2008

How similar are personality scales of the "same" construct? a meta-analytic investigation

Victoria L. Pace
University of South Florida

How Similar are Personality Scales of the "Same" Construct?

A Meta-Analytic Investigation

by

Victoria L. Pace

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Michael T. Brannick, Ph.D.
Co-Major Professor: Walter C. Borman, Ph.D.
Judith Becker Bryant, Ph.D.
Bill N. Kinder, Ph.D.
Stephen Stark, Ph.D.

Date of Approval:
November 19, 2007

Keywords: convergent validity, criterion-related validity, job performance, meta-analysis, nomological network, personality tests, reliability

Dedication

I would like to dedicate this manuscript to my wonderful family members and friends, who provided support and encouragement and firmly believed that I could and should continue to the finish at times when I was not so certain.

Acknowledgements

I would like to acknowledge the help and patience of my major professors, Dr. Michael T. Brannick and Dr. Walter C. Borman. I have learned a great deal about conducting research and writing about it by attempting to follow their examples and suggestions. Through a variety of trials they listened and assisted. I would also like to acknowledge the kind support and insightful comments given to me by committee members, Dr. Judith Becker Bryant, Dr. Bill Kinder, and Dr. Stephen Stark. Thanks also go to Dr. Gregory McColm, who has been supportive since undergraduate years and who graciously agreed to act as chair for my dissertation defense. Additionally, I would like to thank several undergraduate assistants (Hillary Cagle, Jatuporn (Yui) Namyen, and Raquel Hodge) who provided support in the literature search and data entry phases.

Table of Contents

List of Tables

List of Figures

How Similar are Personality Scales of the "Same" Construct?
A Meta-Analytic Investigation

Victoria L. Pace

ABSTRACT

In recent years, meta-analytic reviews have estimated validities for the use of personality scales in the prediction of job performance from an array of empirical studies. A variety of personality measures were used in the original studies, and procedures and decisions concerning the categorization of these measures into Big Five personality factors have differed among reviewers. An underlying assumption of meta-analysis is that the predictors across included studies are essentially the same, as is the criterion. If this is not the case, then problems arise for both theoretical reasons and practical applications. If predictors that are not highly correlated are combined in a meta-analysis, then the theoretical understanding of antecedents and consequents of the predictors will be clouded. Further, combining predictors that are not essentially the same may obscure different relations between predictors and criteria, that is, test may operate as a moderator.

To meet the assumption of similarity, systematic methods of categorizing personality scales are advised. Two indicators of scale commensurability are proposed: 1) high correlations among predictor scales and 2) similar patterns of correlations between predictor scales and job-related criteria. In the current study, the similarity of the most commonly used personality scales in organizational contexts was assessed based on these

two indicators. First, meta-analyses of correlations between scales were conducted. Second, subgroup meta-analyses of criterion-related validity were examined, with specific personality scale and criterion as moderators.

Correlations between criterion-related validity and certain sample characteristics were also conducted to determine if sample characteristics act as moderators of validity. Additionally, an examination of personality scale reliabilities was conducted.

Results reveal that assumptions of similarity among personality measures may not be entirely met. Whereas meta-analyzed reliability and criterion-related validity coefficients seldom differed greatly, scales of the "same" construct were only moderately correlated in many cases. Although these results suggest that previous meta-analytic results concerning reliability and criterion-related validity are generalizable across tests, questions remain about the similarity of personality construct conceptualization and operationalization. Further research into comprehensive measurement of the predictor space is suggested.

How Similar are Personality Scales of the "Same" Construct?

A Meta-Analytic Investigation

Researchers have begun to consider the similarity of personality scales of ostensibly the same construct. In particular, some have complained that many are not similar enough to be grouped together in the same meta-analysis (e.g., Hogan, 2005). What are the convergent correlations among these scales? Do the scores from these scales predict criteria in the same way and to the same degree? For example, are the Hogan Personality Inventory (HPI) scale for Prudence and the NEO PI-R scale for Conscientiousness, both considered to measure conscientiousness, equally predictive of job performance? Do the scales produce equally reliable scores?

To date, researchers have meta-analyzed criterion-related validities of the Big Five personality factors (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) by assuming predictor scales from the included studies were essentially equivalent. Often the determination of equivalence has been based on comparison of definitions of constructs the scales purport to measure. Frequently, the names of test scales at the Big Five level differ. For example, scales that are generally grouped into the Conscientiousness factor also have names such as Work orientation, Prudence, Job involvement, Self-discipline, Forward planning, and Rule consciousness. Although the names differ, the similarity of scores and inferences based upon the scales is an open question. Ultimately, it is an empirical question whether the difference in

names matters. That is, if the measures are so highly correlated that their antecedents and consequents are the same, then the differences in names are of trivial importance. However, if the measures are not highly correlated with one another, or if despite relatively high correlations, the measures show different patterns of relations with other measures, then distinct names and distinct treatments of the measures are warranted.

To assign scales to a Big Five construct, some researchers may have examined the scales at the item level to decide whether each scale appears to measure the same construct, based on face validity. Others have relied on information from previous factor analyses. Still others have consulted categorizations from other researchers, such as the summary of taxonomies given by Hough (1992), for guidance on which scales fall under each of the Big Five constructs. Classification of personality measures into the Big Five continues to progress, and a useful framework for further research appears to have emerged in the work of Hough and Ones (2001). However, there appears to be no clearly quantitative review of scales that examines their commensurability. Hough and Ones have encouraged continued research into how scale constructs relate to criteria so that further refinements to their taxonomy can be made. Based on empirical relationships of these constructs (taxons) to criteria, they hope to be able to merge some taxons and to further differentiate others as needed.

An issue that complicates the assignment of scales to the five factors is the variety of ways in which the personality domain has been divided. Although a five factor structure may be the most widely accepted, there remain many who argue for a greater or fewer number of personality factors. At the low end, Eysenck proposed three factors (Psychoticism, Extraversion, and Neuroticism). At the high end, Cattell proposed 16

personality factors. Accordingly, many personality scales were not developed to measure Big Five factors, but are oriented toward alternative construct sets (Salgado, 1997). Such diversity causes problems because broader scales that may be considered to measure more than one of the Big Five must either not be used in a Big Five meta-analysis or must be grouped according to the Big Five factor with which the scale correlates most highly. Either determination is problematic because eliminating all studies using the broader scale decreases the comprehensiveness of the meta-analysis, whereas assigning the scale to any one of the Big Five introduces construct contamination into that factor. For scales based on taxonomies that include more than five factors, it is likely that more than one scale will be grouped into a single Big Five factor. Because the test developers of such scales clearly had in mind different constructs for each of the scales, this also poses problems.

Scales such as integrity scales, which are often considered to be measures of compound traits, pose the same difficulties with categorization. Therefore, following the example of Hough and Ones (2001), these scales are not categorized into one factor nor are they examined in the current study.

Even among those who are proponents of a five-factor structure, there are different views concerning the facets that make up each of the Big Five. These varied understandings of the exact nature of each of the five factors are reflected in the names of their constituent facets and the relative predominance of each facet within the factor-level measures. To illustrate this point, Costa and McCrae (1992) gave the six facets of Openness to Experience as (Openness to) Feelings, Aesthetics, Fantasy, Actions, Values, and Ideas. The Hough and Ones (2001) taxonomy lists the facets of Openness to

Experience as Complexity, Culture/Artistic, Creativity/Innovation, Change/Variety, Curiosity/Breadth, and Intellect. Aesthetics corresponds to Culture/Artistic, Fantasy to Creativity/Innovation, Actions to Change/Variety, and Ideas to Intellect. However, using the Hough and Ones taxonomy as the organizing structure, the NEO facets of Feelings and Values are not considered pure measures of Openness. The NEO Feelings facet is regarded as a compound measure of Openness and Extraversion (and categorized as a scale of Intraception). The NEO Values facet is described as a compound measure of Openness and Conscientiousness (and categorized as a scale of Traditionalism). Perhaps this type of difference of opinion regarding the construct and components of Openness to Experience accounts for the variety of names for measures grouped into this category (e.g., Creative personality, Culture, Intellectance, Absorption, and Sentience). Although some seem to focus more on the Aesthetic/Artistic/Creative aspects of this construct, others focus more on the Ideas/Curiosity/Cognitive Complexity aspects. Differences in focus are not necessarily problematic if researchers and practitioners recognize that differences may mean one measure is more appropriate for use in certain circumstances than another. For example, when avoidance of adverse impact is a priority, it may not be advisable to select a scale that focuses on cognitive complexity, especially if the primary aim is to predict aesthetic sensibility. Hough, Oswald, and Ployhart (2001) found greater group differences with Openness to Experience measures than measures of other personality factors. They propose that this finding is probably attributable to facet level differences with the intellect facet being more to blame than values or need for variety facets. When comparing measures, it would not be surprising to find a relatively low

mean correlation between scales of Openness that emphasize distinct aspects of the construct.

However, even when measures are substantially correlated with one another within a factor grouping, scales may show differential relations with other measures. For example Pace (2005) found that the observed correlation between the NEO PI-R Openness scale and her Work-specific Openness scale was .72. Despite this correlation, she found the Work-specific Openness scale to be a better predictor of work outcomes of interest than was the NEO PI-R scale. She found that the correlation with supervisory ratings of creativity was .09 for the NEO PI-R Openness scale and .32 for the Work-specific Openness scale.

As McDonald (1999) explained from a psychometric point of view, equivalent test forms are required to display identical relationships with criteria. Although identicalness of relationships is probably too stringent a requirement for inclusion in meta-analysis and not practical, recommendations by Hough and colleagues (Hough & Furnham, 2003; Hough & Ones, 2001; Hough & Oswald, 2005) of following a taxonomy that categorizes personality measures based on relationships between the measured construct and other constructs of interest, i.e. requiring similar nomological networks, seems a reasonable criterion for grouping different measures into a single personality factor. Just how similar the relationships within those nomological networks must be is a question that needs further study.

Systematic Differences between Items in Personality Scales

There are several potential reasons that measures of reportedly the same construct may, in fact, differ markedly in their prediction of important criteria. For example,

whether the scale was developed for clinical or employment-related use may impact its validity for job-related outcomes. Studies by Schmit, Ryan, Stierwalt, and Powell (1995), Bing, Whanger, Davison, and VanHook (2004), and Hunthausen, Truxillo, Bauer, and Hammer (2003) found that a group of scales initially developed for clinical use exhibited significantly improved predictive validities for criteria when the items or instructions were altered to target the criterion context rather than the original general context.

Also, some scales of a particular construct such as Extraversion are more heavily weighted toward one or more facets or subdimensions. Different subdimensions are generally believed to covary, but also to assess somewhat different aspects of the factor. In fact, some contend that there are really more than five factors because facets within a factor may differentially predict criteria. As an illustration, Hough (1992) advised splitting the Extraversion factor into Affiliation and Potency based on the low average correlation between these two subdimensions. According to results by Vinchur, Schippmann, Switzer, and Roth (1998), these Big Five subdimensions differ in their criterion-related validities for both ratings and objective sales criteria of salespeople. Paunonen and Ashton (2001) went further by suggesting that more detailed, facet-level measurement of personality is in order. Their results indicated incremental criterion-related validity for facets over broader factors. Facets that were chosen by judges were able to predict substantial variance in criteria that was not predicted by the broader factors. Criteria used in their study of undergraduate students varied in breadth, but tended to be narrower than overall performance ratings typical of work criteria. Some examples were alcohol consumption, participation in sports, and grade point average.

Ones and Viswesvaran (1996) argued that, for applied use, broad measures of the Big Five are generally more reliable and show higher criterion-related validity than narrower (subdimension or facet) measures when the criterion is broad, such as overall job performance. These authors also provided a convincing argument for their focus on overall job performance rather than on individual performance dimensions. Nevertheless, to enhance theoretical understanding of relationships and for further development of a taxonomy of personality measures, results from fine-grained predictor and criterion measures can also be informative. General consensus about this bandwidth-fidelity trade-off appears to be in favor of matching broad predictors to broad criteria and narrow predictors to narrow criteria.

Another seemingly subtle, but possibly substantive difference between scales thought to measure the same construct was mentioned by Hogan (2005). Based on the tests' construct definitions, other researchers have grouped the NEO Agreeableness scale and the Hogan Personality Inventory (HPI) Likeability scale into the same meta-analyses for the factor Agreeableness. Hogan (2005) contended that the two scales measure different constructs and predict criteria differently. The NEO scale tends to measure passive avoidance of conflict, whereas the HPI scale measures active social charm. Although these systematic nuances in item content may not seem to indicate obviously different constructs, their interpretation by test-takers may elicit very different responses that differentially predict criteria. Avoidance of conflict can be expected to be a useful predictor of employee performance in workplaces where "getting along" (Hogan, Rybicki, Motowidlo, & Borman, 1998) is highly valued, whereas active social charm

might be a more useful predictor when networking and persuasion are necessary components of the job.

Differences in Reliability

To the degree that scale score reliabilities affect predictive validities, scales that produce scores with different reliabilities will differ in prediction. Viswesvaran and Ones (2000) meta-analyzed reliabilities produced by Big Five personality scales and found standard deviations of internal consistency to hover around .10 and standard deviations for test-retest reliabilities to be slightly greater than this across measures of a single Big Five construct. Only minor differences in reliabilities and their standard deviations were observed when comparing the five factors. All coefficients were from technical manuals that reported reliabilities for the normative samples. It is quite possible that reliabilities observed in practice differ to an even greater degree and that some scales consistently produce scores of lower reliability than others. A between-measure comparison of reliabilities will reveal the extent of differences.

Although "reliability is a property of the *scores* (emphasis added) on a test for a particular group of examinees" (Crocker & Algina, 1986, p. 144), rather than a characteristic of the test, differences in the distributions of reliabilities by test could be useful information in a variety of ways. For example, differences in reliability such as those found in the Viswesvaran and Ones (2000) study, ranging from the .40s or .50s to the .90s, would be considered important to most researchers when selecting an instrument to use. If great differences in reliability exist, variables that are associated with these differences can be determined (Vacha-Haase, 1998). Knowledge about differences in

reliability may aid decision-makers in instrument selection and use, as well as interpretation of results.

Additionally, a better understanding of reliability distributions may be particularly important when conducting meta-analyses. In meta-analytic practice, it appears that reliability coefficients and their distributions are commonly taken completely or in part from information in test manuals combined across a variety of scales (e.g., Barrick & Mount, 1991; Dudley, Orvis, Lebiecki, & Cortina, 2006; Hurtz & Donovan, 2000). These distributions are then used to correct for unreliability in the predictor when estimating effect sizes in the population. Although this may be a relatively safe practice, assuming that reliabilities from test manuals are likely to be accurate or a bit high (thus leading to under-correction, rather than over-correction), a more precise look at reliabilities in practice and by test could lead to more accurate corrections.

Consequences of Heterogeneous Scale Groupings

If seemingly similar scales are actually substantially different, readers may wonder what the consequences of this dissimilarity are. This is the well-known "apples and oranges" problem (see Cortina, 2003, or Sharpe, 1997, for further discussion), in which very different elements are combined in a common group and the group's relationship with other variables, such as work outcomes, is assessed. Clearly, if the group elements have differing relationships with the outcome of interest, a group-level effect will obscure these differences and lead to incorrect conclusions. As an illustration, consider a pair of predictor measures (A and B) and an outcome measure C. Assume A is a strong positive predictor of C, and B is a weak positive predictor of C. If A and B are grouped together and we examine only their pooled ability to predict C, we will

underestimate the predictive ability of A and overestimate the predictive ability of B. The situation is worse if one is a positive predictor and the other is a negative predictor. In this case, we may not realize the scales have any predictive ability at all. Therefore, considering the moderating effects of variables such as characteristics of measures or samples allows us to examine whether an "apples and oranges" problem exists.

Meta-Analysis to Determine Average Effect Sizes

Meta-analysis allows for the estimation of effect sizes in populations of interest based on a limited number of results available from existing studies. Methods used in meta-analysis allow for a more precise estimate than would be obtained by taking a simple average across studies. Generally, weights are applied to individual study effect sizes before combining them. This procedure gives greater weight to larger studies or those with less variance due to sampling error. Examination of moderators in meta-analysis is an excellent way to determine whether effect sizes vary according to certain recorded study characteristics such as the specific personality measure used. This information can help to answer questions about the advisability of combining personality scales into a single meta-analysis. Therefore, this study uses meta-analysis to examine personality measures and other moderators.

The issue of whether the grouping of personality scales for meta-analyses is problematic or not deserves careful consideration and empirical testing. If there is not a problem, we can have more confidence in past research conclusions. If there is a problem, we will gain knowledge about differences among personality scales and can implement changes in meta-analytic procedures for evaluating personality construct validities.

A Priori Questions and Expectations

In summary, several questions are raised and answers are sought concerning differences and similarities among scales of the same construct. Specifically, comparisons of scale content based on convergent validity, relationships of scale scores to criteria of interest, and comparisons of scale reliabilities are explored.

Question 1. Are personality scales highly convergent, based on meta-analyzed zero-order correlations between scores from scales that seemingly measure the same construct?

Question 2. Do personality scores from scales of the "same" construct display identical relationships with job-related criteria?

Question 3. Do all widely-used personality scales display the same reliability?

Method

To address the question of whether it is advisable to combine personality scales into a common meta-analysis, substantive ways in which personality measures differ were considered and relevant data were recorded. The degree of difference among scales was then assessed through the use of meta-analysis.

In particular, two indicators of scale similarity were examined for the most commonly used personality scales in organizational contexts: 1) high correlations among predictor scales and 2) similar patterns of correlations between predictor scales and job-related criteria. Past meta-analyses have not explicitly considered both of these indicators.

To address the first indicator, meta-analyses of correlations between scales were conducted. The sizes of these correlations were compared to the average size of correlations between Big Five factors. Correlations between scales of the same construct (different tests) should be much larger than correlations between scales of different constructs from the same test.

To address the second indicator, meta-analyses of criterion-related validity with specific scales as moderators were examined. Correlations of certain sample characteristics with effect sizes were also conducted.

Additionally, meta-analyses of personality scale reliabilities were conducted and compared across measures.

Literature Review

Types of data collected. Examination of scale similarity was limited to measures that have been grouped by Hough and Ones (2001) into each of the Big Five constructs, with the addition of closely-related scales, such as shorter or earlier versions by the same author(s). Compound personality measures that purport to measure more than one Big Five construct, such as those for integrity and customer service orientation, were excluded. Criterion-related validity and reliability data for each of the included personality scales, as well as correlations between these scales, were collected.

Sources of data. Data were extracted from journal articles, dissertations, test manuals, and unpublished studies. Data were found by searching the PsycInfo and ProQuest Thesis and Dissertation databases and through e-mails to test publishers and personality researchers. An extensive list of researchers was generated and contacted based on published literature, participation in Society for Industrial and Organizational Psychology (SIOP) or Academy of Management (AoM) conferences during the past five years, and recommendations by other researchers. Reference sections and tables of recently published personality meta-analyses were also examined for lists of studies they included.

Inclusion criteria. Correlations between scales were taken from studies of adult populations using English language versions of the scales. Scale development articles, other articles by the authors of the scales (e.g. McCrae, Costa, & Piedmont, 1993), and test manuals were one of the primary sources for correlations between similar construct scales from two distinct measures.

Because there were very few scale pairs for which at least six convergent validity correlations could be found, validity coefficients for prediction of job performance were not limited to those personality scales that were included in the convergent validity meta-analyses.

Validity data were recorded from all located studies that used employed samples and English language versions of scales that were included in the Hough and Ones (2001) taxonomy. Only published and unpublished studies from 1990 to present were included in order to minimize overlap with the large and well-known personality meta-analyses by Barrick and Mount (1991) and Tett, Jackson, and Rothstein (1991). Also, personality measures changed relatively little (few revised forms) from 1990 to the present.

Data Coding

The following variables were coded for each study: personality test name, test length (number of items), the setting for which the test was originally developed (work, clinical, other), stated scale construct, corresponding Big Five construct and facet according to Hough and Ones (2001) where applicable, test reliability obtained (internal consistency and test-retest were coded separately when available), correlations with other personality scales (of the same Big Five construct), criterion-related validity coefficients, criterion construct(s), criterion measure(s) and their reliability, sample characteristics (N, type of job, applicants/employees, percent female, percent minority), and published/unpublished status.

Classification of Scales into Big Five Constructs

In an early meta-analysis, Barrick and Mount (1991) used trained subject matter experts to categorize personality measures. Based on categorizations by researchers and

their own combined experiences in grouping these measures and examining criterion-related validity, Hough and Ones (2001) developed a working taxonomy that lists measures that are considered to assess each of the Big Five constructs, as well as some of their facets. In the current meta-analysis, personality scales were categorized following the system from Hough and Ones.

According to Salgado (1997) and Hough, Eaton, Dunnette, Kamp, & McCloy (1990), the most well-known and used personality instruments include the California Psychological Inventory (CPI), Eysenck Personality Questionnaire (EPQ), Guilford-Zimmerman Temperament Survey (GZTS), Myers-Briggs Type Indicator (MBTI), Comrey Personality Scales (CPS), Edwards Personal Preference Schedule (EPPS), Gordon Personal Profile-Inventory (GPPI), Jackson Personality Inventory (JPI), Minnesota Multiphasic Personality Inventory (MMPI), Omnibus Personality Inventory (OPI), Personality Research Form (PRF), and the Sixteen Personality Factor Questionnaire (16PF). Each of these is represented in the Hough and Ones (2001) taxonomy, along with others. A few measures that were deemed to be closely related to scales in this taxonomy were also included. For example, the Eysenck Personality Inventory (an earlier version of the categorized Eysenck Personality Questionnaire) was included. Also, the NEO-FFI (a shortened version of the NEO PI-R) and the NEO-PI (an earlier version of the NEO PI-R) were included. Additionally, Saucier's Mini-Markers (a shortened version of Goldberg's Five Factor Markers) and Goldberg's IPIP (arguably considered a statement version descendant of Goldberg's adjectival Five Factor Markers) were included. Table 1 provides a list of measures that were included in this study.

Table 1
Tests Included in Meta-Analyses (Scale Names in Parentheses)

Agreeableness

ABLE (Cooperativeness)
Adjective Check List (Nurturance)
California Psychological Inventory (Amicability)
Comrey Personality Scales (Empathy)
Edwards Personal Preference Schedule (Nurturance)
Goldberg Big-Five Factor Markers (adjectives) both uni-polar and bi-polar (Factor II: Agreeableness)
Goldberg Big-Five Factor Markers from the International Personality Item Pool, 50 item and 100 item versions (Factor 2)
Hogan Personality Inventory (Likeability)
NEO-FFI (Agreeableness)
NEO-PI (Agreeableness)
NEO PI-R (Agreeableness, Tender-Mindedness)
Personal Characteristics Inventory (Agreeableness)
Personality Research Form (Nurturance)
Saucier's Mini-Markers (Factor II: Agreeableness)

Conscientiousness

Adjective Check List (Achievement, Endurance, Order)
California Psychological Inventory (Achievement via Conformance, Work Orientation)
Comrey Personality Scale (Orderliness)
Edwards Personal Preference Schedule (Achievement, Endurance, Order)
Goldberg Big-Five Factor Markers (adjectives), bi-polar and uni-polar (Factor III: Conscientiousness)
Goldberg Big-Five Factor Markers from the International Personality Item Pool, 50 item and 100 item versions (Factor 3)
Guilford-Zimmerman Temperament Survey (Restraint)
Hogan Personality Inventory (Prudence)
Jackson Personality Inventory (Organization, Responsibility, Risk Taking)
Multidimensional Personality Questionnaire (MPQ) (Harm Avoidance)
NEO PI-R (Achievement Striving, Conscientiousness, Self Discipline)
NEO-FFI (Conscientiousness)
Occupational Personality Questionnaire (Conscientious, Decisive)
Omnibus Personality Inventory (Impulse Expression)
Personal Characteristics Inventory (Conscientiousness)
Personality Research Form (Achievement, Endurance, Harm Avoidance, Impulsivity, Order)
Saucier's Mini-Markers (Factor III: Conscientiousness)
Sixteen Personality Factors (16PF) (Factor G, global Self Control, Q3)

Emotional Stability

Adjective Check List (Ideal Self, Personal Adjustment)
Eysenck Personality Inventory (Neuroticism)
Goldberg Big-Five Factor Markers (adjectives), bi-polar and uni-polar (Factor IV: Emotional Stability)
Goldberg Big-Five Factor Markers from the International Personality Item Pool, 50 item and 100 item versions (Factor 4: Emotional Stability)
Hogan Personality Inventory (Adjustment)
Inwald Personality Inventory (Phobic Personality, Unusual Experiences)
Jackson Personality Inventory (Anxiety)
Minnesota Multiphasic Personality Inventory (MMPI) (Anxiety, Depression, Ego Strength, Hypochondriasis, Obsessiveness, Psychasthenia, Schizophrenia)
MMPI-2 PSY 5 (Neuroticism)
Multidimensional Personality Questionnaire (Stress Reaction)
NEO-FFI (Neuroticism)
NEO PI (Depression, Neuroticism, Vulnerability)
NEO PI-R (Neuroticism)
Occupational Personality Questionnaire (Relaxed)
Personal Characteristics Inventory (Emotional Stability)
Saucier's Mini-Markers (Factor IV: Emotional Stability)
Sixteen Personality Factors (16PF) (Anxiety, Factor C, Emotional Stability)
State Trait Personality Inventory (STPI) (Anxiety)

Extraversion

ABLE (Dominance)
Adjective Check List (Affiliation, Exhibition)
California Psychological Inventory (Sociability, Social Presence)
Comrey Personality Scale (Extraversion)
Edwards Personal Preference Schedule (Dominance)
Eysenck Personality Inventory (Extraversion)
Eysenck Personality Questionnaire (Extraversion)
Goldberg Big-Five Factor Markers (adjectives), uni-polar and bi-polar (Factor I: Surgency)
Goldberg Big-Five Factor Markers from the International Personality Item Pool, 50 item and 100 item versions (Factor 1)
Guilford-Zimmerman Temperament Survey (Ascendancy, General Activity, Sociability)
Hogan Personality Inventory (Sociability)
Inwald Personality Inventory (Loner Type)
Jackson Personality Inventory (Energy Level)
Myers Briggs Type Indicator (Introversion, Extraversion)
MMPI (Social Introversion)
MMPI-2 PSY 5 (Extraversion)

Multidimensional Personality Questionnaire (Social Potency)
NEO-FFI (Extraversion)
NEO PI (Extraversion)
NEO PI-R (Extraversion)
Occupational Personality Questionnaire (Active)
Omnibus Personality Inventory (Social Extroversion)
Personal Characteristics Inventory (Extraversion)
Personality Research Form (Dominance, Exhibition)
Saucier's Mini-Markers (Factor I: Surgency)
Sixteen Personality Factors (16PF) (Factor F, global Extraversion)

Openness to Experience

Adjective Check List (Creative Personality, Change)
Edwards Personal Preference Schedule (Change)
Goldberg Big-Five Factor Markers (adjectives), bi-polar and uni-polar (Factor V: Intellect)
Goldberg Big-Five Factor Markers from the International Personality Item Pool, 50 item and 100 item versions (Factor 5)
Hogan Personality Inventory (Intellectance)
Jackson Personality Inventory (Breadth of Interest, Complexity)
Multidimensional Personality Questionnaire (Absorption)
NEO-FFI (Openness to Experience)
NEO PI (Openness to Experience)
NEO PI-R (Openness to Experience)
Occupational Personality Questionnaire (Conceptual, Innovative)
Personal Characteristics Inventory (Openness)
Personality Research Form (Change, Sentience, Understanding)
Saucier's Mini-markers (Openness)

Scales that were listed as global measures of a Big Five construct or as facets of that construct were grouped as measures of that particular Big Five construct. If a study included validity coefficients (for the same criterion) or reliability coefficients from more than one facet of a Big Five factor, administered to the same group of participants, one of these coefficients were chosen at random after an attempt was made to retain representation of a variety of scales. The choice of one coefficient was made to avoid interdependence among effect sizes. Measures from studies that were included in the

meta-analysis were coded for Big Five construct, as well as for facet according to Hough and Ones (2001) where applicable.

Criterion-Related Validity

Following the example set by Barrick and Mount (1991), criterion-related validity of personality scales were recorded for job proficiency (such as job task, technical, and overall performance ratings as well as productivity data), training proficiency, and personnel data (such as salary changes, tenure, and turnover). Criterion type (objective or subjective) was also coded.

Turnover data, intention to turnover, and absences were incorporated into the withdrawal criterion in the current study. Adequate numbers of effect sizes for meta-analysis were also available for other-rated organizational citizenship behavior (OCB), which could be sub-categorized as individual- or organization-directed in some cases. Contextual performance ratings were also recorded and placed into this category. Counterproductive work behaviors (self-rated, other- rated, and objective), including deviance, formed another criterion category.

Characteristics of Samples

Several characteristics of the sample were coded. The N for effect size, as well as percent female and percent minority in the sample, was recorded when provided. Correlations between these sample characteristics and effect sizes were computed. Other recorded sample characteristics were job type and whether the sample consisted of applicants or incumbents. However, due to the number of subgroup analyses being conducted for personality construct, criterion type, and personality test (leading to ever decreasing *K* for each), no subgroup analyses were computed for job type or

applicant/incumbent status. Additionally, the number of samples consisting of applicants was very small compared to the number that consisted of incumbents.

Analyses

Analyses of Scale Correlations (Convergent Validity). Meta-analyses of correlations between pairs of scales were calculated across all scales, without regard for specific test, for each of the Big Five factors. In other words, independent convergent validity coefficients comparing any two agreeableness scales were meta-analyzed to determine the "average" (using the term loosely) convergence and its distribution, along with other statistics.

Next, meta-analysis of each specific scale's correlations to all other scales of the same construct were conducted when possible. As an illustration, the Intellectance scale (categorized as openness to experience) from the Hogan Personality Inventory was analyzed for its convergent validity with other scales of openness (without respect to the specific tests from which they came).

Thirdly, meta-analyses of correlations between particular pairs of scales of ostensibly the same Big Five construct were conducted when at least six correlations could be found for a given pair of scales. For example, I was able to obtain and analyze correlations between the NEO and CPI scales of Conscientiousness.

Factor-Level Analyses of Criterion-Related Validity and Reliability. Validity for several work criteria (task/technical/overall performance, training performance, counterproductive work behavior, organizational citizenship behavior, and withdrawal) was meta-analyzed for each of the Big Five factors, across scales. Reliabilities were meta-analyzed similarly. In other words, five meta-analyses of each criterion-related

validity type and five meta-analyses of reliability were conducted (one for each of the five personality factors).

Scale-Level Analyses of Criterion-Related Validity and Reliability. Criterion-related validities of each personality scale (at the global factor or facet levels, as categorized by Hough & Ones, 2001) were computed separately by job performance criterion when at least six validity coefficients were found. These can be considered subgroup meta-analyses. Studies using the most popular personality tests and task/technical/overall job performance ratings by supervisors as a criterion were the most available, so these types of meta-analyses were most numerous.

Scale reliabilities were also meta-analyzed for certain specific tests of Big Five global constructs. As much as possible, these tests were chosen to parallel those for which criterion-related validities could be analyzed.

Correlational Analyses. To examine other possible moderators, correlations with criterion-related validity effect sizes were calculated for several sample characteristics. The sample characteristics that were examined were sample size on which the effect size was based, percentage of the sample that was female, and percentage of the sample that was minority (in most cases, African-American, Hispanic, Asian, or other). Criteria for which there were adequate numbers of effect sizes to allow this type of analysis were task/technical/overall performance, organizational citizenship/contextual performance, and counterproductive work behaviors/deviance.

An examination of the correlation between study sample size and effect size was made to determine the extent of publication and presentation bias (studies with smaller N must normally have larger effect sizes to reach significance and be accepted for

publication or presentation at conferences). Although unpublished studies were solicited, these are likely to have been underrepresented due to the difficulty of obtaining them.

The correlation between percent of minority within the sample and effect size was examined to check for implications of differential impact of personality testing based on minority status.

Different norms for male and female samples are often reported by test publishers, but are seldom considered in organizational research. The correlational analysis by gender composition that is calculated here is meant as a preliminary look at whether gender differences should be further examined when using personality as a predictor in the workplace.

Meta-Analytic Procedures

Independence of Effect Sizes. As recommended by Hunter and Schmidt (2004) and Lipsey and Wilson (2001), each meta-analysis was computed using only independent effect sizes in that each sample contributed only one effect size to any particular analysis. This was done because the formulas used to estimate and correct for sampling error assume statistical independence of effect sizes. When the assumption of statistical independence is violated, sampling error variance is underestimated, and the resulting distribution of effect sizes has greater variance than justified. However, as Hunter and Schmidt pointed out, if the number of dependent effect sizes contributed by each study in a meta-analysis is small relative to the total number of effect sizes used in the analysis, error in the estimate of sampling error variance will be reasonably small and not a great concern. According to Hunter and Schmidt, these violations of independence do not bias

the mean effect size found in the meta-analysis, but they may affect confidence intervals around the mean and lead to different interpretations of results.

Outlier Analysis. Outliers for each distribution were carefully examined. Analyses with and without these outliers were conducted if these data points were suspected of having too great an influence on the mean effect size or variance of effect sizes. This is most often a concern when studies that are much larger than the rest (large N) produce effect sizes that are largely discrepant from the remaining studies. In this study, studies were considered outliers based on sample size if they had more than twice as many participants than the next largest study. In a few cases, two or perhaps three studies were considered outliers because they each contained more than twice as many participants as the next largest study. An example would be a set of studies for which there are many sample sizes under 500 and one or two studies with sample sizes in the thousands. Because one of the goals of this study is to gain a clearer, more accurate understanding of the mean and distribution of validity coefficients by measure, it is important to retain the full range of effect sizes that can be expected from a representative sampling of studies. However, because the sample of studies may not be entirely representative, the data were analyzed both with and without possible outlier studies, so that comparisons could be made of the meta-analytic effect sizes and distributions in both cases. In some situations, the inclusion or exclusion of very large studies did not have an appreciable effect on results and their interpretation. However, there were cases for which the decision to eliminate outliers would change the study conclusions. In these cases, caution should be exercised and the best conclusion might be that more studies of all sizes should be conducted and re-analyzed to arrive at a more stable result.

Correcting for Statistical Artifacts. Two approaches to dealing with artifacts were used: 1) "Bare-Bones" meta-analysis in which only sampling error is corrected and 2) Schmidt-Hunter methods (2004) in which the best set of corrections available from the study data were used. In both cases, means, confidence intervals, and credibility intervals are reported as recommended by Hunter and Schmidt (2004).

In the second approach, expected statistical artifacts addressed in meta-analyses of validity coefficients are corrections to individual effect sizes (correlations) for attenuation due to unreliability as well as subject-level sampling error. Reasons that these corrections may be appropriate follow.

Because this study did not seek to closely examine differences in criterion measures other than to consider categories of certain criteria, it was desirable to eliminate what can be considered nuisance variability or measurement error stemming from unreliability due to the criterion. To correct for unreliability, actual criterion reliabilities were used to the extent that these statistics were included in the studies.

Corrections were also made for unreliability in the predictor when using the Schmidt-Hunter approach. Although predictor reliability was examined separately in this study and was considered as a potentially substantive difference among studies, its impact on effect size variability was removed in the Schmidt-Hunter corrections approach, but remained in the Bare-Bones approach. Studies that did not include adequate reliability data were excluded from the corrected meta-analyses.

Hunter and Schmidt (2004) expressed their view that a meta-analysis that does not correct for all possible artifacts is an unfinished meta-analysis. Those who prefer a bare-bones approach to statistical artifacts might argue that it seems unrealistic to imagine that

personality predictors or criterion measures can ever be perfectly reliable; therefore an estimation of effect sizes in an ideal world in which no statistical artifacts remain is less practically useful than an understanding of effect sizes in the observed world. The current study aimed to compute, compare, and discuss results produced by the two approaches to meta-analysis.

Weighting and Combining Effect Sizes. Effect sizes were combined using the weights recommended by Hunter and Schmidt (2004). Sample size weights were used for the bare-bones approach. Adjusted weights were used when correcting for artifacts.

Fixed Effects, Mixed Effects, and Random Effects Models. For each analysis, a random- or mixed-effects model was assumed. When the estimated random-effects variance component for the analysis is zero, this yields a result equivalent to the assumption of a fixed-effects model. Hunter and Schmidt (2004) consider mixed/random effects models preferable to fixed models in nearly all cases.

Although some moderators that were likely to have significant impact were included in the present study, additional factors that are associated with variance in effect sizes probably remain. Therefore, a mixed effects model was tested under the assumption that variability beyond that expected due to sampling error was present but only partially systematic and examined as differences between studies on the specified variables (scale name, for example).

Research has convincingly shown that choice of the appropriate model (fixed, mixed, or random) can have important consequences (e.g., Overton, 1998). Each model carries with it certain assumptions about the type of variance expected in effect sizes. The fixed-effects model assumes that variance in effect sizes between studies is attributable

only to sampling error and/or fixed moderators. The mixed-effects model assumes this variance is attributable to sampling error and fixed moderators, but also to random effects between studies. The random-effects model assumes this variance is attributable to a combination of sampling error and random effects between studies. When these assumptions are not met, confidence intervals can be seriously affected, leading to incorrect conclusions about the significance of the mean effect size or moderator effect. Specifically, when a fixed-effects model is used and random-effects variance is present (a violation of model assumptions), the confidence interval is too narrow and the test is very susceptible to Type I error (too liberal). When mixed- or random-effects models are used and random effects are not present, the opposite problem is likely—an overly wide confidence interval and lower than desired power for detecting real effects (too conservative).

Results

Convergent Validity

Sample-size-weighted mean correlations (estimated mean rho, denoted ρ (est)) between scales, along with the number of correlations on which these means are based (*K*), the total number of participants involved (*N*), weighted variance of the observed correlations (S_r^2), sampling error variance or squared standard error of the observed correlation (SE_r^2), standard deviation of the estimated mean rho (σ_ρ), as well as 95% confidence and credibility intervals are listed in Tables 2 through 6 and Table 8. Table 7 presents a partial correlation matrix for specific scales of extraversion. Effect sizes could not be corrected for unreliability in the personality measures due to the very few cases for which sample-specific reliabilities were reported.

Table 2
Bare-Bones Meta-Analytic Convergent Validities of Specific Agreeableness Scales (with remaining agreeableness scales)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
ACL	10	1330	.41	.025	.005	.141	.31 to .51	.13 to .68
CPI	10	1389	.31	.008	.006	.048	.26 to .37	.22 to .41
Goldberg, Saucier, or IPIP	9	2931	.54	.020	.002	.135	.45 to .64	.28 to .81
HPI	7	985	.48	.017	.004	.113	.39 to .58	.26 to .70
NEO	19	4479	.52	.020	.002	.133	.46 to .59	.26 to .78
PRF	8	1196	.34	.008	.005	.049	.28 to .40	.24 to .44
All Tests	36*	8280	.31	.105	.004	.319	.20 to .42	-.31 to .94
	35	6320	.47	.028	.003	.156	.42 to .53	.17 to .78

Note. All Tests = all tests included in the dataset for this research; ACL = Adjective Check List (Gough); CPI = California Psychological Inventory; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; HPI = Hogan Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; PRF = Personality Research Form
*Includes large (outlier N) studies

Table 3
Bare-Bones Meta-Analytic Convergent Validities of Specific Conscientiousness Scales (with remaining conscientiousness scales)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
ACL	12*	1542	.44	.019	.005	.119	.36 to .52	.20 to .67
	11	1132	.44	.026	.006	.141	.34 to .53	.16 to .71
CPI	27*	5224	.31	.022	.004	.132	.25 to .34	.05 to .57
	26	3685	.27	.027	.006	.143	.21 to .34	-.01 to .56
Goldberg, Saucier, and IPIP	8	2307	.47	.063	.002	.246	.30 to .65	-.01 to .96
HPI	7	1189	.36	.048	.004	.207	.20 to .52	-.04 to .77
NEO	31*	8161	.49	.026	.002	.153	.43 to .54	.19 to .79
	30	6622	.51	.029	.002	.162	.45 to .57	.19 to .83
PRF	8	1259	.34	.039	.005	.184	.20 to .48	-.02 to .70
All Tests	56*	11407	.42	.044	.003	.202	.37 to .48	.03 to .82
	55	9868	.43	.051	.004	.217	.37 to .49	.004 to .85

Note. All Tests = all tests included in the dataset for this research; ACL = Adjective Check List (Gough); CPI = California Psychological Inventory; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; HPI = Hogan Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; PRF = Personality Research Form
*Includes large (outlier N) studies

Table 4
Bare-Bones Meta-Analytic Convergent Validities of Specific Openness Scales (with remaining openness scales)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
ACL	8*	1076	.34	.008	.006	.049	.27 to .40	.24 to .43
	7	666	.28	.006	.009	0	.22 to .34	.28
Goldberg,	6*	1607	.51	.010	.002	.090	.43 - to .59	.33 to .69
Saucier, and IPIP	5	1006	.47	.013	.003	.098	.37 to .57	.28 to .67
HPI	6*	1087	.26	.033	.005	.167	.12 to .41	-.07 to .59
	5	683	.38	.014	.005	.094	.28 to .48	.20 to .56
NEO	19*	5522	.41	.017	.002	.120	.36 to .47	.18 to .65
	18	3562	.48	.015	.003	.107	.42 to .53	.27 to .69
All Tests	26*	6710	.40	.025	.003	.150	.34 to .46	.11 to .70
	25	4750	.45	.029	.003	.161	.38 to .51	.13 to .76

Note. All Tests = all tests included in the dataset for this research; ACL = Adjective Check List (Gough); Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; HPI = Hogan Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI
*Includes large (outlier N) studies

Table 5
Bare-Bones Meta-Analytic Convergent Validities of Specific Extraversion Scales (with remaining extraversion scales)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
ACL	14	2218	.37	.011	.005	.081	.32 to .43	.22 to .53
CPI	32	9365	.57	.016	.002	.121	.53 to .61	.33 to .81
EPI	7*	1017	.66	.010	.002	.090	.59 to .74	.48 to .84
	6	548	.63	.018	.004	.117	.53 to .74	.41 to .86
Goldberg, Saucier, and IPIP	8	2307	.60	.004	.001	.050	.56 to .65	.50 to .70
HPI	8*	1298	.41	.065	.004	.246	.23 to .59	-.07 to .89
	7	894	.58	.005	.004	.032	.53 to .63	.51 to .64
MBTI	42*	11577	.63	.014	.001	.111	.59 to .66	.41 to .85
	41	10359	.65	.013	.001	.107	.61 to .68	.43 to .86
MMPI	19*	7382	.55	.026	.001	.159	.48 to .62	.24 to .86
	17	4080	.58	.025	.002	.152	.50 to .65	.28 to .87
NEO	45	14780	.58	.017	.001	.124	.54 to .62	.34 to .83
PRF	7	1046	.56	.019	.003	.125	.46 to .66	.32 to .81
16PF	14	3289	.61	.015	.002	.116	.55 to .68	.39 to .84
All Tests	103	28521	.56	.023	.002	.145	.53 to .59	.28 to .85

Note. All Tests = all tests included in the dataset for this research; ACL = Adjective Check List (Gough); CPI = California Psychological Inventory; EPI = Eysenck Personality Inventory; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; HPI = Hogan Personality Inventory; MBTI = Myers-Briggs Type Indicator; MMPI = Minnesota Multiphasic Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; PRF = Personality Research Form; 16PF = Sixteen Personality Factors
*Includes large (outlier N) studies

Table 6
Bare-Bones Meta-Analytic Convergent Validities of Specific Emotional Stability Scales (with remaining emotional stability scales)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
ACL	8	1888	.46	.041	.003	.196	.32 to .60	.08 to .85
Goldberg, Saucier, and IPIP	7	2161	.64	.004	.001	.056	.59 to .69	.53 to .75
MMPI	19*	6544	.35	.030	.002	.167	.29 to .43	.02 to .67
	17	3242	.32	.059	.004	.234	.21 to .44	-.14 to .78
NEO	23*	8184	.55	.027	.001	.160	.48 to .61	.23 to .86
	21	4882	.66	.014	.001	.113	.61 to .71	.44 to .88
16PF	7	1333	.66	.010	.002	.092	.58 to .73	.48 to .84
All Tests	35*	11019	.51	.048	.002	.215	.43 to .58	.08 to .93
	33	7717	.51	.059	.002	.239	.43 to .59	.04 to .98

Note. All Tests = all tests included in the dataset for this research; ACL = Adjective Check List (Gough); Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; HPI = Hogan Personality Inventory; MMPI = Minnesota Multiphasic Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; 16PF = Sixteen Personality Factors
*Includes large (outlier N) studies

Table 7
Bare-Bones Meta-Analytic Correlation Matrix of Extraversion Scales

	CPI	NEO	16PF
MBTI	.63	.69	.66
	(7, .07)	(13, .02)	(8, .10)
MMPI		.54	
		(7, .15)	

Note. K, SDrho are in parentheses
CPI = California Psychological Inventory; MBTI = Myers-Briggs Type Indicator; MMPI = Minnesota Multiphasic Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; 16PF = Sixteen Personality Factors

Table 8
Bare-Bones Convergent Validities of Some Specific Test Pairs

Test Pair	Construct	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
NEO and MMPI	Emotional Stability	6*	4205	.41	.010	.001	.097	.33 to .49	.22 to .60
		4	903	.54	.026	.002	.155	.38 to .70	.24 to .84
NEO and CPI	Conscientiousness	7*	2714	.40	.006	.002	.061	.35 to .46	.28 to .52
		6	1175	.42	.012	.004	.095	.33 to .51	.23 to .60
NEO and MMPI	Extraversion	7	4346	.54	.022	.001	.146	.43 to .65	.25 to .82
NEO and MBTI	Extraversion	13	3510	.69	.001	.001	.022	.67 to .72	.65 to .74
CPI and MBTI	Extraversion	7	3671	.63	.005	.001	.066	.58 to .68	.50 to .76
16PF and MBTI	Extraversion	8	1999	.66	.011	.001	.100	.58 to .73	.46 to .85

Note. CPI = California Psychological Inventory; MBTI = Myers-Briggs Type Indicator; MMPI = Minnesota Multiphasic Personality Inventory; NEO = NEO-PI, NEO PI-R, and NEO-FFI; 16PF = Sixteen Personality Factors
*Includes large (outlier N) studies

Estimated mean convergent validities are below .50 in most cases, with convergent validity appearing to be highest among extraversion scales, followed by emotional stability scales. Bare-bones estimates of convergent validities by test ranged from .31 to .54 for agreeableness (see Table 2), .27 to .51 for conscientiousness (see Table 3), .26 to .51 for openness to experience (see Table 4), .37 to .66 for extraversion (see Table 5), and.32 to .66 for emotional stability (see Table 6). Conscientiousness, a current favorite construct in Industrial/Organizational psychology, fares no better than most constructs, with rho estimated to be .42 or .43 over all tests, indicating substantial overlap, but also substantial differences between scales of this construct. Judging by credibility intervals, it appears that there is some convergence among personality tests of the same construct, but it is often unlikely to be above a desired level of .70. However, credibility intervals are generally quite wide, indicating that additional moderating factors may exist and also that more studies may be helpful. These intervals tend to narrow somewhat when it is possible to meta-analyze the convergence of a specific test compared to all others, and they narrow further when examining specific pairs of tests, indicating that specific test name is a moderator of convergence.

Comparisons of convergent validities can be made with results from studies that reported correlations between different factors (e.g. Digman, 1997; Ones, Viswesvaran, & Reiss, 1996; Spector, Schneider, Vance, & Hezlett, 2000). Relevant findings from these studies are included in Table 9. Results from Ones et al. are based on previous meta-analytic research by Ones and are estimated population correlations. Correlations based on the Digman article are unit-weighted, uncorrected mean correlations from nine

adult studies included in his analyses. Results from Spector et al. are based on a single study with N ranging from 332 to 407.

Assuming that results reported in Ones, Viswesvaran, and Reiss (1996) are the most stable due to the large number of studies they are based upon and a large combined N, it is clear that convergent validities are substantially larger than these discriminant validity correlations. Nevertheless, convergent validities vary by test and are lower than the ideal minimum of .70.

Table 9
Mean Correlations among Big Five Personality Dimensions from the Literature

Personality Dimension	1	2	3	4
1. Agreeableness				
2. Conscientiousness	.27 .28 *			
3. Emotional Stability	.25 .42 *	.26 .38 .46		
4. Extraversion	.17 .13 *	.00 .20 .32	.19 .25 .49	
5. Openness to Experience	.11 .12 *	-.06 .13 .27	.16 .12 .30	.17 .40 .40

Note. Results given in or based on the following articles are provided, in order, from top to bottom: Ones, Viswesvaran, & Reiss, 1996; Digman, 1997; Spector, Schneider, Vance, & Hezlett, 2000.
* Not provided

Criterion-Related Validity

A number of meta-analyses were conducted to examine criterion-related validities of the Big Five for training performance, withdrawal (turnover, turnover intentions, absences), organizational citizenship behavior and contextual performance (both overall and separately for OCB-I and OCB-O), counterproductive work behavior and deviance, and task/technical/overall performance. Based on results obtained for each analysis, Tables 10 through 18 have been included. Sample-size-weighted mean validity coefficients (estimated mean rho, denoted ρ (est)), along with the number of correlations on which these means are based (K), the total number of participants involved (N), weighted variance of the observed correlations (S_r^2), sampling error variance or squared standard error of the observed correlation (SE_r^2), standard deviation of the estimated mean rho (σ_ρ), as well as 95% confidence and credibility intervals are listed. For a graphic summary of selected results from these tables, please see Appendix C for preliminary nomological net diagrams for selected tests, based on bare-bones meta-analyses.

When adequate numbers of studies provided both predictor and criterion score reliabilities in their study samples, corrections were made for unreliability as well as for sampling error. Bare-Bones analyses corrected for sampling error only.

It may be noted that in some cases, the standard deviation of rho is zero; therefore the credibility interval is a single value. Although interpretation of this is cautioned here due to the often small numbers of studies included in individual meta-analyses, the interpretation on the face of such results is that sampling error accounts for all the variance in effect sizes and no additional moderators are present.

Training Performance. Estimated mean effect sizes from this study (see Table 10) can be compared to observed mean correlations and estimated true correlations (fully corrected for range restriction as well as sampling error and unreliability using distributions) from Barrick and Mount (1991). Their often-cited meta-analysis found mean correlations (corrected in parentheses) of .06 (.10), .13 (.23), .04 (.07), .15 (.26), and .14 (.25) for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively.

Unfortunately, inadequate numbers of correlations were available for further subgroup (by test) analyses. Therefore, these test-specific analyses were not calculated.

Withdrawal. Although not entirely parallel, results from Table 11 can be compared to results for turnover/tenure from Barrick and Mount (1991). They found mean correlations (corrected in parentheses) of .06 (.09), .09 (.12), .01 (.02), -.03 (-.03), and -.08 (-.11) for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively. These results indicated a tendency for those higher in agreeableness, conscientiousness and emotional stability to stay rather than leave organizations. The current study found small negative correlations between most of the five factors and withdrawal, indicating tendencies not to withdraw, but these effect sizes cannot be considered significant based on credibility intervals.

Inadequate numbers of correlations were available for further subgroup (by test) analyses. Therefore, these test-specific analyses were not calculated.

OCB and Contextual Performance. Table 12 presents results for bare-bones meta-analyzed validity coefficients of each of the Big Five factors for the overall OCB/Contextual Performance criterion. Although estimated rho statistics appear to reveal

several personality constructs as meaningful predictors, credibility intervals are wide enough to include zero, with the exception of agreeableness. Because these credibility intervals include zero even when results from specific tests can be meta-analyzed (with the exception of PCI conscientiousness), it is likely that moderators exist beyond the specific test used.

However, correcting for additional statistical artifacts can strengthen the estimated mean effect size, rho, sometimes pushing the credibility interval upwards so that it no longer includes zero. When corrections for unreliability were made, agreeableness, conscientiousness, and emotional stability emerged as significant predictors of this criterion (see Table 13).

As shown in Table 14, categorizing effect sizes according to whether they focused on citizenship behaviors toward individuals (OCB-I) or toward organizations (OCB-O) revealed significant effects for emotional stability, conscientiousness, and agreeableness with some hint of differential prediction for the two criteria. For example, agreeableness may predict OCB-I better than OCB-O, whereas conscientiousness may better predict OCB-O.

CWB and Workplace Deviance. Confidence intervals for mean effect sizes based on agreeableness and conscientiousness scores indicate that these two factors are potentially meaningful predictors of this criterion (see Tables 15 and 16). However, because none of the mean rho estimates (either Bare-Bones or corrected for unreliability as well) was significant based on credibility intervals, further examination of additional studies and potential moderators is suggested.

Task, Technical, and Overall Performance. Results from this study indicate bare-bones mean correlations of .06, .14, .07, .05, and .03 for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively across all tests (see Table 17). Outlier studies are not included in these statistics. These results are similar in pattern and somewhat similar in size to the results found by Barrick and Mount (1991). For a similar criterion, job proficiency, Barrick and Mount found mean corrected correlations (uncorrected in parentheses) of .06 (.04), .23 (.13), .07 (.04), .10 (.06), and -.03 (-.02) for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively. When only studies that reported sample-specific reliabilities for both predictor and criterion measures are included in the analysis (see Table 18), the current study found corrected mean effect sizes (bare-bones for this sample of studies in parentheses) of .16 (.14), .20 (.18), .11 (.09), .11 (.10), and .03 (.03) for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively. These are similar in pattern (except for a relatively stronger effect size for agreeableness), but stronger or equally strong when compared with the Barrick and Mount results for all of the Big Five factors except conscientiousness. Nevertheless, conscientiousness remains the strongest predictor.

With the exception of openness to experience, the 95% confidence intervals for these effects sizes in the current study did not include zero, indicating that variance around the mean effect size was small enough to produce relatively precise estimates of the mean. However, nearly all 95% credibility intervals for estimated rho included zero. The only exception was for agreeableness when examining studies that included predictor and criterion reliabilities. These wide credibility intervals indicate that the amount of

variance in effect sizes that was attributable to sampling error (and unreliability, in the corrected cases) was relatively small in relation to the overall variance, and the presence of other moderators is likely. Therefore, because the true effect sizes (rho) may vary greatly due to these unexamined moderators, there is less confidence that these true effect sizes have been estimated precisely. In fact, these credibility intervals indicate that the true effect size (validity) of a particular personality construct for prediction of task/technical/overall performance has a greater than .05 chance of being zero in some situations (the reason the credibility interval includes zero).

This provided a good justification for meta-analyzing effect sizes grouped according to the specific personality test used. In doing this, this study considered test as a moderator with the expectation that variance among effect sizes would decrease and thus credibility intervals would narrow. Results shown here (Tables 17 and 18) indicate that this was the case for only some tests. Others continued to show a great deal of variability. Because of the relatively small numbers of studies that these results are based upon, interpretation should be made with caution until further studies can be added to these subgroup analyses.

Table 10
Bare-Bones Criterion-Related Validities for Training Performance

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Agreeableness)	6	1850	.08	.003	.003	.016	.04 to .13	.05 to .11
All Tests (Conscientiousness)	9*	12048	.01	.006	.001	.072	-.04 to .06	-.13 to .15
	8	2255	.02	.031	.004	.167	-.10 to .14	-.30 to .35
All Tests (Emotional Stability)	8*	11951	.03	.003	.001	.049	-.01 to .06	-.07 to.12
	7	2158	.14	.001	.003	0	.12 to .16	.14
All Tests (Extraversion)	9*	12010	.01	.002	.001	.032	-.02 to .04	-.05 to .07
	8	2217	.02	.010	.004	.077	-.05 to .09	-.13 to .17
All Tests (Openness)	6	1850	.10	.004	.003	.036	.05 to .15	.03 to .17

Note. All Tests = all tests included in the dataset for this research
*Includes large (outlier N) studies

Table 11
Bare-Bones Criterion-Related Validities for Withdrawal

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests	10*	5510	-.04	.004	.002	.047	-.08 to -.00	-.13 to .05
(Agreeableness)	8	1023	-.04	.013	.008	.073	-.12 to .04	-.18 to .10
All Tests	15*	3338	-.10	.018	.004	.115	-.16 to -.03	-.32 to .13
(Conscientiousness)	13	1631	-.06	.026	.008	.135	-.15 to .02	-.33 to .20
NEO (Conscientiousness)	6	732	-.07	.023	.008	.120	-.19 to .05	-.31 to .16
All Tests	12*	2041	-.09	.014	.006	.092	-.15 to -.02	-.27 to .09
(Emotional Stability)	11	1432	-.14	.012	.007	.064	-.21 to -.08	-.27 to -.02
All Tests	15*	7248	-.03	.006	.002	.060	-.06 to .01	-.14 to .09
(Extraversion)	12	1663	-.001	.015	.007	.085	-.07 to .07	-.17 to .17
All Tests	10*	1732	.02	.005	.006	0	-.03 to .06	.02
(Openness)	9	1123	-.002	.006	.008	0	-.05 to .05	-.002

Note. All Tests = all tests included in the dataset for this research; NEO = NEO-PI, NEO PI-R, and NEO-FFI
*Includes large (outlier N) studies

Table 12
Bare-Bones Criterion-Related Validities for OCB and Contextual Performance (including OCB-I and OCB-O)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests	28*	17069	.14	.003	.002	.041	.12 to .16	.06 to .22
(Agreeableness)	26	5119	.15	.010	.005	.071	.11 to .19	.01 to .29
NEO (Agreeableness)	11	2090	.17	.016	.005	.103	.10 to .25	-.03 to .38
NEO PI-R (Agreeableness)	6	1325	.22	.019	.004	.120	.11 to .33	-.02 to .45
All Tests	31*	13746	.16	.009	.002	.083	.13 to .19	-.001 to .32
(Conscientiousness)	30	5674	.14	.021	.005	.124	.08 to .19	-.11 to .38
NEO (Conscientiousness)	10	1966	.12	.011	.005	.077	.05 to .18	-.03 to .27
PCI	6*	1198	.17	.008	.005	.053	.10 to .24	.06 to .27
(Conscientiousness)	5	658	.10	.006	.007	0	.04 to .17	.10
All Tests (Emotional	23*	12473	.14	.005	.002	.053	.11 to .17	.03 to .24
Stability)	22	4401	.08	.008	.005	.054	.04 to .12	-.02 to .19
NEO (Emotional Stability)	7	1373	.07	.01	.005	.077	-.01 to .15	-.08 to .22
All Tests	26*	16714	.11	.006	.002	.069	.08 to .14	-.02 to .25
(Extraversion)	24	4764	.06	.017	.005	.107	.01 to .11	-.15 to .27
NEO (Extraversion)	9	1736	.09	.025	.005	.140	-.02 to .19	-.19 to .40
All Tests (Openness)	18	3533	.02	.010	.005	.070	-.02 to .07	-.12 to .16

Note. All Tests = all tests included in the dataset for this research; NEO = NEO-PI, NEO PI-R, and NEO-FFI; PCI = Personal Characteristics Inventory
*Includes large (outlier N) studies

Table 13
Criterion-Related Validities for OCB and Contextual Performance, Corrected for Predictor and Criterion Unreliability (Bare-Bones in Parentheses for Comparison)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Agreeableness)	11	2285	.21 (.17)	.014 (.013)	.006 (.005)	.089 (.089)	(.11 to .24)	.03 to .38 (-.003 to .35)
NEO (Agreeableness)	7	1377	.24 (.20)	.017 (.016)	.006 (.005)	.101 (.104)	(.11 to .29)	.04 to .44 (-.003 to .41)
All Tests (Conscientiousness)	16	2967	.18 (.16)	.014 (.010)	.007 (.005)	.084 (.071)	(.11 to .21)	.02 to .35 (.02 to .30)
NEO (Conscientiousness)	7	1377	.11 (.10)	.015 (.012)	.006 (.005)	.094 (.084)	(.02 to .18)	-.07 to .30 (-.07 to .26)
All Tests (Emotional Stability)	9	1976	.12 (.11)	.007 (.005)	.006 (.004)	.028 (.017)	(.06 to .15)	.07 to .18 (.07 to .14)
All Tests (Extraversion)	10	2055	.05 (.05)	.014 (.010)	.006 (.005)	.085 (.073)	(-.01 to .11)	-.11 to .22 (-.10 to .19)
NEO (Extraversion)	6	1147	.16 (.14)	.020 (.013)	.007 (.005)	.117 (.091)	(.05 to .23)	-.07 to .39 (-.04 to .32)
All Tests (Openness)	7	1452	.05 (.05)	.025 (.017)	.007 (.005)	.134 (.113)	(-.05 to .15)	-.21 to .31 (-.17 to .27)

Note. All Tests = all tests included in the dataset for this research; NEO = NEO-PI, NEO PI-R, and NEO-FFI
*Includes large (outlier N) studies

Table 14
Bare-Bones Criterion-Related Validities for OCB-I and OCB-O

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Agreeableness) For OCB-I	13	3519	.18	.008	.003	.071	.13 to .23	.04 to .32
All Tests (Agreeableness) For OCB-O	16* 14	14386 2436	.13 .11	.002 .008	.001 .006	.027 .049	.11 to .15 .06 to .15	.08 to .18 .01 to .20
All Tests (Conscientiousness) For OCB-I	17	4085	.11	.018	.004	.119	.04 to .17	-.13 to .34
All Tests (Conscientiousness) For OCB-O	17* 16	10855 2783	.18 .19	.003 .013	.001 .005	.042 .085	.16 to .21 .14 to .25	.10 to .27 .03 to .36
All Tests (Emotional Stability) For OCB-I	11	3005	.10	.005	.004	.035	.06 to .14	.03 to .17
All Tests (Emotional Stability) For OCB-O	13* 12	10144 2072	.15 .08	.004 .013	.001 .006	.051 .084	.12 to .19 .02 to .15	.05 to .25 -.08 to .25
All Tests (Extraversion) For OCB-I	12	3289	.07	.025	.004	.145	-.02 to .16	-.21 to .36

Table 14 (Continued)
Bare-Bones Criterion-Related Validities for OCB-I and OCB-O

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests	16*	14385	.11	.006	.001	.072	.07 to .15	-.03 to .25
(Extraversion) For OCB-O	14	2435	-.02	.016	.006	.100	-.08 to .05	-.21 to .18
All Tests (Openness) For OCB-I	10	2631	.01	.009	.004	.075	-.05 to .07	-.13 to .16
All Tests (Openness) For OCB-O	9	1578	.05	.006	.006	.025	.002 to .11	.004 to .10

Note. All Tests = all tests included in the dataset for this research; NEO = NEO-PI, NEO PI-R, and NEO-FFI; PCI = Personal Characteristics Inventory
*Includes large (outlier N) studies

Table 15
Bare-Bones Criterion-Related Validities for CWB and Workplace Deviance

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests	15*	16851	-.13	.02	.001	.153	-.21 to -.05	-.43 to .17
(Agreeableness)	14	8779	-.05	.03	.002	.181	-.15 to .05	-.41 to .31
All Tests	17*	13769	-.12	.055	.001	.232	-.23 to -.01	-.57 to .34
(Conscientiousness)	16	5697	.04	.089	.003	.294	-.10 to .19	-.53 to .62
Goldberg, Saucier, IPIP (Conscientiousness)	7	1379	-.21	.046	.005	.203	-.37 to -.05	-.61 to .19
All Tests	14*	13060	-.02	.039	.001	.195	-.12 to .09	-.40 to .36
(Emotional Stability)	12	2820	-.06	.009	.004	.069	-.11 to -.004	-.19 to .08
All Tests	16*	17590	-.01	.003	.001	.040	-.03 to .01	-.09 to .07
(Extraversion)	13	3472	.04	.005	.004	.041	.002 to .08	-.04 to .12
All Tests	12*	4770	.01	.006	.003	.059	-.03 to .06	-.10 to .13
(Openness)	11	2602	-.02	.009	.004	.068	-.07 to .04	-.15 to .11

Note. All Tests = all tests included in the dataset for this research; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, International Personality Item Pool (50 and 100 item versions)
*Includes large (outlier N) studies

Table 16
Criterion-Related Validities for CWB and Workplace Deviance, Corrected for Predictor and Criterion Unreliability (Bare-Bones in Parentheses for Comparison)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Agreeableness)	10	2034	-.27 (-.22)	.042 (.029)	.007 (.004)	.190 (.157)	(-.33 to -.12)	-.65 to .10 (-.53 to .08)
All Tests (Conscientiousness)	12	2612	-.27 (-.22)	.045 (.029)	.006 (.004)	.197 (.159)	(-.32 to -.12)	-.65 to .12 (-.53 to .09)
Goldberg, Saucier, IPIP (Conscientiousness)	6	1160	-.25 (-.22)	.082 (.054)	.007 (.005)	.275 (.222)	(-.40 to -.03)	-.79 to .29 (-.65 to 22)
All Tests (Emotional Stability)	9	1903	-.09 (-.08)	.017 (.011)	.007 (.005)	.102 (.082)	(-.15 to -.01)	-.29 to .11 (-.24 to .08)
All Tests (Extraversion)	9	1903	.03 (.03)	.013 (.009)	.007 (.005)	.082 (.068)	(-.04 to .09)	-.13 to .19 (-.11 to .16)
All Tests (Openness)	9	1903	-.07 (-.06)	.010 (.006)	.007 (.005)	.048 (.037)	(-.11 to -.01)	-.16 to .02 (-.13 to .02)

Note. All Tests = all tests included in the dataset for this research; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, International Personality Item Pool (50 and 100 item versions)
*Includes large (outlier N) studies

Table 17
Bare-Bones Criterion-Related Validities for Task, Technical, and Overall Performance

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests	62*	22108	.06	.007	.003	.061	.04 to .08	-.06 to .18
(Agreeableness)	60	10158	.06	.012	.006	.075	.03 to .09	-.09 to .21
Goldberg, Saucier, IPIP (Agreeableness)	9	1474	.01	.008	.006	.048	-.05 to .07	-.08 to .10
HPI (Likeability, or Agreeableness)	7	1526	.02	.008	.005	.058	-.05 to .08	-.10 to .13
NEO-FFI (Agreeableness)	9	973	.04	.024	.009	.120	-.06 to .14	-.20 to .27
NEO PI-R (Agreeableness)	11	1793	.09	.015	.006	.094	.01 to .16	-.10 to .27
PCI (Agreeableness)	20	2787	.09	.006	.007	0	.06 to .13	.09
All Tests	93*	33236	.07	.013	.003	.101	.05 to .10	-.12 to .27
(Conscientiousness)	91	15371	.14	.021	.006	.123	.11 to .17	-.10 to .38
Goldberg, Saucier, IPIP (Conscientiousness)	14	2403	.14	.014	.006	.089	.08 to .20	-.03 to .32
HPI (Prudence, or	7*	1526	.11	.018	.004	.115	.01 to .21	-.11 to .34
Conscientiousness)	6	917	.19	.015	.006	.097	.09 to .29	-.004 to .38
NEO-FFI (Conscientiousness)	13	1530	.15	.017	.008	.095	.08 to .22	-.04 to .33

Table 17 (Continued)
Bare-Bones Criterion-Related Validities for Task, Technical, and Overall Performance

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
NEO PI-R (Conscientiousness)	14	2301	.21	.008	.006	.046	.16 to .25	.12 to .30
PCI (Conscientiousness)	26*	4474	.18	.010	.005	.070	.14 to .22	.04 to .32
	25	3661	.18	.013	.006	.078	.14 to .23	.03 to .34
All Tests (Emotional Stability)	68*	28525	.03	.004	.002	.044	.01 to .05	-.06 to .12
	66	10660	.07	.008	.006	.043	.05 to .10	-.01 to .16
Goldberg, Saucier, IPIP (Emotional Stability)	11	1683	.08	.006	.006	0	.04 to .13	.08
HPI (Adjustment, or Emotional Stability)	7	1526	.02	.006	.005	.031	-.03 to .08	-.04 to .08
NEO-FFI (Emotional Stability)	10	1097	.09	.011	.009	.047	.02 to .15	-.003 to .18
NEO PI-R (Emotional Stability)	9	1439	.02	.012	.006	.074	-.05 to .09	-.13 to .16
PCI (Emotional Stability)	20	2787	.10	.005	.007	0	.07 to .13	.10
All Tests (Extraversion)	75*	32960	.02	.006	.002	.058	.005 to .04	-.09 to .13
	72	11217	.05	.015	.006	.093	.02 to .08	-.13 to .23
Goldberg, Saucier, IPIP (Extraversion)	12	1960	.11	.008	.006	.047	.06 to .17	.02 to .21

Table 17 (Continued)
Bare-Bones Criterion-Related Validities for Task, Technical, and Overall Performance

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
HPI (Sociability, or Extraversion)	6	1383	-.08	.004	.004	0	-.13 to -.03	-.08
NEO-FFI (Extraversion)	10	1072	.10	.017	.009	.086	.02 to .18	-.07 to .27
NEO PI-R (Extraversion)	10	1591	.04	.010	.006	.062	-.02 to .11	-.08 to .16
PCI (Extraversion)	19	2708	.04	.015	.007	.086	-.01 to .10	-.12 to .21
All Tests (Openness)	58*	10823	.05	.014	.005	.095	.02 to .08	-.13 to .23
	57	9123	.03	.014	.006	.090	-.004 to .06	-.15 to .20
Goldberg, Saucier, IPIP (Intellect, or Openness)	11	1637	.02	.011	.007	.062	-.04 to .08	-.10 to .15
HPI (Intellectance, or	7*	3083	.06	.015	.002	.112	-.03 to .15	-.16 to .28
Openness)	6	1383	-.07	.005	.004	.015	-.12 to -.01	-.09 to -.04
NEO-FFI (Openness)	9	973	-.01	.032	.009	.152	-.12 to .11	-.30 to .29
NEO PI-R (Openness)	10	1484	-.04	.011	.007	.062	-.10 to .03	- .16 to .09
PCI (Openness)	18	2618	.08	.006	.007	0	.04 to .11	.08

Note. All Tests = all tests included in the dataset for this research; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, International Personality Item Pool (50 and 100 item versions); HPI = Hogan Personality Inventory; PCI = Personal Characteristics Inventory
*Includes large (outlier N) studies

Table 18
Criterion-Related Validities for Task, Technical, and Overall Performance, Corrected for Predictor and Criterion Unreliability (Bare-Bones in Parentheses for Comparison)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Agreeableness)	17	2633	.16 (.14)	.011 (.010)	.008 (.006)	.055 (.061)	(.09 to .18)	.06 to .27 (.02 to .25)
Goldberg, Saucier, IPIP (Agreeableness)	6	647	.11 (.09)	.007 (.006)	.012 (.009)	0 (0)	(.03 to .15)	.11 (.09)
All Tests (Conscientiousness)	25	4093	.20 (.18)	.020 (.015)	.007 (.006)	.110 (.096)	(.13 to .23)	-.01 to .42 (-.01 to .37)
Goldberg, Saucier, IPIP (Conscientiousness)	8	1037	.19 (.17)	.014 (.010)	.010 (.007)	.064 (.052)	(.10 to .24)	.07 to .32 (.07 to .27)
NEO-FFI (Conscientiousness)	6	723	.18 (.15)	.041 (.032)	.011 (.008)	.174 (.156)	(.005 to .29)	-.16 to .52 (-.16 to .46)
All Tests (Emotional Stability)	16	2117	.11 (.09)	.014 (.011)	.010 (.007)	.064 (.057)	(.04 to .14)	-.02 to .23 (-.02 to .20)
Goldberg, Saucier, IPIP (Emotional Stability)	8	888	.14 (.12)	.011 (.008)	.012 (.009)	0 (0)	(.06 to .18)	.14 (.12)
All Tests (Extraversion)	18	2379	.11 (.10)	.018 (.013)	.010 (.007)	.092 (.075)	(.05 to .15)	-.07 to .29 (-.05 to .25)
Goldberg, Saucier, IPIP (Extraversion)	8	983	.15 (.13)	.018 (.014)	.010 (.008)	.086 (.077)	(.05 to .21)	-.02 to .32 (-.02 to .28)

Table 18 (Continued)
Criterion-Related Validities for Task, Technical, and Overall Performance, Corrected for Predictor and Criterion Unreliability (Bare-Bones in Parentheses for Comparison)

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
All Tests (Openness)	18	2211	.03 (.03)	.023 (.016)	.012 (.008)	.109 (.091)	(-.03 to .09)	-.18 to .24 (-.15 to .21)
Goldberg, Saucier, IPIP (Intellect, or Openness)	9	999	.03 (.02)	.015 (.012)	.012 (.009)	.057 (.051)	(-.05 to .09)	-.08 to .14 (-.08 to .12)

Note. All Tests = all tests included in the dataset for this research; Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, International Personality Item Pool (50 and 100 item versions)
*Includes large (outlier N) studies

Reliability

Bare-Bones meta-analyses of reliabilities, reported by scale in Table 19, indicate satisfactory reliabilities for research purposes (over .70) across Big Five dimensions. Furthermore, reliability did not appear to differ much across commonly-used tests. However, only 2 scales, NEO PI-R Conscientiousness and Emotional Stability, surpassed the minimum reliability of .90 suggested by Nunnally and Bernstein (1994) for important decisions such as those related to employee selection.

Table 19
Bare-Bones Meta-Analysis of Reliability

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
PCI (Agreeableness)	9	2034	.79	.004	.001	.055	.75 to .83	.68 to .90
PCI (Conscientiousness)	14	3595	.81	.004	.000	.061	.77 to .84	.69 to .93
PCI (Emotional Stability)	9	2034	.85	.001	.000	.028	.82 to .87	.79 to .90
PCI (Extraversion)	9	2034	.85	.000	.000	.012	.84 to .87	.83 to .88
PCI (Openness)	9	2034	.81	.003	.001	.045	.78 to .84	.72 to .90
NEO PI-R (Agreeableness)	10	2116	.86	.005	.000	.072	.82 to .91	.72 to 1.0
NEO PI-R (Conscientiousness)	11	2021	.91	.002	.000	.040	.89 to .93	.83 to .99
NEO PI-R (Emotional Stability/Neuroticism)	10	1919	.90	.001	.000	.029	.89 to .92	.85 to .96
NEO PI-R (Extraversion)	13	2416	.86	.003	.000	.051	.83 to .89	.76 to .96
NEO PI-R (Openness)	8	1480	.85	.005	.000	.068	.80 to .90	.75 to .98
NEO-FFI (Agreeableness)	13	2300	.74	.002	.001	.037	.71 to .77	.67 to .82
NEO-FFI (Conscientiousness)	18	2996	.81	.003	.001	.050	.78 to .84	.71 to .91
NEO-FFI (Emotional Stability/Neuroticism)	12	2014	.80	.003	.007	.043	.77 to .83	.71 to .88
NEO-FFI (Extraversion)	15	2538	.79	.004	.001	.053	.76 to .82	.69 to .89

Table 19 (Continued)
Bare-Bones Meta-Analysis of Reliability

Test	*K*	*N*	ρ (est)	S_r^2	SE_r^2	σ_ρ	95% Confidence interval	95% Credibility interval
NEO-FFI (Openness)	11	1789	.75	.005	.001	.058	.71 to .79	.64 to .87
Goldberg/Saucier/IPIP (Agreeableness)	17	3112	.79	.006	.001	.073	.76 to .83	.65 to .94
Goldberg/Saucier/IPIP (Conscientiousness)	25	4538	.82	.002	.001	.031	.81 to .84	.76 to .88
Goldberg/Saucier/IPIP (Emotional Stability)	18	3274	.80	.005	.001	.065	.76 to .83	.67 to .92
Goldberg/Saucier/IPIP (Extraversion)	18	3369	.82	.004	.001	.057	.79 to .85	.71 to .93
Goldberg/Saucier/IPIP (Intellect or Openness)	19	3333	.79	.006	.001	.073	.75 to .82	.64 to .93
All Tests	54*	13729	.80	.004	.000	.063	.79 to .82	.68 to .93
(Agreeableness)	53	9851	.80	.006	.001	.074	.78 to .82	.65 to .95
All Tests (Conscientiousness)	77	15218	.81	.009	.001	.089	.79 to .83	.64 to .98
All Tests (Emotional Stability)	53	9741	.82	.011	.001	.100	.79 to .85	.62 to 1.0
All Tests (Extraversion)	61	16355	.83	.004	.000	.058	.81 to .84	.71 to .94
All Tests (Openness)	51	8911	.78	.006	.001	.070	.76 to .80	.65 to .92

Note. Goldberg, Saucier, and IPIP = Goldberg Big Five Factor Markers, Saucier Mini-Markers, and International Personality Item Pool; PCI = Personal Characteristics Inventory
*Includes large (outlier N) studies

Correlational Analyses

Correlations of certain sample characteristics with effect sizes were calculated. These characteristics were sample size, percent female, and percent minority. Criteria for which adequate numbers of studies included this information were counterproductive work behaviors/deviance, organizational citizenship behavior/contextual performance, and task/technical/overall performance.

For counterproductive work behavior/deviance, agreeableness appeared to be a stronger predictor as percentages of females increased in study samples (see Table 20). (For this construct only, validities are already negative, so a negative correlation here further strengthens the validity, whereas a positive one weakens it.) This was also the case for conscientiousness and openness. On the other hand, openness appeared to lose predictive ability as the percentage of minorities in samples increased. However, none of these zero-order correlations were statistically significant.

It should be noted that sample size and percent female were strongly and significantly correlated for agreeableness, extraversion, and openness validities, and this correlation neared significance for the remaining dimensions. In other words, larger study samples tended to include greater percentages of females.

Table 20
Zero-Order Correlations between Sample Characteristics and Personality Validities for Counterproductive Work Behavior/Deviance

Personality Dimension Validity	Sample Size	Percent Female	Percent Minority
Agreeableness	.18	-.40	-.07
	.47	.14	.84
	18	15	10
Conscientiousness	.09	-.43	.28
	.70	.07	.36
	23	19	13
Emotional Stability	.13	-.40	.15
	.60	.14	.66
	18	15	11
Extraversion	-.14	.14	.22
	.59	.60	.49
	19	16	12
Openness	.33	-.39	.52
	.24	.18	.15
	15	13	9

Note. Correlation listed with significance (*p* value) and sample size, in order from top to bottom.

For the criterion OCB/Contextual Performance, only openness validities and sample size were significantly related (see Table 21). Results indicated that openness validities tended to be smaller as sample size increased. This is suggestive of publication/presentation bias in which smaller samples with weaker effects are less likely to be published or accepted for conferences.

Not significant but nevertheless interesting are the results showing that extraversion tended to be a weaker predictor for samples with higher percentages of females and minorities. Also, there is a tendency for conscientiousness validities to become less predictive especially as minority percentages increase.

It should be noted that among the studies that reported both percent female and percent minority for this group of analyses, the correlation between those two characteristics was strong and significant, ranging from .57 to .82.

Table 21
Correlations between Sample Characteristics and Validities for Organizational Citizenship Behavior/Contextual Performance

Personality Dimension Validity	Sample Size	Percent Female	Percent Minority
Agreeableness	-.02	-.04	-.24
	.91	.84	.33
	37	33	18
Conscientiousness	.09	-.15	-.29
	.53	.35	.18
	52	39	23
Emotional Stability	.09	.12	-.19
	.64	.59	.48
	29	23	16
Extraversion	.11	-.32	-.44
	.50	.08	.08
	41	31	17
Openness	-.40	.11	-.11
	.03	.64	.72
	28	19	13

Note. Correlation listed with significance (*p* value) and sample size, from top to bottom.

Table 22 presents results that relate personality validities for task/technical/overall performance to sample characteristics. Although none of the zero-order correlations were significant, two that approached significance may be of particular interest: both emotional stability and openness appeared to become stronger predictors of performance as samples included a larger percentage of females.

For the emotional stability validities, it should be noted that a significant correlation of .46 existed between the percent female and percent minority sample characteristics for the 40 studies that reported both.

Table 22
Correlations between Sample Characteristics and Validities for Task/Technical/Overall Performance

Personality Dimension Validity	Sample Size	Percent Female	Percent Minority
Agreeableness	.07	.18	-.09
	.49	.16	.58
	89	60	34
Conscientiousness	-.10	-.12	.01
	.20	.23	.93
	170	101	64
Emotional Stability	-.13	.21	-.14
	.18	.08	.41
	107	69	40
Extraversion	-.11	-.14	-.11
	.19	.21	.45
	138	82	50
Openness	.14	.21	-.16
	.15	.09	.38
	113	63	33

Note. Correlation listed with significance (*p* value) and sample size, in order from top to bottom.

Discussion

An underlying assumption of previous meta-analyses involving the prediction of job performance using Big Five personality factors is that all personality scales that ostensibly measure the same factor are similar enough to group into a common meta-analysis. To assess the degree of similarity among personality scales that are commonly used in organizational studies, two indicators were examined: 1) high correlations among predictor scales (i.e., evidence of a single factor) and 2) similar patterns of correlations between predictor scales and job-related criteria (i.e., similar nomological nets). Results of this study indicated that the assumption of similarity may not be entirely met, particularly with regard to correlations among predictor scales.

Convergent validities were lower than might be expected, indicating that substantial differences between tests exist. For both the agreeableness and conscientiousness constructs, convergent validities with a variety of other tests were highest for the NEO and the Goldberg families of tests and lowest for the CPI and PRF.

One explanation for these differences may be that the NEO and Goldberg tests were intended as measures of Big Five factors, whereas the CPI and PRF were based on other models of personality. Research by Salgado (2003) showed greater criterion validity of measures that were based on the Five Factor Model compared to those that were not based on this model. He also contended that convergent validity should be lower across measure types than among Big Five measures exclusively. In the current study,

both NEO and Goldberg scales measured global agreeableness, conscientiousness or facets of these factors such as tender-mindedness (agreeableness), achievement striving, or self discipline (conscientiousness) from the NEO PI-R. However, the CPI was not intended to measure the Big Five. Its amicability scale, which was classified by Hough and Ones (2001) as global agreeableness, is a special purpose scale from that inventory. Although the description of the amicability scale seems very similar to those for agreeableness scales, the current research indicated that substantial differences in operationalization of the concept and/or focus of the items were likely. The relatively low convergence of this test with others was almost certainly due to conceptual differences rather than to simple format differences, such as its use of true-false response options as compared to many other scales' use of Likert-type scale options.

Another explanation for relatively low convergent validity, especially applicable to the PRF, is that the Hough and Ones (2001) taxonomy did not classify any of the PRF scales into global agreeableness or conscientiousness. Rather, the included scales from the PRF for these factors were "nurturance," classified into the agreeableness facet of the same name, "achievement" which was classified into the conscientiousness facet of the same name, "harm avoidance" and "impulsivity" which were classified into the cautiousness/impulse control vs. risk taking/impulsive facet of conscientiousness, "order," classified into the conscientiousness facet of the same name, and "endurance" which was classified into the persistence facet of conscientiousness. Recognition that some tests do not measure the global five factors, but rather select facets of them is a great step toward understanding the similarities and differences among personality tests

and why certain measures may be more useful in specific circumstances and when attempting to predict certain criteria.

Clearly, continued development of facet taxonomies and categorization of measures into facets and global factors is needed (Roberts, Chernyshenko, Stark, & Goldberg, 2005). Results from the Roberts et al. study suggested facets of industriousness, order, self-control, responsibility, traditionalism, and virtue as constituents of conscientiousness. Because their study combined data across thirty-six scales from seven different personality inventories, it is possible that some of the inventories neglected to measure a particular facet, whereas others may have focused heavily on that facet. Further work is needed to clarify constituent facets of agreeableness and the other three factors. Perhaps the earlier mentioned CPI amicability scale would fit better into a facet of agreeableness, or other uncategorized scales from the CPI would be more appropriate measures of global agreeableness.

For extraversion, the ACL exhibited particularly low convergent validity. The ACL was developed to measure needs such as exhibition and affiliation. In fact, the scales for exhibition and affiliation, included in this study, were not classified as measures of global extraversion by Hough and Ones (2001), but rather as measures of the facets of dominance and sociability, respectively. It is likely that these scales measure certain aspects of some of the other Big Five factors in addition to elements of extraversion (Piedmont, McCrae, & Costa, 1991). Overlap with several factors would tend to decrease the convergence with any one factor.

Also, the MMPI scales that were categorized into emotional stability displayed fairly low convergent validity. In this case, this is probably due to differences in test

development procedures and goals. The MMPI was originally developed empirically to predict membership in specific clinical groups, whereas most other tests were developed rationally to measure normal personality.

Unreliability of measurement could explain lower convergent validities to a degree. However, it is unlikely to be the entire explanation because reliability was shown to be uniformly satisfactory among tests.

Relatively low convergent validity does not mean that tests with this quality necessarily differ in their usefulness for prediction. Those with higher convergent validities are more similar, and are presumably measuring something closer to a generally understood concept of the construct, whereas those with lower convergent validities may be measuring less commonly included aspects of the construct. If these less commonly included aspects add to the criterion-related validity of the test, it could be helpful to identify them and include them in other tests as well. However, if they are not useful, elimination of the discrepant aspects might be advisable. A closer look at criterion validities, particularly at the facet and item levels, would clarify this issue.

In the current study, the overall pattern of criterion validity results according to Big Five construct was consistent with previous literature. This indicates that the group of studies examined in this paper is not very different in nature from those examined in previous meta-analyses. The numbers of studies included in these meta-analyses (*K*) are, in many cases, similar to the numbers in previous published analyses. This study made new contributions by examining subgroup validities for specific tests when possible.

Some differences in validities by test were found. When examining validities for task/technical/overall performance, for example, we see that the NEO PI-R appears to be

a superior predictor among conscientiousness scales, based on its greater validity and narrower credibility interval. Some of the difference in validity may be explained by the greater comprehensiveness in construct coverage by the NEO PI-R, especially as compared to the NEO-FFI, a shortened version. Perhaps some of the more predictive items in work contexts were eliminated for the shortened form. On the other hand, comprehensiveness comes with a price; administration of the NEO PI-R costs more than the NEO-FFI in time, effort, and money.

If we square rho (correcting only for sampling error) as an indicator of the amount of variance in performance that is accounted for by personality scores, we find that NEO PI-R conscientiousness scores account for over 4% of this variance. In contrast, variance accounted for by conscientiousness tests in general is about 2%. These seem like disappointingly low amounts of variance to consider, and there is certainly much room for improvement. In practical terms, however, any improvement in decision making can lead to competitive advantage. As mentioned by Hogan and Roberts (2001), when only half of the applicant pool has acceptable levels of a desirable quality, a validity coefficient of .20 (slightly lower than estimated for the NEO PI-R conscientiousness scale) improves the probability of a correct hiring decision from 50% to 60% when used as the sole predictor. Of course, higher validity coefficients and the inclusion of additional valid predictors further increase decision-making accuracy.

Using the standards of greater validity and narrow credibility interval (not including zero), the PCI conscientiousness scale also appears to be a consistent predictor, while some other tests appear to be less consistent. This instability could be a function of

few effect sizes, so additional studies could enhance our understanding of the relationships between tests.

To a lesser degree, the remaining four constructs may have some predictive ability for task/technical/overall performance, but tests again appear to vary with some being more consistent predictors than others.

Regardless of credibility intervals, variation in mean level of validity among tests is of practical interest. Informed test consumers can be expected to prefer tests with higher validity. Knowledge of test content and test statistics are important for both test selection and interpretation of scores.

Interestingly, criterion-related validities did not differ as much as one might expect, given the only moderate convergent validities that were observed. Instead, it appears that many personality scales are predicting a portion of variance in criterion scores. But the portions of variance in scores may not be entirely overlapping. If the most effective aspects of the variety of tests currently in use can be determined and combined, perhaps substantial gains in validity for job criteria could be realized.

Further investigation into predictive validities of specific tests for other work-related criteria is advised as studies that report the necessary information accumulate.

Reliabilities of individual scales did not appear to differ significantly across measures. These results, though informative, do not suggest any changes in current procedures regarding reliability. Reliability distributions from the literature that are used by some to correct for statistical artifacts in meta-analysis are probably adequate, assuming these reliabilities are consistent with those reported here.

Correlations of validity coefficients with sample characteristics revealed some potentially interesting results. For several constructs, predictive validity for various work criteria appeared to increase or decrease as samples included larger percentages of females or minorities. We can also think of these effects as affecting increasingly male or heterogenous majority groups in exactly the opposite way. Although many organizations track the effects of hiring decisions by gender and ethnic/racial group, it is seldom clear that researchers examine validities by comparing these subgroups. Clearly, many test developers are aware that norms for personality scores can differ by gender and many report these norms separately as well as combined. A strong encouragement for researchers to report results by subgroup is in order.

Directions for future study include gathering more extensive data to add to the meta-analysis of validities by test. This could be accomplished by inclusion of future studies that report these correlation coefficients, as well as studies conducted and published prior to 1990. Continued vigorous search for unpublished studies could also add to the number of studies to be meta-analyzed. Although additional study results may be available from test developers, it is preferable that independent sources supply the bulk of the included coefficients, rather than including a predominance of effect sizes from studies conducted by the publishers/developers themselves.

Although differences were found among scales in terms of what they measure, based on convergent validities, the specific nature of these differences is not fully known. Comparing factor analyses that yield constituent facets of each of the most commonly used tests of a particular Big Five factor may aid the effort to define content differences between the tests.

Differences in predictive validities for job-related criteria were observed, but were not so extreme that these differences should dictate choice of measures. Selection of measures may be better based on practical considerations such as cost, ease of administration, or personal preference.

Researchers can have reasonable confidence in the generalizability of past personality research into validity. However, questions remain about exactly what aspects of the different tests are predicting job outcomes effectively and whether these predictive "pieces" overlap among tests or are somewhat different. If different, they could be combined to produce a better functioning measure while less predictive aspects of the current measures are eliminated. Continuing efforts toward the improvement of personality testing for prediction of work criteria are encouraged.

References

References marked with an asterisk indicate studies included in the meta-analyses.

*Allworth, E., & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment, 8*(3), 137-147.

*Anderson, N., & Ones, D. S. (2003). The construct validity of three entry level personality inventories used in the UK: Cautionary findings from a multiple-inventory investigation. *European Journal of Personality, 17*(Suppl. 1), 39-66.

*Ashton, M. C., & Lee, K. (2005). Honesty-humility, the Big Five, and the Five-Factor Model. *Journal of Personality, 73*(5), 1321-1353.

*Bacha, R. R. (2003). Specifying personality and self-monitoring effects on overall, task, and contextual job performance. *Dissertation Abstracts International, 64* (7-A), 2004, pp. 2552. (UMI No. 3098510)

*Baer, M., & Oldham, G. R. (2006). The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology, 91*(4), 963-970.

*Bajor, J. K., & Baltes, B. B. (2003). The relationship between selection optimization with compensation, conscientiousness, motivation, and performance. *Journal of Vocational Behavior, 63,* 347-367.

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

*Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78,* 111-118.

*Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*(3), 261-272.

*Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology, 78*(5), 715-722.

*Barrick, M. R., Parks, L., & Mount, M. K. (2004, April). *Self-monitoring as a moderator of the relationships between personality and performance.* Poster presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

*Bauer, T. N., Erdogan, B., Liden, R. C., & Wayne, S. J. (2006). A longitudinal study of the moderating role of extraversion: Leader-member exchange, performance, and turnover during new executive development. *Journal of Applied Psychology, 91*(2), 298-310.

*Beaty, J. C., Jr., Cleveland, J. N., & Murphy, K. R. (2001). The relation between personality and contextual performance in "strong" versus "weak" situations. *Human Performance, 14*(2), 125-148.

*Bessmer, M. A., & Ramanaiah, N. V. (1981). Convergent and discriminant validity of selected need scales from the Adjective Check List and Personality Research Form. *Psychological Reports, 49*(1), 311-316.

*Bibeau-Reaves, D. (2002). Personality traits and preference types of clergy and their effects on church attendance and membership. *Dissertation Abstracts International, 63* (10-B), 2003, pp. 4949. (UMI No. 3068405)

*Bing, M. N., & Lounsbury, J. W. (2000). Openness and job performance in U.S.-based Japanese manufacturing companies. *Journal of Business and Psychology, 14*(3), 515-522.

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.

*Bishop, N. B. (1996). The Big Five and personnel selection: Factors impacting responses and criterion validities. *Dissertation Abstracts International, 57* (3-B), 1996, pp. 2195. (UMI No. 9623208)

*Black, J. (2000). Personality testing and police selection: Utility of the "Big Five." *New Zealand Journal of Psychology, 29*(1), 2-9.

*Bozionelos, N. (2004). The relationship between disposition and career success: A British study. *Journal of Occupational and Organizational Psychology, 77*(3), 403-420.

*Briggs, S. R. (1992). Assessing the Five-Factor Model of personality description. *Journal of Personality, 60*(2), 253-293.

*Burke, L.A., & Witt, L.A. (2002) Moderators of the openness to experience-performance relationship. *Journal of Managerial Psychology, 17*(8), 712-721.

*Burke, L. A., & Witt, L. A. (2004). Personality and high-maintenance employee behavior. *Journal of Business and Psychology, 18*(3), 349-363.

*Bushe, G. R., & Gibbs, B. W. (1990). Predicting organization development consulting competence from the Myers-Briggs Type Indicator and stage of ego development. *Journal of Applied Behavioral Science, 26*(3), 337-357.

*Byravan, A. (1996). Structure of personality disorders from a Five-Factor Model perspective, and the relative superiority of the MMPI-2 PSY-5, NEO-PI-R, and the 16 PF Fifth Edition scales for predicting personality disorders. *Dissertation Abstracts International, 57* (10-B), 1997, pp. 6635. (UMI No. 9708742)

*Byrne, Z. S., Stoner, J., Thompson, K. R., & Hochwarter, W. (2005). The interactive effects of conscientiousness, work effort, and psychological climate on job performance. *Journal of Vocational Behavior, 66,* 326-338.

*Caligiuri, P. M. (2000). The Big Five personality characteristics as predictors of expatriate's desire to terminate the assignment and supervisor-rated performance. *Personnel Psychology, 53*(1), 67-88.
*Canivez, G. L., & Allen, T. J. (2005, August). *Convergent and factorial validity of the 16PF and the NEO-PI-R.* Paper presented at the Annual Convention of the American Psychological Association, Washington, DC.
*Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the Five Factor Model. *Journal of Applied Psychology, 81*(6), 694-704.
*Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*(3), 233-254.
*Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*(4), 847-860.
*Church, A. T. (1994). Relating the Tellegen and Five-Factor Models of personality structure. *Journal of Personality and Social Psychology, 67*(3), 898-909.
*Clevenger, J., Pereira, G. M., Weichmann, D., Schmitt, N., & Schmidt Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*(3), 410-417.
*Colbert, A. E., Mount, M. K., Harter, J. K., Witt, L. A., & Barrick, M. R. (2004). Interactive effects of personality and perceptions of the work situation on workplace deviance. *Journal of Applied Psychology, 89*(4), 599-609.
*Colbert, A. E., Mount, M. K., Harter, J. K., Witt, L. A., & Barrick, M. R. (2004). Interactive effects of personality and perceptions of the work situation on workplace deviance. Unpublished results.
*Collins, J. M., & Schmidt, F. L. (1993). Personality, integrity, and white collar crime: A construct validity study. *Personnel Psychology, 46*(2), 295-311.
*Conte, J. M., & Gintoft, J. N. (2005). Polychronicity, Big Five personality dimensions, and sales performance. *Human Performance, 18*(4), 427-444.
*Conte, J. M., & Jacobs, R. R. (2003). Validity evidence linking polychronicity and Big Five personality dimensions to absence, lateness, and supervisory performance ratings. *Human Performance, 16*(2), 107-129.
Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods, 6,* 415-439.
*Costa, P. T., Jr., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of the MMPI factor scales with measures of the Five Factor Model of personality. Journal of Personality Assessment, 50(4), 640-650.
*Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's Needs and the Five-Factor Model. *Journal of Personality and Social Psychology, 55*(2), 258-265.
*Costa, P. T., Jr. & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual.* Odessa, FL: Psychological Assessment Resources.

*Costa, P. T., Jr. & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*(1), 21-50.

*Costa, P. T., Jr. & McCrae, R. R. (1998). Six approaches to the explication of facet-level traits: examples from conscientiousness. *European Journal of Personality, 12,* 117-134.

*Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*(9), 887-898.

*Craig, R. J., & Bivens, A. (2000). Psychological needs associated with MMPI-2 scales in a nonclinical sample. *Journal of Personality Assessment, 74*(3), 439-446.

*Craig, R. J., Loheidi, R. A., Rudolph, B., Leifer, M., & Rubin, N. (1998). Relationship between psychological needs and the Five-Factor Model of personality classification. *Journal of Research in Personality, 32,* 519-527.

*Crant, J. M. (1995). The Proactive Personality Scale and objective job performance among real estate agents. *Journal of Applied Psychology, 80*(4), 532-537.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Winston.

*Cucina, J. M, Vasilopoulos, N. L., & Leaman, J. A. (2003, April). *The bandwidth-fidelity dilemma and situational judgment test validity.* Poster presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

*Cutchin, G. C. (1998). Relationships between the Big Five personality factors and performance criteria for in-service high-school teachers. *Dissertation Abstracts International, 59* (7-A), 1999, pp. 2263. (UMI No. 9900173)

*Day, D. V., Bedeian, A. G., & Conte, J. M. (1998). Personality as predictor of work-related outcomes: Test of a mediated latent structural model. *Journal of Applied Social Psychology, 28*(22), 2068-2088.

*Dean, M. A., Conte, J. M., & Blankenhorn, T. R. (2006). Examination of the predictive validity of the Big Five personality dimensions across training performance criteria. *Personality and Individual Differences, 42,* 1229-1239.

*Deluga, R. J., & Masson, S. (2000). Relationship of resident assistant conscientiousness, extraversion, and positive affect with rated performance. *Journal of Research in Personality, 34*(2), 225-235.

*Detrick, P., Chibnall, J. T., & Rosso, M. (2001). Minnesota Multiphasic Personality Inventory-2 in police officer selection: Normative data and relation to the Inwald Personality Inventory. *Professional Psychology: Research and Practice, 32*(5), 484-490.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

*Draves, P. R. (2003). An examination of potential moderating effects of personality on the relationship between job attitudes and organizational citizenship behaviors. *Dissertation Abstracts International, 64* (12-B), 2004, pp. 6363. (UMI No. 3116423)

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.

*Duncan, S. (1997). Personality and cognitive characteristics of students in science and engineering. *Dissertation Abstracts International, 58*(9-A), 1998, pp. 3415. (UMI No. 9809035)

*Enright, B. P. (2004). Personality measurement in the prediction of positive and negative police officer performance. *Dissertation Abstracts International, 65* (6-B), 2004, pp. 3154. (UMI No. 3135831)

*Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology, 86*(6), 1270-1279.

*Fannin, N., & Dabbs, J. M. (2003). Testosterone and the work of firefighters: Fighting fires and delivering medical care. *Journal of Research in Personality, 37,* 107-115.

*Ferris, G.R., Witt, L.A., & Hochwarter, W.A. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology, 86,* 1075-1082.

*FormyDuval, D. L., Williams, J. E., Patterson, D. J., & Fogle, E. E. (1995). A "Big Five" scoring system for the item pool of the Adjective Check List. *Journal of Personality Assessment, 65*(1), 59-76.

*Furnham, A. (1996). The Big Five versus the Big Four: The relationship between the Myers-Briggs Type Indicator (MBTI) and the NEO-PI Five Factor Model of personality. *Personality and Individual Differences, 21*(2), 303-307.

*Furnham, A., & Bramwell, M. (2006). Personality factors predict absenteeism in the workplace. *Individual Differences Research, 4*(2), 68-77.

*Furnham, A., Jackson, C. J., & Miller, T. (1999). Personality, learning style and work performance. *Personality and Individual Differences, 27*(6), 1113-1122.

*Furnham, A., Moutafi, J., & Crump, J. (2003). The relationship between the Revised NEO-Personality Inventory and the Myers-Briggs Type Indicator. *Social Behavior and Personality, 31*(6), 577-584.

*Furnham, A., & Stringfield, P. (1993). Personality and work performance: Myers-Briggs Type Indicator correlates of managerial performance in two cultures. *Personality and Individual Differences, 14*(1), 145-153.

*Gaynor, V. M. (1981). Astrology and psychology: Keys to understanding human personality. *Dissertation Abstracts International, 43* (3-B), 1982, pp. 869-870. (UMI No. 8215835)

*Gellatly, I. R., & Irving, P. G. (2001). Personality, autonomy, and contextual performance of managers. *Human Performance, 14*(3), 231-245.

*Gerbing, D. W., & Tuley, M. R. (1991). The 16PF related to the Five-Factor Model of personality: Multiple-indicator measurement versus the a priori scales. *Multivariate Behavioral Research, 26*(2), 271-289.

*Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology, 81*(6), 746-756.

*Gough, H. G. (1996). *California Psychological Inventory manual* (3rd ed.) Palo Alto: Consulting Psychologists Press.
*Gough, H. G., & Heilbrun, A. B., Jr. (1983). *The Adjective Check List manual.* Mountain View, CA: CPP, Inc.
*Griffin, B., & Hesketh, B. (2004). Why openness to experience is not a good predictor of job performance. *International Journal of Selection and Assessment, 12*(3), 243-251.
*Griffith, T. L. (1991). Correlates of police and correctional officer performance. *Dissertation Abstracts International, 52* (5-A), 1991, pp. 1690. (UMI No. 9130949)
*Halfhill, T., Nielsen, T. M., Sundstrom, E., & Weibaecher, A. (2005). Group personality composition and performance in military service teams. *Military Psychology, 17*(1), 41-54.
*Hayes, T. L., Roehm, H. A., & Castellano, J. P. (1994). Personality correlates of success in total quality manufacturing. *Journal of Business and Psychology, 8*(4), 397-411.
*Hinkle, P. N. (1982). Ego-strength, depression, and anxiety: A multitrait-multimethod functional analysis. *Dissertation Abstracts International, 43* (7-B), 1983, pp. 2317. (UMI No. 8228149)
*Hirschfeld, R. R. (1996). Evidence for the discriminant validity of conscientiousness and work orientation as principal components of global trait work motivation. *Dissertation Abstracts International, 57* (10-A), 1997, pp. 4443. (UMI No. 9709133)
*Hochwarter, W. A., Witt, L.A., & Kacmar, K. M. (2000). Perceptions of organizational politics as a moderator of the relationship between conscientiousness and job performance. *Journal of Applied Psychology, 85*(3), 472-478.
*Hogan, R. (1986). *Hogan Personality Inventory manual*. Minneapolis, MN: National Computer Systems.
Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18,* 331-341.
*Hogan, J., & Brinkmeyer, K. (1997). Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology, 50*(3), 587-599.
*Hogan, J., Rybicki, S. L., Motowidlo, S. J., & Borman, W. C. (1998). Relations between contextual performance, personality, and occupational advancement. *Human Performance, 11*(2/3), 189-207.
Hogan, R., & Roberts, B. W. (2001). Introduction: Personality and industrial and organizational psychology. In B. W. Roberts and R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 3-16). Washington, DC: American Psychological Association.
Hough, L. M. (1992). The "Big Five" personality variables—construct confusion: Description versus prediction. *Human Performance, 5,* 139-155.
Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In I. B. Weiner (Ed.) & W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Vol. Eds.), *Comprehensive handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 131-169). New York: Wiley.

*Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

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, pp. 233–277). London: Sage.

Hough, L. M., & Oswald, F. L. (2005). They're right, well… mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance, 18,* 373-387.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

*Hunthausen, J, M., Truxillo, D. M., & Bauer, T. N. (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.

*Inceoglu, I., & Bartram, D. (2006, May). *The validity of personality measures depends on the quality of the criterion measure.* Paper presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

*Jackson, C. J., & Corr, P. J. (1998). Personality-performance correlations at work: Individual and aggregate levels of analyses. *Personality and Individual Differences, 24*(6), 815-820.

*Jacobs, R. L. (1992). Moving up the corporate ladder: A longitudinal study of motivation, personality and managerial success in women and men. *Dissertation Abstracts International, 52* (11-B), 1992, pp. 6124. (UMI No. 9203138)

*Jacobs, R.R., Conte, J.M., Day, D.V., Silva, J.M., & Harris, R. (1996). Selecting bus drivers: Multiple predictors, multiple perspectives on validity, and multiple estimates of utility. *Human Performance, 9,* 199-217.

*Jelley, R. B. (2004). Rater personality and performance-evaluation leniency. *Dissertation Abstracts International, 65* (11-B), 2005, pp. 6080. (UMI No. NQ96832)

*Johnson, J. A. (1994). Clarification of factor five with the help of the AB5C model. *European Journal of Personality, 8,* 311-334.

*Judge, T. A., Martocchio, J. J., & Thoresen, C. J. (1997). Five-Factor Model of personality and employee absence. *Journal of Applied Psychology, 82*(5), 745-755.

*Kamdar, D., & Van Dyne, L. (2007). The joint effects of personality and workplace social exchange relationships in predicting task performance and citizenship performance. *Journal of Applied Psychology, 92*(5), 1286-1298.

*King, E. B., George, J. M., & Hebl, M. R. (2005). Linking personality to helping behaviors at work: An interactional perspective. *Journal of Personality, 73*(3), 585-607.
*Kopischke, T. L. (2001). Content analysis, graphology, and personality assessment: A cross-validation. *Dissertation Abstracts International, 62* (2-B), 2001, pp. 1131. (UMI No. 3003873)
*Kowert, P. A., & Hermann, M. G. (1997). Who takes risks? Daring and caution in foreign policy making. *Journal of Conflict Resolution, 41*(5), 611-637.
*Kraus, E. (2002). Personality and job performance: The mediating roles of leader-member exchange quality and action control. *Dissertation Abstracts International, 64* (1-B), 2003, pp. 452. (UMI No. 3076648)
*Krautheim, M. D. (1997). The development and validation of a customer service orientation scale for university resident assistants. *Dissertation Abstracts International, 58*(9-B), 1998, pp. 5176. (UMI No. 9809963)
*Kudrick, T. R. (1999). Personality and the subjective experience of mental effort. *Dissertation Abstracts International, 60* (5-B), 1999, pp. 2401. (UMI No. 9928737)
*Ladd, D., & Henry, R. A. (2000). Helping coworkers and helping the organization: The role of support perceptions, exchange ideology, and conscientiousness. *Journal of Applied Social Psychology, 30*(10), 2028-2049.
*LaHuis, D. M., Martin, N. R., & Avis, J. M. (2005). Investigating nonlinear conscientiousness-job performance relations for clerical employees. *Human Performance, 18*(3), 199-212.
*Lee, K., Ashton, M. C., & de Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and Five-Factor models of personality structure. *Human Performance, 18*(2), 179-197.
*Liao, H., Joshi, A., & Chuang, A. (2004). Sticking out like a sore thumb: Employee dissimilarity and deviance at work. *Personnel Psychology, 57*(4), 969-1000.
Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.
*Love, K. G. & DeArmond, S. (2007). The validity of assessment center ratings and the 16PF personality trait scores in police sergeant promotions: A case of incremental validity. *Public Personnel Management, 36*(1), 21-32.
*MacDonald, D. A., Anderson, P. E., Tsagarakis, C. I., & Holland, C. J. (1994). Examination of the relationship between the Myers-Briggs Type Indicator and the NEO Personality Inventory. *Psychological Reports, 74*(1), 339-344.
*Martinez, M. T. (2005). A correlational study between the MMPI-2 PSY-5 and the 16PF global factors. *Dissertation Abstracts International, 66* (9-B), 2006, pp. 5137. (UMI No. 3189177)
*Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology, 82*(5), 764-773.
*McCrae, R. R., & Costa, P. T., Jr. (1985). Comparison of EPI and psychoticism scales with measures of the Five-factor Model of personality. *Personality and Individual Differences, 6*(5), 587-597.

*McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five-Factor Model of personality. *Journal of Personality, 57*(1), 17-40.
McCrae, R. R., Costa, P. T., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the Five-Factor Model. *Journal of Personality, 61,* 1-26.
McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
*Melia-Gordon, M. L. (1994). The measurement and meaning of personal projects creativity. Master's thesis accessed through ProQuest Dissertations and Theses database. (UMI No. MM98512)
*Milner, L. M. (1992). A comparative validation study of three personality inventories designed to assess the Five-Factor Model of personality. *Dissertation Abstracts International, 53*(8-B), 1993, pp. 4410. (UMI No. 9236371)
*Mitchell, T., & Serra, S. (2005, April). *Efficacy of a multi-method test battery for selecting correctional officers.* Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.
*Monnot, M. J., Griffith, R. L., English, A., Gujar, A., & Steelman, L. (2004, July). *Predicting typical and maximum job performance.* Poster presented at the 112th Annual Conference of the American Psychological Association. Honolulu, HI.
*Mooradian, T. A., & Nezlek, J. B. (1996). Comparing the NEO-FFI and Saucier's Mini-Markers as measures of the Big Five. *Personality and Individual Differences, 21*(2), 213-215.
*Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology, 58*(3), 583-611.
*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.
*Mount, M. K., & Barrick, M. R. (1995). Manual for the Personal Characteristics Inventory. Libertyville, IL: Wonderlic Personnel Test, Inc.
*Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*(2/3), 145-165.
*Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79*(2), 272-280.
*Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management, 25*(5), 707-721.
*Mount, M., Ilies, R., & Johnson, E. (2006). Relationship of personality traits and counterproductive work behaviors: The mediating effects of job satisfaction. *Personnel Psychology, 59*(3), 591-622.
*Mount, M. K., Witt, L. A, & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the Five Factor personality constructs. *Personnel Psychology, 53*(2), 299-323.

*Myers, I. B., & McCauley, M. H. (1985). *MBTI Manual*. Palo Alto, CA: Consulting Psychologists Press.
*Myers, I. B., McCauley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI Manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
*Neuman, G. A., & Kickul, J. R. (1998). Organizational citizenship behaviors: Achievement orientation and personality. *Journal of Business and Psychology, 13*(2), 263-279.
*Neuman, G. A., & Wright, J. (1999). Team Effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*(3), 376-389.
*Nguyen, N. T. (2004, February). *Response instructions and construct validity of a situational judgment test.* Paper presented at the 11th Annual Conference of the American Association of Business and Behavioral Sciences. Las Vegas, NV.
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
*Oakes, D. W., Ferris, G. R., Martocchio, J. J., Buckley, M. R., & Broach, D. (2001). Cognitive ability and personality predictors of training program skill acquisition and job performance. *Journal of Business and Psychology, 15*(4), 523-548.
Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609-626.
Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660–679
Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354-379.
Pace, V. L. (2005). *Creative performance on the job: Does Openness to Experience matter?* Unpublished master's thesis, University of South Florida, Tampa.
*Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74*(2), 538-556.
Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524-539.
*Paunonen, S. V. & Jackson, D. N. (1996). The Jackson Personality Inventory and the Five-Factor Model of personality. *Journal of Research in Personality, 30,* 42-59.
*Pelo. D. K. (2005). Predicting perseverance of missionary expatriates on overseas assignments: Personality revisited. *Dissertation Abstracts International, 67*(7-B), 2007, pp. 4145. (UMI No. 3223352)
*Piedmont, R. L., McCrae, R. R., & Costa, P. T. (1991). Adjective Check List scales and the Five-Factor Model. *Journal of Personality and Social Psychology, 60*(4), 630-637.
*Piedmont, R. L., McCrae, R. R., & Costa, P. T. (1992). An assessment of the Edwards Personal Preference Schedule from the perspective of the Five-Factor Model. *Journal of Personality Assessment, 58*(1), 67-75.

*Piedmont, R. L., & Weinstein, H. P. (1993). A psychometric evaluation of the new NEO-PIR facet scales for agreeableness and conscientiousness. *Journal of Personality Assessment, 60*(2), 302-318.
*Piedmont, R. L., & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO Personality Inventory. *The Journal of Psychology, 128*(3), 255-265.
*Pollard, C. K. (1988). The relationship of GPA to personality factors of students as measured by the MBTI and the 16PF. *Dissertation Abstracts International, 49* (10-A), 1989, pp. 3006. (UMI No. 8900602)
*Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. L. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the revised NEO Personality Inventory. *Psychological Assessment, 15*(3), 311-325.
*Raja, U., Johns, G., & Ntalianis, F. (2004). The impact of personality on psychological contracts. *Academy of Management Journal, 47*(3), 350-367.
*Reid-Seiser, H. L., & Fritzsche, B. A. (2001). The usefulness of the NEO PI – R Positive Presentation Management Scale for detecting response distortion in employment contexts. *Personality and Individual Differences, 31*(4), 639-650.
Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58,* 103-139.
*Robertson, I. T., Baron, H., Gibbons, P., Maclver, R. & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology, 73,* 171-180.
*Roman, M. A. (1997). Exploring the upper bound of predictive validity of personality testing in selection decisions. *Dissertation Abstracts International, 58* (5-B), 1997, pp. 2733. (UMI No. 9733542)
*Ryan, A. M., Ployhart, R. E., & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83*(2), 298-307.
Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82,* 30-43.
Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76,* 323-346.
*Sarris, A. (2006). Personality, culture fit, and job outcomes on Australian Antarctic stations. *Environment and Behavior, 38*(3), 356-372.
*Saville, P., Sik, G., Nyfield, G., Hackston, J, & MacIver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in the measurement of job competencies across time and in separate organizations. *Applied Psychology: An International Review, 45*(3), 243-262.
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.
Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17,* 881-901.

*Seigler, I. C., Zonderman, A. B., Barefoot, J. C., Williams, R. B., Costa, P. T., Jr., & McCrae, R. R. (1990). Predicting personality in adulthood from college MMPI scores: Implications for follow-up studies in psychosomatic medicine. *Psychosomatic Medicine, 52*, 644-652.

*Skarlicki, D. P., Folger, R., & Tesluk, P. (1999). Personality as a moderator in the relationship between fairness and retaliation. *Academy of Management Journal, 42*(1), 100-108.

*Small, E. E., & Diefendorff, J. M. (2006). The impact of contextual self-ratings and observer ratings of personality on the personality-performance relationship. *Journal of Applied Social Psychology, 36*(2), 297-320.

*Smetana, D. K. (2001). The Five Factor Model: Dimensions or categories? *Dissertation Abstracts International, 62* (11-B), 2002, pp. 5420. (UMI No. 3033877)

Spector, P. E., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. *Journal of Applied Social Psychology, 30,* 1474-1491.

*Stewart, G. L. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology, 81*(6), 619-627.

*Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology, 84*(6), 959-968.

*Stewart, G. L., Carson, K. P., & Cardy, R. L. (1996). The joint effects of conscientiousness and self-leadership training on employee self-directed behavior in a service setting. *Personnel Psychology, 49*(1), 143 – 164.

*Stewart, G. L., & Nandkeolyar, A. K. (2006). Adaptation and intraindividual variation in sales outcomes: Exploring the interactive effects of personality and environmental opportunity. *Personnel Psychology, 59*(2), 307-332.

*Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and linking. *Journal of Occupational and Organizational Psychology, 74*(5), 637-657.

*Strickland, S., & Towler, A. (2005, April). *How openness to experience and charismatic leadership predict creative behavior: The mediating effect of creative self-efficacy.* Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

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

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

*Thoresen, C. J., Bradley, J. C., Bliese, P. D., & Thoresen, J. D. (2004). The Big Five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology, 89*(5), 835-853.

*Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2006). A field study of the role of Big Five personality in applicant perceptions of selection fairness, self, and the hiring organization. *International Journal of Selection and Assessment, 14*(3), 269-277.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58,* 6-20.

Vinchur, A. J., Schippmann, J. S., Switzer, III, F. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83,* 586-597.

Viswesvaran, C., & Ones, D. S. (2000). Measurement error in the "Big Five Factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement, 60,* 224-235.

*Wallace, C., & Chen, G. (2006). A multilevel integration of personality, climate, self-regulation, and performance. *Personnel Psychology, 59*(3), 529-557.

*Wallace, J. C., & Vodanovich, S. J. (2003). Workplace safety performance: Conscientiousness, cognitive failure, and their interaction. *Journal of Occupational Health Psychology, 8*(4), 316-327.

*Wanberg, C. R., & Kammeyer-Mueller, J. D. (2000). Predictors and outcomes of proactivity in the socialization process. *Journal of Applied Psychology, 85*(3), 373-385.

*Weaver, K. M. (1999). The use of the California Psychological Inventory in identifying personal characteristics of effective beginning counselors. *Dissertation Abstracts International, 60* (12-A), 2000, pp. 4334. (UMI No. 9956780)

*White, L. A., Hunter, A. E., & Young, M. C. (2006, May). *Social desirability effects on the predictive validity of personality constructs.* Paper presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

*Wilkerson, A. G. (1990). Jung and Rorschach: A comparative study of introversion/extraversion and introversive/extratensive type. *Dissertation Abstracts International, 51* (4-B), 1990, pp. 2079. (UMI No. 9023689)

*Williams, M. (1999). When is personality a predictor of performance? The moderating role of autonomy. *Dissertation Abstracts International, 60* (7-B), 2000, pp. 3607. (UMI No. 9023689)

*Witt, L.A., Andrews, M. C., & Carlson, D. S. (2004). When conscientiousness isn't enough: Emotional exhaustion and performance among call center customer service representatives. *Journal of Management, 30*(1), 149-160.

*Witt, L. A., & Carlson, D. S. (2006). The work-family interface and job performance: Moderating effects of conscientiousness and perceived organizational support. *Journal of Occupational Health Psychology, 11*(4), 343-357.

*Witt, L.A., & Ferris, G. R. (2003). Social skill as moderator of the conscientiousness-performance relationship: Convergent results across four studies. *Journal of Applied Psychology, 88*(5), 809-820.

*Witt, L. A., Kacmar, K. M., Carlson, D. S., & Zivnuska, S. (2002). Interactive effects of personality and organizational politics on contextual performance. *Journal of Organizational Behavior, 23,* 911-926.
*Wohl, J., & Palmer, A. B. (1970). Correlations between Adjective Check List and Edwards Personal Preference Schedule measures of Murray's needs. *Psychological Reports, 27*(2), 525-526.
*Zeiger, E. A. (1996). The NEO Five-Factor Inventory: Validity considerations. *Dissertation Abstracts International, 57*(5-B), 1996, pp. 3450. (UMI No. 9630645)

Appendices

Appendix A: Studies Included in Meta-Analyses

Table A1
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Allworth & Hesketh, 2000	Task/Technical/Overall Performance	C, E, A
Bacha, 2003	Task/Technical/Overall Performance, OCB/Contextual Performance	C, E, A, N
Baer & Oldham, 2006	Task/Technical/Overall Performance	O
Bajor & Baltes, 2003	Task/Technical/Overall Performance	C
Barrick & Mount, 1993	Task/Technical/Overall Performance	O, C, E, N
Barrick & Mount, 1996	Task/Technical/Overall Performance, Withdrawal	O, C, E, A, N
Barrick et al, 1993	Task/Technical/Overall Performance	O, C, E, A, N
Barrick et al, 2004	OCB/Contextual Performance	O, E, A, N
Bauer et al, 2006	Task/Technical/Overall Performance, Withdrawal	E
Beaty et al, 2001	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Bing & Lounsbury, 2000	Task/Technical/Overall Performance	C, E, N
Bishop, 1996	Task/Technical/Overall Performance	O, C, E, A, N
Black, 2000	Task/Technical/Overall Performance, Training Performance	O, C, E, A, N
Bozionelos, 2004	Task/Technical/Overall Performance, Training Performance	C, E, N
Burke & Witt, 2002	Task/Technical/Overall Performance	O, C, E, A, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Burke & Witt, 2004	CWB/Deviance	O, C, E, A, N
Bushe & Gibbs, 1990	Training Performance	E
Byrne et al, 2005	Task/Technical/Overall Performance	C
Caligiuri, 2000	Task/Technical/Overall Performance	C, A, N
Cellar et al, 1996	Training Performance	O, C, E, A, N
Chan & Schmitt, 2002	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Christiansen et al, 1994	Task/Technical/Overall Performance	C, E, N
Clevenger et al, 2001	Task/Technical/Overall Performance	C
Colbert et al, 2004	CWB/Deviance	O, C, E, A, N
Colbert et al, 2004-unpublished results	CWB/Deviance	O, C, E, A, N
Collins & Schmidt, 1993	CWB/Deviance	E
Conte & Gintoft, 2005	Task/Technical/Overall Performance	O, C, E, A, N
Conte & Jacobs, 2003	Task/Technical/Overall Performance, Withdrawal	O, C, E, A, N
Crant, 1995	Task/Technical/Overall Performance	O, C, E, A, N
Cucina et al, 2003	Training Performance	O, C, E, A, N
Cutchin, 1998	Task/Technical/Overall Performance	O, C, E, A, N
Day et al, 1998	Withdrawal	C, E
Dean et al, 2006	Training Performance	O, C, E, A, N
Deluga & Masson, 2000	Task/Technical/Overall Performance	C, E
Draves, 2003	OCB	C
Enright, 2004	Task/Technical/Overall Performance, CWB/Deviance	C, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Erez & Judge, 2001	Task/Technical/Overall Performance	C, N
Fannin & Dabbs, 2003	Task/Technical/Overall Performance	O, C, E, A, N
Ferris et al, 2001	Task/Technical/Overall Performance	O, C, E, A, N
Furnham & Bramwell, 2006	Withdrawal	O, C, E, A, N
Furnham et al, 1999	Task/Technical/Overall Performance	E, N
Furnham & Stringfield, 1993	Task/Technical/Overall Performance	E
Gellatly & Irving, 2001	OCB/Contextual Performance	C, E, A
Goffin et al, 1996	Task/Technical/Overall Performance	C, E
Griffin & Hesketh, 2004	Task/Technical/Overall Performance	O
Halfhill et al, 2005	Task/Technical/Overall Performance	C, A
Hayes et al, 1994	Task/Technical/Overall Performance	O, C, E, A, N
Hirschfeld, 1996	Task/Technical/Overall Performance, Withdrawal	C
Hochwarter et al, 2000	Task/Technical/Overall Performance	C
Hogan & Brinkmeyer, 1997	CWB/Deviance	O, C, E, A, N
Hogan et al, 1998	OCB/Contextual Performance	O, C, E, A, N
Hough et al, 1990	Task/Technical/Overall Performance, OCB/Contextual Performance, CWB/Deviance	C, E, A, N
Hunthausen et al, 2003	Task/Technical/Overall Performance	O, C, E, A, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Inceoglu & Bartram, 2006	Task/Technical/Overall Performance	O, C, E, A, N
Jackson & Corr, 1998	Task/Technical/Overall Performance	O, C, E, A, N
Jacobs, 1992	Task/Technical/Overall Performance	C, E
Jacobs et al, 1996	Task/Technical/Overall Performance, Withdrawal, CWB/Deviance	O, C, E, A, N
Judge et al, 1997	Withdrawal	O, C, E, A, N
Kamdar & Van Dyne, 2007	Task/Technical/Overall Performance, OCB/Contextual Performance	C, A
King et al, 2005	OCB/Contextual Performance	E, A, N
Kraus, 2002	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Krautheim, 1997	Task/Technical/Overall Performance, OCB/Contextual Performance	A
Ladd & Henry, 2000	Task/Technical/Overall Performance, OCB/Contextual Performance	C
LaHuis et al, 2005	Task/Technical/Overall Performance	C
Lee et al, 2005	CWB/Deviance	O, C, E, A, N
Liao et al, 2004	CWB/Deviance	O, C, E, A, N
Love & DeArmond, 2007	Task/Technical/Overall Performance	C, E, N
Martocchio & Judge, 1997	Training Performance	C

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Mitchell & Serra, 2005	Task/Technical/Overall Performance, Training Performance	O, C, E, A, N
Monnot et al, 2004	Task/Technical/Overall Performance	O, C, E, A, N
Morgeson et al, 2005	Task/Technical/Overall Performance, OCB/Contextual Performance	C, E, A, N
Motowidlo & Van Scotter, 1994	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Mount et al, 1994	Task/Technical/Overall Performance	O, C, E, A, N
Mount et al, 1998	Task/Technical/Overall Performance	O, C, E, A, N
Mount et al, 1999	Task/Technical/Overall Performance	C
Mount et al, 2000	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Neuman & Kickul, 1998	OCB/Contextual Performance	C, E, A
Neuman & Wright, 1999	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Nguyen, 2004	Task/Technical/Overall Performance	O, C, E, A, N
Oakes et al, 2001	Task/Technical/Overall Performance, Training Performance	C, E, N
Pelo, 2005	Withdrawal	C, E, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Piedmont & Weinstein, 1994	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Raja et al, 2004	Withdrawal	C, E, N
Reid-Seiser & Fritzsche, 2001	Task/Technical/Overall Performance, CWB/Deviance	O, C, E, A, N
Roman, 1997	Task/Technical/Overall Performance, Withdrawal	O, C
Ryan et al, 1998	Task/Technical/Overall Performance	O
Sarris, 2006	Withdrawal	O, C, E, A, N
Saville et al, 1996	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Skarlicki et al, 1999	CWB/Deviance	A
Small & Diefendorff, 2006	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Stewart, 1996	Task/Technical/Overall Performance	C, E
Stewart, 1999	Task/Technical/Overall Performance	C
Stewart et al, 1996	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Stewart & Nandkeolyar, 2006	Task/Technical/Overall Performance	O, C
Strauss et al, 2001	Task/Technical/Overall Performance	C, E, N
Strickland & Towler, 2005	Task/Technical/Overall Performance	O, C, E, A, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A1 (Continued)
Studies Contributing Correlations for Meta-Analysis of Criterion-Related Validity

Study	Criteria	Personality Factors
Tett et al, 2003	Task/Technical/Overall Performance	O, C, E, A, N
Thoresen et al, 2004	Task/Technical/Overall Performance	O, C, E, A, N
Truxillo et al, 2006	Withdrawal	O, C, E, A, N
Wallace & Chen, 2006	Task/Technical/Overall Performance, CWB/Deviance	C
Wallace & Vodanovich, 2003	CWB/Deviance	C
Wanberg & Kammeyer, 2000	Withdrawal	O, C, E, A, N
Weaver, 1999	Task/Technical/Overall Performance	C, E
White et al, 2006	Task/Technical/Overall Performance, Withdrawal, OCB/Contextual Performance, CWB/Deviance	E, A
Williams, 1999	Task/Technical/Overall Performance, OCB/Contextual Performance	O, C, E, A, N
Witt et al, 2002	OCB/Contextual Performance	O, C, E, A, N
Witt & Ferris, 2003	Task/Technical/Overall Performance, OCB/Contextual Performance	C
Witt & Carlson, 2006	Task/Technical/Overall Performance	C, N
Witt et al, 2004	Task/Technical/Overall Performance	C

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism

Table A2
Studies Contributing Correlations for Meta-Analysis of Convergent Validity

Study	Tests	Personality Factors
Anderson & Ones, 2003	HPI, OPQ	O, C, E, N
Ashton & Lee, 2005	Goldberg/Saucier, NEO	A, N
Bessmer & Ramanaiah, 1981	ACL, PRF	C, E, A
Bibeau-Reaves, 2002	MBTI, NEO	E
Briggs, 1992	Goldberg/Saucier, NEO	O, C, E, A, N
Byravan, 1996	MMPI, NEO, 16PF	C, E, N
Canivez & Allen, 2005	NEO, 16PF	O, C, E, N
Cattell, 1996 (as cited in Canivez & Allen, 2005)	NEO, 16PF	C, E, N
Church, 1994	MPQ, NEO	O, C, E, N
Costa et al, 1986	NEO, MMPI	E, N
Costa & McCrae, 1988	NEO, PRF	O, C, E, A
Costa & McCrae, 1992	ACL, CPI, MBTI, NEO, PRF	O, C, E, A, N
Costa & McCrae, 1995	CPI, HPI, NEO	O, C, E, A, N
Costa & McCrae, 1998	CPI, NEO, PRF	C
Costa et al, 1991	NEO, CPI	C, E
Craig & Bivens, 2000	ACL, MMPI	E, N
Craig et al, 1998	ACL, NEO	O, C, E, A, N
Detrick et al, 2001	Inwald, MMPI	E, N

Note. See end of table for abbreviation key.

Table A2 (Continued)
Studies Contributing Correlations for Meta-Analysis of Convergent Validity

Study	Tests	Personality Factors
Duncan, 1997	CPI, MBTI	E
FormyDuval et al, 1995	ACL, NEO	O, C, E, A, N
Furnham, 1996	MBTI, NEO	E
Furnham et al, 2003	MBTI, NEO	E
Gaynor, 1981	EPI, MBTI, MMPI	E
Gerbing & Tuley, 1991	NEO, 16PF	C, N
Gough, 1996	CPI, EPI, Goldberg/Saucier, HPI, MBTI, NEO, PRF, 16PF	C, E, A
Gough & Heilbrun, 1983	ACL, CPI, MMPI	C, E, N
Griffith, 1991	Inwald, MMPI	N
Hinkle, 1982	MMPI, 16PF	N
Hogan, 1986	HPI, MMPI	E, N
Jacobs, 1992	CPI, EPPS	C, E
Jelley, 2004	NEO, PRF	A
Johnson, 1994	HPI, NEO	O
Kopischke, 2001	MBTI, NEO	E
Kowert & Hermann, 1997	MBTI, NEO	E
Kudrick, 1999	EPI, NEO	E, N

Note. See end of table for abbreviation key.

Table A2 (Continued)
Studies Contributing Correlations for Meta-Analysis of Convergent Validity

Study	Tests	Personality Factors
MacDonald et al, 1994	MBTI, NEO	E
Martinez, 2005	MMPI, 16PF	E, N
McCrae & Costa, 1985	ACL, EPI, NEO	E, N
McCrae & Costa, 1989	MBTI, NEO	E
Melia-Gordon, 1994	ACL, NEO	O
Milner, 1992	Goldberg/Saucier, HPI, NEO	O, C, E, A, N
Mooradian & Nezlek, 1996	Goldberg/Saucier, NEO	O, C, E, A, N
Mount & Barrick, 1995	Goldberg/Saucier, HPI, NEO	O, C, E, A, N
Mount et al, 1994	Goldberg/Saucier, HPI, NEO, PCI	O, C, E, A, N
Myers & McCauley, 1985	ACL, CPI, EPI, MBTI, MMPI	E
Meyers et al, 1998	ACL, CPI, MBTI	E
Paunonen, 1998	NEO, PRF	O, C, E, A, N
Paunonen & Jackson, 1996	JPI, NEO	O, C, E, N
Piedmont et al, 1991	ACL, EPPS, NEO	O, C, E, A, N
Piedmont et al, 1992	EPPS, NEO	O, C, E, A
Piedmont & Weinstein, 1993	ACL, NEO	C, A
Pollard, 1988	MBTI, 16PF	E
Quirk et al, 2003	MMPI, NEO, 16PF	E, N

Note. See end of table for abbreviation key.

Table A2 (Continued)
Studies Contributing Correlations for Meta-Analysis of Convergent Validity

Study	Tests	Personality Factors
Robertson et al, 2000	NEO, OPQ	C
Siegler et al, 1990	MMPI, NEO, 16PF	O, E, N
Smetana, 2001	MBTI, NEO	E
Wilkerson, 1990	MBTI, MMPI	E
Wohl & Palmer, 1970	ACL, EPPS	O, C, A
Zeiger, 1996	MMPI, NEO, 16PF	E, N

Note. O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Emotional Stability/Neuroticism
ACL = Adjective Check List; CPI = California Psychological Inventory; EPPS = Edwards Personal Preference Schedule; EPI = Eysenck Personality Inventory; Goldberg/Saucier = Goldberg Big Five Factor Markers, Saucier Mini-Markers, or International Personality Item Pool; HPI = Hogan Personality Inventory; Inwald = Inwald Personality Inventory; JPI = Jackson Personality Inventory; MBTI = Myers-Briggs Type Indicator; MMPI = Minnesota Multiphasic Personality Inventory; MPQ = Multidimensional Personality Questionnaire; NEO = NEO-FFI, NEO-PI, or NEO PI-R; OPQ = Occupational Personality Questionnaire; PRF = Personality Research Form; 16PF = Sixteen Personality Factors

Appendix B: SAS Code for Meta-Analysis
(Bare-Bones, and Corrected for Unreliability in Predictor and Criterion)

Thanks to Dr. Michael T. Brannick for original code that was later customized for this project.

```
data d1;
input rxx ryy r n;
cards;
    .67     .93     .33     68
    .71     .95     .23     114
    .75     .89     .18     105
    .78     .86     .31     136
    .78     .96     .06     99
    .78     .99     -.01    95
    .79     .89     .22     143
    .80     .90     -.05    131
    .81     .82     .06     160
    .83     .93     .05     174
    .83     .93     .12     422
    .84     .95     .15     58
    .86     .91     .04     22
    .86     .91     .13     83
    .87     .86     .18     254
    .89     .50     .32     146
    .89     .88     .25     146
    .91     .90     .50     131
    .91     .94     .29     150
    .92     .87     .23     214
    .92     .90     .02     412
    .92     .91     .34     230
    .93     .91     .17     144
    .94     .90     .27     130
    .98     .98     .23     326
proc iml;
*Schmidt and Hunter rxx and ryy corrections as well as for sampling error;
************************************************;
use d1;
read all into x;
************************************************;
rxx = x[,1];        *Reliability of x;
ryy = x[,2];        *Reliability of y;
obsr = x[,3];       *observed correlations;
n = x[,4];          *sample size N;
```

Appendix B (Continued)

SAS Code for Meta-Analysis (Continued)

```
k = nrow(X);          *Number of studies;
sumn=n[+];            *sum of N;
aven = sumn/k;         *average N;
*************************************************************************;
*Bare-Bones first as reference
*************************************************************************;
nr= obsr`*n;          *sum weighted r;
aver=nr/sumn;          *weighted mean;
varr1= obsr - aver;    *deviation from weighted mean;
varr2=n`* varr1##2;    *sum weighted squared deviations;
varr=varr2/sumn;       *weighted variance of obs r (s-squared sub r);
samperr = (1-aver**2)**2/((sumn/k)-1); *sampling error variance;
resr=varr-samperr;     *residual variance (variance of rho);
if resr < 0 then resr = 0;  *keep boundary on residual variance;
sdrho=resr**.5;        *print sdrho;
CI95L = aver-1.96#sqrt(varr/k);
CI95U = aver+1.96#sqrt(varr/k);
CR95L = aver-1.96#sqrt(resr);
CR95U = aver+1.96#sqrt(resr);
*****************************************;
Print '************Schmidt-Hunter Bare Bones Analysis************';
Print 'Number of studies is' k;
Print 'Average sample size is' aven;
Print 'Total sample size is' sumn;
Print 'Estimated population mean is' aver;
Print 'Observed Variance is' varr;
Print 'Sampling Error Variance is' samperr;
Print 'SDrho is' sdrho;
Print '95 percent confidence interval for mean is' CI95L CI95U;
Print '95 percent credibility interval is' CR95L CR95U;
*************************************************************************;
* S-H corrections for unreliability and sampling error
*************************************************************************;
*Disattenuate r;
RC1= j(k,1,-9);
VEsimple = j(k,1,-9);
ve = vesimple;
```

SAS Code for Meta-Analysis (Continued)

```
RC = RC1;
  do count1= 1 to k;
     RC1[count1,1] = obsr[count1,1]/sqrt(rxx[count1,1]#ryy[count1,1]);
   RC[count1,1]= RC1[count1,1];
  end;
A = obsr#1/RC;         * Find Compound Attenuation factor;
w = n#A#A;             * Find weights;
nr= obsr`*n;           *sum weighted r;
aver=nr/sumn;          *weighted mean for uncorrected r;
*Find simple sampling error for uncorrected correlations;
do c1 = 1 to k;
VEsimple[c1,1] = (1-aver#aver)#(1-aver#aver)/(n[c1,1]-1);
end;
VE1 = VEsimple#1/(A#A);  *first approximation to error variance of corrected
correlation;
*****************************************************************;
*  If you want intermediate results, remove the asterisk on the
*  print statement following this comment.
*****************************************************************;
*print obsr rc a w ve1;
rbarc = w`*rc/w[+,];     *meta-analytic mean of corrected correlations;
means=j(k,1,rbarc);      *column of means;
diffsq = (rc-means)#(rc-means); *deviations from the mean squared;
Var_rc = w`*diffsq/w[+];        *variance of the corrected correlations;
Ave_ve1 = w`*ve1/w[+];           *variance of error;
Var_rho= Var_rc-Ave_ve1;     *residual variance (variance of rho);
if Var_rho < 0 then Var_rho = 0;  *keep boundary on residual variance;
sdrho=Var_rho**.5;         *print sdrho;
*CI95L = aver-1.96#sqrt(varr/k);
*CI95U = aver+1.96#sqrt(varr/k);
CR95L = rbarc-1.96#sdrho;
CR95U = rbarc+1.96#sdrho;
*****************************************;
Print ' ';
Print ' ';
Print '**************Schmidt-Hunter Fully Corrected Estimates*****';
Print 'Estimated population mean is (corrected for unreliability in pred&crit and
samperror)' rbarc;
Print 'Corrected Variance is' Var_rc;
Print 'Corrected (refined) Sampling Error Variance is' Ave_ve1;
```

SAS Code for Meta-Analysis (Continued)

```
Print 'SDrho is' sdrho;
Print '95 percent credibility interval is' CR95L CR95U;
quit;
run;
```

Appendix C: Preliminary Nomological Net Diagrams for Selected Tests, Based on Bare-Bones Meta-Analyses

Figure 1. Nomological Net for NEO Agreeableness

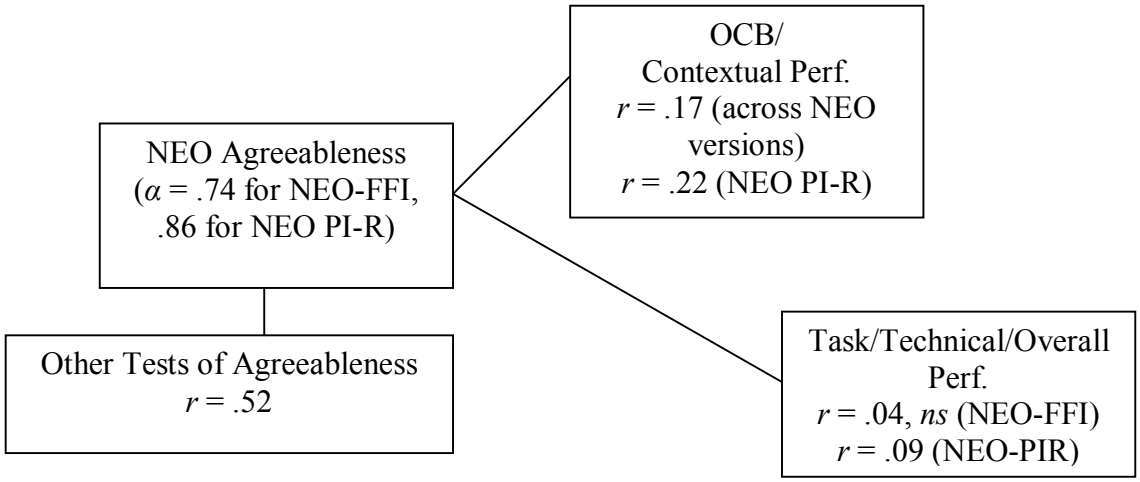


Figure 2. Nomological Net for Goldberg/Saucier/IPIP Agreeableness

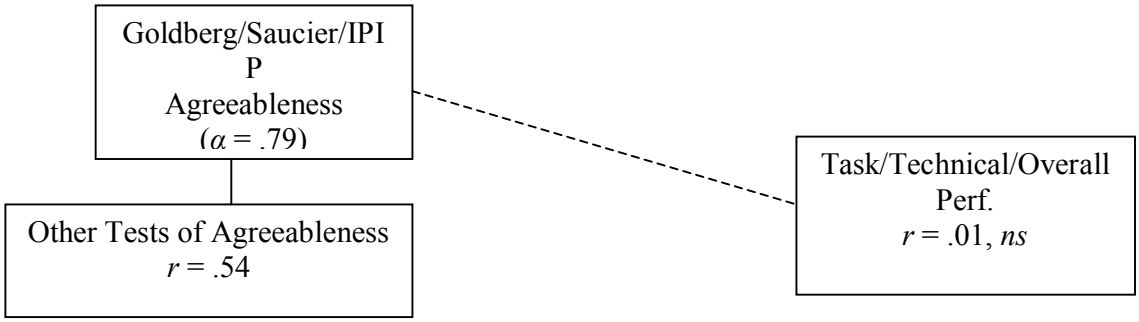


Figure 3. Nomological Net for HPI Likeability

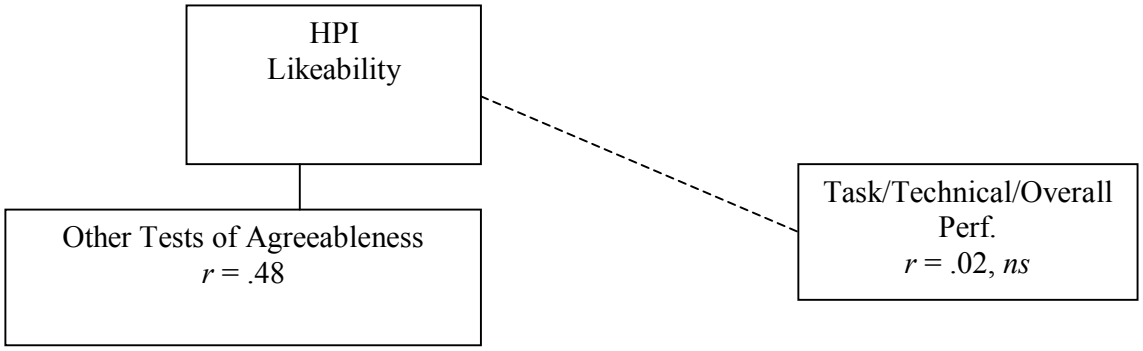


Figure 4. Nomological Net for PCI Agreeableness

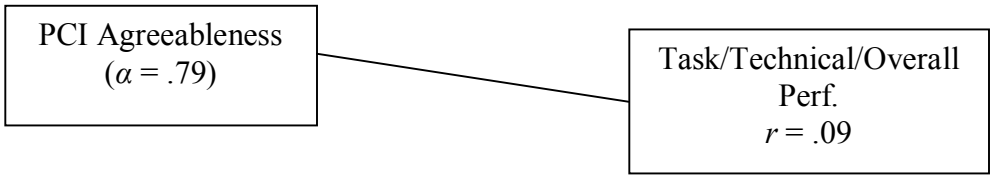


Figure 5. Nomological Net for NEO Conscientiousness

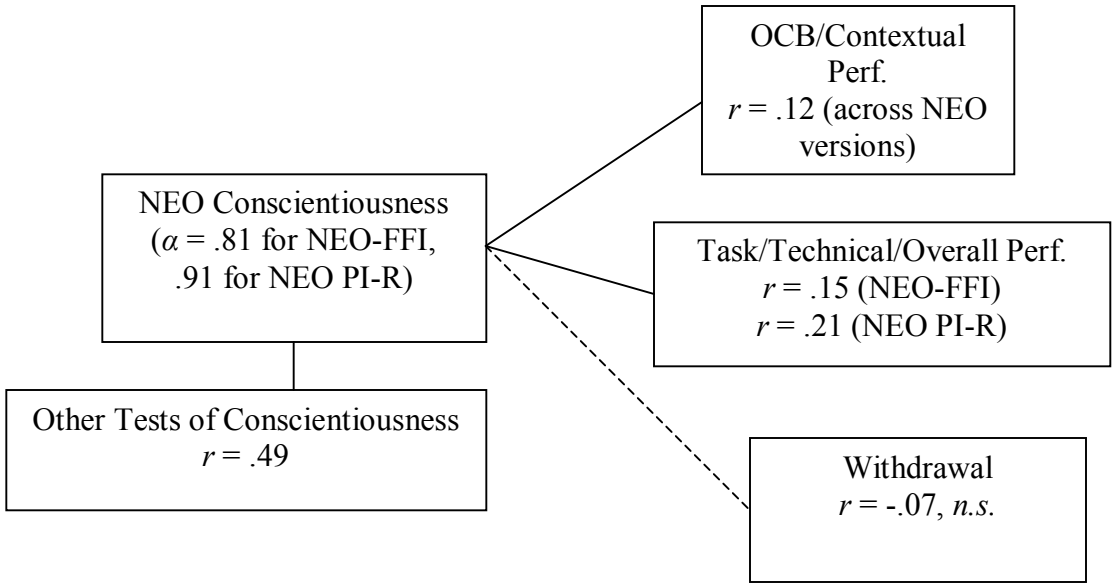


Figure 6. Nomological Net for Goldberg/Saucier/IPIP Conscientiousness

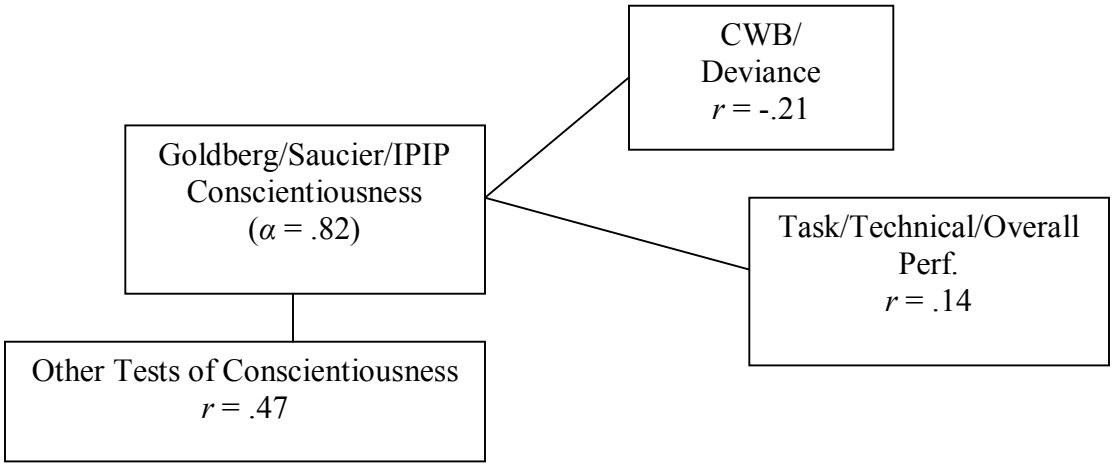


Figure 7. Nomological Net for HPI Prudence

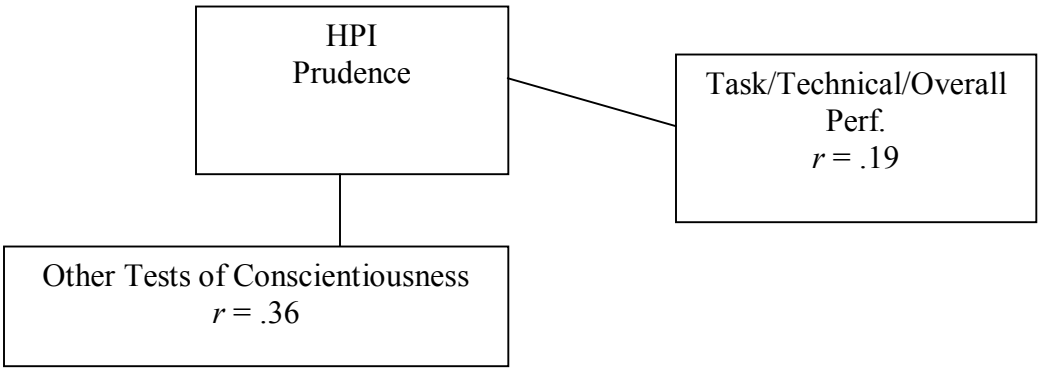


Figure 8. Nomological Net for PCI Conscientiousness

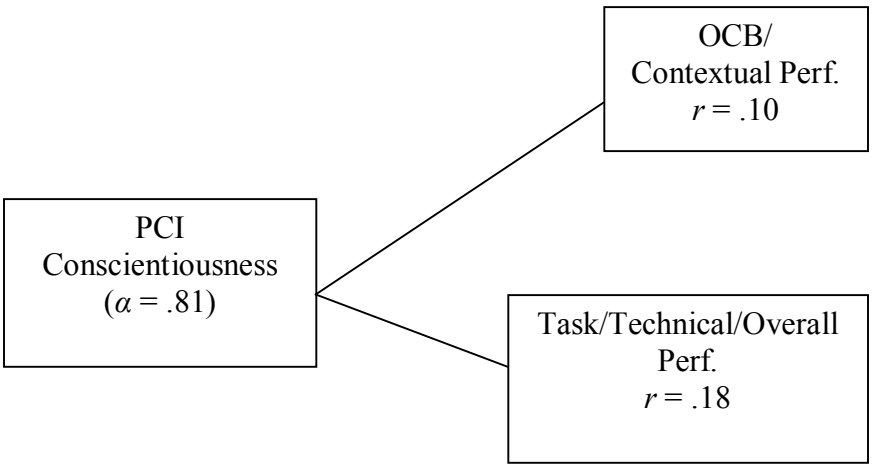


Figure 9. Nomological Net for NEO Neuroticism (*effect sizes recoded to indicate Emotional Stability)

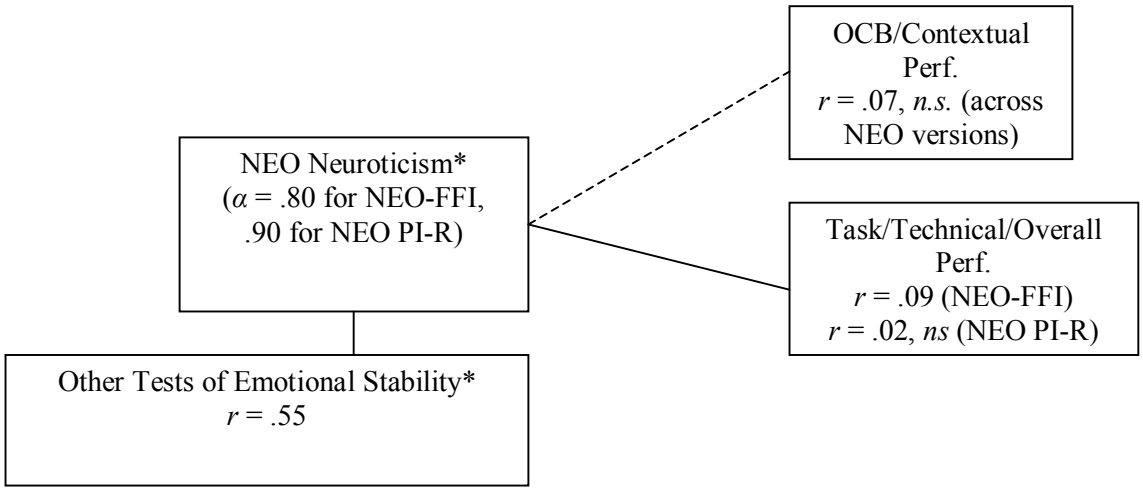


Figure 10. Nomological Net for Goldberg/Saucier/IPIP Emotional Stability

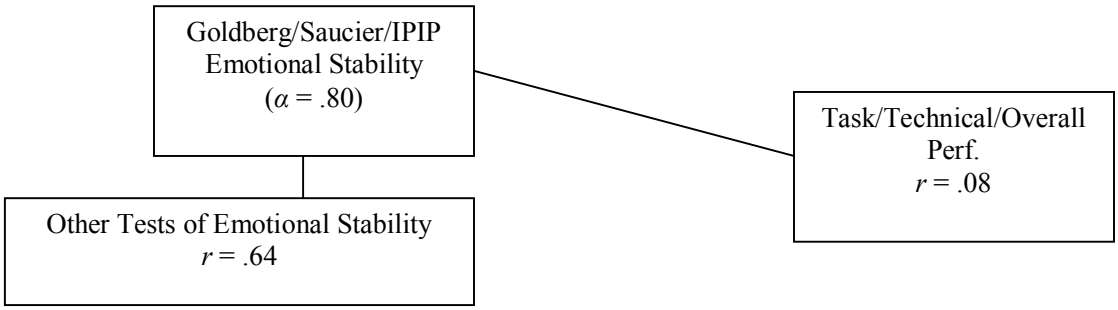


Figure 11. Nomological Net for HPI Adjustment

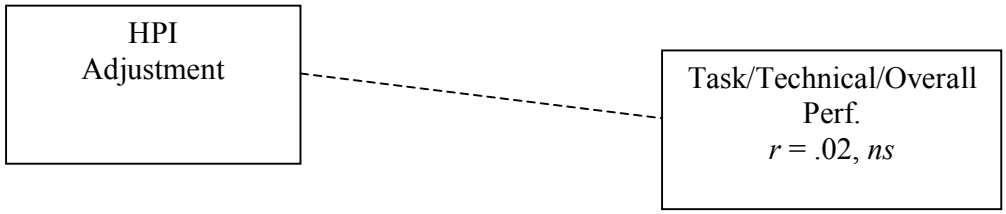


Figure 12. Nomological Net for PCI Emotional Stability

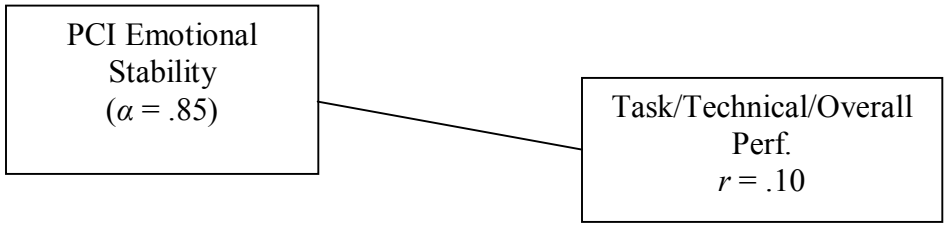


Figure 13. Nomological Net for NEO Extraversion

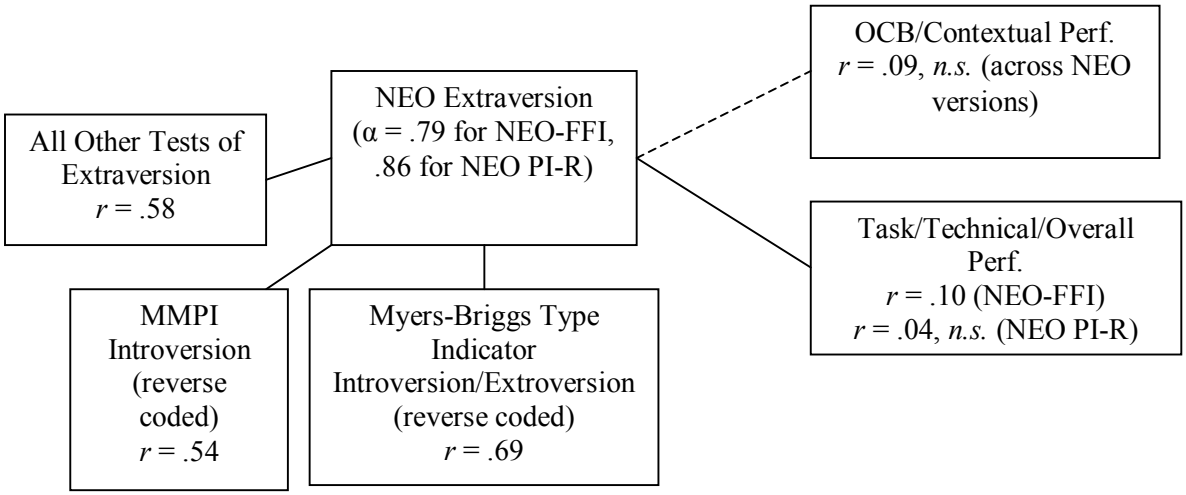


Figure 14. Nomological Net for Goldberg/Saucier/IPIP Extraversion

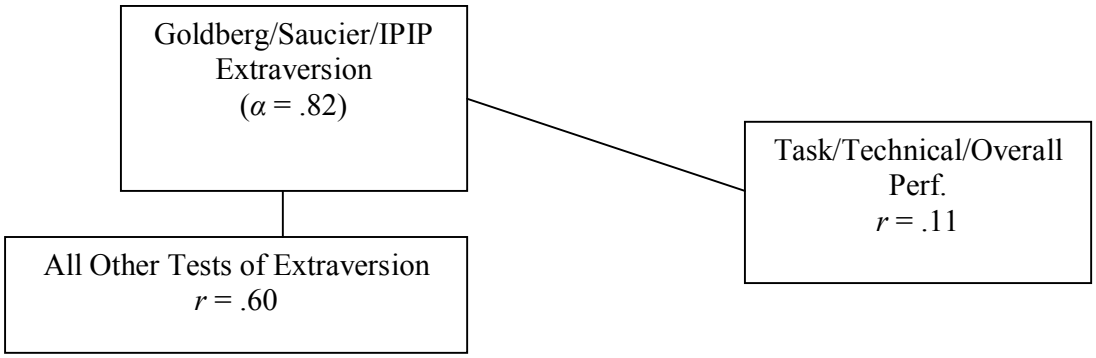


Figure 15. Nomological Net for HPI Sociability

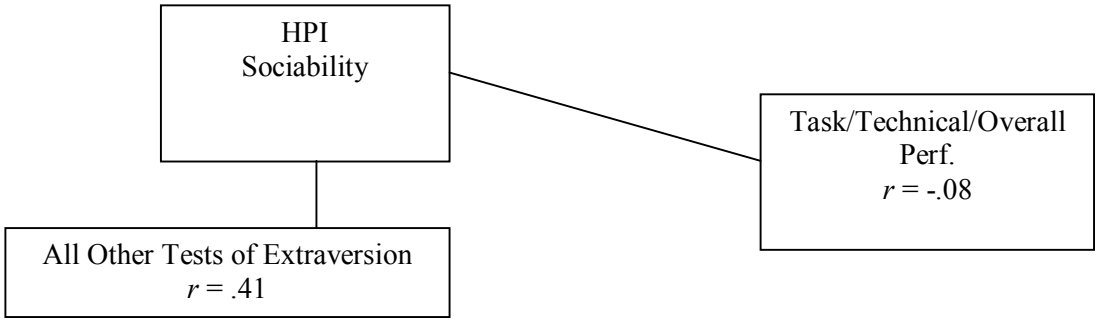


Figure 16. Nomological Net for PCI Extraversion

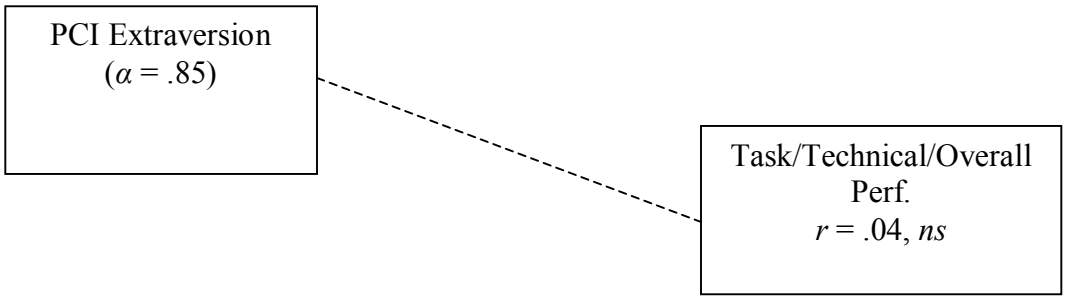


Figure 17. Nomological Net for NEO Openness to Experience

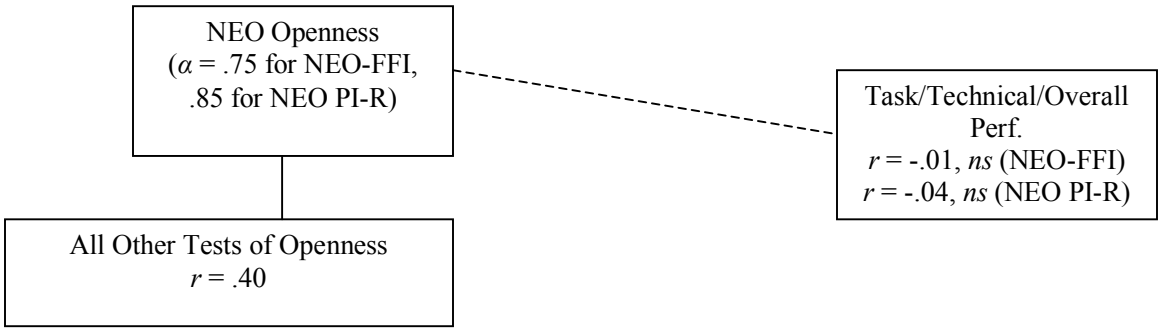


Figure 18. Nomological Net for Goldberg/Saucier/IPIP Intellect

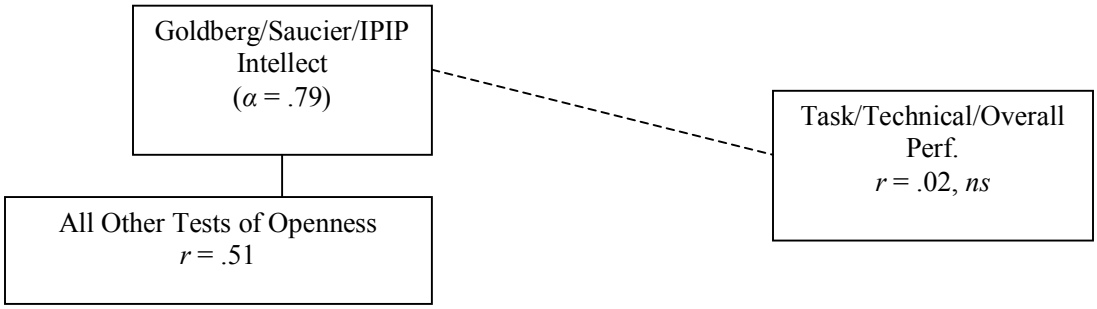


Figure 19. Nomological Net for HPI Intellectance

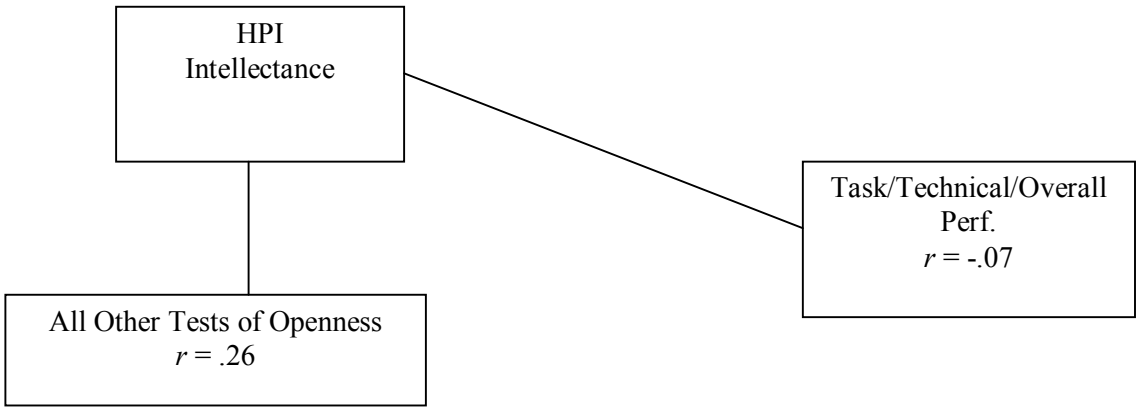
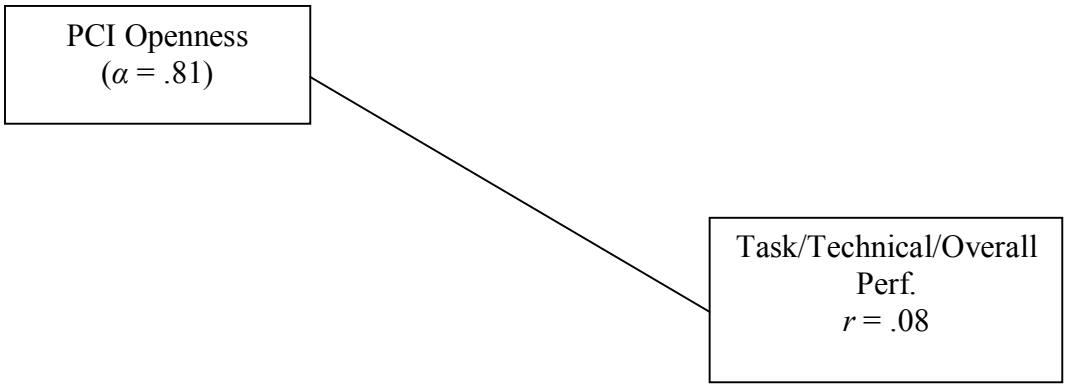


Figure 20. Nomological Net for PCI Openness



About the Author

Victoria Pace earned an A.A. from Florida School of the Arts, B.A. degrees in Art and Mathematics from Florida State University and University of South Florida, respectively, and M.A. and Ph.D. in Industrial/Organizational Psychology at the University of South Florida. While an undergraduate at USF, the first I/O psychologist she ever met was Paul Spector and the first I/O course she took was taught by Wally Borman. From that prestigious introduction to the field, it was inevitable that she would pursue a career in I/O. She has presented papers at Society for Industrial and Organizational Psychology and Southern Management Association conferences, and co-authored a chapter in *A Closer Examination of Applicant Faking Behavior* (R. L. Griffith & M. H. Peterson, Eds.) and an article in *Applied Psychology: An International Review.* She worked at Personnel Decisions Research Institutes and has taught numerous undergraduate and graduate courses and labs.