Spring January 18th, 2017

The Psychometric Evaluation of a Personality Selection Tool

James R. Longabaugh
Seattle Pacific University

The Psychometric Evaluation of a Personality Selection Tool

James Longabaugh

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

in

Industrial-Organizational Psychology

Seattle Pacific University

January, 2017

Acknowledgments

The question of whether it is the journey or the destination that is more important has never been so clear; it is the journey. There have been so many people who have helped and supported me along the way, and I only hope that I can acknowledge as many of them as possible. It is with great gratitude that I extend thanks to each and every one who has helped me attain this high honor, but more so for their contributions of inspiration and motivation along the way.

First and foremost, my advisor, my mentor, and my dissertation chair, Dr. Dana Kendall. I express my appreciation for her guidance, knowledge, and above all her mentorship. She has walked alongside me for the last six years, bestowing her knowledge, providing undue support and inspiration that I shall forever be grateful. She sets the gold standard for what it means to be a mentor.

I would also like to thank my committee members, Dr. Jerry Kehoe and Dr. Lynette Bikos. Jerry truly is one of our field's foremost thought leaders and experts. He has been a solid foundation for which I have developed my applied and consulting acumen in relation to I-O psychology, as well as teaching me to keep fishing for knowledge. I hope he understands how invaluable the projects that we had worked on together have and will be for my success. Dr. Bikos has been instrumental in my penchant for data and statistics. Her unwavering guidance and support, from statistics courses to this dissertation, has been instrumental throughout this journey.

I would also like to thank my professors, Dr. Rob McKenna, Dr. Paul Yost, and Dr. Joey Collins. I express my gratitude for your genuine support, motivation, inspiration, and most of all

pushing me towards my edge and challenging my thoughts. I have gained valuable lessons from each, and I am honored to be part of the Seattle Pacific University I-O Psychology program.

In respect to the SPU I-O program, I have so many colleagues who I also consider to be close friends that have been my support and guiding light along this journey. This has definitely been a journey that one cannot go alone. While so many I could list, I am especially grateful for Dr.'s Bobby Bullock, Katie Kirkpatrick-Husk, Joshua Weaver, Emily Pelosi, Hilary Roche, Daniel Hallak, John Terrill, and soon-to-be Dr.'s Roble Kirce, Kira Wenzel, Serena Hsia, Nathan Iverson, Jared Daniel, and J'Aimee Mission. Thank you!

From my time at IBM Kenexa, there are two people that I have especially looked to for guidance and who have helped me get across the finish line, Dr. Kevin Impelman and Dr. Jeff Labrador. You both helped me think through important pieces of this dissertation that added great value, and I continue to learn from you each and every day.

I would also like to thank my family and close friends who have shown their support, provided motivation, and most of all for their understanding. To my mother, for her undying love, my father for teaching me to give it everything you got to achieve your goals, Cameron for his encouragement, and also the grace from my grandparents and extended family. To my friends, for their understanding when I could not come out to play, and their support to keep moving forward. Last but certainly not least, my wife Melinda, for her praise, for giving me the space and time to study, providing assurance that all will be right, and most of all her unconditional love. To her, I hope that I was present throughout this journey.

Table of Contents

List of Tables

List of Figures

List of Appendices

James Longabaugh

483 words

Abstract

Personality is one of the primary ways that people are distinguished from one another on the basis of their unique tendencies and behavioral patterns. Decades of empirical research have yielded five primary personality traits which have consistently emerged, becoming known as the Five Factor Model (FFM). In particular, the FFM has been widely used in the employee selection realm. However, there have been mixed reviews as to how well the FFM of personality accomplishes that objective, with some research drawing into question the strength of the relationship between personality and job performance.

The purpose of the current investigation is to address these gaps by exploring how a unique personality instrument, developed for commercial applications, may add value for predicting job performance. Archival and anonymous data was collected from the Company for the purpose of investigating the psychometric properties of the proprietary personality selection instrument. The study employed two sets of data, one was a random sample of 10,000 cases for conducting factor analytics and the other was a data set of 1,986 matched cases for investigating reliability. Hypotheses were tested using multiple statistical techniques, such as confirmatory factor analysis (CFA), exploratory factor analysis (EFA), test-retest reliability, Cronbach's alpha, and Pearson correlations.

Results indicated that the proposed factor structure of the Personality Instrument in which the 15 narrow facets are represented by three broad factors could not be supported ($\chi 2$ [3897] = 94548.14, $p < .001$; *GFI* = .704; *CFI* = .433; *RMSEA* = .048) through CFA. Multiple alternative models were also tested, such as a single-factor model of personality ($\chi 2$ [3900] = 96837.26, $p <$

.001; *GFI* = .694; *CFI* = .418; *RMSEA* = .049), the theoretical model poor ($\chi 2$ [3910] = 102,525.28, $p < .001$; *GFI* = .692; *CFI* = .383; *RMSEA* = .050), and also a multitrait multimethod approach ($\chi 2$ [3710] = 56336.08, $p < .001$; *GFI* = .863; *CFI* = .671; *RMSEA* = .038), each of which fell short of showing even moderate model fit to the data. The results of the test-retest reliability analyses showed statistically significant positive relationships between Time 1 and Time 2 Personality Instrument scores (r (905) = .79, $p < .01$), when administered between 0 and 12 weeks apart, and thus supporting the stability of the Personality Instrument. Compared to test- retest reliability coefficients when taken between 0 and 12 weeks, the test-retest reliability for cases in which the Personality Instrument was taken between 13 weeks and 52 weeks was still statistically significant and positive (r (480) = .67, $p < .01$), yet a smaller positive relationship as presumed. Test-retest reliability proved to be a better estimate of stability compared to Cronbach's alpha, which ranged from as low as .05 for Tolerance for Ambiguity to as high as .58 for Affiliative. Lastly, the results investigating as to whether contextualized items within the same substantive facet correlate more so with each other than do generically-worded items.

CHAPTER I

Introduction and Literature Review

Aside from physical appearance, personality is one of the primary human characteristics that distinguish individuals from one another, defining their uniqueness through their behavior, emotions, and the ways they think about the world. For the purposes of this investigation, personality is defined as a "complex pattern of deeply embedded characteristics that are expressed automatically in almost every area of psychological functioning" (pg. 2; Millon, Grossman, Millon, Meagher, & Ramnath, 2004). Components of personality reflect the habitual ways in which people think, feel and act, and these characteristics are quantified and treated by researchers and practitioners as stable individual differences, or traits (Gatewood, Feild, & Barrick, 2011). This allows distinctions to be made among individuals for the purpose of matching their various traits with roles in organizations in such a way that best utilizes their unique skills and strengths. Employees who have a greater fit with the organization and their roles have been shown to have greater satisfaction with their jobs, increased commitment to their organization, and they are also much less likely to quit (Kristof-Brown, Zimmerman, & Johnson, 2005). Not only does this suit the organization in terms of increased employee commitment, but this may in turn lead to greater overall performance and reduced turnover (Kristof-Brown et al., 2005). Consequently, the assessment of personality has grown in importance, and is now widely accepted in both research and applied settings (Hough & Johnson, 2013).

In applied settings, personality is often a trait of choice used by organizations when selecting among applicants because it has been found to be positively correlated with work-related outcomes, specifically job performance (Barrick & Mount, 1991; Judge, Rodell, Klinger, Simon, & Crawford, 2013). Nevertheless, concerns have been raised in the literature as to how

much variance personality actually explains in job performance to a sufficient extent as to justify its use as a predictor. Even when corrected for artifacts such as range restriction or unreliability of measures, often small correlations are found and therefore concerns are raised as to how well personality really predicts proficiency criteria (Barrett, 2008). As an example, in a review of 13 meta-analyses it was found that personality has a median uncorrected validity of .10, and a median corrected validity of only .18 (Morgeson et al., 2007b). Although the amount of variance explained by personality in job performance is not negligible, it is important for researchers to continue refining and developing better measures (Judge et al., 2013; Morgeson et al., 2007a; Schmitt, 2014), especially in the realm of personnel selection.

The Company's[1] Personality Instrument[2] is a recently developed personality-based assessment tool which has the potential to address the weaknesses of existing measures in two primary ways, (a) having items composed in a work context, and (b) personality traits assessed at the narrower facet-level which in combination represent other attributes important for work success. First, many of its items ask participants to self-rate their personality in the work setting, as opposed to across all settings. Second, the Personality Instrument allows for both broad and narrow personality traits to be assessed, with the goal of increasing its predictive precision and utility. Thus, compared with the current personality-based selection tools, the Personality Instrument may have opportunities to be more effective in predicting an applicant's future job performance. Consequently, selection decisions may be more accurate, and the applicants will experience better fit in their roles once hired. Now, I will turn to a description of past validation work of the Personality Instrument.

[1]A pseudonym, Company, is used in place of the actual company's name for confidentiality.
[2]The pseudonym of Personality Instrument is used in place of the instrument's name to retain confidentiality.

Extending Validation Evidence of the Personality Instrument

In 2013, validation work was performed on the Personality Instrument with the purpose of developing a technical manual for providing empirical support for the use of the Personality Instrument in selection contexts. Reported in the technical manual are a number of psychometric attributes of the instrument, such as the means and standard deviations of the fifteen narrow personality facets with the addition of the average raw scores for each narrow facet across a variety of industries. The narrow facets were developed with the intention to be somewhat complex in that they represent aspects of more than a single personality facet of the five factor model (FFM), unlike most other personality instruments designed to measure distinct aspects of personality as related to FFM. The narrow facets are labeled as (a) achievement oriented, (b) adaptable, (c) affiliative, (d) assertive, (e) conventional, (f) cooperative, (g) creative, (h) dependable, (i) detail orientation, (j) independent, (k) optimistic, (l) resilient, (m) self-regulated, (n) stimulation/risk seeking, and (o) tolerance for ambiguity. Also included in the technical manual are the internal consistency estimates for each of the 15 narrow facets, as well as correlations among each. Lastly, and of primary interest, the factor structure of the Personality Instrument was proposed as the result of an exploratory factor analysis (EFA), showing that the 15 narrow facets represent three second-order broad factors. Since 2013, an additional 150,000 respondents have taken the Personality Instrument, and thus the most recent data will be used to test the hypotheses of this study.

Purpose of this Study

As a means to bolster the validity evidence of the Personality Instrument as a selection tool, further work was deemed necessary beyond the initial validation study performed in 2013. Although there are some early indicators suggesting the potential of the Personality Instrument as a value-add to practitioners for assessing personality in selection, the stability and factor structure of the instrument warranted further investigation as a means for refinement and validation evidence. Therefore, in this study the plan was to build upon and extend the psychometric evidence from 2013 by investigating the providing support for the proposed factor structure of the Personality Instrument, as well as investigate the stability of the tool as a means to establish reliability. Through a review of the research literature, it would appear that several currently available methods for assessing personality fail to predict job performance as strongly as some argue it should. Consequently, researchers should continue to focus efforts on refining the measurement of personality, such as by evaluating new assessments like the Personality Instrument under research in this study.

In the following sections of the introduction below, the history and progression of thought on how personality should be addressed and conceptualized is outlined, including the FFM and the use of personality in employee selection. Second, the various questions that arise in the literature surrounding the concept of personality such as its stability over time will be reviewed, and also the item composition of personality measures, work-specific versus general context items. Third, evidence will be presented in support of the Personality Instrument as potentially being more effective in assessing personality than what is currently available for use in selection contexts. Finally, the hypotheses and statistical analyses that will be performed in assessing the psychometric properties of the Personality Instrument will be described.

History of Personality Assessment

The lexical approach to capturing personality. Cattell (1946), was an influential behavioral researcher who posited that human language contains useful and descriptive information for understanding the complexity of personality. In languages across the world, numerous words, or adjectives, exist for describing a person's traits in relation to how they act and behave in their environment. Allport and Odbert (1936) had reasoned that the greater the number of words describing a specific trait, the more importance for that particular trait to be expressed as a single word, and therefore support being provided for that trait to be a distinct concept or dimension of personality. As an early adopter of the lexical approach to studying personality, Cattell's work has been the impetus for other researchers and the influence for the way in which personality is generally conceptualized.

Cattell's trait theory. Cattell (1946; 1965) used what is known as the lexical approach for studying and identifying the various dimensions of personality. Initially, he began with an extensive list of over 4,500 adjectives describing ways in which people may act as found in the English language and acquired from the earlier work of Allport and Odbert (1936; Cattell, 1946). That list of words was used by Cattell (1946; 1965) to identify what he would determine to be the most important factors of personality. His goal was not only to determine a finite number of personality traits, but also to develop a measure for assessing those factors.

Using factor analysis to investigate the dimensionality of personality. In order to determine which factors of personality are deemed to be most important, a statistical technique known as factor analysis is often employed. Factor analysis is a procedure for reducing the redundancy of a set of inter-correlated items to a reduced number of latent, unobserved variables referred to as factors (Field, 2013; Gorsuch, 1983). Specifically, this analytic method allows for the testing of the lexical theory of personality and investigating its multidimensionality. Cattell (1946) was an early adopter of factor analysis, likely as a consequence of him studying under Charles Spearman (1904), who invented the technique. Cattell used the technique of factor analysis in order to investigate and identify the most important dimensions of personality.

Cattell's Model. Cattell developed a personality assessment titled the 16 Personality Factor Questionnaire (16PF; Cattell, 1965). The 16PF was administered to diverse populations (i.e., diverse in terms of race, ethnicity, gender, culture, language, and geographic location) until he had achieved sufficient data to perform a factor analysis (Cattell, 1946; 1965). He used the factor analytic technique to distill Allport and Odbert's (1936) original list of 4,500 words into the sub-facets of word clusters that were identified as most essential for describing an individual's personality. Cattell eventually narrowed the list to less than 200 trait names considered to be common in the English language (e.g., reserved, caring, imaginative, accepting, etc.; Cattell & Mead, 2008). He then sampled data from three primary sources which he believed necessary for establishing personality factors: (a) life data (L-data; an individual's everyday life behaviors), (b) experimental data (T-data; individual's reactions to experimental lab situations), and (c) questionnaire data (Q-data; self-reported behavior and feelings). For a personality factor to be labeled as important, it was required to emerge from factor analyses,

which sampled data from each of the three sources (Cattell, 1965). The studies conducted by Cattell eventually led him to adopt a multidimensional model of personality.

Source traits versus global factors. In his review of hundreds of studies conducted by performing factor analyses, Cattell (1946; 1965) concluded from the factor analyses that the 200 original descriptive terms clustered into 16 primary personality categories that he believed to be the most important factors for understanding personality. The factors found by Cattell were designated as source traits. When he factor-analyzed them, he discovered a hierarchical, or multidimensional, structure to personality in which the 16 source traits comprised five overarching surface traits he termed global factors. For example, five of the source traits, (a) warmth, (b) liveliness, (c) social boldness, (d) privateness, and (e) self-reliance comprised the second-order global factor of identified as Extraversion. Based upon Cattell's (1965) work, this 5-dimensional model has become an enduring basic organizing framework for personality traits, and it resembles what is most commonly known today as the FFM of personality (Cattell & Mead, 2008). The five source traits were labeled as introversion/extraversion, low anxiety/high anxiety, receptivity/tough mindedness, accommodation/independence, and lack of constraint/self-control (Cattell & Schuerger, 2003). Thus, Cattell was rather influential in setting the stage for conceptualizing personality by utilizing factor analysis, and laying the framework for others to later develop the FFM, a generally accepted model of personality. The FFM has been widely used as the theoretical basis for several personality assessments (e.g., IPIP and NEO-PI-R), including the Personality Instrument, and now will follow a brief overview of the history of the FFM.

History and Development of the Five Factor Model

The FFM was used as the framework for developing the Personality Instrument, and subsequently it is comprised of both first-order personality facets which in combination represent higher-order personality factors. Following Cattell's discovery of the sixteen source traits and five global factors, subsequent researchers (Costa & McCrae, 1985; Goldberg, 1981; Gough & Bradley, 1996; Hogan & Hogan, 1992; Jackson, 1994) sought to replicate his work by similarly employing the lexical approach for conceptualizing human personality using their own proprietary personality assessments.

These later researchers have contributed considerable support for the multidimensionality of personality, most importantly leading to the establishment of the FFM. For example, McCrae and Costa (1987) began reviewing existing theories and evolving personality structures, and as a result their findings indicated that the appropriate number of first-order facets (i.e., source traits) were less consistent across studies than the number of second-order factors (i.e., global factors). Accordingly, there has been some question as to how many first-order personality facets should fall in line under the five broad factors of the FFM. Research by Saucier and Ostendorf (1999) identified 18 narrow facets emerging from their review of past research as being replicable across both English and German languages, and thus their work may indicate 18 as being an ideal number of narrow personality facets encompassing the FFM of personality. This is in contrast, and nearly half the number of narrow traits, to Costa and McCrae research which identified 30 first-order facets comprising the five broad factors of the NEO-PI model (Costa & McCrae, 1985).

In essence, whereas the number of lower-order facets of the FFM remains open to debate and varies across personality instruments, there is a general consensus that five second-order

factors are consistently demonstrated empirically (Dilchert, Ones, Van Rooy, & Viswesvaran, 2006; Hough & Dilchert, 2010; Judge et al., 2013; Wiggins & Trapnell, 1996). Although recent evidence suggests the possibility of a sixth factor (i.e., Honesty/Humility; Ashton, et al., 2004), it is generally not used in the context of selection.

Like many other personality assessments used for selection purposes, the Personality Instrument is similarly grounded in the FFM framework. Although, the Personality Instrument is comprised of 15 narrow personality facets which map onto each of the five broad FFM personality factors, it does not signify that the Personality Instrument is any more or less suited for assessing personality. Interestingly, and unlike many other personality instruments, the narrow facets of the Personality Instrument were designed to be somewhat complex in that they are not equally distributed across the five broad factors. This distinction of the Personality Instrument will be further described later, but the following will move into descriptions of each of the five broad personality factors of the FFM.

FFM traits. The FFM remains highly regarded as the primary model of personality, with Google Scholar searches turning up more than 18,000 plus citations to the FFM (Judge et al., 2013). This general framework for conceptualizing personality is understood to be comprised of five primary personality traits which have been shown to consistently emerge across research studies and have consequently adopted the general nomenclature for each of the five broad factors (Borkenau & Ostendorf, 1990; Funder & Colvin, 1988; Hogan, 1983; McCrae & John, 1992; Saucier, Hampson, & Goldberg, 2000; Wiggins & Pincus, 1989; Zuckerman, Bernieri, Koestner, & Rosenthal, 1989). While the five broad FFM factors are mostly generalizable across personality instruments, there is much less consistency in the number of narrow FFM facets and their relative names.

Neuroticism, also named for its inverse of emotional stability, is the extent to which an individual experiences negative affectivity (e.g., sadness, anxiety, fear, hostility; Watson & Clark, 1984), whereas someone low in neuroticism is calm and relaxed, usually being seen as even-tempered and stable (Costa & McCrae, 1992a). Neuroticism is often comprised of lower-order facets representing traits such as vulnerability, impulsiveness, anxiety, hostility, depression, and self-consciousness (Judge et al., 2013). While the FFM framework as exemplified by Judge et al. (2013) utilizes six narrow facets for each broad FFM traits, other FFM models may have fewer or even more facets. Extraversion, like neuroticism in the FFM model as described by Judge et al. (2013), is described as having six narrow facets representing such traits as warmth, assertiveness, positive emotions, excitement-seeking, gregariousness, and activity. The personality trait of extraversion may be one of the more observable personality traits and can be described as contrasting warm, outgoing, and cheerful with reserved, solitary, and somber. An individual low in extraversion may be characterized as quiet, shy, or even

withdrawn, and an individual high on the trait is generally regarded as talkative and social (Costa & McCrae, 1992a).

A third FFM trait is often labeled openness to experience which describes a person who is imaginative, curious, and exploratory, contrasting with traits representing the inverse such as rigid, practical, and traditional (Costa & McCrae, 1992b). A person who is seen as being high on openness to experience may be imaginative and have a need for a variety of stimuli (e.g., work tasks and responsibilities), whereas someone who is low on openness may not be as curious, artistic, and would be much more conservative regarding their values. The personality trait of openness can be comprised of narrow facets such as values, ideas, fantasy, aesthetics, feelings, and actions (Judge et al., 2013). A fourth FFM trait, agreeableness can be seen as contrasting generosity, honesty, and modesty with selfishness, aggression, and arrogance. The facets of trust, compliance, straightforwardness, altruism, tenderness, and modest often comprise this broad FFM factor (Judge et al., 2013). An individual who would be described as being high on agreeableness may display behaviors consisting of friendliness and compliance, and is easy to get along with, whereas someone scoring low on this trait may be described as being cold and self-centered, often times making it difficult to get along with them (Digman, 1990).

Lastly, the FFM trait of conscientiousness which has been found in meta-analytic studies to be the best predictor of performance across jobs (Barrick, Mount, & Judge, 2001), contrasts traits representing the aspects of being disciplined and purposeful with laid-back, unambitious, and weak-willed (McCrae & Costa, 2008). An individual scoring high on conscientiousness will generally demonstrate behaviors consistent with neatness, thoroughness, and organization, and in contrast someone scoring low will display behaviors consistent with laziness, lack of attention to detail and disorganization. Accordingly, this broad FFM trait of conscientiousness is represented

by narrow facets such as competence, order, dutifulness, achievement striving, self-discipline, and deliberation (Judge et al., 2013).

Varying levels of these five broad personality traits distinguish individuals from one another in terms of how they think, act and feel. Researchers and practitioners therefore often measure personality using the general framework of the FFM, allowing for a more accurate picture of an individual and their likelihood to succeed in certain jobs. This leads to a discussion of the empirical evidence providing support for the FFM of personality for predicting certain criteria, and most importantly job performance.

Empirical Support for the Predictive Value of Personality for Work-Related Outcomes

Each of the five aforementioned broad personality factors have been found to correlate with various outcomes, most of which predict work-related outcomes in various jobs (i.e., conscientiousness, extraversion, agreeableness, and emotional stability; Barrick, Mount, & Judge, 2001). For example, researchers have consistently observed relationships between personality and other work-related outcomes such as job satisfaction (Judge, et al., 2002), organizational citizenship behaviors (Borman et al., 2001), leadership emergence and effectiveness (Colbert, Judge, Choi, & Wang, 2012), and turnover (Zimmerman, 2008), as a few examples. Empirical support has also been found for the relationship between personality and other life outcomes, such as subjective personal well-being, social acceptance, relationship conflict, criminality, unemployment, physical health, mental health, and occupational satisfaction, among others (Ozer & Benet-Martinez, 2006). Additionally, there is a vast amount of research supporting the use of personality for predicting job performance (Barrick & Mount, 1991; Barrick et al., 2001; Hough & Dilchert, 2010; Hough & Ones, 2001; Judge et al., 2013;

Ones, Dilchert, Viswesvaran, & Judge, 2007), and below a brief review supporting empirical evidence will be presented.

Among the many types of assessment tools used for selection purposes (e.g., skills tests, cognitive ability, situational judgment, etc.), personality as described previously is often captured in order to predict an applicant's future job performance (Barrick & Mount, 1991; Barrick, et al., 2001; Hough & Ones, 2001; Schmidt & Hunter, 1998; Schmitt, 2014). Because each individual is unique in terms of their personality, practitioners can explain a greater variation in behavior and performance (Hough et al., 2001), in comparison to other assessments which are often unitary and measure only a very specific knowledge set or skill required to perform a job. One reason for why personality is popular with regard to selection is that most jobs are rather complex in that they have numerous unique requirements for satisfactory performance. As a consequence, job performance is not a unitary construct (Motowidlo & Van Scotter, 1994), but rather quite complex.

Compared to selection assessments used to measure other constructs (e.g., job skills, biodata, integrity, etc.), the FFM of personality may be an ideal predictor because it is multifaceted and similar to the criterion of job performance in terms of its complexity (Costa & McCrae, 1995). For example, the factors that contribute to an individual's success as a police officer are most likely different than what makes someone successful as a production line worker for a manufacturing organization. In other words, a certain combination of personality traits may designate one individual over another as being better-suited for a particular job. Therefore, the FFM provides a multifaceted approach for predicting future job-related outcomes, differing from other commonly used unitary selection instruments such as integrity tests or job-related skills tests that may only predict one aspect of performance.

The FFM can potentially predict multiple aspects of performance as it is comprised of the broad factors which have each been shown to predict job-related outcomes across a number of industries and jobs (Barrick & Mount, 1991; Barrick et al., 2001). For example, the trait of conscientiousness has been found to be the strongest and best predictor of job performance, generalizing across industries and jobs, compared to the other four FFM broad factors, predicting 4% of the variance in job performance (Barrick et al., 2001; Hurtz & Donovan, 2000). Barrick, Mount and Judge (2001) found that neuroticism negatively predicted performance across jobs, whereas extraversion positively predicts performance in highly social jobs such as sales and law enforcement (Barrick, et al., 2001). Barrick, et al. also found that agreeableness best predicts performance in teams, and that the trait of openness to experience seems to be not relevant to performance, but it does predict training proficiency. Thus, various factors of the FFM may be more predictive of performance and various criterion than others across different jobs.

Although there has been empirical support for personality as being a valid assessment for selection contexts, there are also critics who pose challenges to using personality for this purpose. A review of these challenges to using personality as a means to select among individuals will be presented next, rationale for how the FFM traits are relevant in this study, including how the Personality Instrument in this study addresses these challenges.

Challenges to Assessing Personality for Selection Purposes

In review of the empirical research surrounding personality when used for employee selection, it would seem logical that the Personality Instrument would be a useful tool. However, there are some potential challenges to address, one of which is whether personality has utility in a selection context. Hurtz and Donovan (2000) challenged the claim of personality as having utility for predicting performance as they argue there is only a low to moderate relationship

between personality and job performance. For conscientiousness, the personality factor purported to be the strongest predictor of performance (Barrick & Mount, 2001; Judge et al., 2013), Hurtz and Donovan contend that even when corrected for artifacts such as range restriction, this FFM trait only accounts for 4% of the variance in job performance. Thus, applying meta-analytic corrections may be misrepresentative and lead to an overestimate of the relationship between personality and performance criteria, resulting in potential legal challenges when personality-based assessments are used in practice (Barrett, 2008).

Moreover, Morgeson et al. (2007a) similarly argue that the relationship between an individual's personality and outcomes, specifically work-related, are weak at best. They claim that instead of observed validity coefficients being reported in empirical studies, specifically meta-analyses, corrections such as range restriction, predictor and criterion unreliability are instead reported. Thus, corrections often made for personality validity coefficients in research may lead to an overestimation of the utility for the test. As recommendations, Morgeson et al. propose that (a) personality tests be combined with other measures such as cognitive ability as in many cases incremental validity has been shown to be higher than when either is used separately, (b) job-related and customized personality measures may have greater face validity and therefore more acceptable to candidates and organizations, as well as (c) future research focusing on criteria more likely to be predicted by personality measures. These suggestions align with the job-relatedness of the items comprising the Personality Instrument, and in specific combinations of the personality facets, they are designed to predict specific worker attributes. Thus, the Personality Instrument may address some of these challenges to using personality for selection.

Researchers and practitioners should strive to identify the shortcomings of personality assessments, and further improve upon their utility in selection contexts. For example, Hough

and Oswald (2008) attribute the often weak relationship between personality and performance to the FFM traits being too broad for measuring personality, and instead insist that the narrow first-order personality facets which comprise the five traits of the FFM are better suited for predicting performance. Thus, while the second-order factors may be too broad, perhaps specific lower-order personality facets in various combinations may more accurately measure personality and therefore be better predictors of performance.

Narrow versus broad personality traits.

Currently, there is less agreement regarding the number of first-order personality facets and their respective labeling (Saucier & Ostendorf, 1999), often changing across organizing frameworks of personality and instruments, and therefore left open to debate. While there may be inconsistency, Judge et al. (2013) offers a generally accepted framework for conceptualizing the narrower facet-level of personality, and also include descriptions of each narrow facet. Their framework is based upon the work of McCrae and Costa (1992) and their effort to develop the Neuroticism Extraversion Openness Personality Inventory (NEO-PI). While the narrow facet labels and sometimes the trait descriptions may change across various personality instruments, they are generally similar across various personality inventories based upon the FFM.

In contrast to McCrae and Costa's (1992) NEO-PI framework and similar other personality frameworks which have an even distribution of narrow facets for each of the five broad FFM traits, the Personality Instrument was developed to have an uneven distribution of narrow facets corresponding to the three respective second-order factors of (a) Potency/Getting Ahead, (b) Social Adjustment, and (c) Discipline/Getting it Right that were identified in the 2013 Personality Instrument technical manual. An in-depth description and review of the Personality

Instrument factor structure will be presented in the following sections, further describing the intentional design and composition of the instrument.

Of relevance to this study, the Personality Instrument by the same token as other personality-based instruments measures personality using the narrower facet-level of the FFM, consequently addressing the concern related to measuring personality by the trait is too broad. Empirical research has supported the merit of using the lower-order facets of the FFM and has brought to light differing correlations with performance criteria than the broader factor level, revealing predictor-criterion relationships that may otherwise be masked (Hough, 1992; Hough & Oswald, 2008). Ones and Viswesvaran (1996) describe instruments using lower-order facets as being a narrow bandwidth personality measure that is more concrete and having clear behavioral connotations, while broad bandwidth measures of personality using the broad FFM factors are more general and often abstract (Allen & Ebbesen, 1981; Hampson, John, & Goldberg, 1986). In light of this, the Personality Instrument best falls into the category of being of narrow bandwidth as a consequence of measuring personality at the narrow facet level. As a result of using specific first-order personality facets, it may be possible to have greater predictive power for certain criteria such as job performance. By linking personality trait predictors and criteria to obtain increased predictive validity, it requires matching specific narrow facets of personality with specific job performance constructs (Barrick et al, 2001; Hough & Furnham, 2003). In doing so, it can result in increased relationships between aspects of personality and criteria, as well as greater understanding of the relationship between personality and performance (Hough & Johnson, 2013).

Thus, by measuring personality at the lower-level facets as does the Personality Instrument, personality assessments may have greater predictive validity and increased utility in

selection contexts. In this study, it is hypothesized that by assessing personality at the narrow facet level of the FFM such as measured by the Personality Instrument, the results may have implications for future measures of personality, potentially leading to greater predictive validity for performance. Similar to this challenge of whether the broad FFM factors or the narrow facets should be used when attempting to predict an individual's success in a job, another debate has persisted for decades regarding whether personality is a trait or a rather a state.

Personality Stability

Although the stability of personality over one's life is not directly relevant to the current study, it does have implications for one of the hypotheses to be tested in this study regarding instrument stability, and therefore it will be briefly discussed here. A question remains as to whether human personality is a stable and enduring trait, or if it changes over a person's life and environment. Mean-level change is referred to here as personality development that generalizes to most people, or rather normative trends regarding personality change in a particular population of people, as personality has been found to change differentially across populations (Roberts, Walton, & Viechtbauer, 2006). Costa and McCrae (1994) found support through longitudinal studies on the NEO-PI that personality traits "reach mature form in adulthood; thereafter they are stable" (p.72), generally after age 30. In later cross-cultural studies that grouped participants into age bands, McCrae et al. (1999) found evidence suggesting modest decreases in extraversion, openness to experience, and neuroticism, while modest increases in agreeableness and conscientiousness continue throughout adulthood. Although this may seem counter to their previous claim, they argue that mean-level changes in personality over time are actually relatively small. In contrast, others like Cattell (1976) strongly believe that personality develops and changes more substantially, fluctuating over a person's lifespan. Cattell believed the change

in personality was a result of interactions with others, relationships, and social experiences from family, social, cultural, biological, and genetic influences (Cattell & Mead, 2008). The theory of whether personality is stable or changes over a person's life remains in question, with evidence supporting both sides of the debate.

Personality traits as malleable. On one side of the debate, researchers have argued against the perception that personality, an individual's patterns of thoughts, feelings and behavior, is an enduring trait over a person's life (Helson, Jones, & Kwan, 2002; Srivastava, John, Gosling, & Potter, 2003). Researchers in this camp believe that personality is malleable and there is a significant variance of change throughout different stages of the human lifespan. It is possible that such changes in personality may be unique to differing populations and as the result of life events in those populations.

Longitudinal studies by multiple researchers have found evidence for mean-level changes of personality traits over various stages and age levels of a person's lifespan (Helson & Moane, 1987; Roberts, Caspi, & Moffitt, 2001; Robins, Fraley, Roberts, & Trzesniewski, 2001). These researchers believe that normative changes to personality, changes shared by a population of interest in terms of maturational or historical processes (McCrae et al., 2000), may be the result of life tasks and roles shared by the population in general (Roberts, Wood, & Smith, 2005). Helson and Wink (1992) also postulate that normative changes may be the result of biological origins, such as when adolescence generally begins in a given population. Although personality may change over time, there is still much left to be explained regarding how much it changes and what factors influence such change. If personality is in fact malleable, it may be expected that an individual's scores on the Personality Instrument may differ when taken at two points in time,

such that the duration of time between test administrations may also affect the strength of the relationship of the scores at two time points.

Personality traits as stable. However, a recent meta-analysis by Roberts, Walton, and Viechtbauer (2006) investigating mean-level changes of personality over a person's lifespan, produced results which supported previous studies that also found empirical support for the theory that personality is relatively stable. For decades, Costa and McCrae (1980) have been studying personality change and their results would suggest that an individual's personality earlier in their life is a good predictor of what his or her personality would be at age 80. Costa and McCrae (1994) investigated the NEO-PI through longitudinal studies and found support for personality traits reaching mature form in adulthood; thereafter they are stable, generally after age 30 (Roberts, Walton, & Viechtbauer, 2006). In addition, Roberts et al. acknowledge that personality may be more malleable for people under 30 years of age, most likely because younger individuals are more likely to be influenced by the environment, and after maturing in adulthood personality traits stabilize throughout the rest of a person's life.

Once in adulthood, McCrae and Costa et al. (1999) argue that although traits may develop throughout adolescence and early adulthood, traits generally remain stable and impervious to influence by environmental factors once in adulthood, and this has been found to hold across cultures (McCrae & Costa, 1994; McCrae et al., 2000). Evidence also exists to support the notion that personality may slightly change in adulthood, yet remaining fairly stable at the mean-level. For example, empirical research has shown that extraversion, openness, and neuroticism modestly decrease throughout adulthood, whereas agreeableness and conscientiousness modestly increase (McCrae et al., 1999). All the same, while McCrae and Costa found evidence showing a change in personality traits in adulthood, the change was

modest at best, and thus one would expect little change when looking at the mean-level over an individual's whole lifetime.

In summary, research on personality stability is germane to the current study as an integral hypothesis will be to investigate test-retest reliability in order to determine the stability of the Personality Instrument over time. Given that personality has been shown to be relatively stable in the short-term, it can be anticipated that if an individual takes the Personality Instrument at one point in time, and then takes the Personality Instrument again at least a week later, the two scores on the Personality Instrument at different time points will be highly correlated. However, the strength of the relationship may diminish as the duration of time between the first and subsequent administrations of the Personality Instrument increases. For example, McCrae et al. (1999) found lower retest correlations for longer durations between test administrations, and researchers have generally come to consensus that there is an exponential decay of consistency over time (Conley, 1984). Thus, although personality may be malleable at certain points in a person's life, the mean-level of personality becomes increasingly stable later in adulthood.

Although not nearly as long-standing as the question of personality stability, it has recently come to light that the context in which personality assessment items are composed may have an impact on their ability to predict performance (Bing, Whanger, Davison, & Van Hook, 2004; Bowling & Burns, 2010), and therefore a review of this literature will be presented next focusing on item specificity as a hallmark of the Personality Instrument's potential usefulness.

Considering Situation Factors: The Importance of Item Context-Specificity

One significant rationale for why personality tends to not strongly predict future job performance is the common use of generically-worded, context-free items that comprise a majority of the measures used in a selection context (Bing et al., 2004). In this section, theory

and evidence will be presented for the use of personality assessments comprised of items that are framed in a work-related context for predicting job performance and other work-related outcomes.

From a theoretical perspective, trait activation theory proposes that the situational context for which someone is experiencing is relevant when trait-consistent behaviors are appropriate for specific situations, influencing how they may behave (Tett & Burnett, 2003; Tett & Guterman, 2000). Therefore, creating items that ask respondents to answer the question when thinking about their behaviors *at work* is consistent with trait activation theory because Tett and Burnett (2003) contend that different environments can elicit different personality traits in the same individual. Thus, the extent to which a particular personality trait predicts performance is likely to be contingent on the context and situation (Judge & Zapata, 2015). Consequently, it may be assumed that the importance of a personality trait corresponds with the extent to which its manifestation is required in given work situations. Thus, it follows that an individual must possess a specific personality trait or they will not behave appropriately in a specific situation when a varying level of the particular trait of interest is required for satisfactory performance. In sum, situations present cues by which relevant traits are activated, directing an individual to behave in particular ways (Tett & Burnett, 2003).

Judge and Zapata (2015) investigated trait activation and the implications for which it relates to job performance. For many of the broad FFM traits, performance was found to be predicted in job contexts that activated specific traits, supporting the theory for which the situation exerts both and specific effects on how well personality predicts important metrics for success on the job. For example, extraversion was found to better predict performance in jobs requiring interpersonal skills and being social, whereas openness to experience had a stronger

relationship with performance for jobs having an emphasis on innovation and creativity (Judge & Zapata, 2015). Thus, the situational or context-specificity of personality assessments may be an important factor for understanding the relationship between personality and job performance. Specific situations as composed in personality-based assessment items may activate certain traits when individuals are primed to use a specific frame of reference for responding to an item. As a consequence, the item context may elicit self-report responses about context-specific behaviors or actions that reflect the personality attribute most elicited by the work-specific context of an item. Together, this evidence suggests that assessing personality in a specific context may allow for increased relationships between aspects of personality and specific work-related criteria, and therefore increasing the utility of using personality for selection purposes.

Priming participants to respond with a specific context in mind. Another theoretical perspective in relation to measurement item-specificity is the frame of reference (FOR) theory, which suggests that items composed in a specific context may influence how participants will respond to an item using a specific FOR. This is in comparison to traditional measures of personality which instruct participants to report how they think, feel and behave generally in their everyday lives (Schmit, Ryan, Stierwalt, & Powell, 1995). As an example, an item may state "I am always on time." One individual completing a personality assessment with this item may respond to it by thinking about how they would behave at work, while another person may use a reference in regards to how they would behave in their home with family. Thus, their responses may be influenced by the situations in which they envision their behavior to occur, using multiple situations for reference throughout their responses on a single assessment.

Therefore, when two individuals approach the same assessment item that was composed in a general context but respond using different FORs, it makes direct comparison between

individuals difficult or impossible. In order to remedy this issue, personality assessment items should be composed in specific contexts, such as in relation to work, so that each individual responds to items using the same FOR rather than using their own.

Recently, one solution has emerged that demonstrates potential for increasing the predictive validity of personality assessments is the addition of the FOR effect in regards to item composition (Bowling & Burns, 2010). Schmit and colleagues (1995) found that by providing the same FOR to all applicants, specifically by contextualizing items in a work-context, the ability to predict job-related outcomes may be improved. Therefore, the primary implication of this theory for providing a FOR by contextualizing personality assessment items is that the ability to predict a person's future job performance can be improved (Lievens, De Corte, Schollaert, 2008).

Contextualization in relation to the FOR theory refers to the specific context in which a particular item, or sets of items are intended to represent. For example, items may be contextualized to be specific to a work or school setting. By contextualizing items, the FOR effect primes respondents to think about specific situations when responding to a question, rather than using their own FOR when presented with a generic item which leaves open the multiple possibilities for interpretation. A personality instrument based on the FOR effect using contextualized items is thought to reduce between-group variability since all groups are primed to conceptualize the items in the same manner (Lievens et al., 2008). For example, respondents taking a personality assessment may be divided into different groups depending on how they respond to items. Whereas some individuals may respond to generic items using the same FOR throughout, others may use different FORs across the range of generic items. Thus, by

introducing the component of contextualization, between-group variability may be reduced as each group utilizes a similar FOR when responding to test items.

In addition, varying degrees of contextualization may be used when developing personality assessments based upon the FOR effect. For instance, an item may instruct a respondent to rate the extent to which they agree or disagree with the phrase "I believe most people lie to their teammates." This degree of contextualization makes reference to "teammates," in which a respondent may infer that the item is framed in the work context. Although some respondents may interpret the item as referencing work, others may reference a situation in which they are playing on a recreational sports team. Unfortunately, this form of contextualization can result in loose, non-standard interpretation across respondents, and consequently some individuals may use one FOR when responding to the example item whereas other individuals may use a different FOR.

Another type of item composition based on FOR has been referred to in research literature as tagged contextualized items (Holtrop, Born, de Vries, & de Vries, 2014), meaning that a generic personality item is modified with the addition of a tag such as "*at work*," or "*at school*." For example, an item may ask a respondent to rate the extent to which they agree or disagree with the phrase "At work, I maintain detailed notes of my conversations with clients." An item which is tagged may be more likely to prime individuals to think about how they think, feel or behave while at work, and thus respond using a FOR specific to work. Many traditional personality assessments are composed of generic items, whereas others are composed of items strictly in a work-specific context. Still others encompass a combination of both generic and work-specific items, and the Personality Instrument falls into this final category.

Currently, there is still a need for research to guide practitioners on whether it is best to have an instrument composed of all generic or contextualized items, or whether an instrument with a combination is best for predicting success. The Personality Instrument in particular is comprised of about 50% of items in a work-context and the remainder in a general everyday context, and accordingly this will be the focus of one of the hypotheses of this study. Therefore, the Personality Instrument may be a value-add to practitioners. Next, a review of empirical support for describing the advantages of using context-specific as opposed to generically-worded items will be presented.

Empirical support for the use of contextualized assessment items. Through the use of FOR, several researchers have found empirical support for improving the predictor-criterion relationship of personality assessments (Bing et al., 2004; Bowling & Burns, 2010; Heller, Ferris, Brown, & Watson, 2009; Hunthausen et al., 2003; Lievens, et al., 2008; Roenicke, 2013; Schmit et al., 1995; Wang, 2011). In addition to increased predictive validity, contextualized items have also demonstrated incremental validity over non-contextualized, generic items (Bing et al., 2004), as well as incremental validity above and beyond general mental ability more so than general context items (GMA; Hunthausen et al., 2003). Furthermore, researchers found that when personality items appear to be context-specific, respondents perceive the assessments more favorably in terms of their job-relatedness, increasing face validity (Schaffer & Postlethwaite, 2012). The key implication is that the ability to predict a person's behavior, specifically work-related outcomes, can be improved when people are given a specific context, or FOR, by which to respond to items.

Also, the use of FOR not only increases reliability, but also reduces the between-person inconsistency in item interpretation (Bing et al., 2004; Holtz, Ployhart, & Dominguez, 2005).

The increase in the reliability and criterion-related validity is assumed to be a result of increased between-person consistency (Bing et al., 2004; Lievens et al., 2008). When each respondent uses the same FOR, the between-person consistency is increased, leading to greater validity. An alternative hypothesis posits that while between-person consistency may contribute to validity, an increase in reliability may result from the reduction of within-person variability. Within-person inconsistency results when an individual responds to some items with a specific FOR, and other items with a different FOR, therefore inconsistently responding to various items depending on the FOR they envision themselves. Holden, Fekken and Cotton (1991) conceive that within-person inconsistencies relate to schematic theories of item responding, in which a person associates each item with a cognitive schema and responds accordingly (Aronson & Reilly, 2006). Therefore, respondents use schemas to process relevant information, and will choose autobiographical information from their memory which will best serve for responding to a particular item (Sanitioso, Kunda, & Fong, 1990). Depending on the item, it may activate differing schemas for which they respond.

As a result, it is plausible that a person responding to generic personality items may use differing schemas to answer various items throughout the assessment, therefore resulting in within-person inconsistency. However, in a study by Lievens et al. (2008) which hypothesized that reliability would likely be impacted when half of the items of a personality instrument are rated with one FOR while the other half of the items are rated using a different FOR. Their results showed no significant differences between the internal consistencies of the contextualized and non-contextualized items, therefore supporting the alternative hypothesis that reliability would not be affected if within-person consistency was increased. Lievens et al. also found that by imposing a FOR and reducing the between-person variability, beneficial effects can be

realized when contextualized items for conceptually relevant broad factors and relative narrow facets match the criterion. Thus, one could infer the benefits of composing personality assessment items in a work-specific context when used for selection.

However, there remains limited research regarding whether personality assessments should be composed of all contextualized items, or if they should have a combination of both work-specific and generic. Of relevance to this study, the Personality Instrument is comprised of nearly 50% work-specific items, while the remaining 50% are composed in a generic context. This study is a stepping stone for addressing this issue as the inter-item consistency of personality facets comprising the Personality Instrument will be evaluated, specifically the consistency between work-related and generic items under the same facet. In this study the aim is to investigate the within-person variability by evaluating the internal consistency of the Personality Instrument with regard to its contextualized and generic items. In summary, there are several factors which may be contributing to the low variance explained in job performance such as whether personality is measured at the broad factor or narrow facet levels, the context in which assessment items are presented to respondents, and the stability of personality over time.

Advantage of the Personality Instrument: Context-specific items. Whereas the Personality Instrument is theoretically grounded in the FFM of personality as other similar commercially available instruments, its hallmark value-add may be its item specificity. Using personality assessments composed of items specific to the workplace environment can potentially result in greater predictive validity of important work-related behaviors (Hough, 1992; Hough & Johnson, 2013; Hough & Ones, 2001), and this is because they have greater fidelity, or accuracy, of measurement for a specific construct (Hogan, Hogan, & Roberts, 1996). As such, the fidelity of an instrument, such as that of the Personality Instrument, may be increased as a result of having context-specific items as a means to match a particular criterion. As a unique attribute of the Personality Instrument, nearly 50% of the instrument is comprised of contextualized, or work-specific items. This unique aspect of the Personality Instrument has the potential to be a more effective instrument to be used by practitioners for assessing personality as a means to predict an individual's performance specifically in relation to how they may perform on the job. Thus, it is proposed that it will be valuable to further investigate the Personality Instrument in terms of its stability over time, factor structure, and also the contextualization of its items. The results may lend support for enhancing the assessment of personality when used for selection, and also provide direction for future research. Next, each of the three primary hypotheses in this study will be highlighted.

Research Hypotheses

Figure 1 is provided below, identifying the proposed model of the Personality Instrument's factor structure, including each of its fifteen arrow facets and the three proposed factors which they represent. The first two hypotheses address the factor structure of the Personality Instrument, followed by hypotheses addressing the stability and reliability of the

Personality Instrument, as well as item functioning within the narrow facets. Each of the hypotheses are described below, and a few are followed by a figure of the relationship between variables being tested.

Hypothesis 1a: Factoral validity of the Personality Assessment. Based upon past empirical research of the Personality Instrument regarding the nature of its factor structure, it was proposed that the Personality Instrument will be best captured by a model with fifteen first-order personality facets and three second-order factors. In 2013, the Company (2013) performed an EFA and discovered a proposed factor structure which was best described as having three second-order factors for which the 15 narrow facets group within. As a means to support the proposed factor structure as identified in the EFA analysis, a CFA was warranted to support the proposed factor structure by assessing the fit of the model to the data. As a byproduct of taking a model generating approach to structural equation modeling (SEM; Jöreskog, 1993), model fit indices as identified in the analyses may suggest the removal of various items from the instrument as a means for increasing its overall fit. While a shorter form of the Personality Instrument may be recommended through a model generating approach for developing a better fitting model, the intent of a potentially shorter form of the Personality Instrument would need to retain sufficient reliability and validity for use as a selection instrument.

Hypothesis 1a: Supporting the factor structure of the Personality Instrument. The Personality Instrument will yield a factor structure of 15 first-order facets and three second-order factors (see Figure 1), as described from previous validation work of the Personality Instrument.

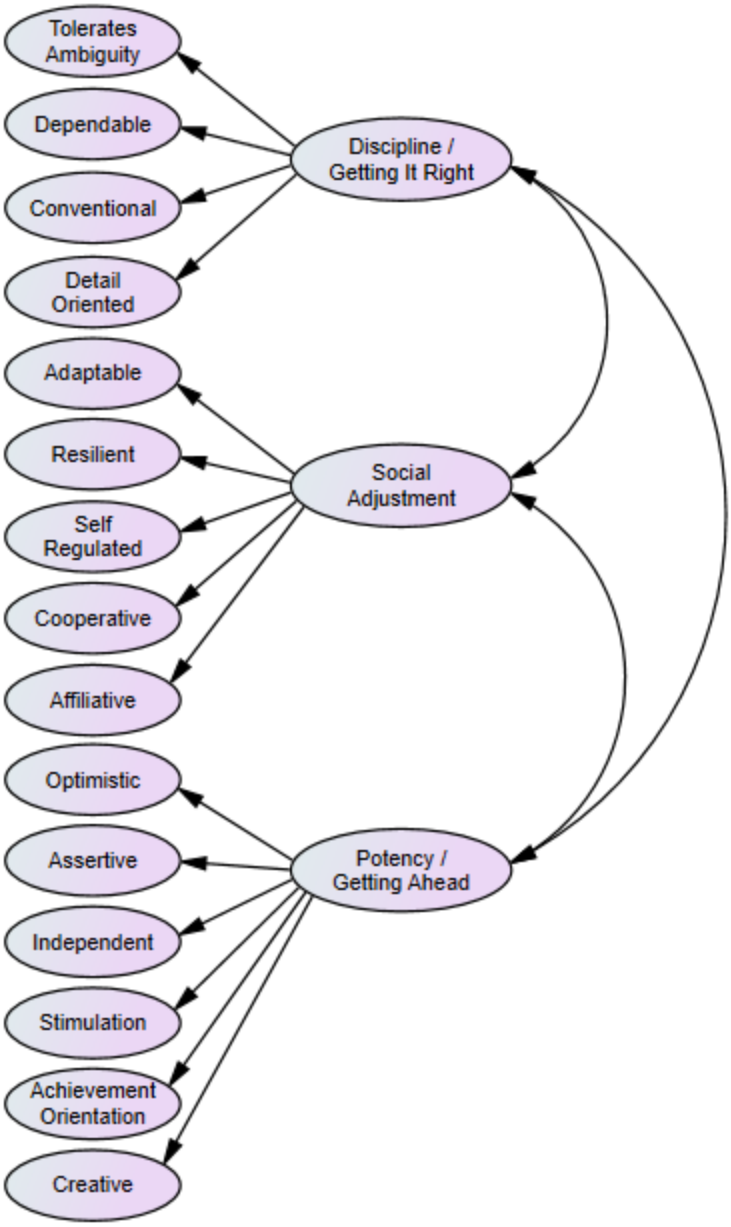

Figure 1. Proposed Empirical Factor Model of the Personality Instrument as Derived from a Previous EFA.

Hypothesis 1b: By taking a model generating approach to support the Personality Instrument's factor structure, a shorter form of the Personality Instrument will be developed by eliminating items which result in a better fitting model.

Hypothesis 2: Stability of the Personality Assessment. Decades of research have provided evidence for the stability of personality (Costa & McCrae, 1990: McCrae & Costa, 1999). Based upon a review of the literature, it is proposed that an individual's scores on the Personality Instrument will be strongly correlated with one another when taken at two points in time. Because the Personality Instrument is based upon the FFM of personality, it was expected that a strong positive and statistically significant test-retest correlation will be established.

Furthermore, the duration of time will influence the relationship between Personality Instrument scores at Time 1 and Time 2, such that a stronger positive relationship would be observed when the Personality Instrument is taken within a short, as opposed to long, duration of time between test administrations. The longer the duration of time between test administrations, the greater the probability that the strength of the test-retest correlation will decline over time (Walsh & Betz, 1985), often due to measurement error or variation in the trait of interest (Furr & Bacharach, 2014). Thus, cases in which participants took the Personality Instrument at two points in time, 12 weeks or less, would have a greater test-retest correlation than individuals who take the Personality Instrument with a duration of more than 12 weeks.

Hypothesis 2a: Scores on the Personality Instrument taken at Time 1 and Time 2 will display a strong positive and statistically significant correlation with each other (see Figure 2).

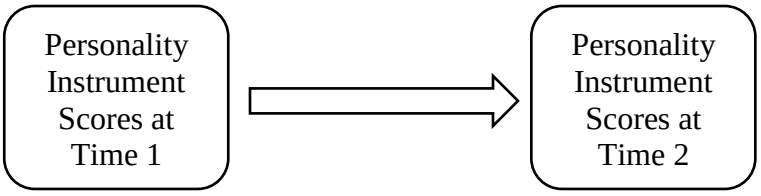

Figure 2. Hypothesized Relationship between Personality Instrument Score at Time 1 and Time 2.

Hypothesis 2b: The duration of time between administrations of the Personality Instrument will influence the strength of the relationship between Personality Instrument scores at Time 1 and Time 2, such that a shorter duration of time between test administrations will have a stronger test-retest correlation than when there is a longer duration of time between test administrations.

Hypothesis 3: Item functioning of the narrow facets. As a unique characteristic of the Personality Instrument, nearly 50% of the items are contextualized in a work context, while the remaining half of the items are generically composed. The purpose of employing contextualized items is that individuals are more likely to respond to items using the cognitive schema for which you primed the person (Aronson & Reilly, 2006), rather than pulling any schemas which first come to their mind. As such, it is likely that individuals may respond to contextualized items using specific schemas as primed by the tagging of the item, while using their own schema from memory to respond to generically worded items, which may vary across many domains of a person's life. Being that a person may respond differently to contextualized and generic items, it is proposed that the contextualized items of the Personality Instrument will be more highly correlated with each other than they will correlate with generic items representing the same narrow facet.

Hypothesis 3: Items contextualized in a work-context will correlate with each other more so than with generic items representing the same narrow facet.

CHAPTER II

Method

This study utilized archival data which was collected by the Company who owns the Personality Instrument, a proprietary personality-based instrument which is under study here. Below, information is presented on how data was collected and screened, the measure that was used and under investigation psychometric investigation in this study, and also the analyses that were applied to test the hypotheses.

Participant characteristics

Participant demographic characteristics such as gender, age, race/ethnicity, and occupation are unknown in this study as this level of data was not collected for study by the Company.

Sampling procedure. Archival data was collected from the Company and provided in two separate data sets to be used for differing hypotheses being investigated in this study. Both sets of archival data contained de-identified participant data.

Evaluating the factor structure. To evaluate that the factor structure of the Personality Instrument fit the data, a CFA, as well as follow-up EFA, were to be conducted. To investigate the factor structure of the Personality Instrument, and to ensure the capability of performing an asymptotic distribution-free (ADF) estimation if needed, 10,000 cases were randomly selected from an archival dataset of 20,000 cases which was provided directly by the Company. The dataset was comprised of item-level data corresponding to each of the 15 narrow facets of the Personality Instrument.

Investigating the stability of the Personality Instrument over time. In order to understand the stability of the Personality Instrument, test-retest reliability was conducted to examine the consistency of rank order between scores of the Personality Instrument at Time 1 and scores at Time 2. For investigating the stability of the Personality Instrument, a dataset of 1,986 complete cases was utilized from the Company's archival database of test takers who have taken the Personality Instrument at two different points in time. Data was de-identified prior to being obtained for analysis. This is a separate dataset from the dataset provided for investigating the factor structure of the Personality Instrument. Data ranged from administrations of the Personality Instrument as early as March 2009 to as most recent as of November 2015. It is possible that individual's may appear in both datasets, both for investigating the factor structure of the Personality Instrument and also the stability of the instrument.

Overall Personality Instrument scores were obtained by summing the scores of each of the 15 narrow facets. In addition, each of the three broad factor scores were obtained by summing the scores for each of their respective narrow facets. As a foundation for the overall Personality Instrument and the three broad Personality Instrument factors, each of the narrow facets were scored by summing their six respective items. Cases for inclusion in the analyses had a duration of time of 0 to 12 weeks between administrations of the Personality Instrument at Time 1 and Time 2, as well as cases which ranged from 13 weeks between test administrations to as great as 318 weeks, or approximately six years.

Sample size, power, and precision. In order to perform the different analyses for investigating the psychometric properties of the Personality Instrument, the analyses each had separate requirements for finding statistical relationships. Sample size is one of the requirements that is positively related to power, and power is defined as the probability of finding significant relationships when they actually exist, or the act of rejecting the null hypothesis when it should be rejected (Field, 2009). Following, sample sizes will be estimated for the both the CFA technique and the test-retest reliability analyses.

Evaluating the factor structure. For Hypothesis 1a, a CFA was performed on a randomly selected sample of 10,000 cases from the Company's archival database consisting of 20,000 cases. Raykov and Marcoulides (2000) suggest at the very least, sample sizes for CFAs should be greater than 10 times the number of estimated parameters. Moreover, in the event of non-normality for item-level data, it will be necessary to employ the asymptotic distribution-free (ADF) estimation method (Browne, 1984). Byrne (2010) maintains that as a rule of thumb, the ADF estimator performs best with very large sample sizes (e.g., 1,000 to 5,000; West, Finch, & Curran, 1995), which greatly reduces the likelihood of distorted estimated values and standard errors (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; West et al., 1995). Using Raykov and Marcoulides' (2000) suggested sample estimate calculation, with 90 items and needing 10 times the number of estimated parameters, 900 cases would have been sufficient. However, a random sample of 10,000 complete cases was utilized, far exceeding the minimum estimation.

Investigating the Personality Instrument's stability over time. A power analysis was conducted using Cohen's (1992) guidelines and G*Power 3.1.7 (Faul, Erdfelder, Buchner, & Lang, 2009) calculations, specifically for Hypothesis 2 and indicating that to detect a moderate

correlation (r = .30; Cohen, 1988) between scores of the Personality Instrument administered at two time points and at a .95 power, a sample size of approximately 115 participants would be necessary. Although I expect a strong, positive correlation between two administrations of the Personality Instrument, and although a small number of participants would be sufficient to detect a strong correlation, the full dataset of 1,986 cases as provided by the Company was utilized, which will be more than sufficient.

Measures and covariates

To test the proposed hypotheses, a single measure was employed which is the measure under study and to be investigated to evaluate its psychometric properties. The following section describes the Personality Instrument and its unique characteristics.

The Company's Personality Instrument. The foundational personality based scale of the Company's Personality Instrument is a 90-item measure designed to assess an individual's personality characteristics, specifically as they are displayed in the workplace context for the purpose of being used as a predictor of important work outcomes in the context of employee selection (Company, 2013). In 2013, the Personality Instrument underwent extensive validation research, which will be described later for supporting its use as a selection instrument.

Whereas the technical manual provides valuable empirical evidence to support the use of the Personality Instrument in personnel selection and also its psychometric characteristics, additional validation studies are warranted. The 2013 study focused on validity evidence based upon the Personality Instrument scale scores' prediction of work performance criteria, whereas the psychometric evaluation in this study is based upon support for the factor structure, reliability and an investigation into the relationship between contextualized and generic items. Thus, this study is not designed to replicate the 2013 study. This study adds value by building on past

validation evidence and providing further support for its reliability and validity in a selection context by assessing the Personality Instrument's stability and factor structure.

Composition of the Personality Instrument. The Personality Instrument is comprised of 15 narrow personality traits (6 items each), which are represented as narrow facets (e.g., Adaptable, Creative & Resilient; refer to Appendix C for narrow facet definitions). The narrow facets of the Personality Instrument are designated as narrow first-order facets representing aspects of the FFM traits, and are based upon over 25 years of published personnel selection research regarding the meaning of personality and how specific personality attributes are related to job success (Barrick & Mount, 2012; Hough & Ones, 2001; Ones, Dilchert, Viswesvaran, & Judge, 2007). While not evenly represented, the Personality Instrument narrow facets are intended to comprise in various combinations all five personality factors of the FFM (Company, 2013); seven of the narrow facets capture aspects of conscientiousness, six capture aspects of emotional stability, four capture aspects of extraversion, another four capture aspects of openness to experience, and the final one captures agreeableness.

In the Personality Instrument technical manual, the author of the scale has described this distribution of Personality Instrument narrow facets across the FFM as representing the strength of the research evidence demonstrating the importance of each of the personality factors for predicting important work behaviors (Company, 2013). For example, research has generally shown that conscientiousness has the strongest correlation with a candidate's future job performance (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Dudley, Orvis, Lebiecki, & Cortina, 2006, Hough & Ones, 2001; & Schmidt & Hunter, 1998), and as a result the personality factor of conscientiousness is represented more so than the other FFM personality factors. In addition to the narrow facets, an exploratory factor analysis is described in the

technical manual and suggests three overarching second-order personality factors (Potency/Getting Ahead, Social Adjustment, & Discipline/Getting it Right), or global facets as Cattell (1965) would describe them. While the Personality Instrument is represented by each of the five FFM broad factors, empirically the items statistically cluster to form only three broad factors. Potentially, this is the result of items comprising each of the narrow facets as being somewhat heterogeneous, and further clarification of implications for this will be highlighted in the discussion section.

Item development. Decades of personality research has resulted in varying frameworks, such as the FFM of personality which remains the most utilized model of personality in the realm of assessment and selection. Research regarding the FFM has created a framework, or common language, for understanding and measuring the complex construct of personality. The development of the Personality Instrument utilized this framework for developing specific questions that measure each of the Personality Instrument's narrow facets. Each of the Personality Instrument's narrow facets are a sum of the item-level scores that capture a small number of related FFM facet scales. While some of the Personality Instrument's narrow facets are intended to essentially have the same meaning as FFM facets, some of the narrow facets are somewhat more complex and represent multiple FFM facets. Specifically, the International Personality Item Pool (IPIP; Goldberg, 1999) was used as a basis for item development as the IPIP stands at over 3,000 items, many of which align to IPIP scales that measure constructs that are the same to those existing in other personality inventories (Goldberg et al., 2006).

The IPIP is a valuable public resource for understanding and measuring personality based upon the FFM, and other personality frameworks. Since its inception, over 450 IPIP personality scales have been developed to measure over 270 constructs such as altruism, achievement-

striving, competence, dominance, etc. (Goldberg et al., 2006). The IPIP website (http://ipip.ori.org/) is intended to be a public domain available for use by researchers and practitioners as a means to aid the development of new personality instruments, for comparing their own instruments against, and ultimately as an international effort to develop and refine existing personality inventories. Therefore, and as with many other personality instruments, the IPIP has been a strong influence for the framework of the Personality Instrument, primarily because of its rigor and historical evidence.

As noted earlier, the Personality Instrument represents facets of each of the five broad factors of the FFM. The Personality Instrument pulls from item stems as described in the IPIP, which are intended to represent a number of different personality instruments with similar frameworks of personality. Specifically, items of the Personality Instrument are derived from item stems of IPIP scales corresponding with personality facets comprising the (a) Multidimensional Personality Questionnaire (MPQ; Tellegen, 1995; 2003), (b) Costa and McCrae's (1992) NEO-PI-R Facets, (c) Temperament and Character Inventory (TCI; Cloninger, et al., 1994), (d) California Psychological Inventory (CPI; Gough, 1996), (e) Hogan and Hogan's (1992) Hogan Personality Inventory (HPI), (f) Cattell's 16 Personality Factor Questionnaire (16PF; Conn & Rieke, 1994), (g) Values in Action Character Survey (VIA; Peterson & Seligman, 2004), and (h) Lee and Ashton's (2004) HEXACO-PI. Whereas item stems were derived from several corresponding instruments, there is an oversampling of item stems specifically covering constructs measured by the NEO-PI.

One of the features of the Personality Instrument is that it is intended to measure personality at the lower-order facet level for capturing complex combinations of narrow attributes important for the prediction of specific work behaviors. Therefore, the Personality

Instrument potentially explains more variance in the dependent variable by keeping the predictors specific and narrow versus broad. Each of the narrow facet items are written as a self-descriptive statement in which individuals respond to statements on a 4-point Likert-type scale ranging from *strongly disagree* to *strongly agree*. Examples of items are "I think it is important to have fun at work," "I show up for work on time, every day," and "I like to keep to myself at work." A distinction of the Personality Instrument from other commercially available personality assessments is that approximately 50% of the Personality Instrument items are written in a work context, and the remaining items are of a general context more similar to other personality assessments. This is a clear distinction of the Personality Instrument, warranting further research to refine and shorten the instrument for better predictive ability and potentially a shorter version, resulting in greater utility. Thus, the Personality Instrument displays distinguishing characteristics which render it unique amongst other currently available personality instruments.

Personality Instrument composites. Besides the feature of having contextualized items, the Personality Instrument items are intended to also create composites, which are narrow facets, and they are intended to represent some aspect of each of the FFM broad traits, as previously described. While eight of the narrow facets were designed to capture aspects of a single broad attribute, and seven were originally designed to capture aspects of two broad attributes. For this study, a more recent rationale for trait mapping of the Personality Instrument items to FFM item stems in the IPIP is presented in Table 2, showing a slightly different distribution of FFM traits being measured by the 15 narrow facets. This more recent item stem mapping is important for multiple hypotheses under investigation in this current study. Furthermore, the result of this recent mapping identified eight of the Core Psychological Traits as measuring aspects of

conscientiousness, seven measuring aspects of neuroticism, six measuring aspects of openness and extraversion, and just three measuring aspects of agreeableness. Consequently, the mapping provides further evidence of the heterogeneous nature of many of the narrow facets, which can have important effects upon the psychometric properties of the Personality Instrument.

Table 1. *Mapping of Personality Instrument Narrow Facets to FFM Traits.*

Personality Instrument narrow facets	Openness to Experience	Conscient-iousness	Extra-version	Agreeable-ness	Neuroticism / Emotional Stability
Achievement Orientation		x			
Adaptable	x		x	x	x
Affiliative			x		x
Assertive			x		
Conventional	x	x			x
Cooperative				x	
Creative	x				
Dependable		x			
Detail Orientation	x	x			
Independent	x	x	x		
Optimistic		x			x
Resilient		x	x		x
Self-Regulated	x				x
Stimulation			x		
Tolerance for Ambiguity	x				

Previous psychometric evaluation of the Personality Instrument. Beyond describing the Personality Instrument's development, the technical manual provides empirical data regarding the Personality Instrument, supporting the appropriateness of its use as a personality selection instrument. The technical manual provides psychometric information such as: (a) means and standard deviations for each of the narrow facets, (b) average raw scores across various industries, (c) internal consistency of each narrow facets, as well as the intercorrelations among narrow facets, and the (d) factor structure. Much of this information can be found in the technical manual provided by the Company.

In 2013, a sample size of 132,040 complete cases was used to investigate the psychometric properties of the Personality Instrument. Internal consistency was assessed by calculating the Cronbach's alpha estimate of reliability for the narrow facets, and was based on the number of corresponding items. It is noteworthy that the range of Cronbach's alphas across the 15 narrow facets, was .11 for the Adaptable dimension to .64 for Affiliative dimension. Implications for this will be discussed in detail, but in short this may stem from the intentional design and resulting heterogeneous composition of several of the narrow facets. For example, Affiliative with the highest Cronbach's alpha coefficient is nearly a unitary construct, which is comprised of five items representing the FFM broad factor of Extraversion and one item representing Neuroticism. Dissimilarly, Adaptable has items representing four broad FFM factors; the factors are openness, extraversion, agreeableness and neuroticism. Therefore, Adaptable likely has a resulting lower Cronbach's alpha as a direct result of its multidimensionality. Again, this will be discussed further, especially pertaining to implications, in the following sections.

Previously, the factor structure of the Personality Instrument's narrow facets were analyzed using exploratory factor analysis (EFA), with the number of factors being determined by the Kaiser Method (i.e., eigenvalue > 1.0; Kaiser, 1960). As a result, three second-order factors emerged among the narrow facets through Varimax-rotated principal components that accounted for approximately 50% of the variation in the covariance matrix. The three broad factors were named (a) Potency/Getting Ahead, (b) Social Adjustment, and (c) Discipline/Getting it Right. This is an interesting finding because the Personality Instrument's framework is based upon the FFM, in which it could have been assumed that five second-order factors would have been expected to emerge. Hypothetically, and of which will be further

discussed, this may be a result of the differential focus on FFM personality factors across the Personality Instrument's narrow facets, in which FFM factors being more important than others for predicting important work behaviors are more represented than other FFM factors (e.g., Conscientiousness; Barrick, Mount, & Judge, 2001; Dudley et al., 2006, Hough & Ones, 2001). Thus, the Personality Instrument is a unique and interesting personality assessment, and therefore additional research should be conducted to further support its use as an instrument in the realm of personnel assessment and selection.

Research Design and Statistical Analysis

Research design. This study employed a non-experimental associational approach because there is no treatment, and both independent and dependent variables vary widely, ranging from low to high and therefore indicative of continuous variables (Gliner, Morgan, & Leech, 2009). No independent variables were manipulated for this study to investigate a cause and effect, nor compare groups of participants. Archival data was used to evaluate the psychometric properties of the Personality Instrument, specifically a) the factor structure using CFA techniques, b) establish reliability by performing test-retest reliability coefficients, and c) investigating the relationship between contextualized and generic items comprising the same narrow facet.

Statistical analyses. The purpose of this study was to test the factor structure of the Personality Instrument, as well as assess its reliability for measuring personality. SPSS AMOS v.23 (IBM Corporation, 1983; 2015) was employed to determine whether the empirical measurement model of the Personality Instrument that is comprised of its 90 items under the 15 narrow facets and their respective three broad factors, as identified in the 2013 psychometric evaluation study, fit the data by performing a CFA. SPSS AMOS would allow for testing the fit of the measurement model by identifying the chi-squared ($\chi 2$) model fit indices (e.g., *GFI, CFI, RMSEA*). An EFA was also performed as follow-up to the CFA analyses using IBM SPSS v.23 (IBM Corporation, 2015). An EFA allowed for exploring the factor structure of the Personality Instrument, and identifying whether the same proposed factor structure that emerged in the 2013 validation would emerge again using new and more recent data.

To estimate the stability, or reliability, of the Personality Instrument, SPSS v.23 (IBM Corporation, 2015) was employed. A simple test-retest reliability analysis was performed using

Pearson bivariate correlations between test administrations of the Personality Instrument at Time 1 and Time 2. This analysis was performed for each of the Personality Instrument's narrow facet scales, the three broad factors, and also at the overall Personality Instrument.

Lastly, SPSS was again used to investigate whether contextualized items correlate more so with each other than do generic items representing the same narrow facet. Simple Pearson bivariate correlations were computed between contextualized and generic items, allowing for an average correlation among each type of item to be calculated.

CHAPTER III

Results

To evaluate and find support the factor structure of the Personality Instrument, SPSS Amos v.23 (IBM Corporation, 1983; 2015) was employed to perform a CFA as a means to support the empirical factor structure of the Personality Instrument as revealed in the technical manual. AMOS was used to test the factorial validity of the measurement model, in which the model fit was compared against the proposed model (see Figure 1), as well as other models as the proposed empirical model did not fit the data. To the extent that the proposed model did not show at least an adequate fit as indicated by the fit indices (e.g., *χ2, RMSEA, CFI, GFI)*, other alternative factoral models of the Personality Instrument were also evaluated.

Preliminary Analyses

Prior to testing any hypotheses, the scale composites were calculated and preliminary analyses were performed to assess the data for normality, identify outliers, check for missing data, compute descriptive statistics such as means and standard deviations, and also alpha coefficients.

Normality and reliability. As a result of the large sample sizes, item and scale normality are often not required for assessment using skewness and kurtosis values as Field (2009) describes samples of 200 or more cases giving rise to small standard error. This often results in significant z-scores for skewness, potentially even when data is normally distributed. However, skewness and kurtosis values were still computed as an exploratory investigation. Aside from evaluating skewness and kurtosis values, a visual inspection of histograms revealed adequately normal distributions for study variables. Skewness values were acceptable indicating a normal distribution and ranged from -.481 to .094, and kurtosis values similar ranged from -.073 to .462. No transformations were made to any items or scales.

Descriptive statistics such as means and standard deviations for each of the narrow facets were calculated using SPSS v.23, along with correlations between each of the narrow facets, as well as Cronbach's alpha coefficients. A summary of descriptive statistics for the narrow facets are presented in Table 3, and their reliability estimates and correlations among other narrow facets is provided in Table 4. Similarly, descriptive statistics and correlations among the broad Personality Instrument factors and the overall Personality Instrument scale are provided in Table 5.

Table 2. *Narrow Facets Means and Standard Deviations.*

Variables	Mean	SD
Achievement Orientation	6.59	4.33
Adaptable	1.99	2.99
Affiliative	8.07	5.13
Assertive	4.24	4.32
Conventional	1.78	4.19
Cooperative	6.40	4.19
Creative	1.23	3.86
Dependable	10.99	3.98
Detail Orientation	8.52	4.06
Independent	5.31	3.57
Optimistic	9.57	4.19
Resilient	5.83	4.37
Self-Regulated	6.47	3.75
Stimulation	3.81	5.41
Tolerance for Ambiguity	1.60	3.46

N = 10,000.

Table 3. *Bivariate Correlations and Reliabilities among Narrow Facets.*

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Achievement Orientation	(.29)														
2. Adaptable	.07**	(.06)													
3. Affiliative	.26**	.25**	(.58)												
4. Assertive	.23**	.12**	.32**	(.38)											
5. Conventional	.12**	-.08**	.06**	.02	(.23)										
6. Cooperative	.16**	.29**	.41**	.12**	.05**	(.39)									
7. Creative	.13**	.14**	.19**	.28**	-.07**	.09**	(.30)								
8. Dependable	.21**	.08**	.28**	.02*	.21**	.35**	-.06**	(.43)							
9. Detail Orientation	.21**	.03**	.16**	.09**	.24**	.24**	-.04**	.40**	(.39)						
10. Independent	.29**	.05**	.14**	.34**	.02	.04**	.29**	.10**	.10**	(.22)					
11. Optimistic	.31**	.19**	.36**	.31**	.07**	.30**	.18**	.34**	.30**	.29**	(.46)				
12. Resilient	.23**	.23**	.33**	.30**	.05**	.29**	.18**	.23**	.16**	.22**	.36**	(.30)			
13. Self-Regulated	.12**	.22**	.25**	.06**	.06**	.32**	.04**	.26**	.29**	.04**	.29**	.20**	(.13)		
14. Stimulation	.24**	.16**	.27**	.38**	-.05**	.12**	.37**	-.09**	-.11**	.25**	.25**	.23**	.06**	(.48)	
15. Tolerance for Ambiguity	.05**	.20**	.17**	.14**	-.14**	.10**	.19**	-.05**	-.11**	.08**	.07**	.22**	.05**	.23**	(.05)

Note. Values in parentheses on diagonal are Cronbach's alpha coefficients. * $p < .05$. ** $p < .01$. $N = 10{,}000$.

Table 4. *Means, Standard Deviations, and Correlations among Overall Personality Instrument and Broad Factors.*

Variable	Mean	SD	1.	2	3	4
1. Personality Instrument	82.47	29.66	---			
2. Discipline / Getting it Right	28.79	13.46	.66**	---		
3. Social Adjustment	30.78	16.27	.82**	.38**	----	
4. Potency / Getting Ahead	19.69	8.06	.81**	.44**	.44**	---

Note. **Correlation significant at the .01 level (2-tailed). N = 10,000.

Confirmatory Factor Analysis to Evaluate Model Fit

In order to perform a CFA using AMOS, there must be no item-level missing data; the dataset used in this study met the requirement because only complete cases were pulled from the Company's archival database. In conducting a CFA, I employed a model-generating approach (Jöreskog, 1993) to develop a measurement model of the Personality Instrument derived from the EFA as previously performed for the technical manual in 2013. The proposed model explained nearly 50% of the variance and was comprised of 15 narrow facets, (i.e., narrow facets), and three broad factors (Company, 2013). The purpose of the CFA was to assess the validity of the factor structure of the Personality Instrument identify a shorter but also well-fitting model.

Testing the original proposed Personality Instrument model. To test Hypothesis 1a, evaluating the empirical model which was found in the 2013 EFA, a model in AMOS was created that included 15 latent variables (narrow facets) and their respective observed variables, or items. Additionally, the three second-order factors were created representing the broad factors of the Personality Instrument of which the narrow facets comprise. Items were loaded onto their respective dimensions as reflective indicators, such that causality flowed from latent variable to observed indicator (Byrne, 2010). The measurement model is shown in Figure 3 as it appeared in AMOS.

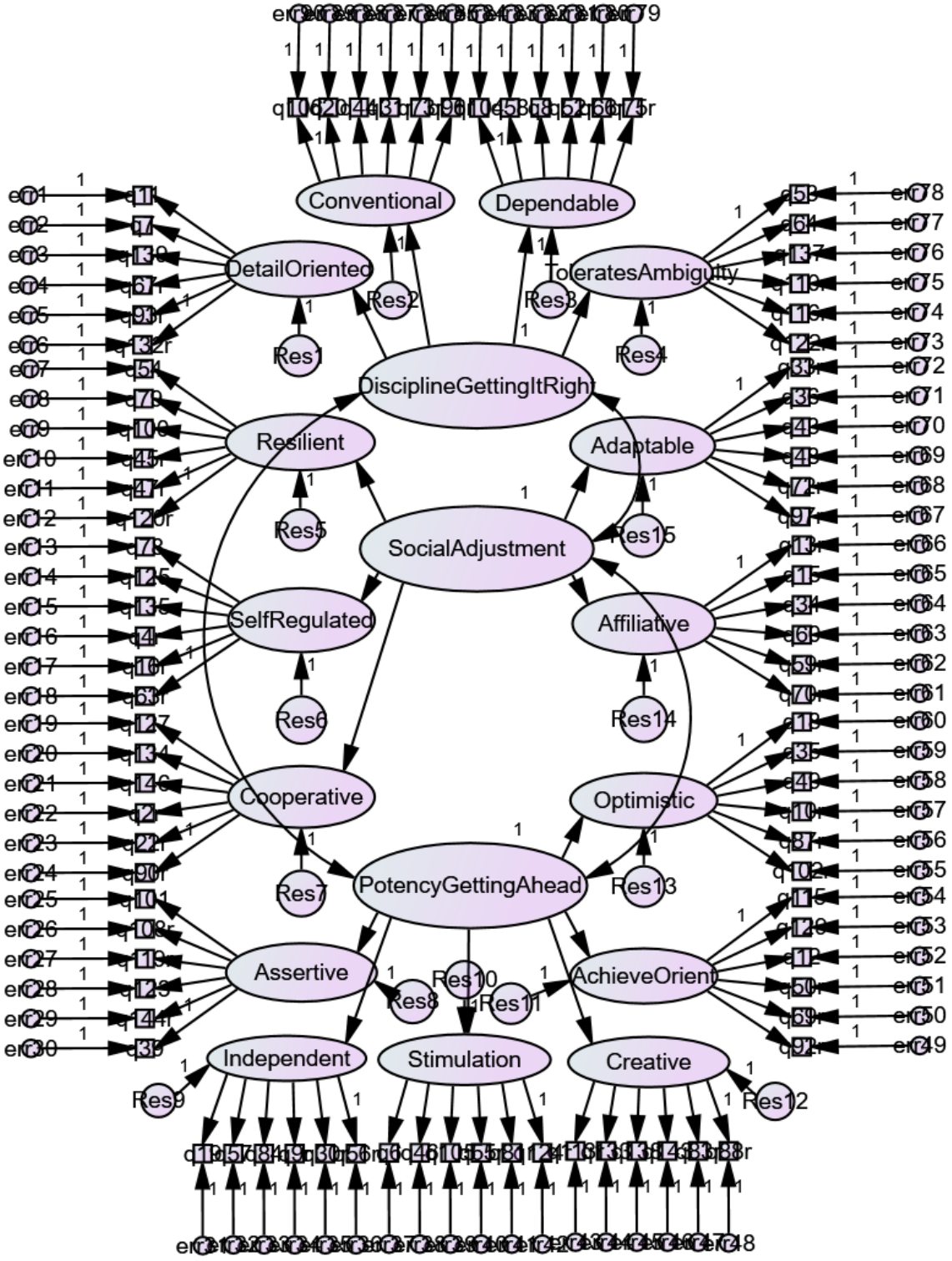

Figure 3. Original Empirical Measurement Model as Constructed in AMOS. $\chi 2$ (3,897) = 94,548.14, $p < .001$; *GFI* = .704; *CFI* = .433; *RMSEA* = .048.

Using the model generating approach (Jöreskog, 1993), the measurement model was evaluated by assessing multiple indicators of fit such as the chi-square difference test, standardized regression weights, modification indices (MIs), and also the fit indices. For example, the primary fit indices assessed were the goodness of fit index (GFI), comparative fit index (CFI; Bentler, 1990), and root mean square of approximation (RMSEA; Byrne, 2010). For clarification, each of the model fit indicators are briefly described below.

The chi-square difference test measures the difference between the covariances predicted by the model and the population covariance matrix and assesses overall model fit. The chi-square difference test can also be used to determine whether one model shows statistically significant improvement in fit in comparison to another model (Byrne, 2010). This test can be described as an accept-support test where the null hypothesis represents the confidence that the model is correct (Kline, 2016). By failing to reject the null hypothesis as a result of the absence of statistical significance (i.e., $p < .05$), the model is therefore supported. Thus, the logic is backwards compared to typical tests of statistical significance where rejecting the null hypothesis supports the theory. However, it is important to note the chi-square difference test is sensitive to large sample sizes, resulting in greater chances that even changes in model fit seem negligible they are likely to result in statistical significance. Another indicator is the GFI which is a measure of the relative amount of variance and covariance in the sample data jointly explained by $\sum$, or the population covariance matrix (Byrne, 2010). This is an absolute index of fit as it compares the hypothesized model with no model at all (Hu & Bentler, 1995). GFI values range from 0 to 1.0, with 1.0 being indicative of good fit.

To balance the sensitivity to sample size of in regards to the chi-square difference test, the CFI is less sensitive not only sample size, but also the number of variables in the model. The

CFI compares the amount of departure from close fit for the hypothesized model against that of the independence (null) model (Kline, 2016). CFI values range from 0 to 1.0, where 1.0 is the best result; a value greater than .95 indicates a strong fit (Byrne, 2010; Hu & Bentler, 1999).

The third goodness-of-fit index is RMSEA, which is a badness of fit index with values ranging from 0 to 1.0, with values less than .05 indicating a good fit (Brown & Cudeck, 1993). RMSEA therefore measures the departure from close or approximate fit (Kline, 2016). However, RMSEA tends to penalize for lack of parsimony and is influenced by large samples, as well as large numbers of variables and degrees of freedom (Erdogan, Ok, & Marcinkowski, 2012). Kenny (2015) states that models with low sample sizes and small degrees of freedom can have artificially large RMSEA values, an opposite problem with the data set used for this study, as well as the complexity of the Personality Instrument itself.

The analysis of the original empirical Personality Instrument model illustrated in Figure 3 and subsequent modifications are presented in Table 6. At first blush, overall model fit was poor ($\chi 2$ [3897] = 94548.14, $p < .001$; *GFI* = .704; *CFI* = .433; *RMSEA* = .048). Upon further review of the results, two regression paths were found to be non-significant and nearly 20% of the standardized regression weights were negative. Additionally, three of the narrow facets (i.e., Detail Orientation, Tolerates Ambiguity, & Adaptable) portrayed negative residuals. This is an indication that some of the exogenous variables had an estimated covariance matrix that was not positive definite, rendering the solution in-admissible. Jöreskog and Sörbom (1984) suggest that this occurrence, often referred to as Heywood cases, is the result of either the model being incorrectly-specified, or potentially a sample size that is too small. Heywood cases are usually identified when impossible values are found, such as correlations larger than 1, or as in the case of the empirical Personality Instrument model, negative variances (Kenny, 2011). Heywood

cases may be caused by a number of issues, such as outliers, non-convergence and underidentification, empirical underidentificaiton, and structurally misspecified models (Chen, Bollen, Paxton, Curran, & Kirby, 2001). To rule out sample size as reason for the Heywood case, Hoetler's sample size estimate was reviewed in the results which recommended a sample size of 435 cases for a probability of $p < .01$, which is well below the current sample size used (N = 10,000).

One potential solution for preventing negative variance estimates and the occurrence of inadmissible solutions is to restrict the search for a solution to parameter values that are admissible, such as 1.0 (Byrne, 2010). For Model 2, the variances of the three residuals were accordingly constrained to a value of 1.0 and the model was re-ran. The result was an increase in chi-square, and an increase in degrees of freedom, while the fit indices also declined. As a third iteration, Model 3, the two non-significant regression paths were removed for the items q79 (Resilient) and q110 (Tolerates Ambiguity), and the model re-ran. This led to a reduction in chi-square and degrees of freedom, and also slightly better fit indices. Next, the MIs were reviewed and it was identified that if error variances err55 (q102) and err56 (q87) of Optimistic were allowed to covary, chi-square would be reduced by 1986.60. For Model 4, the error variances of Optimistic were allowed to covary, resulting in the chi-square to be reduced, along with the fit indices improving slightly.

For Model 5, the MIs were observed and it was determined that if the error variances err88 (q44) and err89 (q20) of Conventional were allowed to covary, chi-square would again be reduced by 1602.91. The model was re-ran and the chi-square was reduced, but little to no change in the modification indices was achieved. After further review of the output, it was identified that the covariance matrix of the three broad factors (i.e., Potency/Getting Ahead,

Social Adjustment, and Discipline/Getting it Right) was not admissible. Thus, the result was an indication that the model was not correctly specified and did not fit the data. Therefore, the empirical model of the Personality Instrument can be deemed as a poor fitting model to the data.

Table 5. *Nesting Table Showing Model Comparisons of Empirical Model.*

Model	X^2	*df*	Model Comp-arison	ΔX^2	Δdf	Fit1 (GFI)	Fit2 (CFI)	Fit3 (RMSEA)
M1	94548.14	3897				.704	.433	.048
M2 - Constrain Variance to 1.0 for Residuals 1, 4 & 15	104725.67	3900	2 vs. 1	10177.53**	3	.692	.369	.051
M3 - Removed non-sig. regression paths for q79 & q110	101512.10	3725	3 vs. 2	3213.57**	175	.695	.376	.051
M4 - err55 <-> err56	99229.60	3724	4 vs. 3	2282.5**	1	.705	.391	.051
M5 - err88 <-> err89	97475.88	3723	5 vs. 4	1753.72**	1	.705	.392	.051

**$p < .01$.

Testing a single and broad general factor model. As a consequence of the factor structure of the empirical model failing to be supported by the first model, it was next decided to attempt a single second-order factor structure of the Personality Instrument. This approach was consistent with MacCallum, Wegener, Uchino and Fabrigar's (1993) recommendation to consider alternate factor structures. Whereas a single general factor of personality did not appear in the 2013 EFA (Company, 2013), the CFA attempt highlighted previously to this which tested the empirical model had showed a strong and positive relationship between the three second-order factors. Accordingly, since the three broad factors of the Personality Instrument were correlated strongly with each other, (range of .66 to .82; see Table 5), it seemed reasonable to test this model. A single second-order factor model of the Personality Instrument can be found in Figure 4.

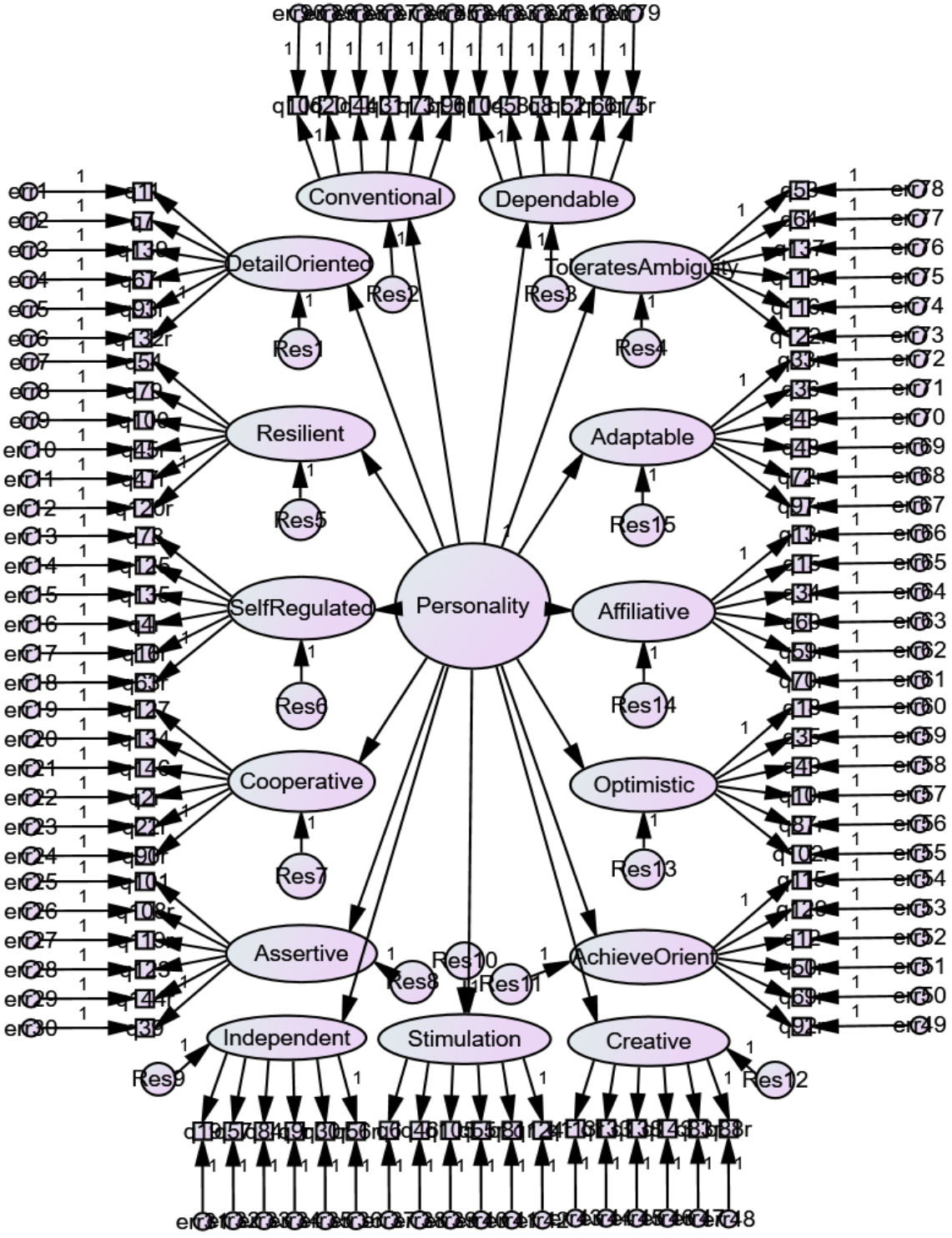

Figure 4. Single Second-order Empirical Measurement Model as Constructed in AMOS. χ^2 (3,900) = 96,837.26, $p < .001$; *GFI* = .694; *CFI* = .418; *RMSEA* = .049.

Similar to the original empirical model, the fit of the single second-order factor model was poor (χ^2 [3900] = 96837.26, $p < .001$; *GFI* = .694; *CFI* = .418; *RMSEA* = .049). Therefore, further review of the results was encouraged, and three non-significant regression paths were found, and 24% of the standardized regression weights were negative. Additionally, two of the narrow facets (i.e., Tolerates Ambiguity, & Adaptable) had residuals which were negative. Like the original empirical model, the estimated covariance matrix was not positive definite, and consequently the solution was not admissible as a result of two Heywood cases. Again, sample size was not an issue because Hoetler's sample size estimate recommendation was 425 for a probability of $p < .01$, which is well below the sample size employed in this study which was 10,000 cases. Thus, the single-second order factor model was also not a good fitting model to the data, and additional model generating modifications were refrained from being ran.

Supplementary analysis for testing the theoretical model as Personality Instrument items represent FFM traits. As another means to test Hypothesis 1a, a third attempt to find support for the factor structure of the Personality Instrument was performed and the theoretical model of the Personality Instrument was tested. The theoretical model was based upon the original framework for which the Personality Instrument was designed. Each of the 90 items corresponding to the narrow facets were mapped onto their respective FFM broad factors of which they were found to rationally align to IPIP narrow personality facets. This process entailed reviewing the item stems as provided on Goldberg's IPIP website, identifying which personality instrument they rationally represented, such as the NEO-PI-R (Costa & McCrae, 1992) or the Hogan Personality Inventory (HPI; Hogan & Hogan, 1992). Therefore, a more current mapping of the Personality Instrument narrow facets to the FFM was imperative to further investigate the factor structure. As a result, each of the 90 items were aligned to their respective FFM trait. The theoretical Personality Instrument model can be found in Figure 5.

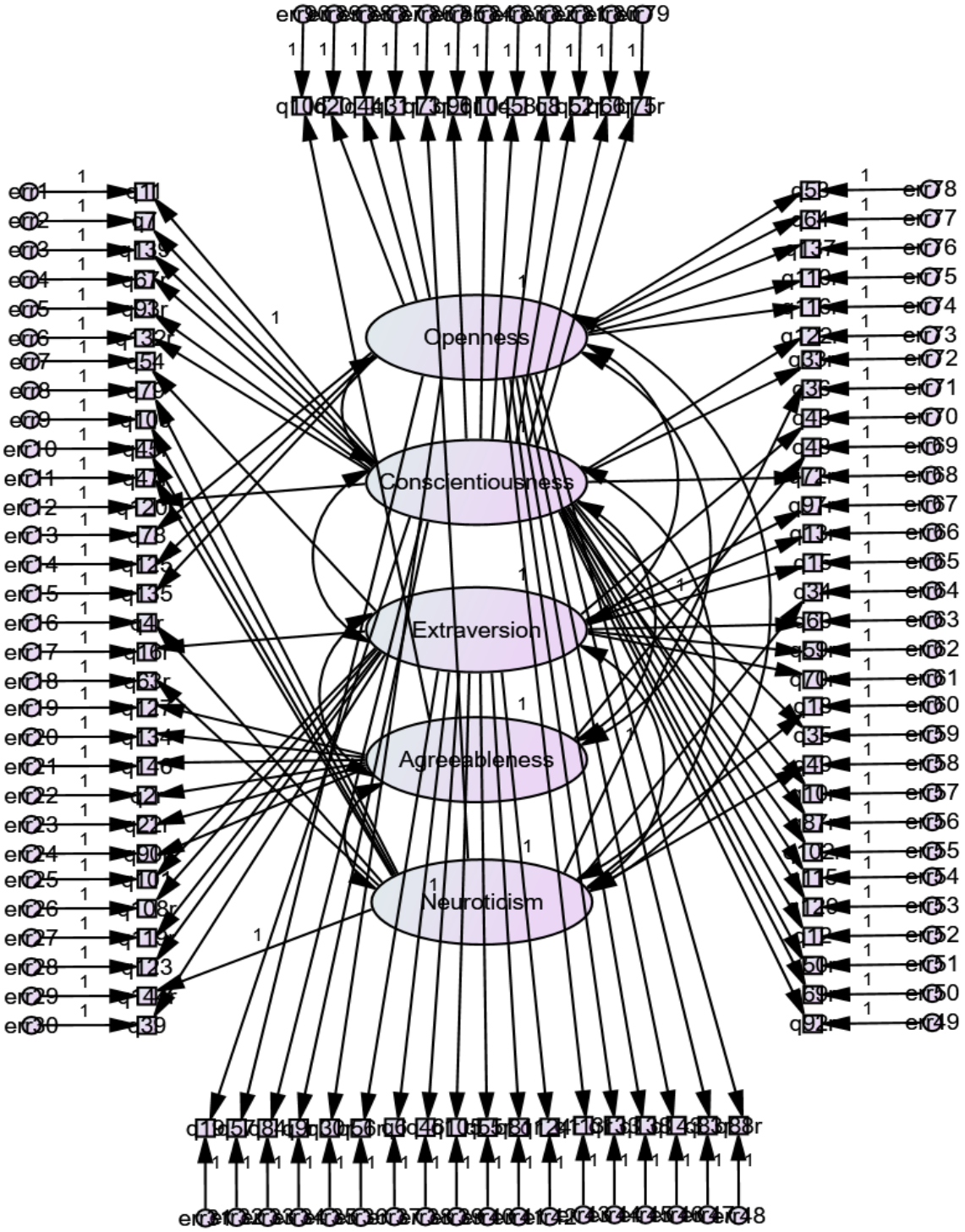

Figure 5. Theoretical Measurement Model as Constructed in AMOS. $\chi 2$ (3,910) = 102,525.28, $p < .001$; *GFI* = .692; *CFI* = .383; *RMSEA* = .050.

As with the original empirical model and the single second-order factor Personality Instrument models, the fit of the theoretical model was poor ($\chi 2$ [3910] = 102,525.28, $p < .001$; *GFI* = .692; *CFI* = .383; *RMSEA* = .050). Albeit poor fit results, the results of the analysis were reviewed and two non-significant regression paths were found, q137 and q79, which could be removed for an improved fit. Additionally, by reviewing the standardized regression paths, about 14% of the standardized regression weights were identified as being negative. For Model 2, the two non-significant regression paths were removed, and the model was re-ran. The results and fit indices of Model 2, and the subsequent alternative models can be found in Table 7. The chi-square and degrees of freedom were greatly reduced, however, there was little change in the fit indices. Next, the regression weights were reviewed and it was found that there were no non-significant regression paths in this instance, which next led to a review of the MIs. In order to improve fit for Model 3, the error variances err55 and err56 which map to Conscientiousness were allowed to covary, and the model was re-ran again. However, only slight improvements to model fit were evident.

For Model 4, the MIs were reviewed and error variances err88 and err89 of openness were allowed to covary, and as a result it a very slight improvement in fit was observed. To further improve the fit of Model 5, the MIs suggested allowing error variances err37 and err42 of extraversion to covary. Once again, there was only a very slight improvement in the fit indices. However, a non-significant regression path to emerge was found to have emerged, and therefore in Model 6 the non-significant regression path of q108 was removed and the model was re-ran again. While the model did improve, it warranted further modification. For Model 7, the MIs suggested allowing error variances err54 and err53 to covary as a means for improving fit. By allowing the two error variances of conscientiousness to covary, little improvement in fit was

experienced. Lastly, for Model 8 the MIs were followed, allowing error variances err10 and err11 of neuroticism to covary. Unfortunately, the result was a message indicating that the covariance matrix was not positive definite as a result of Heywood cases, and thus it was decided to stop any further modifications. Therefore, both the empirical and theoretical models failed to fit the data.

Table 6. *Nesting Table Showing Model Comparisons of Theoretical Model.*

Model	X^2	*df*	Model Comp-arison	ΔX^2	Δdf	Fit1 (GFI)	Fit2 (CFI)	Fit3 (RMSEA)
M1	102525.280	3910				.692	.383	.050
M2 - removed q137 & q79	100742.610	3735	2 vs. 1	1,782.67	175	.690	.387	.051
M3 - e55 < - > 56	98507.808	3734	3 vs. 2	2,234.80	1	.698	.401	.050
M4 - e88 < - > e89	96712.187	3733	4 vs. 3	1,795.62	1	.702	.412	.050
M5 - e37 < - > e42	95810.713	3732	5 vs. 4	901.47	1	.702	.418	.050
M6 - remove q108	93515.699	3646	6 vs. 5	2,295.01	86	.706	.424	.050
M7 - e54 < - > e53	92677.386	3645	7 vs. 6	838.31	1	.708	.427	.049
M8 - e10 < - > e11	91955.662	3644	8 vs. 7	721.72	1	.711	.434	.049

**$p < .01$.

Testing the model fit of the Personality Instrument using the multi-trait multi-method approach to structural equation modeling. As a last effort, the multi-trait multi-method (MTMM) SEM approach was pursued for finding support for the factor structure of the Personality Instrument. MTMM was originally proposed by Campbell and Fisk (1959), and continues to be used as statistical technique for assessing convergent and discriminant validity. For this approach, each measure should load onto its trait and method factors, and it is generally assumed that the trait and method factors are independent. To perform a MTMM SEM CFA, it requires at least three traits and methods for this approach to be identified (Kenny, 2012), which is met with the Personality Instrument model. As a result of the non-positive definite covariance matrices in the empirical and theoretical models, the MTMM approach can help to alleviate Heywood cases (Kenny & Kashy, 1992). Therefore, the MTMM SEM approach for performing a CFA was an appropriate analysis.

The model was a 20-factor correlated trait-correlated method (CT-CM) model in which the FFM factors were modeled as five correlated trait factors, and the narrow facets were modeled as 15 method factors. Trait factor covariances and method covariances were freely estimated; however, no trait-method (e.g., conscientiousness with achievement orientation) covariances were allowed. Each FFM trait factor had multiple manifest indicator variables, such eight for agreeableness, 12 for neuroticism, 19 for openness, 21 for extraversion, and 30 for conscientiousness. For the narrow facet method factors, each had six manifest indicator variables, with a few corresponding all to a single FFM trait factor (e.g., Achievement Orientation, Creative), and others corresponding to multiple FFM trait factors (e.g., Conventional, Optimistic). The MTMM model that was developed can be found in Figure 6.

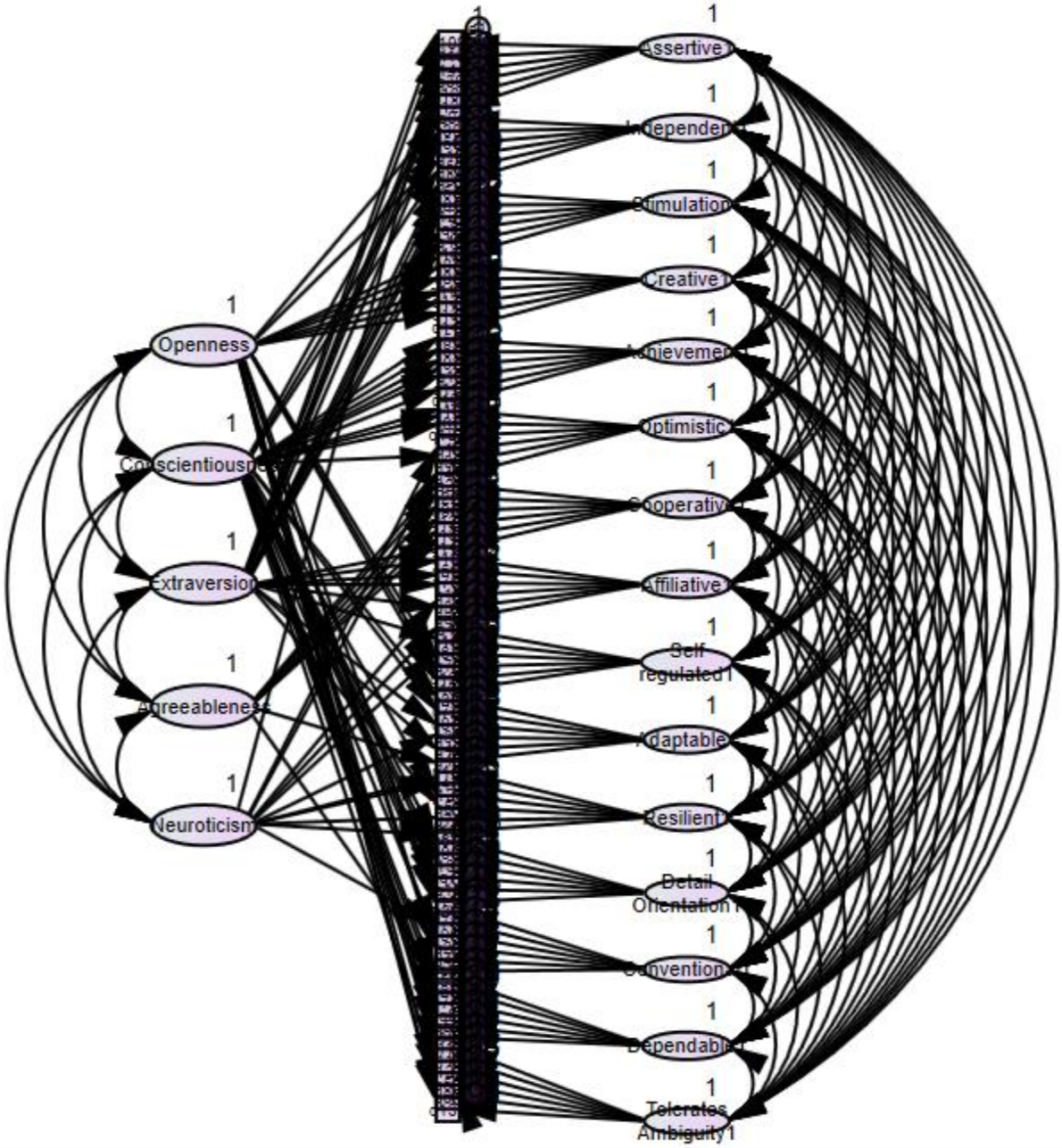

Figure 6. MTMM Measurement Model as Constructed in AMOS. $\chi 2$ (3,710) = 56336.08, $p <$ .001; *GFI* = .863; *CFI* = .671; *RMSEA* = .038.

The fit of the CT-CM model, while better fitting than either of the previous models, was still poor and other issues also came to light ($\chi 2$ [3710] = 56336.08, $p < .001$; *GFI* = .863; *CFI* = .671; *RMSEA* = .038). After reviewing the results of the analysis, it was found that there were three non-significant regression paths related to the narrow facet method factors such as q115, q127, q96, corresponding to Achievement Orientation, Cooperative, and Resilient, respectively, which could be removed for an improved fit. In addition, three non-significant regression paths related to the FFM trait factors were found, q55 and q124, which correspond to extraversion, and q69 of conscientiousness. Also, about 20% of the standardized regression weights proved to be negative.

Lastly, the CT-CM model did not reach an admissible solution due to the covariance matrices for both the FFM trait factors and the narrow facet method factors as a result of not being positive definite. Similar to the previously tested models, the MTMM model was also plagued by Heywood cases. Whereas the fit of the CT-CM model was better than either of the empirical models, and also better fitting than the theoretical model, the CT-CM model was inadmissible. Thus, the CFA posed a number of potential issues and contradictions that warrant deeper analysis, and therefore a lack of evidence to support the proposed factor structure of the Personality Instrument as presented in Hypothesis 3a. Additionally, as a consequence of Hypothesis 3a not being supported, Hypothesis 3b could neither be tested nor supported.

Exploring the factor structure of the Personality Instrument. As a result of the Personality Instrument's factor structure not being supported through the various CFA approaches, it was pertinent to perform an EFA on the Personality Instrument using this new set of data to determine whether an alternative factor structure may have emerged. To begin, the necessary sample size needed to be determined. However, recommendations vary regarding adequate sample size for performing an EFA. For example, Gorsuch (1983) suggested a ratio of five participants to each measured variable, and the sample size should always be greater than 100. However, Nunnally (1978) proposed 10 participants to each variable. Other researchers recommend taking into account important factors affecting sample size such as the extent to which factors are overdetermined, the level of communalities, statistical power, and the nature of the sample (Fabrigar, Wegener, MacCallum, & Strahan, 1999; MacCallum, Widaman, Zhang, & Hong, 1999). As a conservative estimate following the recommendations of Nunnally, 10 times the 90 items of the Personality Instrument would equate to a sample size of 900. Being that the dataset for this study is 10,000 cases, a sufficient sample has been reached.

The maximum likelihood (ML) extraction method was used as the factor analysis fitting method. ML is best suited for normally distributed data such as that obtained for use in this research, and it is ideal because it allows for the computation of a wide range of goodness of fit indices of the model such as statistical significance of the factor loadings, correlations among factors, and confidence intervals (Cudeck & O'Dell, 1994). If for some reason the assumption of normality was violated, then principal axis factoring (PAF) would have been an appropriate secondary method (Fabrigar, et al., 1999). While also another option, principal components analysis (PCA) was ruled out as Fabrigar et al. recommends not using the PCA method when the

objective of the study is to identify latent constructs, such as the case with this study focusing on the factor structure of the Personality Instrument.

Prior evidence suggests the Personality Instrument to be multidimensional (Company, 2013), similar to other measures of personality. However, most of the narrow facets of the Personality Instrument are heterogeneous as they incorporate items which represent multiple FFM traits. Therefore, the factors were rotated to clarify their psychological meaning. Varimax rotation was used as it is intended for factors which are orthogonal, or uncorrelated with each other. An oblique rotation could also have been used, which generates factors that can be correlated or uncorrelated with each other (Furr & Bacharach, 2014). Being that personality is generally found to be multidimensional with uncorrelated factors (Costa & McCrae, 1992), an oblique rotation would have been unnecessary. The purpose of this was to compare the underlying factor structure of the Personality Instrument with the EFA performed in the previous psychometric evaluation study in 2013 that originally supported the 15 first-order narrow facets and the three overarching second-order factors.

In order to identify the number of factors which emerged, the eigenvalues and their relative size were examined, as well as the column of descending eigenvalues to identify a point at which all of the subsequent difference between eigenvalues becomes small (See Figure 7). Furr and Bacharach (2014) refer to the scree plot as the best method for identifying emerging factors, which was also to be examined. The plotted eigenvalues were examined visually to identify a relatively large difference, or leveling off point, signifying the number of factors (Cattell, 1966) for which three factors emerged. Additionally, the rule of thumb related to eigenvalues greater than 1.0 was utilized to identify the number of emerging factors (Kaiser, 1960), as shown in Table 8.

Figure 7. Scree Plot of Personality Instrument Broad Factors.

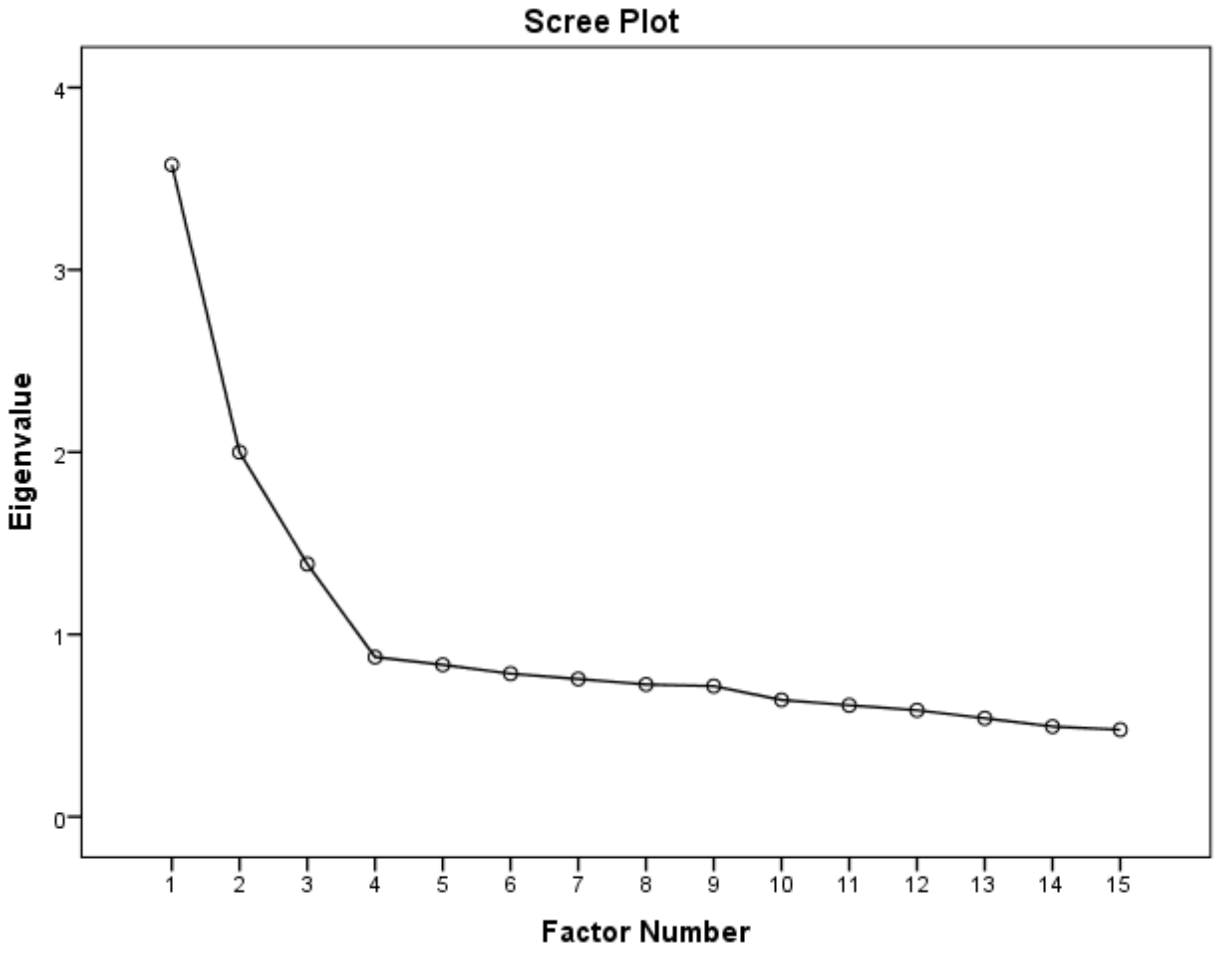

Table 7. Factor Names, Eigenvalues and Variance of Factors.

Personality Instrument Broad Factor	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Potency / Getting Ahead	3.606	24.037	24.037	1.918	12.787	12.787
Social Adjustment	1.950	13.003	37.041	1.734	11.563	24.351
Discipline / Getting it Right	1.381	9.206	46.247	1.304	8.694	33.044

N = 10,000.

An EFA was performed to examine the structure underlying the Personality Instrument with the 15 narrow facets, revealing three second-order factors with eigenvalues exceeding 1.0. These factors together explained approximately 33% of the variance, as shown in Table 8. The scree plot also showed that three broad factors were in sharp decent and then leveled off,

revealing evidence that rotation was necessary for the three factors. The communalities of each of the narrow facets are provided in Table 9. The results corroborated the factor structure that emerged in 2013, which used PCA as the EFA method. Table 10 shows the ordering of narrow facets from top to bottom in the sequence that best represents the commonalities associated with the factors. Although the CFA may have failed to support the factor structure of the Personality Instrument, an EFA using more recent data and an alternative EFA approach produced similar results, suggesting that the Personality Instrument's 15 narrow facets are best described as grouping under three broad factors. In the discussion section below, I will provide some potential explanations for this pattern of results.

Table 8. *Communalities of Personality Instrument Narrow Facets.*

Personality Instrument narrow facets	Initial	Extraction
Achievement Orientation	.203	.264
Adaptable	.167	.254
Affiliative	.332	.402
Assertive	.299	.388
Conventional	.091	.126
Cooperative	.303	.457
Creative	.203	.266
Dependable	.322	.461
Detail Orientation	.266	.387
Independent	.232	.341
Optimistic	.349	.431
Resilient	.267	.312
Self-Regulated	.199	.240
Stimulation	.315	.449
Tolerance for Ambiguity	.123	.178

Extraction Method: Maximum Likelihood. N = 10,000.

Table 9. *Rotated Factor Matrix of Narrow Facets.*

	Factor		
Narrow facets	Potency / Getting Ahead	Social Adjustment	Discipline / Getting it Right
Assertive	.607		
Stimulation	.592		-.254
Independent	.560		
Creative	.472		
Optimistic	.428	.375	.329
Achievement Orientation	.418		.271
Cooperative		.650	
Affiliative	.328	.531	
Adaptable		.481	
Self-Regulated		.423	.239
Resilient	.373	.395	
Detail Orientation		.201	.586
Dependable		.365	.573
Conventional			.354
Tolerance for Ambiguity		.258	-.269

N = 10,000.

Investigating the reliability of the Personality Instrument

Stability of the Personality Instrument over time. To test Hypothesis 2a, which addresses the reliability of the Personality Instrument, test-retest reliability analyses were performed. Individuals completed the Personality Instrument measure twice, with Time 1 and Time 2 separated by a duration of time between 0 and 12 weeks. Table 11 presents Pearson correlations between overall Personality Instrument scores at the two time points. Furthermore, test-retest reliability estimates are presented for each of the three broad Personality Instrument factors and their respective 15 narrow facets.

The results yielded a strong, positive and statistically significant correlation between the two time points, r (905) = .79, $p < .01$; therefore, Hypothesis 2a was supported. For each of the three broad factors of Discipline/Getting it Right, Social Adjustment and Potency/Getting Ahead, test-retest reliability estimates were significant at .68, .74 and .79, respectively. Test-retest

reliability estimates for each of the 15 narrow facets, ranged from .44 to .74, (see Table 11). Table 12 provides the means and standard deviations for each of the three broad factors at Time 1 and Time 2, as well as the overall Personality Instrument score. Included in the table are the correlations between each broad factor and also with the total Personality Instrument score. As can be seen from the results in the table, the test-retest reliability is adequate for the overall Personality Instrument, as well as for each of the three broad factors. However, when looking at each of the narrow facets individually, it is shown that their test-retest reliability is less than that of the broad factor to which they align. In summary, Hypothesis 2a is supported as a strong, positive and statistically significant test-retest correlation that has been observed when the Personality Instrument is taken within twice within 12 weeks.

For testing Hypothesis 2b, which investigated the strength of the relationship between test administrations as a duration of time 13 weeks or more, additional test-retest correlations were computed. These additional test-retest correlations were for individuals who took the Personality Instrument within a longer duration of time, such that the duration between test administrations was 13 weeks or greater. In Table 11, test-retest correlations are provided for participants who completed Time 1 and Time 2 of the Personality Instrument between 13 and 52 weeks. For the longer duration between test administrations, and although there is a decline in the strength of the correlation, a strong, positive and statistically significant correlation, r (480) = .67, $p < .01$, remains.

For each of the three broad Personality Instrument factors of Discipline/Getting it Right, Social Adjustment and Potency/Getting Ahead, the test-retest correlations have correspondingly declined to .55, .60 and .71, respectively. Likewise, the test-retest correlations for each of the 15 narrow facets were also reduced, ranging from as low as .37 to as strong as .60.

The test-retest correlations for the two durations of time (i.e., 0-12 weeks and 13-52 weeks) are computed from two different samples, making the two test-retest correlations independent. In order to test whether the two test-retest correlations are significantly different from one another, the correlations were transformed using Fisher's *r* to *z* transformation calculation, which used an online calculate from http//:www.vassarstats.net. As a result of applying the Fisher's *r* to *z* transformation, the sampling of the transformed coefficient becomes nearly normally distributed. In addition, the variance of correlation coefficients becomes nearly the same regardless of the value of the population correlation (Kenny, 1987). When the resulting *Z* is statistically significant, then it can be concluded that the correlations differ between the two groups. Table 11 provides the *Z*s and relative significance. A review of the table shows that the test-retest correlation coefficients between the 0-12 weeks and 13-52 weeks groups are statistically significantly different for the Overall CPA, and each of the three broad factors. In addition, nine of the 15 narrow facets (e.g., Dependable and Detail Orientation) are also statistically significantly different from one another. For half a dozen narrow facets (e.g., Tolerates Ambiguity and Conventional), the test-retest correlation coefficients between 0 to 12 weeks and 13 to 52 weeks are not statistically significantly different.

These analyses further support hypothesis 2b, in which it was proposed that the duration of time between test administrations would impact the strength of the test-retest correlation. Therefore, in summary, although there is generally a negative association between the length of time and the strength of the reliability estimates, the reliability estimates for the overall Personality Instrument remain strong and positive, regardless of duration. This pattern of results is similar when the Personality Instrument is disaggregated into (a) the three broad factors and (b) the narrow facets themselves.

Table 10. *Test-retest Reliability of CPA, Including Broad and Core Psychological Attributes.*

Scale Level or Trait	0 – 12 Weeks Pearson Correlation	13 – 52 Weeks Pearson Correlation	Fisher's *r* to *z* transformation
Overall CPA	.79**	.67**	4.61*
Discipline/Getting it Right	.68**	.55**	3.72*
Tolerates Ambiguity	.49**	.44**	1.13
Dependable	.64**	.53**	2.97*
Conventional	.53**	.47**	1.41
Detail Orientation	.65**	.53**	3.27*
Social Adjustment	.74**	.60**	4.55*
Adaptable	.44**	.43**	.22
Resilient	.62**	.51**	2.87*
Self-Regulated	.55**	.37**	4.06*
Cooperative	.60**	.53**	1.82
Affiliative	.70**	.58**	3.62*
Potency/Getting Ahead	.79**	.71**	3.26*
Optimistic	.61**	.56**	1.34
Assertive	.72**	.55**	5.11*
Independent	.58**	.46**	2.92*
Stimulation	.74**	.63**	3.69*
Achievement Orientation	.54**	.50**	.97
Creative	.68**	.60**	2.40*

*. Correlation significant at the .05 level (2-tailed). **. Correlation significant at the .01 level (2-tailed). N = 907 for 0 – 12 weeks, and N = 480 for 13 – 52 weeks.

Table 11. *Means, Standard Deviations, and Correlations among Personality Instrument and the Three Broad Factors.*

Variable	Mean	SD	1	2	3	4	5	6	7
1. Personality Instrument T1	78.78	34.61							
2. Personality Instrument T2	81.11	33.63	**.79****						
3. Discipline / Getting it Right T1	21.84	10.25	.62**	.48**					
4. Discipline / Getting it Right T2	22.12	10.49	.43**	.61**	**.68****				
5. Social Adjustment T1	28.93	14.81	.82**	.65**	.46**	.37**			
6. Social Adjustment T2	29.93	14.50	.64**	.82**	.39**	.46**	**.74****		
7. Potency / Getting Ahead T1	28.00	19.44	.83**	.52**	.23**	.13**	.46**	.38**	
8. Potency / Getting Ahead T2	29.05	19.01	.63**	.81**	.18**	.18**	.38**	.43**	**.79****

**. Correlation is significant at the .01 level (2-tailed). Values in bold on diagonal are test-retest reliability correlations. N = 907.

Investigating the item functioning of contextualized and generic items

Internal consistency of the narrow facets. In order to investigate the internal consistency of each of the narrow facets, Cronbach's alphas were calculated. Due to several of the narrow facets being heterogeneous, and potentially because of the small number of items representing each, it Cronbach's alpha coefficients in this case may be better a better index of homogeneity. Each of the alphas for the narrow facets are reported in Table 13. Cronbach's alphas ranged from as low as .05 for Tolerance for Ambiguity to as high as .58 for Affiliative. Consistent with what was observed in the 2013 validation studies, the internal consistency estimates using Cronbach's alpha are likely not a sound indicator of internal consistency as many of the narrow facets were intentionally developed to measure more than one aspect of an FFM broad factor.

Table 12. *Internal Consistency Estimates for the Narrow Facets*.

Personality Instrument narrow facet	Cronbach's Alpha
Achievement Orientation	.29
Adaptable	.06
Affiliative	.58
Assertive	.38
Conventional	.23
Cooperative	.39
Creative	.30
Dependable	.43
Detail Orientation	.39
Independent	.22
Optimistic	.46
Resilient	.30
Self-Regulated	.13
Stimulation	.48
Tolerance for Ambiguity	.05

N = 10,000.

Relationships between contextualized versus generic items. As per Hypothesis 3, one of the goals of this study was to investigate the relationship between contextualized items and generic items under the same narrow facet. Several of the Personality Instrument's narrow facets are comprised of not just a combination of positive and negatively worded statements, but also a combination of contextualized items and generically worded items. In order to determine if there was any substantive difference between how contextualized items correlate with one another versus how they correlate with generic items, simple bivariate correlations were computed between each of the six items as they relate to each narrow facet. For example, the narrow facet of Achievement Orientation is comprised of four contextualized items and two generic items, whereas the Optimistic dimension scale is comprised of three contextualized and three generic items. However, there are a couple narrow facets which have only a single contextualized item.

In order to assess whether contextualized items correlate with each other more so than with generic items under the same narrow facet, bivariate correlations were computed between each set of contextualized items, between generic items, and also among all six items comprising the scales. The average correlation was then computed for the group of contextualized as well as generic items, and also all six items, under the same narrow facet in order to understand if any differences exist between the mean correlations. Table 14 provides the results of the mean bivariate correlation analyses.

For some of the narrow facets, such as Achievement Orientation and Adaptable, the analyses show that contextualized items do in fact correlate more so with each other than do the generic items, and also more so than they correlate with the generic items themselves. However, there are other instances in which generic items under the same narrow facet correlate more so with each other than they do with contextualized items, such as for Affiliative and Conventional.

This is an interesting finding, and perhaps an indication that while shared context may influence correlations between items, shared item meaning may have more influence on the correlations between items than the effect of item context.

Five of the narrow facets (i.e., Achievement Orientation, Adaptable, Assertive, Optimistic, & Stimulation) display contextualized items as correlating more so with each other than the relative generic items, whereas there are seven narrow facets (i.e., Affiliative, Conventional, Cooperative, Dependable, Independent, Self-Regulated, and Tolerance & Ambiguity) in which generic items correlate more so with each other than with contextualized items. Unfortunately, there are three narrow facets (i.e., Creative, Detail Orientation, & Resilient) which only have a single contextualized or generic item, and therefore do not lend additional information to the overall picture.

In order to test whether the correlations between the contextualized items and the correlations between the generic items are statistically and significantly different, Lee and Preacher's (2013) calculator for testing the difference between two dependent correlations and available from http://quantpsy.org was used. Whereas the computation of Fisher's *r* to *z* transformation for understanding the differences between test-retest correlation coefficients at different durations of time were presented earlier, this test of difference between correlations is for two dependent samples, or in other words the same individuals responded to both the contextualized and generic items for which correlations were observed. Similarly, the result of Lee and Preacher's calculator is a z-score which can be compared in either 1-tailed or 2-tailed fashion against the unit normal distribution. When the resulting z-score values are greater than 1.96 for a 2-tailed test they are considered significant.

From a review of Table 14, all but one of the narrow facets (i.e., Tolerance for Ambiguity) have statistically significant differences between the correlations for contextualized items and the correlations between generic items. In summary, the results of Hypothesis 3 have failed to be supported, warranting further investigation in future research.

Table 13. *Mean Correlations Among Contextualized Items, Generic Items, and Between Both.*

Narrow facets	Contextualized	Generic	Between	Fisher's *r* to *z* transformation
Achievement Orientation	.102	-.046	.051	10.50*
Adaptable	.156	-.013	-.035	12.05*
Affiliative	.085	.234	.168	-10.78*
Assertive	.090	.059	.108	2.20*
Conventional	-.009	.064	.061	-5.17*
Cooperative	.060	.099	.114	-2.77*
Creative	--	.034	.108	--
Dependable	.097	.169	.141	-5.16*
Detail Orientation	--	.087	.122	--
Independent	.035	.329	.042	-21.62*
Optimistic	.175	-.007	.147	13.00*
Resilient	.096	--	.015	--
Self-Regulated	-.025	.012	.046	-2.62*
Stimulation	.146	.118	.132	2.01*
Tolerance for Ambiguity	.024	.030	-.005	-.42

Note. Between = average of correlations between a generic item and a specific item. N = 10,000.

CHAPTER IV

Discussion

Summary of Results

The psychometric properties of the Personality Instrument were investigated in this study, looking at the factor structure, reliability, and also the unique characteristics of its contextualized item content. Below, the major findings and results of the hypotheses are presented. First, the results indicated that the proposed factor structure of the Personality Instrument in which the 15 narrow facets group under three broad factors could not be found to be supported through multiple CFA models (i.e., empirical, theoretical, and MTMM approach). Neither of the statistical approaches lent support for the factor structure of the Personality Instrument displayed adequate levels of fit, and as a consequence the first hypothesis which predicted the originally proposed empirical model would be found through a CFA was not supported.

Although these analyses were based upon a previous EFA in 2013, the empirical model was tested as proposed, but unfortunately support could not be found. One potential reason for this was due to the analyses being fraught with issues, namely Heywood cases, in which impossible values were found such as negative variances or correlations greater than 1.0. Likely, this is due to non-convergence and under-identification (Boomsma & Hoogland, 2001) as a result of many of the narrow facets scales being heterogeneous and measuring multiple facets representing FFM broad factors. Additionally, there were multiple negative regression paths within each of the CFA approaches tested, which also likely created issues of their own.

Because the Personality Instrument's factor structure failed to be supported with a CFA, a post-hoc EFA was performed to determine if perhaps the data being used in this study may be influencing the factoral structure of the Personality Instrument. Instead of using PCA with

Varimax rotation as previously used in the 2013 validation study, an EFA employing ML with Varimax rotation was deemed to be a better choice as previously described. However, the results suggested the same factor structure of the Personality Instrument; the EFA best described the Personality Instruments factor structure to have three second-order factors in which the 15 narrow facets were grouped. In summary, although the current EFA replicated the results from 2013, the CFA failed to do so.

Though the factor structure of the Personality Instrument failed to be supported through various CFA approaches, perhaps one of the most important psychometric properties to date was established, reliability of the instrument. Test-retest reliability was shown to be strong and positive (r = .79) when the Personality Instrument is administered twice within an interval of 0 and 12 weeks, therefore supporting the second hypothesis of the study. This is a substantial finding as prior reliability, or internal consistency, was primarily based upon Cronbach's alpha, and which because of the heterogeneous nature associated with most of the 15 narrow facets, is unlikely to be the best estimate of reliability. Additionally, test-retest reliability was calculated for individuals who took the Personality Instrument with a duration between 13 weeks and 52 weeks. The resulting test-retest reliability (r = .67), while considered to be less than sufficient, was only slightly below the .70 rule of thumb for practice (Field, 2009).

Existing longitudinal studies show similar test-retest correlations, from .36 to .55 in childhood, and .70 to .79 in adulthood (Hampson & Goldberg, 2006). This is similar to Roberts and Del Vecchio (2000) that found test-retest correlation coefficients from .31 in childhood, and increasing to .54 during college years, .64 at age 30, and reaching a plateau of .74 between 50 and 70 years of age. While the sample in this study does not have age identifiers to parse out and look at the test-retest correlations among different age ranges, the test-retest correlations are

somewhat similar to findings from other studies, such Caspi, Roberts and Shiner (2005). Caspi et al. found similar test-retest correlation coefficients at comparable age ranges, and also made important observations such as test-retest correlations over time are often moderate, increase with older age ranges, yet the test-retest correlations decrease as the duration of time between test administrations increase. If demographic information was available regarding participants who completed the Personality Instrument in this study, the data could have been cut to investigate similar age ranges and their respective test-retest correlation coefficients for comparison. As a result of the test-retest correlations computed in this study, both of the second hypotheses related to reliability were supported as it was found that the Personality Instrument encompasses adequate test-retest reliability, and as suspected, the reliability decreased over time.

Establishing reliability by computing test-retest reliability coefficients was a significant addition over the psychometric evaluation performed in 2013, especially in regards to the heterogeneous nature of many of the narrow facets. As a consequence of their intentional design to measure multiple facets of FFM traits, Cronbach's alpha is unlikely to be a trustworthy indicator of stability, but rather an index for whether item under the same narrow facet are homogeneous, or measuring a unitary construct. Alpha levels ranged from as low as .05 for Tolerance for Ambiguity, and as high as .58 for Affiliative. Neither of the narrow facets reached what is often considered the sufficient rule of thumb of .70 (George & Mallery, 2003) when using Cronbach's alpha as coefficient of internal consistency. A likely contributor to this finding is the multidimensionality of the Personality Instrument and the heterogeneous nature of several of the narrow facets. Additionally, there is a small number of items representing each of the narrow facets, which is also a significant limitation for using alpha as an estimation of reliability (Cho & Kim, 2015).

Lastly, the hypothesis which investigated whether contextualized items correlate more so with each other than do generic items which also represent the same narrow facet remains inconclusive, and no support for the hypothesis was found. In five of the narrow facets, contextualized items grouped more so with each other than did generically worded items. Conversely, seven of the narrow facets showed that generic items more so correlated with each other than did the contextualized items representing those same narrow facets. The remaining three narrow facets only had a single contextualized or generic item, and therefore provided no additional insight. As a result, the hypothesis was not supported, but further investigation may bring additional insight to how contextualized and generic items interact when comprising a common factor. From the results, it would appear that the correlations between the contextualized and generic items are dependent upon the narrow facet, and not a trend throughout the Personality Instrument. Additionally, the results may reflect the heterogeneous nature of several of the narrow facets, and therefore it may be possible that shared meaning matters more so than whether the items are of a contextualized or generic composition.

In summary, the findings from this study shed light on the psychometric characteristics of the Personality Instrument, most importantly by establishing reliability. Further research is warranted to better understand the factor structure of the Personality Instrument, perhaps by employing other factor analytic techniques. In addition, continued research should be conducted on the Personality Instrument to determine its relationship with job success metrics, especially in relation to the contextualized and generic items.

Implications for Theory and Practice

The results of this study provide multiple implications for theory and practice which extend beyond the Personality Instrument itself. Studies such as this which investigate

alternative personality-based assessments remain important for the advancement of the predictor-criterion relationship in a selection context, as well as enhancing our understanding for how best to measure personality. This study in particular investigated the Personality Instrument, which has a few unique properties unlike many other commercially available personality instruments commonly used today. For example, the Personality Instrument's 15 narrow facets are not equally dispersed amongst each of the well-established FFM broad traits, and instead are distributed in relation to the results of historical empirical research supporting their relationship with important work outcomes. Likewise, the Personality Instrument's item content is comprised of nearly 50% of which are composed in a contextualized context referencing work, versus other instruments which are primarily comprised of generically worded items. Thus, there are important nuanced characteristics of the Personality Instruments which relate to theories of personality measurement and scale construction, from reliability, to factor analytic techniques, and also how personality items are composed, which provide insight and practicality for personality assessments beyond the Personality Instrument.

Alternative personality instruments. The core foundation of the Personality Instrument is based upon the FFM, in which the FFM is widely supported as being represented by five broad personality factors. However, multiple attempts at conducting EFAs has resulted in a factor structure of the Personality Instrument which is represented by three broad factors, even though the CFA is comprised of narrow facets representing each of the FFM personality factors. As previously described, each of the narrow facets is comprised of six items, and some of these narrow facets have items representing a single FFM trait, while others represent two to four FFM traits. With this understanding, it is evident that the Personality Instrument is a rather complex personality measure in regard to its multiple heterogeneous narrow facets. Most other personality instruments, while somewhat heterogeneous, have narrow facets comprised of items which represent a single FFM broad trait. Here in lies the distinction between the Personality Instrument and other personality-based instruments.

As such, the Personality Instrument more so represents those traits which have been found through empirical research to have the greatest relationship with important work criteria. Potentially, it is this unequal representation of the broad FFM traits which has influenced the factor structure of the Personality Instrument, and as a result is better represented by three broad factors. While perhaps not as mainstream as personality measures representing the FFM of personality (e.g., NEO-PI or CPI), other measures of personality exist in parallel which are represented by varying number of personality traits, such as Eysenck's (1991) three factor model or Ashton et al.'s (2004) proposed Hexaco model. Thus, while the FFM remains a predominate foundation for the traits representing personality, other models have not been completely ruled out.

Reliability and the Personality Instrument. This study has demonstrated practical importance by providing an improved and more precise understanding of the Personality Instrument's reliability. This is particularly significant as prior validation work primarily centered on supporting the reliability of the Personality Instrument by calculating Cronbach's alpha coefficients for each of the narrow facets, with poor to moderate results at best. Compared to alpha, test-retest reliability in this situation is a better estimate of reliability, and therefore lends support for not always holding Cronbach's alpha in such high regard. In the case of the Personality Instrument, the narrow facets are comprised of only six items, and it is well-known that test length has a relationship with reliability (Lord & Novick, 1968), such that a longer test with redundant items will likely have a greater alpha coefficient, of which is likely artificially inflated.

Cho and Kim (2015) present rationale for why Cronbach's alpha in many cases may not be the best indicator of reliability, especially as alpha is dependent upon average item interrelatedness and the number of items. Cho and Kim provide an example that alpha alone is not an indicator of internal consistency, but that the number of items can have a significant impact on the alpha coefficient value. The Spearman Brown Prophecy (1910) is an excellent example to demonstrate that the number of items can substantially increase alpha, by predicting reliability of a test after changing the test length (Allen & Yen, 1979). For example, if the narrow facet of adaptable has an alpha of .06 with six items currently representing the construct, by holding all else the same in terms of item composition and representation, an alpha of .70 can be achieved if the scale was comprised of 215 items. Similarly, if achievement orientation with an alpha of .28 is represented by the same/similar 36 items, the alpha increases to .70. Therefore,

a consequence of only having six items representing each narrow facet is that alpha may be low, and not a sound indication of reliability.

Cortina (1993) also provided some guidance for when alpha would be appropriate for use as an indicator of reliability: a) the test measures a single factor or construct, b) test items are tau-equivalent statistically, and c) the error scores of items are uncorrelated. Tau-equivalency is said to be when items measuring the same latent variable are on the same scale and using the same precision of measurement, and when this assumption of alpha as an indicator of reliability is violated, alpha is likely to be a lower-bound of reliability (Novick & Lewis, 1967). The first assumption of Cortina's regarding alpha as reliability is clearly violated by several of the Personality Instrument's narrow facets as they are comprised of items representing multiple aspects of FFM traits. In addition, the tau-equivalency assumption is also violated as a result of measuring multiple latent variables. Lastly, the CFA procedures brought to light a few cases in which the error scores of items representing the same narrow facet were correlated. Thus, less reliance on Cronbach's alpha and more reliance on other estimates of reliability, such as test-retest reliability as used in this study, may be better estimates of reliability for the Personality Instrument, and other instruments which are somewhat complex and have scales measuring multiple constructs.

Alternative methods for determining factor structure. Although CFAs tend to be touted as the hallmark statistical analysis for supporting the factor structure of a statistical model, they do not always produce the anticipated results. Kline (2015) contends that CFAs as follow-up analyses to EFAs are neither required, nor advisable. Based upon EFA results, and even when the same number of factors are represented, CFAs generally do not confirm the results of an EFA (Kline, 2015). One potential reason is that indicators in EFAs often have secondary pattern coefficients that are relatively high, which may account for relatively high proportions of variance (van Prooijen & van der Kloot, 2001). Thus, when performing a CFA, constraining secondary pattern coefficients to zero may be too conservative and restrictive and therefore the model may be less likely to fit the data (Kline, 2015).

In the realm of personality research, even some of the most popular personality instruments have been shown to experience similar issues, such as when CFA techniques have been employed to the NEO-PI (Costa & McCrae, 1992a) in an attempt to confirm the factor structure (McCrae et al., 1996). Similarly, Borkenau and Ostendorf (1990) and Church and Burke (1994) have likewise had issues with CFA techniques appropriately accounting for the data and displaying adequate levels of fit. Lee and Ashton (2007) state that even when factor structure models replicate across different samples when performing EFAs and include all of the broad factors underlying the variable set, CFAs tend to reject the proposed factor structure. McCrae et al. (1996) propose that many personality variables are associated with multiple factors, making them somewhat complex and heterogeneous, and therefore not a simple-structured or orthogonal factor structure. This in turn may attribute to poor levels of fit, similarly as found in this study.

Thus, instead of employing CFA techniques, other methods to confirm the factor structure may be more appropriate for instruments based upon a personality framework. EFAs are well-suited for exploring the number of common factors influencing personality instruments, and understanding the relationship among those common factors representing the instrument (Field, 2009). EFAs although more of an a priori approach and exploratory in nature, do provide meaningful insight, especially when an instrument has recently been developed or is being refined (Ruscio & Roche, 2012). Even when using multiple samples and groups (e.g., younger and older, White and Non-White, male and female, etc.), EFAs performed on the NEO-PI-R unceasingly support the FFM of personality (Costa, McCrae, & Dye, 1991). Osborne & Fitzpatrick (2012) provide procedures for evaluating EFA results when the same variables have been replicated over independent samples, and used as a method for supporting the factor structure of a given model. As Kline (2015) would proclaim, CFAs are not inherently superior to EFAs, and thus EFA results can be just as meaningful, if not more so, depending on the situation.

Other methods, such as using orthogonal Procrustes rotation may be a better technique for finding support for the factor structure of the Personality Instrument, similarly as McCrae et al. (1996) applied to the NEO-PI to confirm the underlying FFM. Procrustes rotations are called targeted rotations because the intent of this technique forces the data to conform to a predetermined structure (Digman, 1967), and can either be oblique or orthogonal. The more conservative approach, and of suggestion for finding support for the factor structure of instruments employing an underlying personality framework, is the orthogonal Procrustes rotation (Schonemann, 1966). By performing the orthogonal Procrustes rotation, factors are rotated and the matrix is constrained to be orthogonal (Paunonen, 1997). McCrae et al. (1996) proclaims this technique to be well-suited for hypothesis-guided rotation, and by doing so

obtained significant results for supporting the factor structure of the NEO-PI to bear resemblance to the FFM. Although not readily accessible for use in statistical programs such as SPSS, other programs such as R Programming and LISREL offer the capability of performing such a technique.

Finally, multiple group factor analysis (MGFA) is another method for consideration as an alternative to CFAs. As a factor analytic technique, MGFA is often used for investigating the degree to which measures are invariant, or unchanging, across groups (Chen, 2008). Experts of factor analysis have touted its usefulness for hypothesis testing such as for confirming the factor structure of complex personality-based instruments (Gorsuch, 1974; Guttman, 1952). Lodhi, Deo and Belhekar (2002) found that compared to techniques based upon principal components analysis, MGFA only explained 3.4 percent less variance, and therefore may be an equitable option. Thus, MGFA, as well as EFAs and targeted orthogonal Procrustes rotations are each worth further consideration as alternative techniques to CFAs for establishing factoral validity of personality-based instruments which may be somewhat complex and heterogeneous in nature. In summary, other factor analytic techniques are available and worth exploring, not just for identifying the factor structure of the Personality Instrument, but other instruments also based upon a personality framework.

FOR effect and contextualized items. The FOR effect has important implications for pre-employment assessment developers, one being that it has the probability to promote better measurements of a candidate's potential performance and job success. This can be accomplished by contextualizing items comprising personality instruments for the purpose of increasing the predictor-criterion relationship, and as a result making more informed hiring decisions. Additionally, the FOR effect may increase the consistency by which candidates respond to items throughout the assessment, and therefore may be less likely to change the context in their mind when responding to items (Schmidt et al., 1995). Furthermore, having items composed in a specific context may reduce between-person differences in how a group of candidates respond to items as they have each been primed to respond to the items using the same context.

The underpinnings of this result may be due to the cognitive-affective system (CAS) theory of personality, in which a person's behavior is not merely the result of a personality trait, but instead an individual's perceptions of themselves in that particular situation and the variation in which they behave (Mischel & Shoda, 1995). Another similar theory, trait activation theory (TAT), posits that a person's traits are expressed in work behaviors as a response to particular trait-relevant cues (Tett & Guterman, 2000; Tett, Simonet, Walser, & Brown (2013). In a selection context, this may create difficulties by failing to find personality coherence as the psychological characteristics of the situation are failing to be controlled for, which may be indicative of generically composed personality items. However, by employing more specific measures of personality by composing contextualized items, it is possible that predictive validity may be bolstered. Likely, this is a result of prompting candidates to respond using a specific FOR by which answering assessment questions.

In addition, contextualized personality instruments have been shown to have incremental validity over GMA whereas non-contextualized personality instruments have failed to add predictive validity (Bing et al., 2004; Hunthausen et al., 2003). Much of the empirical research on the contextualization of personality items has primarily focused on conscientiousness (Bing et al., 2004, Hunthausen et al., 2003; Schmit et al., 1995) as vast research is available promoting the predictive validity of conscientious compared to other FFM broad traits (Barrick, Mount, & Judge, 2001; Hurtz & Donovan, 2000). Thus, by employing personality assessments with contextualized items, organizations have a greater probability of predicting whether or not a candidate will likely succeed and perform on the job.

Future Research and Limitations

Although the study portrayed a few strengths contributing to our understanding of how the Personality Instrument functions in regards to psychometrics, there are limitations that must be addressed. First, the data used in this study was archival, and most of all lacked demographic variables such as age, race/ethnicity, and gender, which may have provided important information for understanding any potential group differences. Other important variables for comparing groups may include job role, industry, tenure, etc. In order to understand important group differences, future research could investigate how different groups may respond to and score on the Personality Instrument compared to other groups.

Motivations to distort responses. It is reasonable to assume that some candidates may be motivated to complete the assessment multiple times, perhaps to obtain a higher score for having a greater likelihood of being considered for a particular job. Although the Personality Instrument survey platform limits an individual to a single completion of the Personality Instrument, participants may have found ways to get around this by creating a new profile and entering a different email address from their original profile in order to take the Personality Instrument multiple times. As a result, the archival data obtained from the Company for investigating test-retest reliability had limitations in regards to the data that was utilized. In order to match participants, the Company identified people by name and email address to identify those who took the instrument at two different times. For future research, the Personality Instrument could be administered in a controlled research study, rather than a practical scenario, where individuals may be less likely to distort their scores on a second administration, and also have a greater probability of accurately matching participants' Time 1 and Time 2 Personality Instrument scores.

Contextualized versus generically phrased items. Recent research has supported the application of contextualized items comprising personality instruments for personnel selection (Bing et al., 2004; Bowling & Burns, 2010; Heller, Ferris, Brown, & Watson, 2009; Hunthausen et al., 2003; Lievens, et al., 2008; Roenicke, 2013; Schmit et al., 1995; Wang, 2011). However, much is still left to be explored, and further research focused on the Personality Instrument in regards to its contextualized and generic items may be important for furthering our knowledge of how the two item types aid in the measurement of personality traits. Future studies with the Personality Instrument should investigate the predictive validity of the instrument for understanding the differences in relationships between contextualized items and criterion variables, as well as the relationship between generic items and criterion variables. Such research may further support item contextualization as a means to improve the predictor-criterion relationship for personality.

Conclusion

In the current study, several psychometric properties of the Personality Instrument were investigated. In terms of traditional indicators of construct validity, the results were poor. Overall, the results indicated that the factor structure of the Personality Instrument as originally proposed in the EFA from 2013 failed to be supported through multiple CFA technique attempts. However, a post-hoc EFA was conducted again, this time using Maximum Likelihood versus PCA, with a Varimax rotation, also found that the 15 narrow facets were best represented with three broad factors. In addition, shorter form of the Personality Instrument could not be developed as a result of failing to support the Personality Instrument's factor structure through a model generating approach to CFA. Therefore, a shorter form of the Personality Instrument may

be developed through other means such as an item analysis approach, and continued research on the resulting factor structure.

Results however did indicate sufficient test-retest reliability for measurement purposes. This finding was a substantial addition to support the reliability of the Personality Instrument, especially since most of the 15 narrow facets are heterogeneous as they measure more than a single FFM trait. Establishing reliability through test-retest reliability analysis may be a better estimation of reliability compared to Cronbach's alpha, which estimates lower-bound internal consistency, and is fraught with violated assumptions as a consequence of the Personality Instrument's intentional design.

While Cronbach's alpha is often considered an indicator of reliability, the narrow facets failed to show what is considered sufficient internal consistency. This lends to the support and importance relating to the test-retest reliability findings. Likely, the low Cronbach's alpha coefficients are based upon multiple factors, such as a low number of items (i.e., six items per narrow facet), and several of the narrow facets measure multiple FFM broad traits, making them heterogeneous and therefore violating Cortina's (1993) assumption for alpha to be an indicator of reliability. Lastly, the analysis of how contextualized and generic items group together under the same narrow facet failed to support the hypothesis. The results were mixed, as for some narrow facets the contextualized items correlated more so with each other, and in other instances, the generic items correlated more so with each other than did the contextualized items. Although the results were inconclusive and failed to lend much insight, future research could investigate the predictive validity to determine if the contextualized items of the Personality Instrument have stronger and positive relationships with criterion variables than do the generic items.

References

Allen, R. B. & Ebbesen, E. B. (1981). Cognitive processes in person perception: Retrieval of personality trait and behavioral information. *Journal of Experimental Social Psychology, 17*, 119-141.

Allen, M., Yen, W. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*, 1-171.

Aronson, Z. H., & Reilly, R. R. (2006). Personality validity: The role of schemas and motivated reasoning. *International Journal of Selection and Assessment, 14*, 372-380. doi: 10.1111/j.1468-2389.2006.00358.x

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Blas, L. D., Boies, K., & Raad, B. D. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356-366.

Barrett, G. V. (2008). Practitioner's view of personality testing and Industrial-Organizational psychology: Practice and legal considerations. *Industrial and Organizational Psychology, 1*, 299-302.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

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*, 9-30.

Barrick, M. R., & Mount, M. K. (2012). Nature and use of personality in selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel selection and assessment* (pp. 225-251). New York, NY: Oxford University Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150–157.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. Dutoit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future.* Lincolnwood, IL: Scientific Software International.

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11*, 515-524.

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

Bowling, N. A., & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences, 49*, 95-101.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverly Hills, CA: Sage.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56*, 453-484. doi: 10.1146/annurev.psych.55.090902.141913

Cattell, H. E. P., & Mead, A. D. (2008). The sixteen personality factor questionnaire (16PF). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 2 - Personality measurement and testing*. Thousand Oaks, CA: Sage Publications, Inc.

Cattell, R. B. (1946). *The description and measurement of personality*. New York, NY: World Book.

Cattell, R. B. (1965). *The scientific analysis of personality.* Baltimore, MD: Penguin Books.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, I*, 245-276.

Cattell, R.B., Eber, H.W. & Tatsuoka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B. (1976). *Personality and learning theory: The structure of personality in its environment.* New York, NY: Springer Publishing Company, Inc.

Cattell, H., & Mead, A. (2008). The sixteen personality factor questionnaire (16PF). In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 2 — Personality measurement and testing*. (pp. 135-160). London: SAGE Publications Ltd. doi: http://dx.doi.org/10.4135/9781849200479.n7

Cattell, H. E., & Schuerger, J. M. (2003). *Essentials of the 16PF*. New York: John Wiley & Sons.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95,* 1005.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research, 29,* 468-508.

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods, 18,* 207-230.

Cloninger, C. R., Przybeck, T. R., Svrakic, D. M., & Wetzel, R. D. (1994). *The Temperament and Character Inventory (TCI): A guide to its development and use*. St. Louis, MO: Center for Psychobiology of Personality, Washington University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly, 23,* 651-669.

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. *Personality and Individual Differences, 5,* 11-26.

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104. doi:10.1037/0021-9010.78.1.98

Costa, P. T., & McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life span development and behavior* (Vol. 3, pp.65-102). New York, NY: Academic.

Costa, P. T, & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T, & McCrae, R. R. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

Costa, P. T., & McCrae, R. R. (1992a). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*, 343-359. doi:10.1521/pedi. 1992.6.4.343

Costa, P. T., Jr., & McCrae, R. R. (1992b). Trait psychology comes of age. In T. B. Sonderegger (Ed.), *Nebraska symposium on motivation: Psychology and aging* (pp. 169-204). Lincoln: University of Nebraska Press.

Costa, P. T., Jr., McCrae, R. R. (1994). Stability and change in personality from adolescence through adulthood. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 139-150). Hillsdale, NJ: Erlbaum.

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*, 887-898.

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin, 115*, 475-487. doi:10.1037/0033-2909.115.3.475.1994-32085-00

Curran, P. J., West, S. G., & Finch J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis, *Psychological Methods, 1*, 16-29.

Digman, J. M. (1967). The Procrustes class of factor-analytic transformations. *Multivariate Behavioral Research, 2*, 89-94.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

Dilchert, S., Ones, D. S., Van Rooy, D. L., & Viswesvaran, C. (2006). Big Five factors of personality. In J. H. Greenhaus, & G. A. Callanan (Eds.), *Encyclopedia of career development.* (pp. 36-42). Thousand Oaks, CA: Sage.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the

intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*, 40-57.

Erdogan, M., Ok, A., Marcinkowski, T. J. (2012). Development and validation of children's responsible environmental behavior scale. *Environmental Educational. Resources, 18* (4), 507-540.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3?--Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12*, 773-790.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi:10.1037/1082-989X.4.3.272

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th Ed.). New York, NY: Sage Publications.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149-158.

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). SAGE Publications, Inc.: Thousand Oaks, CA.

Gatewood, R. D., Feild, H. S., & Barrick, M. (2011). *Human research selection* (7th ed.). Mason, OH: South-Western.

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2009). *Research methods in applied settings: An integrated approach to design and analysis* (2nd ed.). New York, NY: Routledge.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (ed.), *Review of Personality and Social Psychology* (Vol. 2). Beverly Hills, CA: Sage, pp. 141-165.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several Wve-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Gorsuch, R. L. (1974). *Factor analysis*. Philidelphia, PA: W. B. Saunders Company.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (third ed.). Palo Alto, CA: Consulting Psychologists Press.

Guttman, L. (1952). Multiple group methods for common-factor analysis: Their basis computation, and interpretation. *Psychometrika, 17*, 209-222.

Hampson, S. E., Goldberg, L. R. (2006). A first large-cohort study of personality-trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology, 91*(4), 763-779.

Hampson, S. E., John, O. P., and Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgements of trait implications. *Journal of Personality and Social Psychology, 51*, 37-54.

Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. *Journal of Personality, 77*, 1051–1084.

Helson, R., Jones, C., & Kwan, V.S.Y. (2002). Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology, 83*, 752–766.

Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology, 53*, 176-186.

Helson, R., & Wink, P. (1992). Personality change in women from the early 40s to early 50s. *Psychology and Aging, 7*, 46-55.

Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *1982 Nebraska Symposium on Motivation* (pp. 55-89). Lincoln, NE: University of Nebraska Press.

Hogan, R., & Hogan, J. (1992). *Hogan personality inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477.

Holden, R. R., Fekken, C. G., & Cotton, D. H. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 111-118.

Holtrop, D., Born, M., & de Vries, R. E. (2014). Predicting performance with contextualized inventories, no frame-of-reference effect? *International Journal of Selection and Assessment, 22,* 219–223.

Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of Frame-of-Reference and Pre-Test Validity Information on Personality Test Responses and Test Perceptions. *International Journal of Assessment and Selection, 13,* 75-86.

Hough, L. M. (1992). The "Big Five" personality variables – construct confusion: Description versus prediction. *Human Performance, 5*, 139-155.

Hough, L. M., Dilchert, S. (2010). Personality: its measurement and validity. In J. L. Farr and N. T. Tippins. *Handbook of Employee Selection*. New York: Routledge.

Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In I. B. Weiner (Ed.), Handbook of Psychology. Hoboken, NJ: John Wiley & Sons, Inc.

Hough, L. M., & Johnson, J. W. (2013). Use and importance of personality variables in work settings. In I. B. Weiner (Ed.) *Handbook of Psychology* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Hough L. M., Ones D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. Sinangil Kepir, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 1: Personnel psychology, pp. 233–277). London: Sage.

Hough, L. M., & Oswald, F. L. (2008). Personality testing and Industrial-Organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology, 1,* 272–290.

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351-362.

Hunthausen, J. M., Truxillo, D. M., Bauer, T. B., & Hammer, B. L. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology, 88,* 545–551.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. Journal of Applied Psychology, 85, 869-879. doi: 10:1037//0029010.85.6.869

Jackson, D. N. (1994). *Jackson personality inventory-revised manual*. Port Huron, MI: Sigma Assessment Systems.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp.294-316). Newbury Park, CA: Sage.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Pscyhometrika, 59,* 381-389.

Jöreskog, K. G., & Sörbom, D. (1984). *Lisrel VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Mooresville, Indiana: Scientific Software.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical frameworks. *Journal of Applied Psychology, 98,* 875-925. doi: 10.1037/a0033901

Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal, 58,* 1149-1179.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. doi:10.1177/001316446002000116

Company (2013). Technical manual for the Personality Instrument personality assessment.

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences*. Boston, MA: Little, Brown and Company.

Kenny, D. A. (2011, October 29). Single latent variable model [Web log comment]. Retrieved from http://davidakenny.net/cm/1factor.htm

Kenny, D. A. (2015, November 24). Measuring model fit [Web log comment]. Retrieved from http://davidakenny.net/cm/fit.htm

Kenny, D. A., & Kashy, D. A. (1992). Analysis of multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165-172.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.) The Guilford Press: New York, NY.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58,* 281-342.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39,* 329-358.

Lee, K., & Ashton, M. C. (2007). Factor analysis in personality research. In R. W. Robins, R. C. Fraley, & R. F. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pgs. 424 – 443). New York, NY: The Guilford Press.

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

Lievens, F., DeCorte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268-279.

Lodhi, P. H., Deo, S., & Belhekar, V. M. (2002). The Five-Factor Model of personality: Measurement and correlates in the Indian context. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 227-248). New York, NY: Kluwer Academic/Plenum Publishers.

Lord, F., & Novick, M. R. (1968). *Statistical theories of mental tests*. New York, NY: Addison-Wesley.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-89.

Lodhi, P. H., Deo, S., & Belhekar, V. M. (2002). Personality measurement in India. In R. R. McCrae & J. Allik (Eds.), *The five factor model of personality across cultures.* Springer: New York, NY.

McCrae. R. R., & Costa. P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 51*, 81-90.

McCrae. R. R., & Costa. P. T., (2008). Empirical and theoretical status of the Five-Factor Model of personality traits, in G. J. Boyle, Matthews, G., & D. H. Saklofske (eds.), *Personality theory and assessment: Volume 1 personality theories and models*. Thousand Oaks, CA: SAGE Publications, Inc.

McCrae, R. R., Costa, P. T., Jr., de Lima, M. P., Simoes, A., Ostendorf, F., Angleitner, A., Mausic, I., Bratko, D., Caprara, G. V., Barbarenelli, C., & Chae, J. (1999). Age differences in personality across the adult life span: Parallels in five countries. *Development Psychology, 35*, 466-477.

McCrae, R. R., Costa, P. T. Jr., Ostendorf, F., Angleitner, A., Hrebrikova, M., Avia, M.D., Sanz, J., Sanchez-Bernardos, M. L., Kusdil, M. E., Woodfield, R., Saunders, P. T., Smith, P. B. (2000). Nature over nurture: Temperament, personality, and lifespan development. *Journal of Personality and Social Psychology, 78*, 173-186.

McCrae, R. R., & John, O. P. (1992). An introduction of the five factor model and its applications. *Journal of Personality, 60*(2), 175-215.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102,* 246-268.

Millon, T., Grossman, S., Millon, C., Meagher, S., & Ramnath, R. (2004). *Personality disorders in modern life*. Hoboken, NJ: John Wiley & Sons, Inc.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683-729. doi: 10.1111/j.1744-6570.2007.00089.x

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60,* 1029-1049.

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

Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, DC: American Psychological Association.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Novick, M. R., Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Pscyhometrika, 32*(1), 1-13.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60,* 995-1027.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection, *Journal of Organizational Behavior, 17,* 609-626.

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation, 17,* 1-8.

Ozer, D. J., & Benet-Martinez, V. (2005). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 1-8.

Paunonen, S. V. (1997). On chance and factor congruence following orthogonal Procrustes rotation. *Educational and Psychological Measurement, 57*, 33-59. doi: 10.1177/0013164497057001003

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press/Washington, DC: American Psychological Association.

Raykov, T., & Marcoulides, G. A. (2000). A method for comparing completely standardized solutions in multiple groups. *Structural Equation Modeling, 7*, 292-308.

Roberts, B. W., Caspi, A., & Moffitt, T. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology, 81*, 670–683.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3-25. doi: 10.1037//0033-2909.126.1.3

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 3–27.

Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating five factor theory and social investment perspectives on personality trait development. *Journal of Research in Personality, 39*, 166-184.

Robins, R. W., Fraley, C., Roberts, B. W., & Trzesniewski. K. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*, 617-640.

Roenicke, C. C. (2013). Extending the frame of reference effect beyond conscientiousness (Doctoral dissertation).

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*, 282-292. doi: 10.1037/a0025697

Sanitioso, R., Kunda, Z. & Fong, G. T. (1990). Motivated recruitment of autobiographical memory. *Journal of Personality and Social Psychology, 59*, 229-241.

Saucier, G., Hampson, S. E., & Goldberg, L. R. (2000). Cross-language studies of lexical personality factors. In S. E. Hampson (Ed.), *Advances in personality psychology*, Vol. 1 (pp. 1-36). New York, NY: Psychology Press.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology, 76*, 613-627.

Schmidt F.L., & Hunter J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607-620.

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior*, *1,* 45-65. doi: 10.1146/annurev-orgpsych-031413-091255

Schonemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika, 31*, 1-10.

Shaffer, J. A., Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*, 445-494.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, P. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*, 1041-1053. doi: 10.1037/0022-3514.84.5.1041

Tellegen, A. (1995/2003). *Multidimensional Personality Questionnaire–276 (MPQ-276) test booklet*. Minneapolis: University of Minnesota Press.

Tett, R. P., & Burnett, D. D. 2003. A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500–517.

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

Tett, R. P., Simonet, D. V., Wasler, B., & Brown, C. (2013). Trait activation theory: Applications, developments, and implications for person-workplace fit. In N. D. Christiansen R. P. Tett, *Handbook of personality at work.* New York, NY: Routledge, 71-100.

van Prooijen, J.-W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61*, 777-792.

Walsh, W. B., & Betz, N. E. (1985). *Test and assessments*. Prentice-Hall, Inc.: Englewood Cliffs, CA.

Wang, Q. A comparison of work-specific and general personality measures as predictors of OCBS and CWBS in China and the United States. (Doctoral dissertation).

Watson, D., & Clark, L. A. (1984). Negative Affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465-490.

West, S. G, Finch, J. E., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 305-316.

Wiggins, J. S., Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In: J. S. Wiggins (Ed.). *The five-factor model of personality: Theoretical perspectives*. New York: Guilford.

Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology, 61*, 309-348. doi: 10.1111/j.1744-6570.2008.00115.x

Zuckerman, M., Bernieri, G., Koestner, R., & Rosenthal, R. (1989). To predict some of the people some of the time: In search of moderators. *Journal of Personality and Social Psychology, 57*, 279-293.

Appendices

Appendix A

Personality Assessment narrow facet definitions

Attribute	Meaning
Resilient	Maintains energy and motivation in spite of hardships or negative outcomes.
Assertive	Not intimidated by others and speaks directly. Is willing to initiate action and take leadership.
Dependable	Keeps commitments and follows rules.
Affiliative	Seeks the company of others and enjoys engaging with them.
Tolerates Ambiguity	Is comfortable with uncertainty and maintains effectiveness in spite of it.
Cooperative	Prefers to collaborate with others and sees value in group success.
Independent	Prefers to work alone and trusts one's own thinking and work.
Achievement Orientation	Desires accomplishment and success and is willing to expend effort to achieve it.
Detail Orientation	Is careful to get work right and analyze information thoroughly
Optimistic	Expects positive outcomes and looks forward to the future.
Self-Regulated	Reflects on one's own actions and is has a realistic self-awareness.
Adaptable	Is responsive to changing conditions and sees value in different perspectives.
Stimulation-Risk Seeking	Enjoys excitement and adventure and is willing to risk failure to achieve success.
Conventional	Prefers the traditional way of doing things and believes it is wrong to ignore the past.