Ones (2010) in their meta-analysis on the accuracy of third-party personality assessments. However, neither sample was sufficient for identifying which sources (e.g., strangers, new acquaintances, friends, family, coworkers) provide the best rating for predicting criteria important to organizations (e.g., performance, OCBs, CWBs, absenteeism, tardiness) nor do they answer why? Furthermore, we still do not fully understand how factors such as familiarity or intimacy with the target, access to behavioral information, and bias in other and self-reported ratings contribute to a more or less valid assessment when predicting performance? Consistent with this notion it was stated that:

“Given the promising findings for higher criterion-related validities than are typically found in self-report measures of personality, future research should work toward understanding third-party cognitive, affective and relationship biases when judging personality traits, determining whether those biases are helpful or harmful to criterion-related validity and investigating what situation manipulations or incentives will provide more veridical or valid personality ratings (Oswald & Hough, 2011, p. 172).”

The proposed study will attempt to add clarity to this research paradigm by extending past findings regarding third-party personality accuracy by testing the moderating effect and main effect of two relationship familiarity facets on self-other correlations, discriminability, criterion validity and incremental validity of third-party personality assessments. I see the current study contributing to the

literature on third-party personality accuracy and the use of personality assessment in personnel selection in several ways:

- Answering the call for more research in IO regarding the use of alternative sources for personality assessment.
- Investigating discriminant validity as a criterion variable for third-party personality assessment accuracy
- Separating the effects of depth of the relationship and breadth of observation in regards to accuracy criteria.
- Testing the effects of the aforementioned relational components on the criterion-related validity in a controlled/simulated environment.
- Understanding which facet results in the most incremental prediction beyond the self-report assessment of personality.

It is important to note that there may be legal and other obstacles to using other sources for personality information. With that said, hiring managers currently solicit performance, behavior and trait information from outside sources in the form of references and letters of recommendation (Leising, Erbs & Fritz, 2010), so it does not seem impractical for them to formally attain personality related information. A full discussion of the practical challenges to the use of third-party assessments in industry is outside the scope of this paper, however; some practical implications will be discussed further in the discussion section.

The Development and Validation of the Big Five Personality Dimensions

Before I review the research and theory regarding third-party personality assessment it is important to touch upon research and theory in self-report personality assessment and the psychometric development and validation of the Big Five as this was the approach used to assess individuals'

personalities in the current study. In this portion of the paper I will review briefly the development and validation of the Big Five, talk about its use in the realm of Industrial Psychology and discuss in more detail the ongoing debate concerning whether its use is even warranted. Finally, I will talk about how some of the short-falls of these assessments could be alleviated by the incorporation of third-party assessments.

Digman (1991) notes that despite criticism, practically all personality theorists agree that personality traits are linked to behavior. If this is true, what is personality specifically? The truth is that the definition of personality is still debated even today, however, it is generally defined as “consistent behavior patterns and intrapersonal processes originating within the individual” (Burger, 2011, p. 4). The idea that these behavioral patterns are consistent, is related to the notion that the reliable and valid assessment of them can have vast utility for Psychologists (e.g., more accurate clinical diagnoses, a better understanding of human behavior as influenced internally or externally, better ability to predict behavior for selection and placement into organizations). The biggest advancement in last 100 years of personality research is development and validation of Big Five (Costa & McCrae, 1992; Digman, 1990; Goldberg, 1993).

Several researchers and theorists were credited with the idea of examining language to gain a clearer and more concise understanding of human personality (McDougall, 1932; Klages, 1926; Baumgarten, 1933, in Digman, 1990). Their contentions led to Allport and Odbert’s (1936) initial examination of personality in language. It was thought that if we are able to break down, organize and categorize the way in which people describe their and another’s personality, we can identify its basic structure and components. Using factor analysis Cattell (1943, 1946, 1947, 1948) developed a system that organized the many thousands of words used to describe individual differences. However, this early system (i.e., 16 primary factors and 8 secondary factors) lacked the parsimony demanded by the scientific method (Digman, 1990). Fiske (1949) attempted to replicate Cattell’s work but was only able to find evidence for a five-factor structure. Tupes and Christal (1961) reanalyzed the work of both Cattell and

Fiske and uncovered the emergence of five factors (i.e., *Surgency, Agreeableness, Dependability, Emotional Stability and Culture*). While this study went unnoticed for a long period of time, Norman (1963) was credited with one of the earliest attempts to reduce personality to a consistent and manageable set of factors. He replicated the structure from Tupes and Christal's (1961) study to inform his contentions. Norman's (1963) study identified a taxonomy of personality attributes. Using peer nomination rating techniques he found evidence for five relatively orthogonal, clear and consistent factors of personality. Further substantiating the identification of these five factors Borgatta (1964) found evidence for five stable factors that he interpreted as, *Assertiveness, Likeability, Emotionality, Intelligence and Responsibility*. While the emergence of five-factors is agreed upon in the lexical approach to the study of personality, the labels of those dimensions was debated. Fortunately, Digman (1990) solidified the use of the five-factor model of personality in his review on the structure of personality relying heavily on Norman's (1963) interpretation of the five-factors of personality. Digman's (1990) interpretation of the Five-Factor Model included *Extraversion/Introversion, Friendliness/Hostility, Conscientiousness, Neuroticism/Emotional Stability and Intellect*. In his version, Assertiveness was subsumed under Extraversion, Likability became Friendliness, Emotionality became Neuroticism, Responsibility became Conscientiousness and Intelligence became Intellect, to differentiate the ability component from the personality component. The final labels and definitions of these dimensions were derived from an extensive review of past research (Digman, 1990) and while the specific names of the dimensions are not always consistent, the meaning of them is generally the same. Extraversion pertains to things like social adaptability (Fiske, 1949), surgency (Tupes & Christal, 1961; Norman, 1963), assertiveness (Borgatta, 1964), social activity (Guilford, 1975), sociability (Hogan, 1986) and interpersonal involvement (Lorr, 1986). Likability (which is commonly referred to as Agreeableness) includes things such as conformity (Fiske, 1949), likability (Borgatta, 1964), friendly compliance (Digman, 1988), and love (Peabody & Goldberg, 1989). Conscientiousness includes the will to achieve (Digman, 1988; Fiske, 1949), dependability (Tupes & Christal, 1961), task interest (Borgatta, 1964), prudence (Hogan, 1986), and self-control (Lorr, 1986). Neuroticism pertains to things like emotional control (Fiske, 1949), emotionality

(Tupes & Christal, 1961; Borgatta, 1964; Buss & Plomin, 1984), anxiety (Cattell, 1957), emotional stability (Guilford, 1975; Lorr, 1986), and adjustment (Hogan, 1986). Intellect (which is commonly referred to as Openness) includes inquiring intellect (Fiske, 1949), culture (Norman, 1963; Tupes & Christal, 1961), openness (Costa & McCrae, 1985), and intellectance (Hogan, 1986). Since this piece the the “Big Five” (i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience) has become the common language for researchers and theorists interested in examining personality. Furthermore, it offered a means by which applied psychologists could cheaply and easily examine the personalities of individuals for a multitude of purposes (e.g., clinical diagnosis, human research, personnel selection and placement). Industrial Organizational Psychologists have furthered our knowledge of the use of personality inventories based on the Big Five. However, many are not convinced that personality inventories have utility in selection settings and that we should continue our use and study of them.

Research and Practice of the Five Factor Model (FFM) in IO Psychology

One of the immediate practical implications of an easy, reliable and valid means of assessing personality was for placement and selection of individuals in jobs and organizations. As a result, a proliferation of personality assessment research and practice ensued in the field of Industrial Organizational psychology. Of course this attention was not without skepticism and extensive debate (Guion & Gottier, 1965; Mischel, 1968). Oswald and Hough (2011), for example, have contended that much of the research on personality in IO over the past fifty years has yielded contradictory and at times inconclusive results, however, they also stated that today the collection of evidence suggests that personality traits not only predict outcomes important to organizations (e.g., job performance, turnover, absenteeism), they also predict outcomes important to workers as well (e.g., career success, job satisfaction). In addition, these authors described the benefits of using personality assessment supplemental to assessments of cognitive ability (i.e., they add incrementally to the prediction of performance).

In a seminal meta-analysis of the predictive validity of the FFM of personality, Barrick and Mount (1991) concluded that Conscientiousness was a significant predictor of performance across a diversity of job types, while Extraversion surfaced as a significant predictor in jobs requiring a degree of interpersonal skill (e.g., sales). This study triggered a propagation of research on personality assessment in the field of IO Psychology (Oswald & Hough, 2011). As a result, I/O Psychologists have generally relied on the FFM model of personality for understanding the relationship between personality traits and organizational outcomes. Such research has continued to expand our understanding of the relationship between personality and an array of outcomes important to organizations. Specifically, Oswald and Hough (2011) in their review of personality research in I/O Psychology, cite studies that demonstrate significant relationships between personality and leadership and career success (Judge & Hurst, 2007; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007), job performance outcomes (Barrick, Mount & Judge, 2001; Bartram, 2005, Judge & Illies, 2002; Hough, 1992, Ones, Viswesvaran & Schmidt, 1993), contextual performance (Borman, Penner, Allen & Motowidlo, 2001; LePine, Erez, & Johnson, 2002), and counterproductive work behaviors (Berry, Ones, & Sackett, 2007; Ones, Viswesvaran, & Schmidt, 200). It is important to note that all of this research was based entirely on self-report personality assessment and to date little has been done to understand the role of third-party personality assessment in I/O Psychology. In fact, in the most recent handbook of Industrial Organizational psychology, Oswald & Hough (2011) only cite one study (e.g., Connelly, 2008) pertaining to the subject.

Before I review the literature on third-party personality assessment it is imperative to understand the process and different sources of variance believed to be contained in self-report measures of personality. A review as such may illuminate some of the biases with self-report assessment and enlighten us on the differences between self and other reported personality.

Variance in Self-report Personality Assessment

There is an abundance of research pertaining to the how people perceive their social worlds and this same body of knowledge can be used to understand how we perceive ourselves (Kenny, 2004). Also

referred to as self-perception (Kenny, 2004) the assessment of one's own personality represents a case of perceiver and target assessment where the perceiver is the self. Thus, in this approach we can only attain personality information from a single source. Kenny's (2004) PERSON model; a model used to organize and understand variance in other-person perception can also be used to organize and understand variance in self-perception. Some of the components of this model are relevant to self-perception while others are only relevant to other-person perception. In the next section, I will define each component of the PERSON model and explain what each tells us about self-perception.

Only two components of this model are relevant to self-perception (i.e., P and O) and they will be discussed in depth in this section. The components of the PERSON model that are not relevant to self-perception are E- error R- residual, S- stereotype and N- norm. Error in this model is defined as the part of the judgment not shared with other perceivers and not correlated with the judgment of other target behaviors. Stereotypes in the PERSON model are defined as assumptions based on physical appearance shared by multiple perceivers. Residual is defined as idiosyncratic notions unique to each perceiver or personal stereotypes. Finally, Norm is referred to as inconsistencies in the perception of behavior across acts, which are consensual due to the shared meaning perceivers assign to acts. Kenny (2004) contends that in the case of self-perception the 'number of acts (n) that the self has observed is incredibly large (i.e., tens of thousands of acts observed). When 'n' is very large the other sources of variance in the model (i.e., E, R, N, S) would drop and the self-report would be the aggregation of P (personality) components and O (opinion) components. I will discuss E, R, N and S in greater depth in the section on other-person perception. As implied earlier, the PERSON model dictates how variance is parsed between the components of the model based on characteristics of the perceiver's relationship or experiences with the target being perceived. Opinion in self-perception refers to the unique elements of one's self-perception not shared with others (Kenny, 2004). The classic example of this would be self-serving bias. Self-serving bias involves attributing negative aspects of the self to external causes while taking full credit for success. (Campbell & Sedikides, 1999). Theoretically, other perceivers do not possess the

motivation to perceive us in more positive light that we perceive ourselves (while this makes sense theoretically, I will discuss instances where this might not be true later in this review). The result in terms of self-assessment is an overly bias view of the self. In regards to self-perception, a significant portion of the variance is accounted for by the O component of Kenny's (2004) model. P or the personality component of Kenny's (2004) model is interpreted much the same as it is for outside perceivers. That is, P represents how the target person acts in general (i.e., across time and settings). Statistically, this component is represented by the mean across all of our behaviors across time. It is our "true" personality, if there is such a thing. The fact that the self makes more observations than others implies that the self should contain more P variance than other perceptions. It is important to note here, that P is a theoretical construct that is not possible to pinpoint to a single accurate judgment. Regardless of this notion, I can say that variance in self-perception is made up of variance from the *personality* and *opinion* components of Kenny's (2004) PERSON model. It is also important to note that E in Kenny's (2004) model does not mean variance attributed that is not related to a person's true personality (i.e., error). This more conventional definition of error would be partially subsumed under the O component as it may or may not reflect our true personality. With that said, other types of the traditional definition of error are not accounted for by P and O (e.g., the influence of mood, time of day, weather). In addition, the intentions of the perceiver can also introduce error variance in self-perception. For example, people often portray an ideal image of themselves in personality assessments. This ideal-self deviates from the actual-self and therefore constitutes error. While Kenny's (2004) model may not address error in perception as is typically defined, it is apparent that error contributes variance in self-perception. In particular, this source of variance would move us further from accurate perception. The amount of variance each component accounts for is not completely known but it is likely that it is dependent on the circumstances of perception (e.g., placement and selection, clinical assessment, new introductions), and the amount and quality of our observations over time. In the next section I will review another model meant to explain when other person perception is more or less accurate. This model can be loosely applied to self-perception as well.

Theoretically it makes sense that more observations result in more accurate perception. However, this does not account for the quality and diversity of those observations. Funder (1995) developed a model (i.e., RADU- Relevance of behavior to personality, Availability of behavior to perceiver, Detection of behaviors by the perceiver and the Utilization of the observed behavioral information in forming an impression) that is intended to explain how and when judgments made by an observer, of a specific target, are more or less accurate. His model contends that one can only expect accurate assessment of a target if *relevant* behaviors are made *available* to the observer and the observer is able to *detect* these behavioral cues and appropriately *utilize* them to form their judgment. Additionally, we may not always be aware of our behavior or able to adequately report it. As it turns out the same RADU framework (Funder, 1995) for understanding the accuracy in third-party personality assessment can be used to understand the accuracy of self-rated personality. How each component of the model is interpreted in this context will be quite different. The R, A components are environmental and the D, U components are perceiver sources of variance. The R in this model refers to the *relevance* of behaviors for various personality traits. For example, if an individual has never interacted in a team or group setting they may have very little information regarding their level of agreeableness. To ensure we have relevant information about ourselves for each of the Big Five, we would have to have behaved in a breadth of situations conducive to eliciting the relevant behaviors. The A in this model refers to the *availability* of those behaviors to the perceiver. Because we are physically present during all of our behavioral episodes, all behavior is available to the self. While behavioral information may always be available to the self, whether or not we detect it is an entirely different question. D refers to whether or not we actually notice or detect the behavior. Interestingly there is at least one situation where we may not detect our own behavior. For example, some cognitive and behavioral processes that become so well learned, can be automated and enacting them can occur without our conscious awareness. As an illustration, many of us have had the experience of driving to a destination and not remembering the journey there. Driving is a commonly cited cognitive and behavioral process that is largely automated. Lastly, U refers to the *utilization* of the *relevant, available* behavior *detected* to make a judgment. In terms of self-perceptions utilization can be

affected by the context of assessment (i.e., we may use information differently when being assessed for a job versus for psychology study), and our experience and education (i.e., trained psychologists may have a better understanding as to how certain behaviors reflect certain traits). Using the RADU model one could contend that self-perception is likely to be more accurate when we have behaved in a diversity of settings conducive to eliciting behaviors relevant to each of the Big Five, when those behaviors are under our conscious awareness, when the context for perception does not motivate us to use behavioral cues in a bias way and when we have special training or education in what behaviors represent certain traits. The RADU model is an excellent supplement for understanding the variance in self-perception and particularly when these perceptions are more or less accurate.

The preceding review provides an understanding of the sources of variance in self-perception, how and when those sources account for more or less variance. Additionally, I reviewed the conditions for producing more or less accurate perceptions of the self. In the next section I will review the sources of variance in third-party personality assessment, how and when those sources account for more or less variance and what conditions lead to more accurate third-party ratings.

Variance in Third-Party Personality Assessment

To gain a better understanding of the various sources of variance in third-party personality assessment it is important to refer back to the literature on ‘person perception’. Some of the earliest theoretical work pertaining to this domain was conducted by Cronbach, Gleser, Nanda & Rajaratnam (1972) when they established generalizability theory. In regards to generalizability theory, it is implied that an individuals’ rating of a target includes the aggregation of multiple components (e.g., judge, target and their interaction). An extension of this theoretical work was performed by Kenny and La Voie (1984) when they developed the Social Relations (SR) Model. This model organizes variance in person perception in a social interaction. First, the authors contended that there are actor effects that contribute variance. These effects pertain to the individual’s mean level of behavior in the presence of a variety of other partners. Additionally, there are partner effects – or the amount of behavior a person is able to elicit

from a target. The relationship effect constitutes variance attributed from the unique adjustment in behavior based on the specific relationship the dyad shares. Lastly, instability and/or error represents the noise in the perception of a target. While SR model provides an excellent and simplistic model for beginning to understand the variance in third-party personality assessment, Kenny (1991, WAM; Kenny, 1994.; Kenny, 2004, PERSON) elaborated on this model in terms of person perception. Thus, Kenny's Weighted Average Model (WAM) (1991) and his Personality, Error, Residual, Stereotype, Opinion, and Norm (PERSON) model, will help us gain more insight on the contributions of variance in third-party personality perception.

Kenny's (1991) WAM was developed to better understand the contributions of variance to consensus between two raters of a target's personality. The six parameters of this model include the number of acts raters observed, the proportion of acts raters observed in common, consistency of behavior of the target, similar meaning systems of the raters, weight of unique impression and communication. In general, this model was used to better understand why consensus does not increase as a result of better acquaintance. While this model has demonstrated some utility in assisting researchers interested in studying person perception from a theoretical standpoint (Borkenau & Liebler, 1993; Funder, Kolar & Blackman, 1995), empirical research relying on the model has been limited. The model itself lacks parsimony (Ickes, 1996) and its parameters do not lend themselves to adequate operationalization (Kenny, 2004). To address the shortcomings for research on WAM, Kenny (2004) developed the PERSON Model. This model makes clearer distinctions regarding the apportioning of variance relative to the Social Relations Model. As a result, I will rely on components of PERSON in order to provide an organizing framework for understanding sources of variance in third-party personality assessment.

Generally speaking, the PERSON model helps us understand two questions about person perception relative to the current research; 1.) What level of a given personality characteristic (e.g., neuroticism), do two raters believe a target individual possesses and 2.) To what extent do they agree on that rating? The answers to these questions will outline how variance is partitioned in third-party

personality assessment but this does not mean the model adequately captures every source of variance. Further elaborations on person perception theory in the third-party personality accuracy literature will be discussed later in this paper.

At the broadest level Kenny (2004) breaks the variance in person perception into categorical information and behavioral information. Categorical information is defined as non-behavioral information that a rater learns about a target at the initial meeting. Examples of this type of information include, appearance (e.g., height, weight, attractiveness) and demographic characteristics such as race, gender and age. As can be gleaned from the examples provided this categorical information are commonly referred to as ‘stereotypes’. Essentially, categorical information attributes the greatest portion of variance in person perception early on in the interpersonal interaction. Thus, the longer the interaction the less impact categorical information should have on third-party personality assessment. Behavioral information is defined as the meaning that is attached to the acts of a target (Kenny, 2004). The model does not assume to know the exact meaning attached to a given act but instead presumes that ongoing instances of behavior are partitioned into discrete acts such as ‘telling a riddle’ or ‘yelling at a waitperson’. In this way the model can categorize specific acts and later derive the meaning by examining the third-party rater’s assessment of a target. At the broadest level PERSON dichotomizes sources of variance in person perception into categorical sources and behavioral sources. The proceeding discussion further elaborates on sources of variance subsumed by these two.

In total PERSON identifies six sources of variance in a rater’s assessment of a target’s personality. These sources include, personality, error, residual, stereotype, opinion and norm. Of the six sources stereotypes and residuals are considered categorical while personality, norm, opinion and error are considered behavioral. In this model stereotypes are defined as shared expectations based entirely on physical appearance (Kenny, 2004). A classic example of stereotyping is the many association’s people make about one another based on race. In the United States these associations are widely shared through communication (Thompson, Judd & Park, 2000) and this is a crucial feature of stereotypes. In contrast,

residual is defined as idiosyncrasies or stereotypes that are unique to each perceiver. As an illustration, Kim may believe that people with blond hair are generally smart while Mary believes they are unintelligent. Both stereotypes and residual influence perception early in the interaction, the difference is that residuals are only shared in chance occurrences.

In regards to behavioral sources of variance in person perception, personality is defined in PERSON as the common way in which a target individual is seen (Kenny, 2004). Put another way, personality represents the rating that would be made if all raters were allowed to view all of the target's behaviors. It is important to note that this component is hypothetical as no single rater will be able to see all acts of a target. Norm refers to the unique meaning raters allocate to a given act. This meaning is consensual. For example, if smiling is perceived by people as a sign of being friendly then two raters watching John smile would contend that John is indeed friendly. Norm also has the potential to reflect inconsistencies in the behavior of a target. That is, if the first time I encounter John he smiles but the second time he frowns, my new rating of John contradicts my initial rating. In my initial observation I may have rated John as high in friendliness while in the subsequent observation I rated him low in this dimension. Variance attributed from the norm component of PERSON is an aggregation of each act observed of a target. Again it is important to remember that these acts are ones that have shared meaning in the eyes of the raters. Opinions are defined in PERSON as the unique interpretations a rater attributes to behaviors. For example, if one rater generally has a more favorable view of John, all of John's acts will be colored by this favorable view. So even if John frowns, the rater may still justify that frown as a sign of John's friendliness. Lastly, error in the PERSON model is defined as the part of a rater's assessment of a target that is not shared with other raters and not related to the evaluation of other behaviors of the target (Kenny, 2004). In reference to variance in person perception error is what is left over after accounting for P, N or O. While Kenny's (2004) PERSON model is useful in beginning to understand the sources of variance in third-party personality assessment of the rater and the target during an interpersonal interaction, it provides less information about sources of variance central to the situation or circumstances

under which ratings are being made. It also assumes that ratings occur only when an interaction is occurring. Research has shown that people can formulate impressions of one another via direct observation without real or extensive interaction. For example, many students of psychology indicate their interest in the field started because of their passion or propensity to “people watch”. This practice involves the formulation of an impression of a target based on passive observation without direct interaction. What settings or circumstances would produce a fair and accurate assessment of a target in these instances? In the next section, I will review Funder’s (1995) Realistic Accuracy Model (RAM). This model not only furthers our understanding of sources of variance in person perception within the rater, it elaborates on situational sources of variance in third-party assessment that can either mitigate or enhance the accuracy with which a rater can assess a target’s personality.

The Accuracy Paradigm in Third-party personality Assessment

An understanding of the variance in third-party personality assessment would be incomplete without a review of the unfolding body of literature on third-party accuracy and the moderators to accurate judgment. While it has been debated whether a single accurate rating of a target’s personality can be achieved (Shrauger & Schoeneman, 1979), there has been plenty of interest in studying the phenomena from this perspective (Funder, 1995; Kenny 1991; Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Letzring, 2010, Letzring & Wells, 2005; Connelly & Ones, 2010). This work was born from a desire to move away from simply studying sources of error in perception.

According to Funder (1995) the earliest work in accuracy actually began with a heavy focus on error in human judgment, perception and personality assessment (e.g., Kahneman & Tversky, 1973; Ross, 1977, Nisbett & Ross, 1980; Ross & Nisbett, 1991). This line of research was incredibly influential in depicting the specifics of different types of error. Further work in the error paradigm provided suggestions for reducing it, such as making people aware of the heuristics they rely on when judging another person (Funder, 1987). While Funder (1995) admits that research on error sheds light on accuracy, he further states that it alone can only tell half the story. The “accuracy paradigm” as Funder (1995), refers to it, can

answer many more questions about how to achieve accurate judgment, when accuracy is most likely (i.e., under what circumstances), who are the most accurate judges, what characteristics are most conducive to accurate judgment, and what aspects of people make them easier to judge accurately. Many features of Funder's (1995) Realistic Accuracy Model (RAM) and work regarding moderators were leveraged in the present study.

There are essentially three approaches to the accuracy paradigm; the pragmatic, constructivist and realistic (Funder & West, 1993). The pragmatic approach is concerned with the practical value of an accurate assessment. This approach views the ability to judge others' personalities as a useful mechanism for operating in social environments (McArthur & Baron, 1983; Swann, 1984). According to this perspective, a successful social interaction is the result of accurate personality perception and judgment (Funder, 1995). For example, if you were at a party where you did not know many of the attendees prior, but were able to mingle and navigate the social environment without awkwardness or uncomfortable moments, you would be seen as an accurate judge of personality according to the pragmatic approach. That is, your ability to accurately read the personalities of attendees enabled you to navigate the social environment with ease. According to Kruglanski (1989) the constructivist approach sees judgments of personality as social constructions where the accuracy of these constructions depends on your level of consistency or consensus with other judges. Using the party again as an illustration, imagine you attend the party with a friend and both of you meet the same person and later discuss your perceptions of that person. If there are differences in your perceptions it would be an indication that one or both of you is inaccurate in your perception. If both of you share the same impression(s) of that person it does not necessarily mean you are both accurate but that there is reliability between your observations and perceptions, lending evidence that your judgments may be accurate. While Funder (1995) does not deny the usefulness of these approaches to the accuracy paradigm, a third approach is proposed which he refers to as the Realistic Accuracy Model (RAM). It is based on the assumption that "personality traits are real characteristics of individuals" (Funder, 1995, p. 653). This approach considers the importance of the

construct validity of personality in terms of accurate judgment. That is, while judgment in general is a social psychological issue, accurate judgment of the specific personality attributes of a target requires a deep understanding of personality in addition to being an objective social judge. That is, to be an accurate judge of personality you have to know what personality is as a whole. In addition to its focus on the operationalization of personality, Funder's (1995) approach contends that accurate judgment can only result when the relevancy, availability, detection and utilization (RADU) of behavioral cues is apparent in the judgment of a target. Unique to this approach, is a move away from characteristics of the judge and toward features of the situation and target that lead to accurate judgment.

RAM assumes that personality traits are real and as a result accuracy in assessing traits requires a range of criteria (Funder, 1995). Moreover, it describes a model where accurate judgment is achieved when the judge has *relevant* information about the target's personality *available* to them, and when they *detect* those behavioral cues and *utilize* them appropriately in their judgment. In addition to outlining the conditions for accurate judgment, this model helps identify moderators that should increase accuracy.

The range of criteria to determine the level of accuracy in personality judgment can be organized along the same pragmatic, constructivist and, most recently, realistic approaches to personality judgment. Swann (1984) explains that an accurate judgment from the pragmatic perspective would occur when an individual maintains competency in their social interactions. An accurate judge according to Swann (1984) would be able to read a social situation and behave in an effective manner. According to this approach, the judge's effectiveness in social situations is a manifestation of their ability to accurately judge the situation and the people in it. This pragmatic criterion implies that researchers collect information about the success or failure of a judge in social settings if they are to understand the judge's ability to accurately assess others (Funder, 1995). Obviously, this approach to accuracy is practically difficult to study in a lab setting. Additionally, social competence may not solely be a reflection of accurate social perception, but the result of other constructs and processes (e.g., Extraversion, experience, social role).

The constructivists approach deems reality as a function of human perspective (Kruglanski, 1989). Therefore, criterion measuring accuracy from the constructivist perspective would compare the level of agreement between multiple judges rating the same target or phenomena (e.g., inter-rater reliability, agreement, and consensus). By thinking about these two approaches operationally as opposed to theoretically, it is much easier to understand their strengths and weaknesses. The most apparent feature is that neither incorporates information about the target themselves but instead focuses solely on the makeup of the judge's social functioning and ratings of others. In response to these shortcomings, the realistic approach demands that a broad array of information be obtained from a target in order to make some determination about the actual personality of the target (Funder, 1995). With that said, information about the judge is not thrown out in this approach but instead is a supplement to information about the target when determining level of accuracy. A critical element of this approach comes from Egon Brunswik's (1956) "probabilistic functionalism" in which he contends that while truth exists we cannot directly measure it and therefore must rely on the wide variety of means (e.g., pragmatic, constructivist and realistic) to uncover it. Funder (1995) used this logic when developing his RAM approach to accuracy. Essentially, there is no single way to identify an individual's true personality. Each method (i.e., self, other, and behavioral assessment) is flawed in some way. It is only by integrating and analytically comparing the variety of data sources (e.g., friends, family, coworkers, superiors, new acquaintances) that we are able to begin to understand where someone's true personality lies. Connelly and Ones (2010) used this logic in their meta-analysis regarding the accuracy and predictive validity of observers' ratings. They used self-other correlation, inter-rater reliability and predictive validity (i.e., GPA and Job performance) of third-party ratings as their criteria for accuracy, and while they admit deficiencies with each approach, they contend that other-ratings satisfying these three criteria make a strong case for an accurate assessment. That is, a personality assessment by a judge of a target that is highly correlated with the target's self-perception, has high overlap with other judges rating the same target and is consistent with behavioral reports of that target such as job or academic performance, would make a strong case for an accurate assessment. In their study, they found that the criterion (i.e., self-other

correlations, inter-rater consensus, and correlations with actual behavior) used to rate accuracy, dictated who were the most accurate judges of personality (e.g., strangers, acquaintances, coworkers, friends and family,). Specifically, they found that intimate others (i.e., family, close friends and significant others) contained the most accurate rating of a target when the criterion was self-other correlation. This was explained by the fact that these individuals have special knowledge of the targets traits that is not readily available in their behavior (e.g., beliefs regarding culture and religion, ability to deal with difficult and stressful situations). In fact, their ratings of a target's Neuroticism and Openness was especially pronounced relative to ratings made by strangers and new acquaintances. Essentially, these individuals have special knowledge about how targets view themselves that strangers and new acquaintances do not have available to them. Their findings regarding the inter-rater reliability of third-party ratings demonstrated that raters (regardless of their level of intimacy with the target) are generally consistent when rating Extraversion and Conscientiousness (i.e., the two highly visible traits) but lack consistency when rating Agreeableness, Openness and Neuroticism. They explain this finding by stating that individual ratings are highly idiosyncratic from one rating to the next. Inconsistency between two rater's ratings is then the result of differences in perceptions, observations and interpretations of the targets behavior. Lastly and may be most surprising was their finding that third-party assessments were actually better at predicting academic achievement (i.e., GPA) than self-reports. These differential findings regarding the accuracy of third-party assessments require further investigation. As a result, each of the three criteria (discriminant validity in place consensus) will be examined within my study.

Moderators of Accuracy in Third-Party Personality Assessment

Funder (1995) reviewed four variables that should affect the accuracy of third-party assessments: 1) Characteristics of the judge, 2) the target, 3) the traits assessed and 4) the availability of information.

Judges

According to Funder (1995) accurate judges are better able to detect and utilize cues of behavior when making their ratings. There are three aspects to being a "good judge" of personality. The first would

be strongly related to education, as knowledge about personality and how its manifest in behavior should help an individual better detect behaviors and subsequently utilize relevant behavioral cues. Aside from education an individual's interpersonal experiences may also increase their knowledge of personality. For example, a person who has had a variety of professional, educational and personal experiences overtime would have likely had interactions with a variety of people. As a result, they may be better at differentiating amongst people and in turn better at accurately interpreting their personality traits. In addition, to these experiential variables of individuals, certain individual difference variables (i.e., ability/intelligence, personality, interests) should also be related to the "good judge". Akert and Panter (1988) concluded in their paper that extraverts are better at decoding nonverbal cues in social interaction than introverts because they are more experienced in interpersonal interactions. More recently, Letzring's (2007) found that judges whose assessments had greater overlap with her accuracy criteria (i.e., the mean of the self-rating, a familiar other-rating and a clinical assessment) possessed a more positive view of human nature, were more socially perceptive, conscientious, outwardly focused, socially skilled, and warm and compassionate. Letzring (2014) added to our knowledge of the "good judge" by demonstrating that normative accuracy (i.e., the average ratings across a set of targets for each item on a personality assessment) was related to agreeableness. Lastly, Funder (1995) discusses the effect of motivation on being a "good judge". For example, Flink and Park (1991) found that when motivation was manipulated, consensus between judge's ratings became stronger.

Targets

Good targets are described as individuals who can be judged more accurately than others, even after only a few opportunities to observe their behavior (Allport, 1937; Colvin, 1993; Funder, 1995). According to RAM, information regarding the "good target" is most applicable to the relevance and availability aspects of the model (1995). For example, an individual conducive to accurate judgment is one whose behavior is readily available to the observer and one whose behavioral cues are not misleading. The notion that some individuals are more easily rated than others has several important

implications. First, these individuals may express more behavior by virtue of being more energetic (Funder, 1995). Second, ratings of targets that adjust their behavioral cues may not reflect who they actually are. For instance, Snyder (1987) describes social monitors as individuals that adjust their social behavior based on cues in the environment. Low social monitors are more likely to be “good targets” than high social monitors. In support of this notion, Bem and Allen (1974) found that people who described themselves as consistent across social settings were rated with more agreement by others. Colvin (1993) extended this notion of consistency by contending that the well-adjusted are not only consistent across situations but also in their words, behaviors and thoughts. As a result, these people are rated more consistently and accurately by others. Donahue, Robins, Roberts and John (1993) found that inconsistency is a sign of a fragmented self-perception.

Traits

Good traits are those that are thought to be rated more accurately and with more ease by third-party raters (John & Robbins; Funder; 1995). In terms of RAM “good traits” pertain to the relevance and availability components (Funder, 1995). That is, certain traits are more conducive to behavioral manifestation than others. Funder (1995) discusses two features that influence whether a trait is “good”: visibility and evaluativeness. Specifically, traits high in visibility are those that are expressed outwardly in our behavior (e.g., extraversion). Often when we hear lay people describe one another they first talk about an individual’s outgoingness, talkativeness or lack thereof. This is because an individual’s interpersonal behavior is readily available in many of the situations we are in. By its very nature, extraversion describes the way in which people behave. In contrast, traits low in visibility convey aspects internal to us such as our thoughts, beliefs and feelings (Connelly & Ones, 2010). Emotional stability and Openness to Experience are traits that are thought to be low in visibility (Zillig, Hemenover, Dienstbier, 2002). For example, an individual’s Openness to Experience describe thoughts they have regarding the arts, foreign cultures and religion, while an individual’s Emotional Stability describes someone’s affective reactions to things like work stress, breakups, and punishment. Traits high in evaluativeness are those that

people desire to possess because of social pressure such as compassion, intelligence and bravery (Funder, 1995). That is, these traits tend to be socially desirable. They are considered difficult to rate because of people's motivation to artificially display or hide them in their behavior affecting their relevance and availability to the observer. John & Robbins (1993) found the intellect aspect of Openness and Agreeableness to be the most evaluative of the Big Five traits. The discussion on trait-based moderators of accuracy is an important one for the current study. As you will see later in the paper I will be discussing moderators to accurate judgement. The moderating effect of the variables I am interested in are likely going to operate differently based on the visibility and evaluativeness of the Big Five traits. This notion will be discussed at length in the section regarding my hypotheses.

Information

Lastly, "good information" is described by Funder (1995) as information available to the observer whether he or she perceives and utilizes it in a meaningful way or not. Good information applies to both the sheer amount of information available to observers as well as the quality of that information. Research with newly acquainted individuals has failed to demonstrate that more information in the form of increased observation results in greater accuracy (Kenny, Albright, Malloy & Kashy, 1994). However, Connelly & Ones (2010) demonstrated that the quality of information does appear to have an effect on accuracy. While Funder (1995) does not specify what part of his RAM model "good information" would be most applicable to, one would have to argue that situations that induce relevant behavioral cues for making accurate judgments would constitute "good information" for making an accurate rating. Therefore, when examining the effect of good information a researcher can only look at *relevancy* as *availability* is more difficult to manipulate. Connelly & Ones (2010) define relevancy as an environment that allows a target to express traits. When operationalized this way, it is hard to argue that researchers have adequately manipulated this component of the model. If the behavioral information available to the rater is not relevant to one of the Big Five personality traits then it should not be thought of as high quality information for that trait. Three past papers have elaborated on the moderating effect of

information quality on accuracy. The first was a theoretical piece by Snyder and Ickes (1985) in which they contended that observations of individuals in unstructured situations were more likely to produce accurate portrayals of the target's personality than structured situations. Letzring & Wells (2005) followed similar logic in their empirical investigation of the effects situational strength on accurate judgment. Situation strength pertains to the idea that any social setting contains either implicit cues or explicit rules that influence the degree to which that environment dictates behavior (Hatrup & Jackson, 1996; Meyer, Dalal & Hermida, 2010; Mischel, 1973; Snyder & Ickes, 1985). While Letzring and Wells (2005) did find support for the effect of situational strength on accuracy, they admitted to certain limitations of their research. Specifically, they found that individuals in a "get-to-know you" interaction better predicted accuracy criteria (i.e., the average between the personality rating of the target, a familiar other and a trained clinician), than individuals in a problem solving interaction. While one can successfully argue that better quality information should come from a less formal interaction, you cannot be certain you will receive quality behavioral information for each of the Big Five dimensions. That is, we know such informal interactions may produce quality information for Extraversion, as it is the most visible trait in such interpersonal interactions, but we cannot be certain that good information for any of the other Big Five dimensions will surface as the TAP of the situation was not directly manipulated. In addition, Letzring and Wells (2005) focused solely on an average of personality assessments (i.e., ultimate accuracy criterion) as their dependent variable. This means that a new acquaintance or stranger's third-party assessment of a target would be considered accurate if it is correlated with the average of the target's self-assessment, a familiar other's assessment and a clinician's assessment of that target. What is missing from this approach is the incorporation of actual behavior; something of great interest to applied IO Psychologists. Connelly & Ones (2010) also studied the effect of information quality on accuracy in their meta-analysis. They operationalized information quality using Starzyk, Holden, Fabrigar and MacDonald's (2006) six dimensions of acquaintanceship (i.e., duration, frequency of interaction, knowledge of goals, physical intimacy, self-disclosure and social network familiarity). They reduced these six dimensions to, frequency of interaction and interpersonal intimacy and determined that better

quality information is available to observers when they have both (i.e., family and friends). That is, high quality information is available to individuals who have interacted frequently and have a close personal relationship with a target. It is important to note, that this quality information is not necessarily available to the friend or family member in the form of observation. While this may be true, additionally, by knowing the target intimately you also are likely to have deep knowledge regarding how the target views themselves. That is, a target is more likely to open up to an intimate other about their thoughts, feelings, intentions and beliefs than they will a stranger or new acquaintance. This would provide the intimate other with information about the target that not only, is not available in their behavior (e.g., thoughts and feelings regarding culture) but information that may contradict how they behave in certain situations (e.g., assertiveness at work may contradict their true level of agreeableness). While it is difficult to argue that such close acquaintances would not have an abundance of quality information about a target, one could also argue that their affection and knowledge about how the target views themselves may result in bias similar to that found in self-ratings (i.e., self-report bias). As evidence of this, research has shown, that of the observers examined (e.g., strangers, incidental acquaintances, cohabitators, spouses, family, and friends) those with more intimate relationships with the target demonstrated greater overlap with the target's self-report than those with less intimate relationships (Connelly Ones, 2010).

Depth and Breadth

Unique to the current study is a more in-depth investigation of the notion that intimate others have better quality information regarding a target than non-intimate others. As stated, Connelly and Ones (2010) concluded that two features of acquaintanceship that improve the accuracy of a third-party assessment are frequency of interaction and interpersonal intimacy. It is difficult to argue the benefits of interpersonal intimacy especially for traits low in visibility (i.e., Neuroticism and Openness to Experience). Indeed, individuals described as having a close interpersonal relationship with the target maintain higher self-other overlap than those that do not (Connelly & Ones, 2010). Additionally, it has been found that these individuals predict behavior at least as well as the self (Connelly & Hulsheger,

2012; Connelly & Ones, 2010; Mount, et al., 1994). However, frequency of interaction may matter less in improving third-party accuracy. Remember, research with newly acquainted individuals has failed to demonstrate that more information in the form of increased observation results in greater accuracy (Kenny, Albright, Malloy & Kashy, 1994). Interestingly, several studies have found that coworkers better predicted performance than did intimate others (Connelly & Ones, 2010; Mount, et al., 1994). The explanation for this finding was that coworkers had better insight regarding the context where the behavioral information was assessed (i.e., at work). That is, while our close friends, family and significant others may know us well, they may not have seen us behave at work. Therefore, their assessments of our personalities are less likely to predict behavior in this context relative to someone who works with us every day. As a follow-up to this notion, Connelly and Hulsheger (2012) looked at the differences in predictability between intimate others and coworkers. They referred to this study as the “clearer lens” versus the “narrower perspective”. They found that individuals with a “clearer lens” or those maintaining a close interpersonal relationship with the target, better predicted behavior than those with a narrower perspective. These mixed findings leave one important question that needs to be answered. If frequency of observation and contextual specificity matter less than intimacy in predicting behavior using third-party assessments, then what is it about the intimate relationship that drives this effect? To answer this question I identified two aspects of a close personal relationship that should conceptually drive increased accuracy or predictability: *depth* of the relationship and *breadth* of behavioral observation. Depth is essentially equivalent to interpersonal intimacy but will be operationalized differently than in past research (*see methods*). Essentially, this variable constitutes knowing the target well. Breadth considers the various situations you have observed the target behave in. The notion of breadth in the current article is almost synonymous with the idea of *situational bandwidth* (Speer, Christiansen, Goffin & Goff, 2014). Speer and colleagues (2014) defined *situational bandwidth* as, “the impact of situational demands on the breadth of construct measurement for assessments that rely on behavioral observations” (p. 283). Their notion of situational bandwidth comes from assessment center literature and the idea that a more comprehensive assessment of an individual’s performance across behavioral dimensions is possible to the

degree that behavior was assessed under dissimilar circumstances. In the current paper, I am applying this concept to understanding which third-party raters should have the highest quality information for which to provide ratings of a target's personality. This is a departure from the frequency of observation operationalization as it is possible that someone has observed years of a target's behavior but only in a single context (e.g., work, school, home). Instead, it is thought that the greater breadth of contexts you have observed a target in, the better your assessment of their personality will be able to predict their behavior in any single context. In addition, Speer and colleagues (2014) make a point worth noting regarding breadth of observations when rating a psychological construct that pertains to measurement. They state, "When assessment situations are similar, observations do not measure the underlying theoretical construct as well because they amount to a less representative sampling of entire behavioral domain." (p. 283). Thus, the more diverse the situations are for observation, the more likely you will have opportunities to observe *relevant* information for each of the Big Five traits. This would help explain why individuals with the "clearer lens" better predicted the performance of targets in Connelly and Hulsheger's (2012) study. In the following section leading to the specific hypotheses,

Hypotheses

The following hypotheses will extend past findings regarding third-party personality accuracy by testing the combined effects of two facets of relationship familiarity on self-other correlations, discriminant validity, criterion validity and incremental validity of third-party personality assessments. I see the current study contributing to the literature on third-party personality accuracy and the use of personality assessment in personnel selection in several ways. First, I will statistically parse relationship familiarity into the *depth* of the relationship and the *breadth* of behavioral observations in order to determine how each is related to accuracy criteria differentially. Second, I will test the effects of depth and breadth on the criterion-related validity of third-party personality ratings using GPA as the dependent variable in the model. Finally, I will investigate the roles of breadth and depth in determining the

incremental prediction afforded by third-party ratings beyond self-report assessments of a target's personality.

Past research has firmly established that the third-party personality rating of a target from a familiar other is more accurate than that of acquaintances or strangers. (Connelly & Ones, 2010; Connelly & Hulsheger, 2012). Most recently, Connelly & Hulsheger (2012) expanded our knowledge in this area when they compared the predictive validity of personal acquaintances and coworkers reports of a target's personality. Interestingly, they found that personal acquaintance's (i.e., those with a "clearer lens") reports of a target's personality better predicted job performance than coworkers (i.e., those with a narrower perspective). However, these same researchers called for more "concrete" phrasing regarding the operationalization of the information available to acquaintances/strangers and more familiar others. That is, relationship labels are vague and may be misleading to interpret. As social scientists it is more precise to know, what is it about these relationships that result in improved accuracy? Specifically, I looked at the third-party rating of a target's personality from acquaintances, coworkers, roommates, close-friends and significant others as has been done before. However, one unique feature of the current study is my attempt to expand past work by uncovering the important aspects of the "clearer lens"? That is, I want to parse the operationalization of the personal acquaintance or intimate other in an effort to identify the best acquaintance for personality information in a selection context and to understand theoretically why they demonstrate superior prediction. Thus, what aspects of this relationship drive the increased accuracy or predictability? Is it the close interpersonal intimacy (i.e., *depth* of the relationship) that results in knowing someone better, as suggested by Connelly and Ones (2010) and Connelly and colleagues (2007)? As stated previously, they likened interpersonal intimacy to the *information quality* moderator from Funder's (1995) RADU model. They contended that individuals maintaining a close personal relationship with a target are in a position to have special knowledge regarding the targets personality, especially for low visibility traits (i.e., Neuroticism and Openness) and this would result in higher accuracy when rating those traits. On the other hand, intimate others may also have had a greater

opportunity for cross-contextual observation (or *breadth* of observations), as Funder's (1995) RADU model also addressed, and this could contribute to increased accuracy by making *relevant* information *available* to third-party raters for observation. It is also possible that some combination of the two results in increased accuracy or predictability. That is, *depth* of the relationship and *breadth* of observation may interact in predicting accuracy criteria. In order to investigate these contentions, I collected information from informants regarding the number of different contexts (e.g., Interacting with superiors? In a social gathering? Interacting with family?) they have observed the target in (i.e., breadth of observation) and by asking them about their length of acquaintance and their knowledge of personal details about the target (e.g., When was the target born? What is the target's birthday? What is the target's major in school). The later feature regarding the relationship between the third-party rater and the target will be referred to as "*depth*" throughout the rest of the paper, while the prior will be referred to as "*breadth*".

Self-Other Overlap

As discussed earlier in this paper, one of the first approaches to assessing the accuracy of third-party ratings of personality is to examine self-other agreement. Self-other agreement refers to the overlap of personality reports by the self and by others (Funder & West, 1993). This constructivist¹ approach is a form of consensus² that places special weight on the perspective of the self. Issues regarding self-reported personality are well documented (Morgeson et al., 2007; Murphy & Dzieweczynski, 2005; Ones et al., 2007). However, Kenny's (1991) WAM model considers the self as a source of personality information that has experienced more individual behavioral incidences than any other source. This vast number of self-observations across time and settings provides the self with enough information to make a generally accurate rating relative to others. This logic has driven the abundance of applied use and research using self-report personality assessments. In addition, the self is in a unique position (as a source of personality information) to possess knowledge about their personality that is not readily available to outside

¹ Personality judgments are viewed as social constructions where accuracy is assessed by consensus between two people

² The extent to which judges agree in their ratings of a common target (Kenny, et al., 1994)

observers (e.g., thoughts, feelings, beliefs, and motivations) (Funder, 1995). As such, self-other agreement is considered one index for evaluating the accuracy of other-rated personality.

Two meta-analyses investigating self-other overlap found that the correlation between self and other-ratings of personality was greater when the rater and the target had higher acquaintanceship and greater intimacy (Connelly, Kavanagh & Viswesvaran, 2007; Connelly & Ones, 2010). That is, when a target and the person rating their personality have an existing, long standing close relationship (e.g., significant others, friends, and family) they will have higher self-other agreement than a dyad with less tenure and intimacy (e.g., acquaintances, classmates, coworkers). Again, Kenny's (1991) WAM helps explain why intimate others appear more accurate using self-other agreement as an index of accuracy. First, intimate others have simply had more opportunity to observe the target's behavior. Like the target, they have observed more behavioral instances than someone less intimate with the target. Second, Connelly & Ones (2010) contended that these individuals have special knowledge of target's trait information not readily available (i.e., Emotional Stability and Openness to Experience) to newly acquainted observers. Specifically, by sharing an intimate relationship with the target they have knowledge of the target that is not externally expressed in behavior (e.g., preference for the arts, level of concern for themselves and others). They note that it is not simply the sheer quantity of the observations a rater has made of a target, but also the quality or intimacy of the relationship that results in higher self-other correlations. Funder's (1995) RADU model is consistent with the findings of Connelly & Ones (2010). Specifically, intimate others possess knowledge of the target's personality not readily *available* (i.e., Neuroticism and Openness) in their outward behavior.

The first question I hope to answer is what impact *depth* and *breadth* have on self-other overlap? Connelly and Ones (2010) contend that while most informants will have significant overlap with the target's Extraversion (visible trait) similar levels of overlap with less visible traits (i.e., Openness, Neuroticism) are only possible when the dyad shares a close personal relationship. Therefore, depth rather

than breadth should be most impactful with respect to self-other overlap on ratings of low visibility traits, I hypothesize that:

H1: The relationship between the third-party personality ratings of a target on low visibility traits (i.e., Openness, Neuroticism & Conscientiousness) and the self-reported personality ratings on those same traits will be significantly stronger when the judge and the target possess greater *depth* of relationship with one another than when they possess lesser relationship *depth*.

While depth of the relationship seems necessary for dramatic gains in overlap between the other and self on low visibility traits, such deep intimacy may also contaminate the third-party rating of highly evaluative traits (i.e., Agreeableness). That is, because Agreeableness is often equated to “likability,” and targets are likely to behave in a more agreeable manner towards individuals with whom they share a deeply intimate relationship. Thus, if a third-party rater bases his/her rating of agreeableness purely based on a close personal one-on-one relationship he/she is likely to rate this trait higher than the self-rating which would be based on one’s knowledge of their behavior in a variety of different settings and relationships. If this is true then increases in overlap between the other rater and target’s self-report of Agreeableness can only occur if the third-party rater has had the opportunity to observe the target in situations that might elicit variable levels of agreeable behavior. Thus, relationship *breadth* may strengthen self-other overlap with respect to ratings of highly evaluative traits such as Agreeableness and to a lesser extent on Conscientiousness. As such I hypothesize that:

H2: The relationship between third-party personality ratings of highly evaluative traits and self-reported personality rating on those same traits (i.e., Agreeableness & Conscientiousness) will be significantly stronger when the judge reports greater *breadth* of observation of the target than when they report lesser *breadth* of observation of the target.

As reviewed earlier, self-other agreement is only one of several approaches to understanding the accuracy of third-party ratings of personality. Next, I will shift focus toward the discriminant validity of

third-party rater's assessment of a target's personality to understand which third-party raters discriminate better and most importantly why?

Discriminant Validity

While Beer and Watson (2008) found that third-party raters differentiate less between dimensions of personality than the self, we do not know exactly which third-party raters (i.e., intimate others, colleagues or acquaintances) are most guilty of this. If intimate others possess greater overlap with the self-report than those less intimate with the target then one could contend that intimate third-party raters should differentiate better than less intimate others. If we are to look at discriminant validity as an additional index of accuracy (i.e., better discrimination between dimensions of personality results in less error and therefore more accuracy), then intimate others would be considered more accurate than less intimate others by virtue of having more *relevant* information *available* about the target across dimensions of personality. Essentially, those with the “clearest lens” will differentiate better between dimensions of personality than those with a “narrower perspective”. However, what aspect of the “clear lens” (i.e., depth or breadth) produces this effect. When I consider trait visibility and evaluativeness the way in which the two variables moderate the impact of discriminant validity will be more nuanced. Accuracy in rating low visibility traits should not be helped by *breadth* of observation because much of these traits are internal to targets and not readily *available* in their outward behavior. Conversely, these internal aspects of Openness and Neuroticism will likely be revealed to individuals with a close interpersonal relationship with the target (i.e., greater depth). Specifically, the informant will demonstrate greater discriminant validity (i.e., lower average correlation between Big Five traits) in their rating of the target across the Big Five when they report greater depth of relationship with the target. On the other hand, individuals lacking a deep relationship with the target will have little to go by when rating less visible traits. As a result, they are likely to base their ratings of those traits on information they do have available to them (e.g., outgoingness, vibrancy, gregariousness, other superficial features). That is, less intimate others are likely to fill in the gaps where outward behavior is not available and base their rating

of less visible traits on their perception of the targets extraversion, as this is the most visible trait. The result should be greater halo error in the ratings made by less intimate others.

As a result of this I hypothesize the following:

H3: The discriminant validity (i.e., average difference between Big Five ratings) of third-party assessments of a target's personality will be greater for third-party raters with greater *depth* of the relationship to the target than for those with a lesser depth of relationship.

Criterion Related Validity

In the most recent work regarding relationships between third-party assessments of personality and job performance, both variables were collected from the same source (Hulsheger & Connelly, 2012). In the current study we will rely on academic achievement in the form of GPA as our dependent variable. Past research has demonstrated significant prediction of GPA by Conscientiousness (Hough, 1992; Komarraju, Karau & Schmeck, 2009; Laidra, Pullman & Allik, 2007; Musgrave-Marquart, Bromley & Mahlon, 1997; Nofle & Robins, 2007; O'Conner & Paunonen, 2007; Poropat, 2009), Neuroticism (Laidra, Pullman & Allik, 2007; Komarraju, Karau & Schmeck, 2009; Musgrave-Marquart, Bromley & Mahlon, 1997;), Agreeableness, (Komarraju, Karau & Schmeck, 2009; Laidra, Pullman & Allik, 2007; Musgrave-Marquart, Bromley & Mahlon, 1997; Poropat, 2009), and Openness (Komarraju, Karau & Schmeck, 2009; Laidra, Pullman & Allik, 2007; Musgrave-Marquart, Bromley & Mahlon, 1997; Poropat, 2009). My next sets of hypotheses seek to investigate the degree with which "clearer lens" facets (i.e., *depth* and *breadth*) moderate the predictive validity of third-party ratings of personality with respect to aggregated multi-dimensional behavioral criteria in the form of academic achievement. These hypotheses are based on the notion that ratings of highly evaluative personality traits should be better predictors of performance when they are obtained from those who have had the opportunity to observe a target in a greater *breadth* of situations, whereas ratings of low visibility traits should better predict academic

achievement when they are obtained from those who have a greater depth of relationship with a target.

As a result my hypotheses are as follows:

H4: The relationship between third-party ratings of low visibility traits and performance (GPA) will be stronger when the third-party rater has a greater *depth* of relationship with the target than they will be for third-party raters with a lesser *depth* of relationship to the target.

H5: The relationship between third-party ratings of highly evaluative traits and performance (GPA) will be stronger when the third-party rater has had a greater *breadth* of observation of the target than when they have had a lesser breadth of observations.

It is exciting to gain a better understanding as to how and why certain third-party raters make excellent sources of personality information for targets in a selection context but it is unlikely that organizations will forgo self-report assessments in response to this research. Instead it is probable that businesses and other entities using personality will want to use such assessments in combination with self-assessments. Such an approach enables you to cross-validate self-assessments and provides you with more information to make an informed decision. Before I conclude this study I felt it was important to understand how different third-party assessments add to the prediction of performance beyond self-assessments. That is, which third-party assessments add the most incremental variance in predicting criteria?

Incremental Validity

In a selection environment, having too much overlap between the self and third-party ratings may offer little in terms of incremental validity for predicting performance. Specifically, if the two ratings are highly correlated then the third-party rating will offer little unique information about the target that would add to predicting their performance beyond the self-rating in a regression analysis.

Further complicating things is the notion that highly correlated assessments from an intimate dyad may contain similar types of bias. Precisely, if the intimate other is a significant other, family

member or simply someone reporting the greatest relationship *depth*, their view of the target may possess similar bias (i.e., self-presentation effects) as the target themselves. These individuals may see the target through a brightly colored lens in much the same way as the target views themselves. Connelly and Ones (2010) touch upon this issue when they discuss the distorting effects of “friendship bias” (p. 1111). In their view, these individuals may be less willing to disclose potentially negative information about the target. Additionally, Kenny (2004) contends that familiarity with the target or length of acquaintanceship as it is referred to in the literature, may not be as important as other factors (e.g., breadth of observations) when it comes to making an accurate assessment of another’s personality. If ultimately we want to use third-party assessments in conjunction with self-assessments to predict performance criteria, then we want to make sure to select third-party raters that offer the most incremental validity above and beyond the self-report. Past research has demonstrated that familiarity and/or interpersonal intimacy, strengthen the relationship between self-reports and third-party ratings of personality and this is especially true for low visibility traits (Connelly & Ones, 2010; Connelly et al., 2007). Interestingly, these same studies did not show increased overlap for highly familiar others on highly evaluative traits (i.e., Agreeableness). As a result I expect people who report having a greater *depth* of relationship with the target to offer little unique information about them that the target has not already provided. Conversely, people who report having observed the target in more varying contexts (i.e., breadth) should have more unique information about the target omitted from the target’s self-report. An excellent explanation of this contention is offered in Connelly and Ones’s (2010) citation of Hogan’s socioanalytic theory of personality. This theory contends that personality can be seen as internally held thoughts, feelings, and beliefs, or it can be thought of as our outwardly expressed reputation. Clearly most criteria (i.e., academic performance) are more consistent with the later and third-party raters who have observed the target in multiple contexts have a better perspective on reputation than the self or a close confidant. As a result my final hypothesis states:

H6: The incremental validity, above and beyond self-report assessments, of third-party raters will be significantly greater when the rater reports greater *breadth* when compared to raters who report less *breadth*.

The preceding hypotheses were intended to elaborate on past findings regarding third-party personality accuracy by testing the moderating effect and main effect of two relationship familiarity facets on self-other correlations, discriminant validity, criterion validity and incremental validity of third-party personality assessments. The current study contributes to the literature on third-party personality accuracy and the use of personality assessment in personnel selection in several ways. First, I answered the call for more research on third-party personality assessment in the field of IO Psychology. This study adds further validation to the notion that self-assessments alone greatly underestimate the predictive power of personality in industry. In addition, this study helps applied IO Psychologists identify the type of individuals who would be excellent sources of information regarding a target's personality. Second, I will statistically parse relationship familiarity into the *depth* of the relationship and the *breadth* of behavioral observations in order to determine how each is related to accuracy criteria. This answers Connelly and Ones's (2010) call for a more "concrete" operationalization of relational features and helps explain why those with a "clearer lens" (Connelly & Hulsheger, 2012, p.11) maintain the highest level of personality assessment accuracy. Third, I will test the effects of *depth* and *breadth* on the criterion-related validity in a controlled/simulated environment. This is the first study that I know of that investigated the validity of third-party assessments on performance in a lab setting. Finally, I expanded on our understanding of how relational variables differentially result in incremental prediction beyond the self-report assessment of personality. This last point is important as one might assume (i.e., based on the conclusions from Connelly & Hulsheger, 2012) that individuals with the closest relationship with the target will provide the most accurate information. While this conclusion may be true, the absence of novel information would reduce the propensity of their assessment to add incrementally to that of the target. However, those who

have observed the target across a variety of situations should add the most incremental validity to self-assessments.

METHODS

Participants

This study was conducted with 626 dyads. The targets within these dyads were undergraduate students recruited from a large southeastern university to participate in exchange for course credit. Third party raters within the dyad were undergraduates or friends and family of the targets outside of the University setting. 70% of participants reported being female. 62% reported being white, 20% Hispanic, and 14% black. Participants' ages ranged from 18 to 59 years, with an average age of $m = 20$ years.

Procedures

Target participants who received experimental credit for their participation were asked to bring a partner with them to the study that they had at least some familiarity with (i.e., family member, roommate, significant other, classmate, acquaintance, and friend). After arrival both participants completed a general demographic survey, a survey assessing the depth and breadth of their relationship and self and third-party personality measures.

Measures

Personality. Participants completed the McCrae, Costa, and Martin (2005) NEO-PI-3 for both themselves and their partner. Items were measured using a 5-point Likert scale that ranged from 1 (Strongly Disagree) to 5 (Strongly Agree). Internal consistency for the 12 items representing agreeableness was estimated using *coefficient alpha* ($\alpha = .74$), for neuroticism ($\alpha = .82$), for extraversion ($\alpha = .80$), for openness ($\alpha = .78$) and for conscientiousness ($\alpha = .83$) Responses to these items were averaged to use for the final analyses.

Scholastic Achievement Test (SAT) scores. Participants were asked to report their scores on the SAT exam if they had taken the exam. Scores on this variable ranged from 840 to 2,250 with a mean score of 1,288 ($SD = 293$).

Cumulative GPA. Participants were asked to report their cumulative college grade point average (GPA). This variable ranged from 1.13 to 4.0, with a mean of 3.45 ($SD = 0.50$).

Length of relationship. Both the target participants and their third party raters were asked to report the number of months they had been acquainted. In terms of how long third-party raters knew the target the range was from 0-342 months with $m = 9$ months. Fifty-eight percent of the sample reported they have interacted with the target everyday over the last six months.

Depth. To compute the depth of the dyadic relationship I had the third-party rater answer seven personal questions about the target. These responses were then scored against the targets' responses for accuracy. Depth scores reflected the number of items the third party rater correctly answered (0-7). Higher scores are an indication of greater personal closeness with the target. The mean of relationship depth in the current sample is ($m = 4.3, SD = 1.7$). The questions used to assess depth of relationship are as follows:

1. What is the targets middle name?
2. What is the target's major?
3. What is the target's year in school?
4. What is the target's place of birth?
5. What is the target's job status
6. What is the target's date of birth?
7. What is the target's favorite TV show?

Breadth. To compute breadth of observations of the targets, third party raters were asked a series of questions, in the form of check boxes, regarding the various contexts they have observed the target in (e.g., Have you observed the target interacting with.... supervisors, professors, students, friends, in a

social group, family and you one-on-one?). Participants were asked to check any box that applies. These items were scored dichotomously. That is, a 1 indicated they have observed them and a 0 they have not. There were a total of seven contexts for which they could have observed the target. Therefore, the range for breadth was from 1-7. 1 indicating only one context and 7 indicating observations in every context. The following are the questions asked about the contexts third-party raters have observed the target in:

1. How often have you observed the target interacting with coworkers?
2. How often have you observed the target interacting with authority figures?
3. How often have you observed the target interacting with instructors?
4. How often have you observed the target interacting with students?
5. How often have you observed the target interacting with you one-on-one?
6. How often have you observed the target interacting in a group social setting?
7. How often have you observed the target interacting with family or a significant other?
8. How often have you observed the target interacting with strangers?

Their responses ranged from, 1 = never to 4 = more times than I can count, on these questions. A response of 2 implied they had only observed the target in the specified context once. To compute the final breadth variable responses of 1 or 2 on the above items received a 0 and responses of 3 or 4 received a 1. It was decided that a single observation was not sufficient to receive credit for an observation. Hence, third-party raters needed to indicate they had observed the target in a context more than once for credit on any particular observation. The mean across all third-party raters on breadth of observations was ($m = 5.2$, $SD = 1.73$).

Context consistency. Context consistency was computed to control for third-party raters that had viewed the target in the same context as the performance criteria (i.e., interacting with students).

Responses range from 1 = never observed them interact with students to 4 = observed them interacting with students more times than I can count. A response of 2 indicated that they had only observed them in this context once. Third-party raters answering 3 or 4 to this question received a 1 and a 0 for responses 1 or 2. 72% of respondents reported a 3 or a 4 on this variable.

RESULTS

Descriptive Statistics and Intercorrelations

Table 1 shows the descriptive statistics and intercorrelations for all study variables. Relationship Depth was significantly positively related to self-other overlap for Neuroticism ($r = .083, p = .050$), Extraversion ($r = .134, p = .002$), Openness ($r = .192, p = .000$), Agreeableness ($r = .138, p = .001$), Conscientiousness ($r = .177, p = .000$) and the length of relationship between the dyad ($r = .321, p = .000$). Observational Breadth demonstrated significantly positive relationships with Openness ($r = .123, p = .003$), Conscientiousness ($r = .085, p = .044$) and length of relationship between dyads ($r = .290, p = .000$). The two variables of focus in this study (i.e., Breadth and Depth) are related to length of relationship ($r = .29, p = .000, r = .32, p = .000$) but not so highly that they do not offer something unique in the prediction of accuracy criteria.

Curiously, the gender of the target and the informant were significantly related to a number of study variables. Specifically, the gender of the target was significantly related to self-other overlap for Neuroticism ($r = .183, p = .000$), Agreeableness ($r = .112, p = .007$), Relationship Depth ($r = .151, p = .000$) and Observational Breadth ($r = .088, p = .030$). These results indicate that when the target is a woman higher self-other overlap between the target and third-party rater occur for Neuroticism and Agreeableness and the third-party rater reports higher levels of depth and breadth. The gender of the third-party rater was related to the average discriminant validity of their Big Five ratings of the target ($r = .096, p = .020$) indicating that women maintain higher discriminant scores across the Big Five traits when rating a target compared to men. The third-party rater's gender was also related to self-other overlap for Neuroticism ($r = .110, p = .008$). This implies that when the third-party rater is a woman, her rating of the target is more aligned with the target's self-rating on Neuroticism. While gender was not an initial consideration for the effects of depth and breadth on accuracy criteria these preliminary results demonstrated that gender should be an important control variable for subsequent regression analyses. These preliminary findings for the target's gender are consistent with past meta-analytic research

demonstrating that women are higher than men on Extraversion, Anxiety, Trust, and “especially” tender-mindedness (nurturance) (Fiengold, 1994, p. 445). In addition, the findings for the third-party rater’s gender are at least somewhat consistent with research demonstrating women’s superiority in accurately assessing non-verbal emotional cues (Thayer & Helge Johnson, 2001).

Table 1.
Descriptive statistics and intercorrelations of study variables.

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. S-O Overlap (N)	0.21	0.31	1																							
2. S-O Overlap (E)	0.27	0.3	-.009	1																						
3. S-O Overlap (O)	0.38	0.29	.031	.048	1																					
4. S-O Overlap (A)	0.38	0.3	.049	.112**	.081	1																				
5. S-O Overlap (C)	0.32	0.32	.038	.092*	.033	.160**	1																			
6. Self-Report (N)	2.61	0.646	.115**	.017	-.055	-.021	.035	1																		
7. Self-Report (E)	3.71	0.536	-.046	-.035	-.055	.059	.017	-.283**	1																	
8. Self-Report (O)	3.41	0.505	.016	-.047	-.032	.062	.026	.057	.146**	1																
9. Self-Report (A)	3.73	0.512	.050	.053	.049	-.036	.046	-.170**	.328**	.064	1															
10. Self-Report (C)	3.89	0.571	.117**	-.027	.021	.130**	-.101*	-.228**	.299**	.093*	.268**	1														
11. T-P Report (N)	2.53	0.647	.091*	-.027	-.059	.032	.011	.311**	-.066	.055	-.042	.024	1													
12. T-P Report (E)	3.63	0.577	.035	-.006	.037	.059	.111**	-.177**	.454**	-.013	.145**	.057	-.241**	1												
13. T-P Report (O)	3.25	0.48	.003	-.028	.055	.034	.110**	.021	-.057	.411**	-.011	-.138**	.019	.054	1											
14. T-P Report (A)	3.67	0.605	.091*	.038	.084*	-.016	.050	.034	.075	-.025	.324**	-.021	-.247**	.302**	.087*	1										
15. T-P Report (C)	3.94	0.622	.153**	-.004	.153**	.148**	.015	-.026	-.035	.011	.111**	.363**	-.175**	.172**	.103*	.298**	1									
16. Relationship Depth	4.34	1.71	.083*	.134**	.192**	.138**	.177**	.034	-.049	.026	.013	.020	.155**	.020	.090*	.006	.094*	1								
17. Observational Breadth	5.27	1.73	.039	-.014	.123**	.069	.085*	.058	.013	.078	-.029	.018	.068	.081*	.132**	-.010	.072	.418**	1							
18. T-P Discriminant Validity	0.89	0.33	-.009	.059	.137**	.053	.010	-.182**	-.006	-.027	.083*	.129**	-.652**	.241**	-.032	.347**	.507**	.007	.062	1						
19. Relationship Length	34.54	52.94	.037	.001	.111**	-.019	.074	.057	-.072	-.026	.021	-.030	.027	-.094*	-.041	-.016	.033	.321**	.290**	.041	1					
20. Target Gender	—	—	.183**	.053	-.040	.112**	.061	.175**	.133**	.011	.216**	.178**	.234**	.094*	.017	.098*	.125**	.151**	.088*	-.062	-.034	1				
21. T-P Gender	—	—	.087*	.011	.019	.019	.065	-.010	.022	-.048	-.005	.030	-.069	.055	-.009	.030	.092*	-.023	.009	.113**	-.015	.097*	1			
22. Student Context	0.74	0.438	-.027	-.027	.018	-.031	.025	.028	.017	.006	-.027	.023	-.024	.033	-.002	-.008	-.009	.115**	.498**	.025	.092*	.022	.017	1		
23. GPA	3.45	0.453	-.038	.039	.023	-.007	.029	-.027	.078	.040	.036	.264**	.002	.085	-.026	.022	.159**	.029	-.056	-.004	-.006	.086	-.017	-.060	1	
24. SAT	1308	306	-.078	.032	-.013	-.035	-.010	-.008	-.095	.020	-.132*	-.163**	.011	-.059	.054	-.048	-.139*	.023	.018	-.004	-.039	-.114*	.053	.061	.007	1

Note. Descriptive statistics are based on pairwise deletion of missing cases (Ns range from 526-623. **Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Regression Results for Self-Other Overlap

I tested the effects of relationship depth and observational breadth on self-other overlap for hypotheses 1 and 2 by generating variables that represented the correlation or overlap between the third-party rating and the target's self-rating for each of the Big Five variables. The abridged NEO used for this study relied on twelve items to measure each Big Five dimension. Bivariate correlation coefficients were generated by correlating the target's self-rating across those twelve items with the third-party rater's rating for the twelve items. This was done for each participant in the sample for each of the Big Five traits. The resulting correlation coefficient was used as the dependent variable in the subsequent regression analyses and represented the overlap between the target and the third-party rater. The predictor variables were third-party gender, target gender, relationship depth, observational breadth and length of relationship. That is, to test the proceeding two hypotheses, five regression analyses were run where self-other overlap for each of the Big Five represented the dependent variable and third-party gender, target gender, relationship depth, observational breadth and length of relationship were the predictors. Gender and length of relationship were considered control variables and depth and breadth were my variables of interest. Considering the overlap between these two variables ($r = .42, p = .000$), it was necessary to include them in a single regression model.

Hypothesis 1 stated: The relationship between the third-party personality ratings of a target on low visibility traits (i.e., Openness & Neuroticism) and the self-reported personality ratings on those same traits will be significantly stronger when the judge and the target possess greater *depth* of relationship with one another than when they possess lesser relationship *depth*. Results from the regression analyses found that relationship depth was not a significant predictor of self-other overlap for Neuroticism ($\beta = .054, t = 1.106, p = .269$), however, the model was significant ($r^2 = .038, p = .001$). Concerning the overlap for Openness to experience (other low visibility trait) depth was significant ($\beta = .166, t = 3.446, p = .001$) as was the model ($r^2 = .046, p = .000$). For self-other overlap on Conscientiousness depth was again significant ($\beta = .165, t = 3.360, p = .001$) as was the model ($r^2 = .041, p = .000$). Regarding the

remaining two dimensions relationship depth demonstrated significant prediction for both Extraversion ($\beta = .179, t = 3.674, p = .001$), model significant ($r^2 = .026, p = .013$) and Agreeableness ($\beta = .134, t = 2.778, p = .006$), model significant ($r^2 = .031, p = .004$). The aforementioned results provide partial support for hypothesis 1 as depth was a significant predictor for Openness but not Neuroticism. Interestingly, depth also strengthened the relationship between the self and other report personality assessments for the remaining high visibility traits as well. The gender of the third-party rater was a significant predictor of self-other overlap for Conscientiousness ($\beta = .134, t = 2.778, p = .006$), indicating that when the third-party rater is a woman their rating of the target's conscientiousness is more consistent with the target compared to a third-party rating from a man.

Hypothesis 2 stated: The relationship between third-party personality ratings of highly evaluative traits and self-reported personality rating on those same traits (i.e., Agreeableness & Conscientiousness) will be significantly stronger when the judge reports greater *breadth* of observation of the target than when they report lesser *breadth* of observation of the target. Again the bivariate correlations between the target self-ratings and the third-party rater ratings across the twelve items used to measure each of the Big Five were used as the dependent variables in the proceeding regression analyses. Observational breadth was not a significant predictor of overlap between the self and third-party rater ratings of Agreeableness ($\beta = .033, t = .699, p = .485$), the model, however was significant ($r^2 = .031, p = .004$). Likewise, observational breadth was not a significant predictor of the overlap between self and third-party ratings of Conscientiousness ($\beta = .004, t = .078, p = .938$) but again the model was significant ($r^2 = .041, p = .000$). The only predictor that was significant in these two models was Relationship Depth. Interestingly, breadth of observation and length of relationship did not demonstrate significant prediction for overlap between the self and other for any of the remaining Big Five traits. These results appear to suggest the importance of relationship depth for accuracy regarding self-other overlap.

Regression Results for Discriminant Validity

In order to test hypothesis 3 I had to compute a variable that represented the overall discriminant validity of a third-party rater's Big Five assessment. To do this I first computed an absolute value score for each Big Five trait relative to the others (i.e., 10 discriminant scores). I then computed an average of those absolute values representing the average discriminant score of the third-party rater's ratings on the Big Five. This average was then used as the dependent variable for testing hypothesis 3.

Hypothesis 3 states: The overall discriminant validity (i.e., average difference between each Big Five rating) of third-party assessments of a target's personality will be greater for third-party raters with greater *depth* of the relationship to the target than for those with a lesser depth of relationship. To test this hypothesis, the average discriminant validity score for third-party ratings was regressed onto relationship depth, observational breadth, length of relationship, target gender and third-party rater gender. Results indicate that relationship depth is not a significant predictor of the average discriminant validity of third-party ratings of Big Five traits ($\beta = -.021, t = -.442, p = .659$), however, the model was significant ($r^2 = .027, p = .010$). This evidence disconfirms hypothesis 3. Neither observational breadth ($\beta = .074, t = 1.568, p = .118$) nor length of relationship ($\beta = .023, t = .509, p = .611$) demonstrated significant prediction of the average discriminant validity score for third-party ratings of Big Five traits. Nonetheless, the gender of the third-party rating was significantly related to discriminant validity across the Big Five ($\beta = .137, t = 3.221, p = .001$), model significant ($r^2 = .027, p = .010$). This is an indication that third-party raters that are women discriminate between Big Five traits better than men.

Table 2.

Regression Results for the Predictability of Breadth and Depth on Self-Other Overlap and Discriminant Validity.

Variable	Correlation Between Participant A and B					Average Discriminant Validity
	N	O	E	A	C	
Relationship Depth	.54	.17**	.18**	.13**	.17**	-.021
Observational Breadth	-.02	.05	-.08	.03	.01	.074
Third-party Gender	.06	.04	.02	.04	.08*	.14**
Target Gender	.16**	-.07	.02	.07	.03	-.07
Length of Relationship	.04	.04	-.04	-.07	.02	.028
Equation R ²	.04**	.05**	.03**	.03**	.04**	.01

Note. Table values are standardized beta weights.

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

Moderated Regression Results for Predictive Validity

Hypothesis 4 stated: The relationship between third-party ratings of low visibility traits and GPA will be stronger when the third-party rater has greater *depth* of relationship with the target than they will be for third-party raters with a lesser *depth* of relationship to the target. Hypothesis 5 stated: The relationship between third-party ratings of highly evaluative traits and GPA will be stronger when the third-party rater has had a greater *breadth* of observation of the target than when they have had a lesser breadth of observations. To test hypotheses 4 and 5 I followed the steps for interpreting interactions purported by Aiken & West (1991). That is, I created standardized values for all my predictor variables used in the subsequent regression analysis. After standardizing the predictors I created my product terms representing two-way interactions between relationship depth and third-party ratings of each big five trait, and two-way interactions between observational breadth and these same third-party ratings. I then regressed GPA on third-party ratings of the big five traits, relationship depth, relationship breadth, length of acquaintance, gender of the target, and the aforementioned product terms. At step 1, I regressed GPA onto the third-party ratings of the Big Five, breadth and depth and the control variables (context consistency, length of relationship, target's gender and target's SAT). At step 2, we evaluated the

incremental contribution of the 10 two-way interactions among the predictors. The resulting model used to test hypotheses 4 and 5 was significant ($r^2 = .172, p = .002$).

Table 3 shows the moderated regression results. The variable representing the interaction between relationship depth and third-party ratings of Neuroticism was not a significant predictor of GPA ($\beta = -.070, t = -.915, p = .361$). However, the variable representing the interaction between depth and third-party ratings of Openness was significant predictor of GPA ($\beta = .120, t = 1.680, p = .094$ – two-tailed). The plot of the interaction between third-party ratings of openness and depth (see Figure 1) indicated that when the third-party rater has higher depth their rating of target's openness better predicts GPA (positive direction). That is, higher openness is equated to higher GPA and lower openness is equated to lower GPA. As a result, my results partially supported the hypothesis that relationship depth would strengthen the predictive validity of third-party ratings of Big Five low visibility traits and GPA. The variable representing the interaction between observational breadth and third-party ratings of Agreeableness was a significant predictor of GPA ($\beta = .176, t = 2.171, p = .031$) while controlling for individuals who observed the target in an academic setting ($\beta = -.103, t = -1.456, p = .147$). The plot of the interaction (see Figure 2) indicates that when a third-party rater reports more breadth of observations their rating of the target's agreeableness predicts GPA in the expected direction (i.e., higher agreeableness is equated to higher GPA). However, third-party raters who are low in breadth predict GPA in an unexpected fashion. Specifically, when they are low in breadth and their rating of the target is high on agreeableness, the target maintains a higher GPA and when their rating is low on agreeableness the target maintains a lower GPA. As a result I cannot fully confirm my hypothesis that observational breadth will moderate the relationship between third-party ratings of highly evaluative traits and performance. Likewise, the variable representing the interaction between observational breadth and the third-party rating of Conscientiousness (other highly evaluative trait) was not significant ($\beta = -.007, t = -.101, p = .920$). Surprisingly, the variable representing the interaction between observational breadth and third-party ratings of Extraversion was a significant predictor of GPA ($\beta = .185, t = 2.557, p = .011$). The plot (see

Figure 3), indicates that when a third-party rater reports greater breadth of observations of the target, their ratings of the target's Extraversion better predicts the targets' GPA.

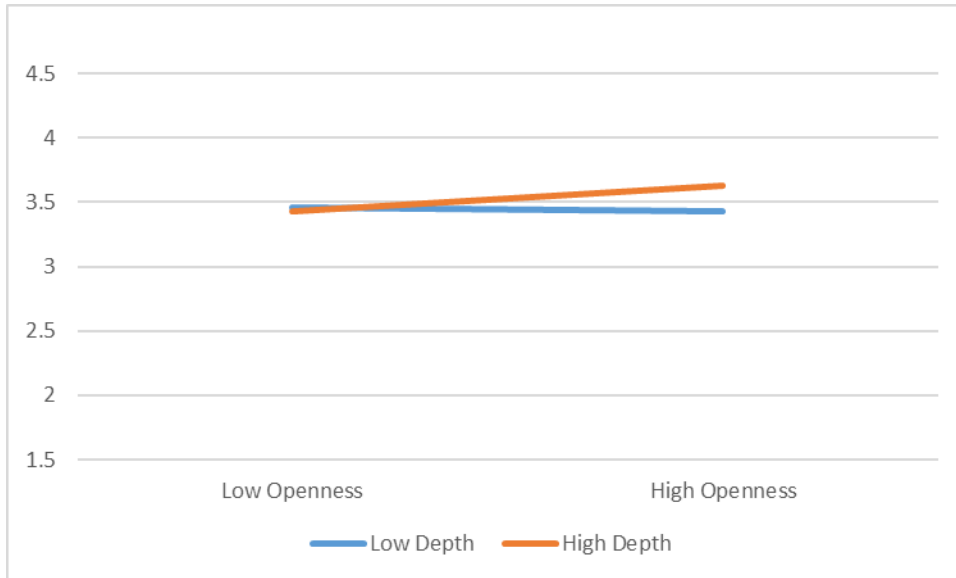


Figure 1.
Interaction between depth and third-party rated openness predicting GPA.

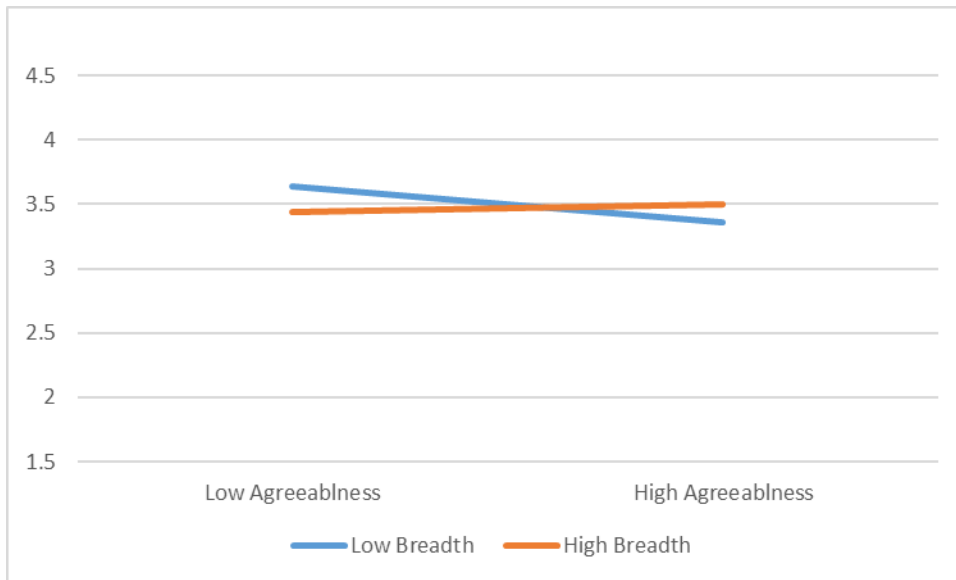


Figure 2.
Interaction between breadth and third-party rated agreeableness predicting GPA.

Table 3.

Moderated Regression Results for the Predictive Validity (GPA) of Third-party Ratings of Big Five Personality Traits and Their Interactive Effects with Breadth and Depth.

Variable	Step 1	Step 2
Third-Party (N)	-.007	-.038
Third-Party (E)	.100	.087
Third-Party (O)	-.045	-.047
Third-Party (A)	-.105	-.119
Third-Party (C)	.192**	.196**
Relationship Depth	.062	.087
Observational Breadth	-.046	-.027
Relationship Length	.062	-.020
SAT	.053	.076
Target Gender	.162*	.172**
Context Consistency	-.104	-.103
Third-Party (N) x Depth		-.070
Third-Party (E) x Depth		-.059
Third-Party (O) x Depth		.120*
Third-Party (A) x Depth		-.023
Third-Party (C) x Depth		.071
Third-Party (N) x Breadth		.086
Third-Party (E) x Breadth		.185**
Third-Party (O) x Breadth		-.056
Third-Party (A) x Breadth		.176*
Third-Party (C) x Breadth		-.007
Equation R ²	.097*	.172**

Note. Tabled values are standardized beta weights

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

Moderated Regression Results Incremental Validity

Hypothesis 6 stated: The incremental validity, above and beyond self-report assessments, of third-party ratings will be significantly greater when the rater reports greater *breadth* of observations when compared to raters who report less *breadth*. To test hypotheses 6, I again relied on the steps for interpreting interactions purported by Aiken & West (1991). That is, I created standardized values for all my predictor variables used in the subsequent regression analysis. After standardizing the predictors I created my product terms representing two-way interactions between relationship depth and third-party ratings of each big five trait, and two-way interactions between observational breadth and these same third-party ratings. I regressed GPA on the predictors in two steps. At step 1, I regressed GPA onto the third-party ratings of the Big Five, self-report ratings of the Big Five, breadth and depth and the control

variables (context consistency, length of relations, target's gender and target's SAT). At step 2, we evaluated the incremental contribution of the 10 two-way interactions between depth and the five trait ratings and between breadth and the five trait ratings.

The model regarding the predictors in step 1 was significant ($r^2 = .144, p = .002$). Likewise the model regarding the product terms was also significant ($r^2 = .206, p = .001$). The model in step 2 was used to test my last hypothesis. The variable representing the interaction between observational breadth and Agreeableness was significant ($\beta = .183, t = 2.237, p = .026$). Consistent with Hypothesis 6, when plotted (see Figure 4) this interaction indicated that when a third-party rater reports higher breadth of observations their rating of the target's agreeableness better predicts GPA in the expected direction. That is, higher ratings of agreeableness equate to higher ratings of target-GPA and lower ratings equate to lower GPA. Interestingly, the opposite was true for individuals low in breadth. Specifically, low ratings of agreeableness were equated to higher GPA and higher ratings of agreeableness were equated to lower GPA. However, the test for simple slopes of both lines was not significant (agreeableness at low breadth – $t = -.0946, p = .924$, agreeableness at high breadth – $t = .0273, p = .978$, breadth at low agreeableness – $t = -.0722, p = .942$, breadth at high agreeableness – $t = .0497, p = .960$). While these findings initially appeared to support my hypothesis regarding the incremental prediction of third-party ratings for highly evaluative traits when they report high levels of observational breadth, I cannot fully confirm that hypothesis as a result of the test for simple slopes. None of the variables representing the interaction between relationship depth and Big Five traits were significant (see

Table 4). To my surprise, the variable representing the interaction between observational breadth and third-party Extraversion was also significant ($\beta = .158, t = 2.141, p = .033$). As shown in Figure 5, when plotted this interaction reveals that third-party ratings of extraversion for raters high in breadth better predict GPA. Specifically, high ratings of extraversion equate to higher target GPA, while lower ratings equate to lower GPA. The opposite was true for individuals low in breadth and the slope of the line was more gradual. In sum, ratings obtained by individuals high in observational breadth (but not so for relationship depth) offer greater incremental prediction of GPA above and beyond self-report Big Five assessments, specifically for agreeableness and extraversion.

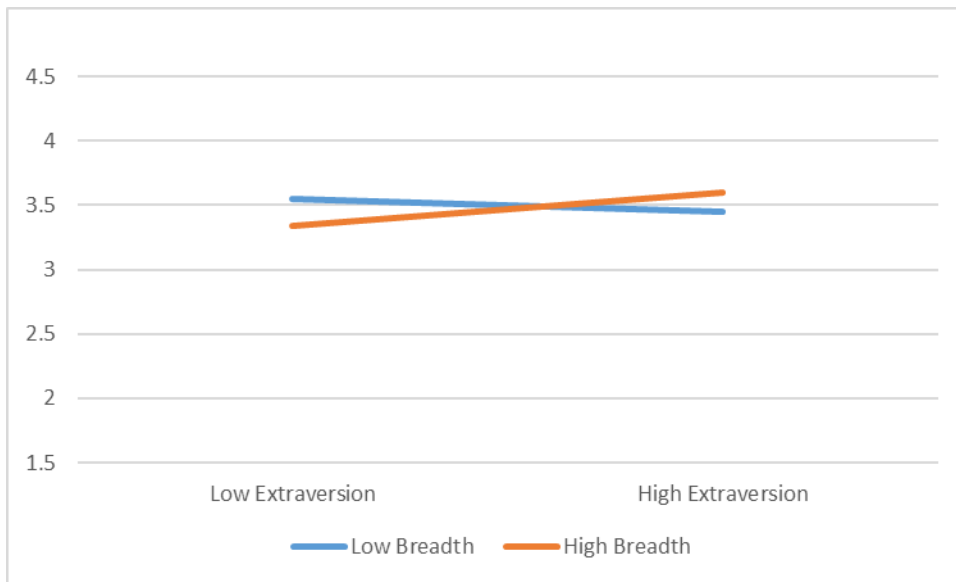


Figure 3. Interaction between breadth and third-party rated extraversion for predicting GPA.

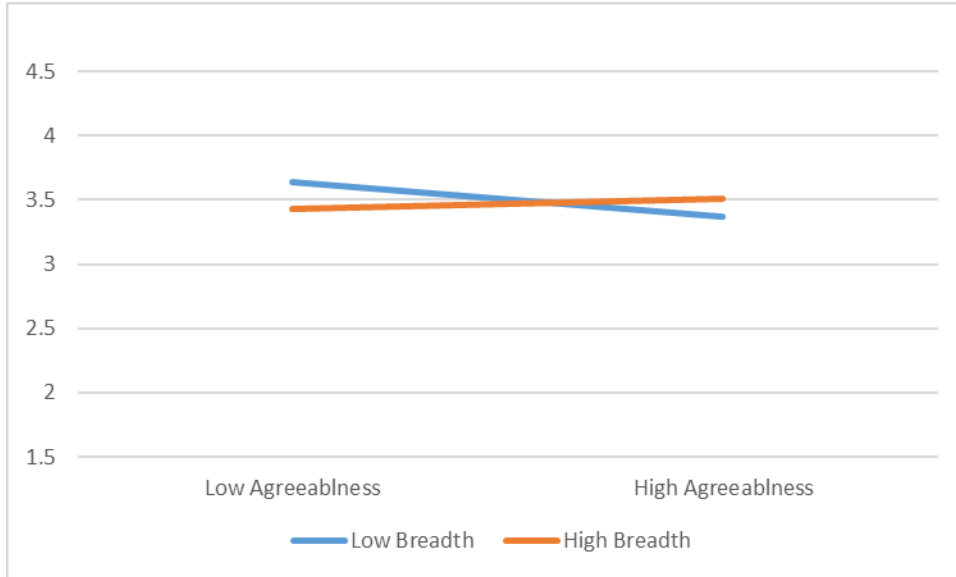


Figure 4.
Interaction between breadth and third-party rated agreeableness for the incremental prediction of GPA.

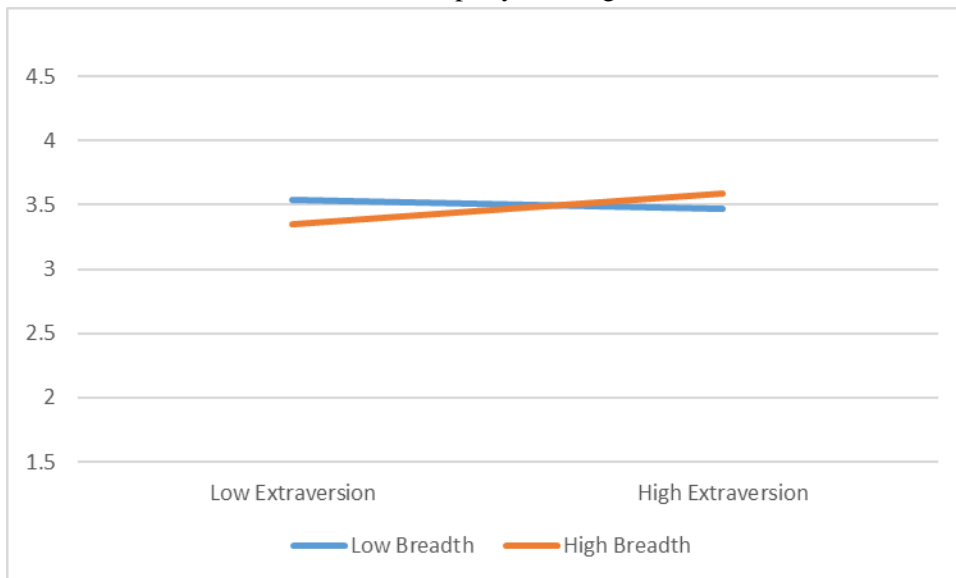


Figure 5.
Interaction between breadth and third-party rated extraversion for the incremental prediction of GPA.

Table 4.

Moderated regression results for the incremental validity (GPA) of third-party reports beyond self-reports of Big Five personality traits and their interactive effects with breadth and depth.

Variable	Step 1	Step 2
Third-Party (N)	-.036	-.056
Third-Party (E)	.088	-.089
Third-Party (O)	.024	.035
Third-Party (A)	-.069	-.105
Third-Party (C)	.069	.096
Target (N)	.048	.044
Target (E)	.016	-.002
Target (O)	-.069	-.094
Target (A)	-.008	-.004
Target (C)	.268**	.231**
Relationship Depth	.064	.082
Observational Breadth	-.050	-.035
Relationship Length	-.031	-.024
SAT	.075	.095
Target Gender	.132*	.153*
Context Consistency	-.082	-.082
Third-Party (N) x Depth		-.081
Third-Party (E) x Depth		-.060
Third-Party (O) x Depth		.095
Third-Party (A) x Depth		-.018
Third-Party (C) x Depth		.063
Third-Party (N) x Breadth		.074
Third-Party (E) x Breadth		.158*
Third-Party (O) x Breadth		-.063
Third-Party (A) x Breadth		.183*
Third-Party (C) x Breadth		-.021
Equation R ²	.144**	.206**

Note. Tabled values are standardized beta weights

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

DISCUSSION

Implications for Self-Other Overlap

The results of the preceding study indicate that indeed both observational breadth and relationship depth are important considerations when using third-party ratings in personality research and applied selection practice. Relationship depth demonstrated strong effects on the self-other overlap index of accuracy criteria for both high and low visibility traits. This is in contrast to the findings of Kertz and Sherker (2003) that relationship quality (i.e., self-report cohesion and conflict) did not increase self-other overlap between a third-party rater and a target. In addition to this finding Leising, Erbs and Fritz (2010) found that ‘liking’ a target only inflated ratings but did not improve self-other overlap. Our finding regarding the effect of depth on self-other overlap has several implications. In research settings it may be more important to determine how well an individual knows a target before assuming their relationship with the target (e.g., friend, family, significant-other) ensures they will provide an accurate assessment. On the other hand, in an applied setting where the concern is predicting future performance of the target, an individual with high depth can offer greater predictive validity (openness to experience) but not incremental validity over and above self-reported personality. In such instances it may be more beneficial to find a third-party rater with greater breadth of observations so as to offer a unique and still relatively accurate assessment.

Implications for Discriminant Validity

Contrary to what was hypothesized, relationship depth did not increase the discriminant validity of third party-ratings across the Big Five. On the other hand, observational breadth demonstrated prediction that neared significance. Again this ran counter to what was hypothesized but may not be completely illogical. Individuals that have observed the target’s behavior across a range of settings may simply have more information and less bias, slightly helping their propensity to discriminate. My general conclusion is that the ability to discriminate has less to do with aspects of your relationship with a target and likely more to do with your education (e.g., psychology, sociology), individual differences (e.g.,

social and emotional intelligence, gender) and experience (e.g., job experience in relevant fields). Our curious findings regarding women's superior ability to discriminate is somewhat consistent with research demonstrating their superior ability to correctly identify non-verbal emotional cues (Thayer & Helge Johnson, 2001) and higher social abilities (Petrides & Furnham, 2000). Essentially, women appear to attend more to social cues that may be manifestations of a target's underlying personality.

Implications for Criterion-Related Validity

Past research has demonstrated the predictive power of third-party assessments of personality (Barrick et al., 2001; Connelly & Ones, 2010; Connelly & Hulsheger, 2012). I wanted to better understand why certain third-party raters provide assessments of personality that offer more predictive power than others. Connelly and Hulsheger (2012) elaborated on this research finding that familiar others (i.e., individuals with close ties to the target outside of work) offer better prediction than individuals who have only observed the target in a single context (i.e., work colleagues with no personal relationship with the target outside of work). I theorized that this increase in prediction had to be a function of either knowing someone well (i.e., relationship depth), or having observed them behave in a variety of settings (i.e., observational breadth). Furthermore, I believed that these two aspects of a target-third-party relationship would have differential effects on trait ratings due to their visibility and evaluativeness. Specifically, relationship depth was expected to matter more where manifestations of traits were not readily available in behavior, and observational breadth would matter more where the social desirability of exhibiting a trait may artificially affect behavior (i.e., highly evaluative traits). As a result of this thinking I expected both relationship depth and observational breadth to moderate the relationship between personality and performance but to do so through different traits. Relationship depth should help the predictability of low visibility traits (i.e., Neuroticism and Openness) and observational breadth should help the predictability of highly evaluative traits (i.e., Agreeableness and Conscientiousness). While these contentions were not fully supported by the research here, the results do demonstrate that the general notion that going beyond relationship descriptions (i.e., strangers, friends, family, significant-

others) and length of relationship when choosing third-party raters may be beneficial to researchers and practitioners alike. Additionally, it is important to note that observational breadth did interact with the third-party rating of the most evaluative trait (i.e., Agreeableness) in predicting performance. The fact that the interaction term of depth by openness was also significant is an indication that my theoretical thinking was sound. That is, depth and breadth would moderate the relationship between third-party ratings and performance through their differential effects on low visibility traits and highly evaluative traits. Admittedly, the plots of these interactions did not make a clear case for exactly how these interactions operate. Certainly more research will be needed to better understand the complexities of these interactions. In general, the implication for selection consultants is to at least consider the type of trait third party raters are asked to evaluate in determining the relative importance of relationship depth and observational breadth. Certain jobs may be more concerned with openness to experience (e.g., expatriate assignments) while others agreeableness (e.g., teaching). Measures assessing depth and breadth may help consultants choose or select the best third-party rater for a given situation. This might better insure accurate trait information for the notoriously difficult to rate low visibility traits and highly evaluative traits. With that said, it is unlikely that a selection consultant would not combine a third-party assessment with the target's own self-assessment when attempting to predict performance. This logic led me to my final conclusions.

Implications for Incremental Validity

It is not advised for a selection consultant to rely on any single assessment of a target's personality to predict performance, even if that assessment is the target's self-report. Instead it is recommended to either combine multiple assessments by averaging them to develop a final assessment or to combine them using a regression model to predict performance. Thus, it is important that consultants attain unique information from each source about the target if they are to maximize predictive power. My results indicate third-party raters that have observed the target's behavior across a greater variety of contexts, are more likely to provide a unique assessment (above and beyond self-assessments) than

someone that knows the target well. This finding follows Hogan's (1986) Socioanalytic Theory of Personality. That is, someone who knows the target well may also be in-tune with the target's internally held beliefs and motives for behavior. These individuals then would provide a third-party rating very similar to the target's self-rating. This is in opposition to a third-party rater who has observed the target in a variety of settings. These individuals should provide a rating that aligns more closely with the target's social reputation. To be clear, an individual with high relationship depth is likely to view the target as the target views themselves. Whereas an individual with greater observational breadth is likely to view the target as other people view them. This deviation from the self-report is more likely to provide unique information that would add incrementally to the prediction of a target's behavior. Again, it is important to note that the complexities of these interactions are not fully clear from the data at hand. More research on these interactions is warranted.

The use of multiple or third-party ratings for personality information about a target is becoming more common. It is now thought that researchers and practitioners alike should be cautious relying solely on single self-assessments for decisions and research conclusions. Recent research by Connelly & Ones (2010) and Connelly and Hulsheger (2012) indicates that the most accurate third-party ratings come from individuals highly familiar with the target (i.e., friends, family and significant-others). This may lead researchers and practitioners that rely on personality assessment to attain third-party ratings from friends, family or those close to the target. The current study should provide some caution with this logic. While some depth is indeed good, too much is likely to result in an assessment that provides no new information about the target that cannot be gleaned from their self-assessment. If we want an assessment that moves closer to the target's outward reputation, we will want third-party raters that have observed the target in a variety of contexts. Such an assessment is more likely to add incrementally to the prediction of behavior beyond the self-report when compared to a rating made by a highly intimate other without a breadth of observations. We know that familiar others provide more accurate assessments than acquaintances or

strangers (Connelly & Ones, 2010; Connelly & Hulsheger, 2012) on most criteria of accuracy. This study helps us understand why and on specifically what criteria this is most applicable to.

Limitations and Directions for Future Research

One of the most apparent limitations of this research is the construct validity of *depth of relationship* and *breadth of observations*. Both concepts are new to the literature and were an attempt to move beyond the effect of relationship categories and their effects on rating accuracy. I intended to add clarity as to why people with more intimate relationships with the target demonstrate differing levels of accuracy across criteria (i.e., self-other overlap, discriminant validity and behavioral prediction). While I can confidently conclude that both depth and breadth provide more sound logic in explaining moderated effects compared to relationship categories or the length of the relationship, work on improving these constructs is warranted. Starting with depth, the selection of the number and type of personal questions to ask third-party raters was somewhat arbitrary. Psychometric research intended to pinpoint what and how many questions to ask would make this a more reliable and valid construct. With that said, it is still beneficial to keep such a measure short as it will often be combined with several other selection instruments. This may be difficult concerning the desire for reliability and validity of the construct. In terms of breadth of observations, the situations that were asked of the third-party rating did constitute differing levels of situational strength (i.e., interacting with superiors, with students, with family), however, research focusing on the psychometric validation of this construct is also warranted. Such measures may help clear up the nature of the interactions between these constructs and third-party ratings when predicting behavior. Psychometrically reliable and valid measures of relationship depth and observations breadth could provide benefit beyond the selection of third-party raters of personality. Practitioners could use them to find good sources for reference checks and letters of recommendation. Researchers relying on self-assessments of any construct (e.g., emotional intelligence, social intelligence, cultural competence), might provide measures of depth and breadth to third-party raters before interpreting their assessments. If I intend on combining multiple assessments to make a selection decision

scores on breadth and depth should help me select the best assessments for making my decision. In sum, there is evidence that depth and breadth are important considerations when using third-party ratings, but more work on the validation of these constructs is needed.

Another limitation is the reliance on self-reported GPA as my dependent variable for assessing predictive and incremental validity. Kuncel, Crede and Thomas (2005) demonstrated that self-reported grades may be less construct valid than what has been conventionally thought. However, they did identify circumstances where self-report GPA is more reliable. They state that college students tend to report more accurate GPAs than high school students (Kuncel et al., 2005). In addition, they do state that self-report grades predict outcomes to a similar magnitude as actual grades (Kuncel, et al., 2005). Considering this and the fact that my findings followed the general logic stated in my hypotheses, I am confident in the conclusions drawn in this research. Therefore, I do not believe that the use of self-report GPA is a major threat the validity of my findings. However, the use of GPA combined with a largely homogeneous sample (i.e., college age students) should caution researchers in generalizing these findings to other people and performance criteria.

There are several limitations that must be discussed regarding the use of multiple regression to analyze my main effects and interactions. First, regression analyses are notorious for having low power (Evans, 1985; Morris, Sherman & Mansfield, 1986) increasing the likelihood of committing type II error. Fortunately, the large sample size used in this study mitigated the potential for type II error even when as many as twenty predictor variables were used to assess interactions and incremental prediction (i.e., acceptable n to k ratios for power analysis). In addition to having an acceptable n to k ratio to validate my research conclusions, measurement error is cited as something that can reduce power when performing regression analyses. It is important to note that one should be cautious generalizing these findings to dissimilar settings. This study was conducted largely with a sample of college students majoring in psychology. While the sample size was large and diverse in terms of gender and ethnicity it was fairly homogenous in terms of age and academic specialization. With that said, it is possible that psychology

students already maintain an above average ability to rate the personality of a target accurately mitigating the moderating impact of my variables of interest (i.e., depth and breadth). It is possible that these two variables would have an even greater impact on the broader population. Follow up research in field settings with a greater range of ages and specializations would further validate the findings of the current study.

Lastly, much if not all of the recent proliferation of research on third-party personality assessments has focused on broad Big Five traits. This research has found that people familiar with targets provide more accurate assessments (Connelly & Ones, 2010; Connelly & Hulsheger, 2012) especially for traits low in visibility (i.e., Neuroticism and Openness). A committee member of the current research recommended an evaluation of the hypotheses at the facet level of Big Five Factors. Unfortunately, the NEO-PI-3 was not designed for facet level investigations of the Big Five (McCrae & Costa, 2007). In fact, Saucier (1998) attempted to develop a scoring system for facets of the Big Five measured using the brief NEO but those scores did not correspond to facets measured by the full NEO-PI-R. As a result, we still do not know is how familiarity affects sub-facets of the various Big Five traits and whether or not sub-facets differ in terms of visibility and evaluativeness. Accordingly, openness is considered a lower visibility trait, however, there may be sub-facets of openness that are more visible (e.g., adventure seeking, creativity). Research investigating the accuracy and predictive validity of third-party ratings using facet dimensions of the Big Five is warranted. This information may also be of interest to selection consultants looking to predict specific behavior for specific contexts.

Conclusions

In sum, relationship depth predicted self-other overlap across four of the five dimensions of personality. This finding helps clear up why our close-friends, family members and significant others exhibit high levels of accuracy on this specific index of accuracy. With that said, depth does not appear to result in better discrimination, as was hypothesized, nor does it result in better behavioral prediction for low-visibility traits. In addition, higher depth is not an asset when combining a third-party rating with a

self-report to predict performance. On the other hand, observational breadth was not related to self-other overlap for any of the Big Five traits, nor was it related to discriminate validity. It was, however, related to better prediction for agreeableness, a highly evaluative trait, and to the incremental prediction of performance beyond self-reports.

This study answers the call for more research on third-party ratings of personality. It increases our understanding of the effects of relational familiarity and the accuracy of third-party ratings. Specifically, this study demonstrates that the breadth of observations and depth of a relationship between a dyad will differentially affect the accuracy of third party ratings based on trait visibility and evaluativeness. Lastly, practitioners should be cautious soliciting third-party ratings with very close ties to a target if they plan on combining those ratings with the target's self-assessment. It is unlikely that such a third-party rater would add uniquely to the prediction of performance.

APPENDIX: UCF IRB APPROVAL LETTER



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Kimberly A. Jentsch**

Date: **May 17, 2013**

Dear Researcher:

On 5/17/2013 the IRB approved the following modifications to human participant research until 5/16/2014 inclusive:

Type of Review:	Submission Response for IRB Continuing Review Application Form
Modification Type:	Consent form revision, Addition of sub-investigators David Mihm and Mitchell Tindall, Removal of co-investigator Mary Jane Sierra, Removal of research associates Adam Goth and Travis Rios
Project Title:	Simulated Environments for Customer Service Training
Investigator:	Kimberly A. Jentsch
IRB Number:	SBE-07-05082
Funding Agency:	Office of Naval Research
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30 days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form cannot be used to extend the approval period of a study. All forms may be completed and submitted online at <https://iris.research.ucf.edu>.

If continuing review approval is not granted before the expiration date of 5/16/2014, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Patria Davis on 05/17/2013 10:07:40 AM EDT

A handwritten signature in black ink, appearing to be "J. B. Smith" or similar, written in a cursive style.

IRB Coordinator

REFERENCES

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213.
- Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- Akert, R. M., & Panter, A. T. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences, 9*(6), 965-972.
- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*(3), 387.
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1), i.
- Allport, G. W. (1937). *Personality* (pp. 173-181). New York: Holt.
- Allport, G. W. (1937). The functional autonomy of motives. *The American Journal of Psychology, 50*(1), 141-156.
- Allport, F. H. (1937). Teleonomic Description In The Study of Personality. *Journal of Personality, 5*(3), 202-214.
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*(3), 518.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*(1), 9-30.

- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). The FFM personality dimensions and job performance: Meta-analysis of meta-analyses. *International Journal of Selection and Assessment*, 9(1), 9-30.
- Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90(6), 1185.
- Baumgarten, W. (1933) Beziehungen zwischen Forstwirtschaft und Berg- und Hüttenwesen im Kommunionharz: ein Beitrag zur Wirtschaftsgeschichte des Harzes. PhD Thesis, Faculty of Forestry, University of Munich, Germany.
- Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, 90(3), 250-260.
- Beer, A., & Watson, D. (2008). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality*, 76(3), 535-560.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 81(6), 506.
- Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin*, 20(4), 367-378.
- Berry, D. S. (1990). What can a moving face tell us? *Journal of Personality and Social Psychology*, 58(6), 1004.

- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*(2), 410.
- Borgatta, E. F. (1964). The structure of personality characteristics. *Behavioral Science, 9*(1), 8-17.
- Borgatta, E. F. (1964). A very short test of personality: The behavioral self-rating (BSR) form. *Psychological Reports, 14*(1), 275-284.
- Borgatta, E. F. (1964). A note on the consistency of subject behavior in interaction process analysis. *Sociometry, 27*(2), 222-229.
- Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*(3), 546.
- Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality, 61*(4), 477-496.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment, 9*(1-2), 52-69.
- Bowers, K. S. (1973). Situationism in psychology: An analysis and a critique. *Psychological review, 80*(5), 307.
- Burger, J. M. (2011). *Personality*, Wadsworth Publishing Company.
- Buss, A. H., & Plomin, R. (1984). Temperament: Early developing personality traits. 857-865.

- Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology, 3*(1), 23.
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology, 38*(4), 476.
- Cattell, R. B. (1943). The description of personality. I. Foundations of trait measurement. *Psychological Review, 50*(6), 559.
- Cattell, R. B. (1943). Fluctuation of sentiments and attitudes as a measure of character integration and of temperament. *The American Journal of Psychology, 195-216*.
- Cattell, R. B. (1946). *Description and measurement of personality*. New York: World Book.
- Cattell, R. B. (1946). Personality structure and measurement: The operational determination of trait unities. *The British Journal of Psychology. General Section, 36*, 88.
- Cattell, R. B. (1946). Simple structure in relation to some alternative factorizations of the personality sphere. *The Journal of General Psychology, 35*(2), 225-238.
- Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika, 12*(3), 197-220.
- Cattell, R. B. (1947). The ergic theory of attitude and sentiment measurement. *Educational and Psychological Measurement, 7*, 221-246.
- Cattell, R. B. (1948). Concepts and methods in the measurement of group syntality. *Psychological Review, 55*(1), 48.
- Cattell, R. B. (1948). Primary personality factors in the realm of objective tests. *Journal of Personality, 16*(4), 459-486.

- Cattell, R. B. (1948). The primary personality factors in women compared with those in men. *British Journal of Statistical Psychology*, *1*(2), 114-130.
- Cattell, R. B. (1948). The meaning of clinical psychology. An introduction to clinical psychology. *Journal of Educational Psychology*, *39*(7), 444-447
- Cattell, R. B. (1948). The integration of factor analysis with psychology: A reply to Professor Godfrey Thomson's review of "The Description and Measurement of Personality."
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book.
- Cattell, R. B. (1957). The conceptual and test distinction of neuroticism and anxiety. *Journal of Clinical Psychology*, *13*(3), 221-233.
- Cattell, R. B. (1957). A universal index for psychological factors. *Psychologia*, *1*, 74-85.
- Colvin, C. R. (1993). Childhood antecedents of young-adult judgability. *Journal of Personality*, *61*(4), 611-635.
- Colvin, C. R. (1993). "Judgable" people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, *64*(5), 861.
- Connelly, B. S. (2008). *The reliability, convergence, and predictive validity of personality ratings: Another perspective*. ProQuest.
- Connelly, B. S., & Hülshager, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of personality*, *80*(3), 603-631.

- Connelly, B. S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092.
- Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review*, *13*(1), 62-72.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R*. Psychological Assessment Resources.
- Costa Jr, P. T., & McCrae, R. R. (1985). Concurrent validation after 20 years: Implications of personality stability for its assessment. *Advances in personality assessment*, *4*, 31-54.
- Costa Jr, P. T., & McCrae, R. R., (1985). Openness to experience. *Perspectives in personality*, *1*, 145-172.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory. *Personality and Individual Differences*, *12*(9), 887-898.
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 653-665.
- Costa, P. T., & McCrae, R. R. (1992). Neo PI-R professional manual.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5.
- Costa, P. T., & McCrae, R. R. (1992). Professional manual: revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources*, 179-198.

- Costa Jr, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*(1), 21-50.
- Costa, P. T., & McCrae, R. R. (1995). Primary traits of Eysenck's PEN system: Three- and five-factor solutions. *Journal of Personality and Social Psychology, 69*(2), 308.
- Costa, J. P. T., & McCrae, R. R. (1995). Positive and negative valence within the five-factor model. *Journal of Research in Personality, 29*(4), 443-460.
- Costa, P. T., & McCrae, R. R. (1995). Trait explanations in personality psychology. *European Journal of Personality, 9*(4), 231-252.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). Theory of generalizability for scores and profiles. The dependability of behavioral measurements.
- Digman, J. M. (1988). Classical theories of trait organization and the Big Five factors of personality. In *annual meeting of the American Psychological Association, Atlanta, GA*.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*(1), 417-440.
- Digman, J. M. (1991). The Big Five: Up, down, and beyond. In *S. Strack (Chair), Beyond the Big Five. Symposium at the 99th annual meeting of the American Psychological Association, San Francisco, CA*.
- Donahue, E. M., Robins, R. W., Roberts, B. W., & John, O. P. (1993). The divided self: Concurrent and longitudinal effects of psychological adjustment and social roles

- on self-concept differentiation. *Journal of Personality and Social Psychology*, 64(5), 834.
- Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A Meta-analysis. *Clinical Psychology: Science and Practice*, 7(4), 435-453.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.
- Brunswik, E. (1956). Historical and thematic relations of psychology to other sciences. *The Scientific Monthly*, 83, 151-161.
- Ekehammar, B. (1974). Interactionism in personality from a historical perspective. *Psychological bulletin*, 81(12), 1026.
- Ekehammar, B. (1974). Sex differences in self-reported anxiety for different situations and modes of response. *Scandinavian Journal of Psychology*, 15(1), 154-160.
- Endler, N. S., & Magnusson, D. E. (Eds.). (1976). *Interactional psychology and personality*. New York, Hemisphere.
- Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, 98(3), 513.
- Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological bulletin*, 116(3), 429.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44(3), 329.
- Flink, C., & Park, B. (1991). Increasing consensus in trait judgments through outcome dependency. *Journal of Experimental Social Psychology*, 27(5), 453-467.

- Forehand, G. A., & Von Haller, G. (1964). Environmental variation in studies of organizational behavior. *Psychological bulletin*, 62(6), 361.
- Fritzsche, B. A., McIntire, S. A., & Yost, A. P. (2002). Holland type as a moderator of personality–performance predictions. *Journal of Vocational Behavior*, 60(3), 422-436.
- Funder, D. C., & West, S. G. (1993). Consensus, Self-Other Agreement, and Accuracy in Personality Judgment: An Introduction. *Journal of personality*, 61(4), 457-476.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological review*, 102(4), 652.
- Funder, D. C. (1995). Stereotypes, base rates, and the fundamental attribution mistake: A content-based approach to judgmental accuracy. In Y.T. Lee, L. Jussim, & C.R.
- Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of Personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, 69(4), 656.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26.
- Guilford, J. P. (1975). Factors and factors of personality. *Psychological Bulletin*, 82(5), 802.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135-164.
- Hatrup, K., & Jackson, S. E. (1996). Learning about individual differences by taking situations seriously. *Individual differences and behavior in organizations*, 507-547.
- Hogan, R. (1986). Manual for the Hogan personality inventory.

- Hogan, J. (1986). Hogan personnel selection series manual. *Minneapolis, MN: National Computer Systems.*
- Hough, L. M. (1992). The 'Big Five' personality variables--construct confusion: Description versus prediction. *Human performance, 5*(1-2), 139-155.
- Hülshager, U. R., & Maier, G. W. (2010). The careless or the conscientious: Who profits most from goal progress?. *Journal of Vocational Behavior, 77*(2), 246-254.
- Ickes, W. (1996). On the deep structure of attribution-affect-behavior sequences. *Psychological Inquiry, 7*(3), 236-240.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*(4), 521-551.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*(4), 652.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797.
- Judge, T. A., & Ilies, R. (2002). Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes, 89*(2), 1119-1139.
- Judge, T. A., & Hurst, C. (2007). Capitalizing on one's advantages: role of core self-evaluations. *Journal of Applied Psychology, 92*(5), 1212.
- Judge, T. A., & Hurst, C. (2007). The benefits and possible costs of positive core self-evaluations: A review and agenda for future research. *Positive organizational behavior, 159-174.*

- Kahneman, D., & Tversky, A. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237.
- Kenny, D. A., & La Voie, L. (1984). The social relations model. *Advances in Experimental Social Psychology*, 18, 142-182.
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98(2), 155.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.
- Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the big five. *Psychological Bulletin*, 116(2), 245.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8(3), 265-280.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43(1), 23.
- Klages, L. (1926). *The Psychological Achievements Nietzsches*.
- Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106(3), 395.
- Kruglanski, A. W. (1989). The social psychology of science. In *Lay Epistemics and Human Knowledge* (pp. 223-245). Springer US.

- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology, 98*(4), 668.
- LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: a critical review and meta-analysis. *Journal of Applied Psychology, 87*(1), 52.
- Letzring, T. D. (2005). *The Good Judge: Personality and Behavior Correlates of Accurate Personality Judgement*. Unpublished Dissertation, University of California Riverside.
- Letzring, T., & Wells, S. (2005). *Quantity and Quality of Available Information Affect the Realistic Accuracy of Personality Judgment*.
- Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*(1), 111.
- Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology, 41*(1), 42.
- Letzring, T. D. (2014). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality*.
- Lorr, M., & More, W. W. (1980). Four dimensions of assertiveness. *Multivariate Behavioral Research, 15*(2), 127-138.
- Lorr, M. (1986). *Interpersonal style inventory (ISI): Manual*. Western Psychological Services.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215.

- McCrae, R. R., & Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6*(5), 587-597.
- McCrae, R. R., Costa, Jr, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO personality inventory. *Journal of Personality Assessment, 84*(3), 261-270.
- McCrae, R. R., Martin, T. A., & Costa, P. T. (2005). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Journal of Personality Assessment, 12*(4), 363-373.
- McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 141-156). Washington, DC: American Psychological Association.
- McDougall, W. (1932). Of the words character and personality. *Journal of Personality, 1*(1), 3-16.
- McDougall, W. (1933). *The energies of men: A study of the fundamentals of dynamic psychology*. New York: Scribner's.
- Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36*(1), 121-140.
- Mischel, W. (1968). Consistency and specificity in behavior. *Personality and Assessment, 13*-19.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*(4), 252.
- Mischel, W. (1973). On the empirical dilemmas of psychodynamic approaches: Issues and alternatives. *Journal of Abnormal Psychology, 82*(2), 335.

- Mischel, W. (1977). On the future of personality measurement. *American Psychologist*, 32(4), 246.
- Mischel, W. (1977). The interaction of person and situation. *Personality at the crossroads: Current issues in interactional psychology*, 333, 352.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60(4), 1029-1049.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, 79(2), 272.
- Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18, 343-357.
- Nisbett, R., & Ross, L. (1980) *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66(6), 574.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 4(6), 681.

- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*(4), 679.
- Ones, D. S., Schmidt, F. L., & Viswesvaran, C. (1993, May). Nomological net for measures of integrity and conscientiousness. In *Eighth annual conference of the Society of Industrial and Organizational Psychology, San Francisco*.
- Ones, D. S., Schmidt, F. L., & Viswesvaran, C. (1993, May). Integrity and ability: Implications for incremental validity and adverse impact. In *Eighth annual conference of the Society of Industrial and Organizational Psychology, San Francisco*.
- Schmidt, F. L., Viswesvaran, C., & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology, 53*(4), 901-912.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*(4), 995-1027.
- Oswald F.L., & Hough L.M. (2010). Personality and its assessment in organizations: Theoretical and empirical developments. In Zedeck S. (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 2: 153-184). Washington, DC: American Psychological Association.
- Paunonen, S. V., & O'Neill, T. A. (2010). Self-reports, peer ratings and construct validity. *European Journal of Personality, 24*(3), 189-206.
- Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57*(3), 552.

- Pervin, L. A. (1985). Personality: Current controversies, issues, and directions. *Annual Review of Psychology*, 36(1), 83-114.
- Petrides, K. V., & Furnham, A. (2000). Gender differences in measured and self-estimated trait emotional intelligence. *Sex roles*, 42(5-6), 449-461.
- Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, 24(2), 239-254.
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313-345.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in experimental social psychology*, 10, 173-220.
- Nisbett, R., & Ross, L. (1991). *The person and the situation*. New York: McGraw Hill.
- Selfhout, M., Denissen, J., Branje, S., & Meeus, W. (2009). In the eye of the beholder: Perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process. *Journal of personality and social psychology*, 96(6), 1152.
- Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self- concept: Through the looking glass darkly. *Psychological Bulletin*, 86(3), 549.
- Snyder, M. (1987). *Public appearances, private realities: The psychology of self- monitoring*. New York: WH Freeman.
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. *Handbook of Social Psychology*, 2, 883-947.

- Starzyk, K. B., Holden, R. R., Fabrigar, L. R., & MacDonald, T. K. (2006). The Personal Acquaintance Measure: a tool for appraising one's acquaintance with any person. *Journal of Personality and Social Psychology, 90*(5), 833.
- Swann, W. B. (1984). Quest for accuracy in person perception: a matter of pragmatics. *Psychological Review, 91*(4), 457.
- Tett, Christiansen, Swanavelder, Meyer & Bartram (Eds.). (2012). A sticky topic: Using personality tests in selection settings. *Proceedings from SIOP 12': The Twenty Eighth Conference for the Society of Industrial and Organizational Psychologists*. Houston, TX.
- Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences, 52*(2), 195-201.
- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*(4), 397-423.
- Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian journal of psychology, 41*(3), 243-246.
- Thompson, M. S., Judd, C. M., & Park, B. (2000). The consequences of communicating social stereotypes. *Journal of Experimental Social Psychology, 36*(6), 567-599.
- Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (No. ASD-TR-61-97). PERSONNEL RESEARCH LAB LACKLAND AFB TX.

- Vazire, S. (2005). Informant Reports: A Cheap, Fast, And Easy Method For Personality Assessment. *Journal of Research in Personality*, 472-481.
- Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281.
- Vazire, S. (2010). Informant reports. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 167–178). Washington, DC: American Psychological Association.
- Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, 41(1), 221-229.
- Weiss, H., & Adler, S. 1984. Personality and organizational behavior. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 4: 1-50). Greenwich, CT: JAI Press.
- Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin*, 28(6), 84.