Personality Test Validation Research: Present-employee and job applicant samples

Kevin Michael Bradley

Dissertation submitted to the Faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Psychology

Neil M. A. Hauenstein, Chair
Roseanne J. Foti
Jack W. Finney
John J. Donovan
Kevin D. Carlson

August 28, 2003
Blacksburg, Virginia

Keywords: Employee-selection; Testing and Assessment; Personality; Validation Research.

Personality Test Validation Research: Present-employee and job applicant samples

Kevin M. Bradley

(ABSTRACT)

In an effort to demonstrate the usefulness of personality tests as predictors of job performance, it is common practice to draw a validation sample consisting of individuals who are currently employed on the job in question. It has long been assumed that the results of such a study are appropriately generalized to the setting wherein job candidates respond to personality inventories as an application requirement. The purpose of this manuscript was to critically evaluate the evidence supporting the presumed interchangeability of present-employees and job applicants. Existing research on the use of personality tests in occupational settings is reviewed. Theoretical reasons to anticipate differential response processes and self-report personality profiles according to test-taking status (present employees versus job applicants) are reviewed, as is empirical research examining relevant issues. The question of sample type substitutability is further probed via a quantitative review (meta-analysis) of the criterion-related validity of seven personality constructs (Neuroticism, Extraversion, Openness to Experience, Agreeableness, Conscientiousness, Optimism, and Ambition). Further, the meta-analytic correlations among these personality constructs are estimated. Test-taking status is examined as a moderator of the criterion-related validities as well as the personality construct inter-correlations. Meta-analytic correlation matrices are then constructed on the basis of the job incumbent and the job applicant subgroup results. These correlation matrices are utilized in a simulation study designed to estimate the potential degree of error when job incumbents are used in place of job applicants in a validation study for personality tests.

The results of the meta-analyses and the subsequent simulation study suggest that the moderating effect of sample type on criterion-related validity estimates is generally small. Sample type does appear to moderate the criterion-related validity of some personality constructs, but the direction of the effect is inconsistent: in some cases, incumbent validities are larger than applicant validities. Alternatively, incumbent validities sometimes are smaller than applicant validities. Personality construct inter-correlations yield almost no evidence of moderation by sample type. Further, where there are between group differences in the

personality construct inter-correlations, these differences have little bearing on the regression equation relating personality to job performance. Despite a few caveats that are discussed, the results are supportive of the use of incumbent samples in personality-test validation research.

Acknowledgements

This research project and the attendant graduate education could not have been completed were it not for the support and assistance of numerous individuals. I would like to thank Neil Hauenstein for his guidance and wisdom as I progressed through my graduate training, as well as for his camaraderie and fellowship over the years. If Neil had not reached out to me during my second year at Virginia Tech and taken me under his wing, it is unlikely that I would have ever completed graduate school. There were times during the completion of this dissertation that I still did not know if it would ever be completed; thanks, Neil, for knowing when to be passively supportive and when to stir me into action.

Thanks also to my dissertation advisory committee: Kevin Carlson, John Donovan, Jack Finney, and Roseanne Foti. Your challenging questions and insightful comments during the prospectus meeting and the final defense helped ensure that the full potential of this line of inquiry would be realized.

This project also would not have been possible were it not for the numerous researchers who responded to my requests for data. I greatly appreciate the conscientious efforts of all those who took time to search their files, re-analyze their existing data, and forward results to me. They are too numerous to name individually, but please be assured that your assistance will not soon be forgotten.

My development as a researcher is also due in large measure to the efforts of my professors in the Virginia Tech Department of Psychology, and I thank them for all they have shown me. More specifically, I would like to thank the past and present faculty of the Industrial and Organizational area: John Donovan, Roseanne Foti, R. J. Harvey, Neil Hauenstein, Jeff Facteau, Sigrid Gustafson, Joe Sgro, and Morrie Mullins.

I would also like to thank the Graduate Research Development Project and the Graduate Student Assembly at Virginia Tech for grant funds supporting this research.

Thanks to my many colleagues, classmates and friends in the Virginia Tech Psychology Department in general, as well as the I/O Psychology area more specifically. While I have benefited tremendously from my interactions with them all, I especially would like to thank my cohort and others with whom I shared advanced seminars: Steve Burnkrant, Shanan Gibson, Dan LeBreton, Kevin Keller, Jean-Anne Hughes Schmidt, Gavan O'Shea, Amy Gershenoff, Andrea

Sinclair, Greg Lemmond, Ty Breland, and Carl Swander.

I also want to single out Gavan O'Shea and thank him for the many good laughs and great times we shared both inside and outside of educational settings. With the exception of Neil, you have been the single most important influence on my development as a researcher, and it has been a privilege going through graduate school with you. You have been a tremendous influence on my personal development, and most importantly, you are a true friend.

To my siblings, Joe, Colleen, and Tom, thanks for many great diversions away from my life as a graduate student. Some of my favorite experiences over the past umpteen years have been our ski trips, Jimmy Buffett concerts, and weekends at the beach. Getting away from graduate school, reconnecting with family, and having a heck of a time helped keep me sane and able to go on.

To my wife, Kristi, words are not able to express my gratitude for your support and understanding over the years, but especially during the last two years. You listened patiently during our walks while I would go on about mind-numbing details. You sacrificed many comforts so that I could devote myself wholly to my dissertation, and you never complained when I spent evenings and weekends coding studies instead of spending time with you. Thank you for your encouragement and optimism during times of uncertainty and doubt.

Finally, to my Mother and Father, the best teachers I have ever had. Your tremendous sacrifices have enabled me to take advantage of opportunities that others can only dream of. Thank you for keeping after me and not accepting mediocrity in my schoolwork. Thank you also for doing whatever it took so that I could pursue this dream. I'm finished my homework – can I go out and play?

Table of Contents

List of Tables

List of Figures

Chapter One: Overview

When faced with the task of providing evidence of the usefulness of psychological tests for choosing from among a number of job applicants, personnel psychologists frequently rely on criterion-related validation studies to demonstrate that the test scores in question are correlated with success on the job. In most instances, the goal of criterion-related validation studies is to estimate how well test scores forecast the future job performance of current job applicants. A desirable scenario for obtaining such an estimate is to (a) administer the test battery to a group of current job applicants; (b) randomly select employees from among those applicants; (c) wait an appropriate time interval to afford an accurate determination of each individual's level of success on the job; and (d) determine the relationship between test scores and success on the job. Such a design is called a predictive validation study with random selection of job applicants (Sussman & Robertson, 1986).

It is often more practical to (a) administer the assessment device in question to a sample of present-employees; (b) obtain measures of success on the job for each of those employees; and (c) determine the relationship between scores on the assessment device and success on the job. This type of study design is known as a concurrent validation strategy. It is assumed that results from a concurrent validation study of present-employees provide an accurate estimate of the validity of test scores that will in practice, be used in a predictive manner – that is, to forecast the future job performance of current job applicants. The purpose of the current study is to test the assumption that results from concurrent studies of present-employees generalize to job applicants. In addition to the examination of criterion-related validity estimates across present-employee and job applicant samples, a more expansive view comparing inter-correlations of personality predictor measures, regression coefficients, prediction equations, and utility estimates across sample types is taken.

In actuality, the true distinction between a predictive validation study and a concurrent validation study concerns the time interval between collection of the predictor scores and the measure of job success, or, criterion data. In a predictive study, there is a discernible time interval between collection of the predictor and criterion data. In a concurrent study, there is little or no time interval between collection of the predictor data and collection of the criterion data. In many research applications the type of design (predictive versus concurrent) is confounded with the sample of individuals participating in the study. Predictive studies commonly sample from

among existing job applicants whereas concurrent studies commonly sample from among present-employees. Because the focus of this paper is on potential differences between present-employees and job applicants, the terms present-employee studies and job applicant studies are used in this manuscript. Present-employee studies are those studies wherein at the time the predictor measure was administered to the study participants, the hiring organization had made an offer of employment to the individuals and completion of the personality inventory was voluntary. Job applicant studies are studies that administered the predictor measure as a required element of the selection system to individuals that were applying for positions with the host organization. Also classified as job applicants are studies that sampled present-employees that were under consideration for promotions. As such, the primary distinction between present-employees and job applicants concerns whether or not the participants completed the personality inventory while under consideration for an occupational appointment.

At times, liberties are taken with the use of the terms present-employee and job applicant studies in reference to previous literature where the original author or authors used the terms concurrent and predictive studies. Where there appears to be potential for misrepresenting what those authors stated, it will be necessary to use the terms predictive and concurrent studies.

The question of the generalizability of present-employee research to job applicants has been investigated and debated before (Barrett, Phillips, & Alexander, 1981; Guion, 1998; Guion & Cranny, 1982). Barrett et al. (1981) reviewed a number of lines of evidence supporting the generalization of results in the context of *cognitive ability* tests. In the context of *personality tests*, similar evidence supporting the generalizability of results is lacking. To be more specific, some researchers have found slightly higher criterion-related validity estimates for personality tests in studies of job applicants (as compared to present-employees; Tett, Jackson, & Rothstein, 1991), while other researchers have found slightly higher criterion-related validity estimates for personality tests in studies of present-employees (as compared to job applicants; Hough, 1998a).

Although there has been mixed evidence concerning the comparability of present-employee and job applicant based validation studies of personality tests in terms of bivariate validity coefficients, a more important issue concerns the generalizability of *prediction equations* derived from studies of these two different groups. When a criterion-related validation study is conducted, multiple regression analysis will often be used to derive an equation for predicting job performance from personality test scores. In future applications of the personality test, scores

can be used to predict how well each individual will perform on the job. Typically, the organization will want to select employees from among those individuals with the highest predicted levels of job performance. If the prediction equations from the two types of samples were to differ, then the rank ordering of future job applicants in terms of their predicted performance levels would also differ. The consequence of this would be that different individuals would be hired by the organization.

The properties of ordinary least squares regression all but assure that two prediction equations derived from two different samples will differ. Ordinary least squares regression derives an optimal solution based on the specific sample of data in the regression analysis. Any effort to cross-validate the resultant regression equation on a new sample of data from the same population will almost always yield a higher degree of discrepancy between predicted and observed values on the outcome variable than had been observed on the initial sample (Pedhazur, 1997). Even if a prediction equation is derived on a sample of job applicants, there will be some degradation of predictive validity, or shrinkage, when this prediction equation is applied to a future sample of job applicants. For example, a selection battery might yield a validation multiple R of .40 with a sample of job applicants. Yet, when the same prediction equation is used to predict performance among a future group of job applicants, it might be observed that predicted performance only correlates with actual performance with a correlation of .30. The key questions for comparing present-employee validation studies with job applicant validation studies are "how much cross-validation shrinkage occurs when results from a present-employee study are applied to future job applicants, and what are the practical implications of this shrinkage?" The current research is explicitly focused on these two questions.

Before proceeding, it is necessary to establish guidelines for the types of personality measures this paper is concerned with. The focus of this paper is on what are often called self-report personality inventories. In a self-report personality inventory, the respondent is presented with a series of adjectives, phrases, or sentences. There are a number of variants in the response formats of personality inventories. In some cases, the test-taker indicates if the adjective, phrase, or sentence is descriptive of them by choosing "true" or "false". Alternatively, the examinee may be asked to use a Likert style scale to indicate the extent to which the stimulus item is descriptive of them. Personality inventories that present a single phrase, adjective, or sentence and ask test-takers to respond using either a true/false or a Likert style scale response are referred to as

single-stimulus inventories. As an alternative to single-stimulus measures are forced-choice inventories. In forced-choice inventories the individual is presented with two or more adjectives, phrases, or sentences and he or she is required to select the one option that is most descriptive of him or herself. There has been renewed interest in the use of forced-choice response formats for personality tests in personnel selection (Jackson, Wroblewski, & Ashton, 2000). Inventories utilizing forced-choice response formats were included in the criterion-related validity analyses described below. However, as will be described below, correlations among personality scale scores were a central aspect of the current investigation. Forced-choice response formats are known to yield lower estimates of the correlations between personality scale scores, particularly when only a few scales are investigated (Baron, 1996). For this reason, inventories utilizing a forced-choice format were excluded from the analyses of correlations among personality constructs.

Also among the types of inventories excluded from this analysis are projective personality tests (McClelland & Boyatzis, 1982), conditional reasoning personality tests (James, 1998), vocational interest and job preference inventories (Holland, 1979), and measures of biographical experiences or biodata (Owens, 1976). Projective personality tests are excluded due to the relative infrequency of their application in employment settings. Similarly, conditional reasoning measures are excluded due to the lack of research on this relatively recent development in personality assessment. Vocational interest inventories are excluded as they are typically used to predict the nature of one's employment or the type of occupation that an individual might prefer as opposed to their actual degree of success on a particular job (De Fruyt & Mervielde, 1999).

In many ways, the arguments presented in this paper apply to biodata inventories as well as self-report personality inventories. The primary reason biodata is excluded here is that biodata items often are not indicators of identifiable psychological constructs (Klimoski, 1993). Often, a single biodata item is indicative of multiple personal attributes. Although this is also true for some empirically derived personality inventories (e.g., the California Psychological Inventory (CPI), Gough, 1989), it is typically possible to classify scales from personality inventories into existing taxonomies of personality. The same cannot be said for biodata measures.

Personality is defined here as *individuals' characteristic patterns of thoughts, feelings, behaviors, attitudes, and motives* (Pervin, 1996). Throughout this paper a variety of terms are

utilized when discussing personality. The terms attributes, traits, and dispositions will all be used interchangeably to refer to personality, while the term personality *factor* is reserved to refer to the five-factor taxonomy of personality.

In the next chapter, evidence concerning the usefulness of personality tests in the prediction of job performance is reviewed. It is noted that in much of the research on personality tests in employee selection, the potential for differences between present-employees and job applicants has not received careful consideration. In the remaining chapters, the methodology and results of a study designed to investigate the generalizability of present-employee based studies to job applicants are described and discussed.

Chapter Two: Literature Review

Evidence Supporting the use of Personality Inventories in Personnel Selection

It is reasonable to expect that an individual's characteristic patterns of thoughts, feelings, behaviors, attitudes, and motives will be related to her or his job performance. Despite the importance of situational constraints on human behavior, most researchers agree that behavior is a function of both the situation and the individual. Perhaps the most obvious example of how personality would be related to job performance is in regard to dispositional achievement motivation. High Need for Achievement individuals derive pleasure from overcoming obstacles and achieving difficult goals (Spangler, 1992). Indeed, projective personality techniques, self-report questionnaires, and conditional reasoning measures of the Need for Achievement and/or the Achievement Motive all have been shown to be related to ratings of job performance (Goffin, Rothstein, & Johnson, 1996; James, 1998; Spangler, 1992). While most researchers (though certainly not all) today agree that personality inventories exhibit useful levels of criterion-related validity, this was not always the case. Indeed, Kluger and Tikochinsky (2001) presented the personality-performance relationship as an example of a "commonsense hypothesis" that had long been accepted as a truism, fell out of favor due to lack of empirical support, and eventually was resurrected. The primary debate over the years has been whether or not personality is related to job performance in all jobs (the validity generalization position), or if personality is only related to job performance in certain settings (the situational specificity position).

One of the earliest reviews of the criterion-related validity of personality inventories was conducted by Ghiselli and Barthol (1953). In order to assess the usefulness of personality as a predictor of job performance, they accumulated studies published between 1919 and 1953.[1] Weighting by sample size and grouping studies according to job type, they found average validity coefficients ranging from .14 for general supervisory jobs to .36 for sales-oriented jobs. Their general conclusion was that *under certain circumstances* (emphasis added), validities were better than might be expected, but that enough studies reported negative results to warrant caution in the use of personality tests.

[1] In Ghiselli and Barthol (1953), as well as in Ghiselli's later research, studies were only included in the review if the personality trait assessed in the study appeared to be important for the job in question.

Locke and Hulin (1962) reviewed the evidence concerning the criterion-related validity of the Activity Vector Analysis (AVA). The AVA is an adjective checklist in which the respondent (a) checks any or all of 81 adjectives that anyone may have ever used to describe him or her and (b) checks any or all of the same adjectives that he or she believes are truly descriptive of him or herself. The goal of their study was to evaluate the AVA "in terms of its demonstrated ability to make better-than-chance predictions of success on a job". They located 18 studies that had examined validity evidence for the AVA. The general conclusion they drew from their analysis was there was little evidence to support the usefulness of the AVA as a predictor of job performance. They argued that only the study by Wallace, Clark, and Dry (1956) met the requirements for a sound validation study (large N, administration of the test before hiring, and cross-validation of findings); that study found AVA scores were not significantly related to performance in a sample of life insurance agents.

Guion and Gottier (1965) extended the inquiry into the validity of personality measures by examining personality inventories other than the AVA. They reviewed manuscripts published in the *Journal of Applied Psychology* and *Personnel Psychology* between the years 1952 and 1963. They found the results from these studies were relatively inconsistent. Therefore they concluded, "there is no generalizable evidence that personality measures can be recommended as good or practical tools for employee selection" and personality measures must be validated in the specific situation and for the specific purpose in which one hopes to use them.

In 1973, Ghiselli published a more comprehensive review of aptitude (including personality) tests in employment hiring, including both published and unpublished studies. He estimated the weighted average criterion-related validity of predictors according to occupational type. His discussion centered on the types of predictors yielding the highest levels of validity within each occupational type. He found that among sales jobs and vehicle operator jobs, personality measures were among the best predictors of performance. Personality inventories were found to be of low to moderate utility in clerical jobs, managerial jobs, and service jobs, and were of no use at all in protective service jobs.

The development of meta-analytic techniques (Schmidt & Hunter, 1977) had a significant influence on reviews of research on personality inventories. Prior to that time, only Ghiselli consistently computed weighted averages of validity coefficients when summarizing the results of studies of personality tests. With the advances in meta-analytic techniques, researchers began

to investigate the possibility that differences in study characteristics (such as sample size, variance on the predictor measure, and measurement error in the criterion) might account for the observed variability in the relationships between personality and job performance. Schmitt, Gooding, Noe, and Kirsch (1984) utilized a bare-bones meta-analytic approach to estimate the average validity of a number of predictors of job performance, and to estimate the extent to which sampling error alone might account for variability in validity coefficients across studies. They estimated that the criterion-related validity (uncorrected for range restriction or measurement error in the criterion or predictor) for personality inventories was .15, and 23% of the variability in validity estimates across studies could be accounted for by sampling error. This study provided additional support to the earlier conclusion drawn by Guion and Gottier (1965) and Ghiselli and Barthol (1953): there is no evidence the validity of personality generalizes across situations.

One possible cause of the observed variability in the validity of personality attributes across settings and studies is differences in the personality attributes measured. Although Ghiselli (1973) attempted to account for this by only including studies in which the personality construct seemed relevant to the job in question, other researchers did not follow this procedure (e.g., Schmitt et al., 1984). Important developments in identifying the structure of personality traits occurred over 50 years ago (Cattell, 1947), but only recently have industrial psychologists incorporated taxonomies of personality traits into their reviews of the validity of personality inventories. Barrick and Mount (1991) classified personality inventories according to the big five (Conscientiousness, Extraversion, Emotional Stability, Agreeableness, and Openness to Experience) personality factors and examined the criterion-related validity of personality constructs accordingly. Barrick and Mount (1991) also corrected observed validities not only for sampling error but also for range restriction on the predictor measures and measurement error on the predictor and criterion measures. This allowed them to estimate the true population correlation between each of the big five personality factors and job performance, and to estimate the extent to which study-to-study differences in statistical artifacts account for differences in the observed correlation coefficients in those studies.

Despite prior research that suggested validities of personality measures did not generalize across jobs, these authors predicted that two of the big five personality factors, Conscientiousness and Emotional Stability, would generalize across settings and criteria. They

located published and unpublished studies conducted between 1952 and 1988, resulting in the inclusion of 117 studies. When data across all occupations and all criteria were examined, the estimated population correlation between Conscientiousness and job performance was $\rho = .22$. Although statistical artifacts could only account for 70% of the variance in the correlations across studies, the estimated true population correlation between Conscientiousness and job performance was positive for every occupational group, and the 90% credibility value for the Conscientiousness-performance correlation across all occupations was .10. On the basis of these results, they concluded that Conscientiousness was a valid predictor for all occupational groups. Regarding the other big five personality factors, the estimated true correlation between personality and job performance was either zero or was negative for at least one occupational group. The Barrick and Mount (1991) study has often been cited as evidence that the validity of Conscientiousness generalizes across settings.

Tett et al. (1991) also meta-analyzed the validity of personality predictors of job performance. A key difference between their work and that of Barrick and Mount (1991) is that Tett et al. explored additional moderators of the validity of personality inventories. One of the primary moderators they tested was the conceptual rationale for including a particular personality trait as a predictor of job performance. They referred to the findings of Guion and Gottier (1965) who submitted that theoretically based studies of relationships between personality and performance generally yielded poorer results than empirically driven studies of the same. One of the primary purposes of the Tett et al. (1991) study was to evaluate the support for this claim. Therefore, the authors focused on the conceptual rationale of the original study as a potential moderator of validity. If the authors of the original study did not provide a theoretical basis for including a specific temperament characteristic, Tett et al. classified it as an exploratory study; if the primary study authors provided a theoretical underpinning for a personality-performance relationship, Tett et al. categorized the study as adopting a confirmatory research strategy (1991). A second difference between the Tett et al. (1991) review and the Barrick and Mount (1991) report is that Tett et al. (1991) argued that there may be situations in which a personality trait is expected to be negatively related to job performance. In such a study, a negative correlation is not a “negative finding”; it is actually a positive finding. As such, they computed the absolute value of the correlation between a predictor measure and a performance criterion for each study, and aggregated the absolute value correlations. The results of their study suggested that

personality is a better predictor of job performance when used in a confirmatory manner, that the big five factor Agreeableness had the strongest relationship with job performance, and that very little of the variance in the validity of personality across studies could be accounted for by differences in statistical artifacts.

Ones, Mount, Barrick, and Hunter (1994) criticized the decision of Tett et al. (1991) to include only studies that utilized a confirmatory approach when estimating the validity of the big five personality factors, arguing instead that all available studies should have been included in the meta-analysis, regardless of research strategy. However, the purpose of the Tett et al. (1991) meta-analysis was to identify moderators of the validity of personality tests as predictors of job performance, and, they identified research strategy as a moderator of validity. More specifically, they found that theoretically derived personality predictors (confirmatory studies) were, in general, superior to empirically derived predictors. Arguing that confirmatory research strategies are superior in terms of professional practice as well as for theory development, they chose to focus on such studies. Further, Tett et al. were not attempting to replicate the findings of Barrick and Mount (1991). Instead, they were attempting to extend the findings of Barrick and Mount (1991).

Ones, Viswesvaran, and Schmidt (1993) reviewed the evidence concerning the validity of a specific type of personality inventories, tests of integrity. They also looked at a number of factors that might moderate the validity of integrity tests, such as the type of integrity test, the nature of the criterion, and the validation sample type. They accumulated 665 validity coefficients based on a total N of 576,400. Their findings suggest that integrity tests are valid predictors of both job performance and counter-productive behaviors across settings, although there are factors that moderate the validity of such tests. For example, they found that the estimated true criterion-related validity of integrity tests as predictors of ***job performance*** was higher when the validation sample consisted of job applicants as compared to present-employees. On the other hand, they found that the estimated true criterion-related validity of integrity tests as predictors of ***counter-productive behavior*** was higher when the validation sample consisted of present-employees as compared to job applicants.

Mount and Barrick (1995) expanded on the Barrick and Mount (1991) study by including a greater number of original studies. The focus of the 1995 study was on the relative merits of a broad personality factor (Conscientiousness) versus more narrow personality traits (achievement

and dependability). Evidence from their review supports the position that when the criterion to be predicted is broad (overall job proficiency), there is relatively little difference between the predictive validity of the broad personality factor and the more narrow personality traits. However, when the criterion to be predicted is specific (e.g., employee effort or employee reliability) ***and*** the criterion is conceptually related to the narrow trait, narrow traits demonstrate higher levels of predictive validity.

Salgado (1997) examined the criterion-related validity of the big five personality factors in the European Community. The purpose of his study was to investigate whether the validity of the big five personality factors generalized across geographic boundaries. He accumulated the results of 36 studies conducted within the European Community between the years 1973 and 1994. The results of his analysis yielded a population parameter estimate of $\rho = .25$ for the correlation between Conscientiousness and job performance. Although statistical artifacts were estimated to account for only 66% of the observed variance in validities, the lower bound of the credibility value was .13, supporting the conclusion that Conscientiousness has a positive correlation with job performance across settings. Salgado (1997) also found that Emotional Stability exhibited generalizable validity across settings, with a population parameter estimate of .19 and a credibility value of .10.

Frei and McDaniel (1998) focused on the criterion-related validity of a specific type of personality related measure, customer service orientation. They gathered 41 validity coefficients with a total N = 6,945. Results from this investigation supported the conclusion that customer service measures have a strong, generalizable relationship with job performance. The true population criterion-related validity estimate (that is, corrected for range restriction and measurement error in the criterion) was $\rho = .50$ and all of the variance in validity estimates could be accounted for by statistical artifacts.

Hough (1992; 1998a) has also examined the validity evidence for personality as a predictor of job performance and other criteria. Although much of the recent research on personality predictors of performance has adopted the five-factor taxonomy, Hough (1998a) utilized an eight-factor taxonomy. The eight factors in her taxonomy are affiliation, potency, achievement, dependability, adjustment, agreeableness, intellectance, and rugged individualism. Mapping her classification system onto the big five would place affiliation and potency as distinct factors that are conceptually similar to Extraversion. Similarly, achievement and

dependability are distinct factors that are conceptually similar to Conscientiousness. Adjustment, agreeableness, and intellectance are conceptually similar to Emotional Stability, Agreeableness, and Openness to Experience, respectively. Rugged individualism, on the other hand, does not map onto the big five taxonomy.

Hough does not adopt the meta-analytic techniques that most others have used. Specifically, she does not attempt to estimate the variance in observed validity coefficients that is due to statistical artifacts. Instead, she simply reports the mean validity estimates across studies. Two more unique features of the Hough (1992; 1998a) analyses deserve mention. First, the studies she gathered were sub-grouped according to the type of validation study design (predictive or concurrent) utilized. Second, she categorized the criterion from each study as job proficiency, training success, educational success, or counter-productive behavior. A noteworthy finding from her investigation was that the mean validity of the eight personality factors varied as a function of study design. Achievement was the best predictor of job proficiency across both study designs, with an estimated validity of .19 in predictive studies and an estimated validity of .13 in concurrent studies. The value of .19 in predictive studies is identical to the average observed *r* for achievement measures in the Mount and Barrick (1995) meta-analysis.

Finally, Hurtz and Donovan (2000) estimated the criterion-related validity of personality measures that were explicitly designed to measure the big five personality factors. These researchers expressed concern about the construct validity of the big five, as utilized in prior meta-analytic reviews. They pointed out that other researchers (R. Hogan, J. Hogan, & Roberts, 1996; Salgado, 1997) had questioned the manner in which earlier quantitative reviews had categorized various personality scales into big five categories. Potential consequences of this are inaccurate estimates of the mean and variance of the validities of each of the big five personality factors. On the basis of 26 studies that met their inclusion criteria, Hurtz and Donovan (2000) found that Conscientiousness exhibited generalizable validity, with an estimated true criterion-related validity of $\rho = .20$, and a 90% credibility value of .03. Emotional Stability also exhibited generalizable validity with an estimated true criterion-related validity of $\rho = .13$, and a 90% credibility value of .06. The estimate of the validity of Conscientiousness is slightly lower in the Hurtz and Donovan study than in the Mount and Barrick ($\rho = .31$; 1995) or the Salgado ($\rho = .25$; 1997) study. On the basis of their study, in concert with numerous other reviews that have indicated low to moderate validities of the big five, these authors suggested that future research

focus on more narrow personality factors that are conceptually aligned with the performance criterion in question.

It is noted here that two issues have received significant attention by reviewers of personality inventories in personnel selection research. The first concerns the degree to which the validity of personality inventories generalizes across settings. Early researchers generally concluded that there was no evidence that validities generalize across situations (Ghiselli, 1953; Guion & Gottier, 1965). More recent evidence utilizing advances in psychometric meta-analysis provide evidence for the generalizability of Conscientiousness, Emotional Stability, customer service orientation, and integrity (Barrick & Mount, 1991; Frei & McDaniel, 1998; Hurtz & Donovan, 2000; Ones et al., 1993; Salgado, 1997). Yet, despite the evidence concerning the generalizability of validity, there is ample evidence that situational moderation of the validity of personality tests does exist (Barrick & Mount, 1993; Helmreich, Sawin, & Casrud, 1988; Stewart, 1996; Tett & Burnett, 2003).

The second issue that has received a great deal of attention has been the expansion of the predictor domain to include specific personality constructs. Early reviews categorized all personality factors as a single predictor category (Ghiselli, 1973; Schmitt et al., 1984). More recently, researchers have expanded their taxonomies to include at least the big five and perhaps more specific personality factors (Schneider, Hough, & Dunnette, 1996). For example, Hough has been one of the most adamant proponents of expanding our view of personality predictors beyond the big five framework. Hough (1998a) cites results of her research in which the validity of achievement was substantially greater than the validity for dependability. As a necessary consequence, she argues, classifying both studies as measures of Conscientiousness would seemingly dilute the predictive power of achievement.

Although these two issues are of extreme importance and warrant continued attention, there is another relatively neglected matter - the issue of sample type. To be sure, most of what we know or purport to know about the usefulness of personality tests in personnel selection comes not from research on job applicants, but rather comes from research on present-employees. As with the age old debate concerning the external validity of research conducted with undergraduate college students, one must question if the same psychological variables and levels of motivation are functioning in present-employees as are functioning in job applicants. Generalizable validity evidence from existing meta-analyses (Barrick & Mount, 1991; Hurtz &

Donovan, 2000) can be interpreted as support that the validity of Conscientiousness (for example) does generalize across present-employee and job applicant populations. Yet, this is simply evidence that the validity is positive in both populations - the actual magnitude of the validity coefficients in the two populations can differ by a practically meaningful amount. As discussed above, Locke and Hulin (1962) found substantial differences in the validity evidence for the AVA between studies that were conducted using present-employees and studies that were conducted with job applicants. In the following section, the potential differences between present-employee and job applicant based studies are discussed, with an eye toward relevant research in this domain.

The Comparability of Present-Employee and Job applicant Designs in Criterion-Related Validation

Selection practitioners are interested in the usefulness of tests in terms of distinguishing between those job applicants who will be successful on the job from those job applicants who would not be successful on the job. The most commonly sought after index of a predictor's usefulness is the criterion-related validity of that predictor in the population of job applicants. For reasons of convenience and time efficiency, present-employees are often sampled in place of job applicants. Sampling job applicants would necessarily entail a longitudinal study, as some time would come to pass before a performance criterion indicator was available for the hired applicants. It is not readily apparent the extent to which present-employee and job applicant studies yield interchangeable results. Present-employee and job applicant samples are likely to enact very different role-taking behavior when responding to personality tests. Role-theory, as applied to personality testing, suggests that test taking is a social situation in which test-takers use the personality inventory to communicate information about themselves and inform the test interpreter how they wish to be regarded (J. Hogan & R. Hogan, 1998; R. Hogan, 1991; Kroger, 1974). A pivotal question is, do incumbents and applicants communicate the same information about themselves? The selection situation can be expected to foster impression management on the part of the applicant, as most applicants want to convey competence and skill in order to secure a job offer (Tedeschi & Melburg, 1984). The empirical evidence on this point is clear; based on the outcome of mean scores on selection tests, applicants present more favorable self-information than incumbents (Green, 1951; Heron, 1956; Hough, 1998b; Robie, Zickar, &

Schmit, 2001; Rosse, Stecher, Miller, & Levin, 1998; Smith, Hanges, & Dickson, 2001; Stewart, 1997).

Although applicants most assuredly engage in behavior designed to convey a favorable image, this does not mean that such self-presentation or impression management is entirely conscious or deceptive. Evidence indicates subtle situational cues such as the perceived purpose of testing, characteristics of the test administrator, and the test title influence test-takers' responses to personality inventories, even when test-takers have been explicitly instructed to respond honestly (Kroger, 1974). Kroger and Turnbull (1975) administered an interest inventory and a personality inventory to undergraduate students; one group of students were told the inventories were designed to assess military effectiveness whereas the other group of students were told the inventories were designed to measure artistic creativity. Although participants had been randomly assigned to groups, and had been instructed to respond to the tests honestly, students in the artistic creativity condition scored higher than students in the military effectiveness condition on interest scales such as Artist, Musician, and Architect. Conversely, students in the military effectiveness condition scored higher than students in the artistic creativity condition on interest scales such as Aviator and Army Officer.

Contextual differences between present-employees and job applicants led many industrial psychologists to be cautious about generalizing the results from present-employees to job applicants (e.g., Locke & Hulin, 1962). In recent years, however, reviews of the validity of personality inventories in selection have not examined the possibility of sample type as a potential moderator of criterion-related validity. For example, Barrick and Mount (1991), Churchill, Ford, Hartley, and Walker (1985), Ford, Walker, Churchill, and Hartley (1987), Frei and McDaniel (1998), Hurtz and Donovan (2000), Mount and Barrick (1995), Salgado (1997), and Vinchur, Schreischeim, Switzer, and Roth (1998) do not investigate sample type as a moderator of personality criterion-related validity in their meta-analyses. On the other hand, only Hough (1998a), Ones et al. (1993), and Tett et al. (1991) distinguish between sample types when conducting their analyses.[2] Lack of attention to sample type could reflect the implicit belief on

[2] Hough (1998a) actually distinguished between predictive and concurrent validation study designs, while Tett et al. (1991) grouped studies according to incumbents versus recruits. In keeping with the conventions of the present manuscript, I use the terms job applicant and present-employee. It is certainly possible that some of the studies contained within Hough's

the part of researchers that sample type does not matter, or it could reflect that the original source studies are typically based on present-employees (Lent, Aurbach, & Levin, 1971). For example, McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001) examined the validity of situational judgment tests. Based on the suggestion of a reviewer, they investigated the possibility that sample type might moderate the validity of situational judgment tests. It is interesting to note that the validity estimate based on concurrent studies (the majority of which were likely present-employee based)[3] was $\rho = .35$ and the predictive validity estimate was $\rho = .18$. What is of greater interest (concern?) here is the fact that 94% of the validation studies included in their meta-analysis were based on concurrent studies, while only 6% were based on predictive studies. Similarly, J. Hogan and Holland (2003) report that 95% of the studies in their analysis were concurrent studies while 5% were predictive (the precise testing conditions are not given, but again, it is likely that the majority of concurrent studies were conducted with incumbents). It is unfortunate that much of the existing evidence concerning the validity of personnel selection measures has neglected to consider the motivational context of the study participants.

To comprehend better the shift in our willingness to rely on present-employee studies, it is necessary to consider arguments put forth by Barrett et al. (1981). These researchers questioned the presumed superiority of job applicant studies, arguing that many of the reasons for this presumed superiority were unfounded. Specifically, they critiqued four frequently cited reasons for the advantage of job applicant based studies: (a) the problem of missing persons in present-employee studies; (b) range restriction in present-employee studies; (c) differences between job applicants and present-employees in motivation and other characteristics; and (d) the possibility that job experience might influence the predictor constructs in present-employee studies. The problem of missing persons suggests poor performers either have been terminated or have left the job, and top performers have been promoted out of the job. Barrett et al. (1981)

---

review were predictive studies of present-employees. And it is evident that some of the samples of recruits in the Tett et al. study were individuals that completed a personality inventory post-hire, during orientation or training.

[3] Ones et al. (1993) conducted a hierarchical moderator analysis investigating validation study design (predictive versus concurrent) and validation study sample (applicants versus incumbents). Sixty-three of the 64 concurrent studies they reviewed utilized present-employee samples.

suggested that the problem of missing persons in present-employee samples is a question of range restriction, essentially leaving only three substantive reasons for preferring job applicant based studies. In turn, they argued job applicant based studies are no less susceptible to range restriction than are present-employee studies. Suppose, for example, an organization is interested in estimating the validity of a measure of Extraversion as a predictor of sales performance. Even if they sample present-employees that have not been selected on the basis of an Extraversion measure, there is likely to be a limited range of extraverts in the sample. This is due to the fact that if Extraversion is indeed related to sales performance, then introverts will have left the job at a disproportionately high rate. If applicants serve as the validation sample, are administered an Extraversion measure and are selected on the basis of some other predictor, it is distinctly possible that the alternative predictor will be correlated with Extraversion. This will result in indirect range restriction on the Extraversion measure among those applicants who are successful. They concluded that job applicant based studies are just as likely to suffer from range restriction as are present-employee studies. In either case, they submit, validity estimates can be corrected for range restriction.

With respect to potential differences between present-employee and job applicant samples, Barrett et al. (1981) argued that it is possible to control for some of these possible confounds (e.g., age). They further suggested that concerns over motivational differences between present-employees and job applicants are unwarranted. Essentially, they argued that it is unknown what effect differential motivation has on validity estimates. The evidence they cited suggesting differential motivation is not a cause for concern came from studies involving cognitive ability as a predictor of job performance. They did not provide evidence supportive of the assumption that motivational differences between present-employees and job applicants do not matter in the context of personality testing.

Finally, Barrett et al. (1981) critiqued the assumption that job experience and training are likely to affect predictor and criterion scores of incumbents, thereby invalidating such results as estimates of validity in job applicants. They espoused the view that because it is possible to control for tenure and experience when conducting validation studies, this is essentially a non-issue. The general conclusion of their paper was that there is no evidence for the presumed superiority of job applicant based studies over present-employee based studies. It should be noted that Barrett et al. did not claim that their arguments necessarily apply to predictors other

than cognitive ability tests.

A second study that likely increased researchers' willingness to accept the results of present-employee based studies as accurately reflecting results of job applicant based studies was the meta-analysis by Schmitt et al. (1984). They compared the criterion-related validity estimates from job applicant studies with those from present-employee studies and found what they interpreted as minimal differences (average r = .30 in job applicant studies without selection on the predictor, average r = .26 in job applicant studies with selection on the predictor, and average r = .34 in present-employee studies). Schmitt et al. concluded that frequently cited reasons for expecting different results between present-employee and job applicant samples (e.g., motivational effects and job experience) might not be that important.

One difficulty in interpreting these results is that Schmitt et al. collapsed across all predictors in their meta-analysis. That is, they did not distinguish between personality predictors and cognitive ability predictors when comparing validity estimates from predictive and concurrent studies. Potentially, the differences between present-employee and job applicant studies could be greater for personality tests than for cognitive ability tests. That is, the possibility remains that lower levels of motivation among present-employees as compared to job applicants can cause present-employee studies to *underestimate* the operational validity of *ability* tests while *overestimating* the operational validity of *personality* tests. Results of a study by Schmit and Ryan (1992) are consistent with this possibility. They found that in a sample of individuals motivated to present themselves favorably (as compared to a sample of individuals who were not similarly motivated), there was a decrement in the validity of personality inventories and a gain in the validity of ability tests.

While the Barrett et al. (1981) and the Schmitt et al. (1984) papers might be viewed as evidence that present-employee and job applicant samples are comparable, there is also reason to question the interchangeability of results from different samples in validation research. First, the findings of Schmit and Ryan (1992) call into question the assumption that motivation exerts a similar influence on validity estimates for cognitive ability test scores and personality test scores. Second, the results of Hough's research (1998a) suggest that studies based on present-employees yielded estimates that were, on average, .07 higher than those studies based on job applicants.[4]

[4] A second study published by Hough in 1998 (1998b) has been cited (Hough & Ones, 2001) as

The third piece of evidence that calls into question the comparability of present-employee and job applicant studies are the results of the Tett et al. (1991) meta-analysis. Although they actually concluded studies of job applicants led to higher validity estimates than studies of present-employees, they incorrectly categorized the Project A data as a study of recruits when in fact, the study they included was a study of incumbents (see Campbell, 1990, p. 234). Given the size of the Project A data, their finding of higher validity for studies of job applicants would likely have been a finding for higher validity among present-employees, had they correctly categorized the Project A study. They pointed out that when the Project A data was omitted from their analyses, there was no significant moderating effect of sample type. Fourth, more recent research based on Project A has found the job applicant validities of the Assessment of Background and Life Experiences (ABLE) composites for predicting "will do" performance factors were lower than the validities from the present-employee sample (Oppler, Peterson, & Russell 1992; Russell, Oppler, & Peterson, 1998). Fifth, the results of the Ones et al. (1993) meta-analysis, while revealing impressive predictive validity estimates in applicant studies, also revealed a differential pattern of the relative magnitude of validity estimates for integrity tests depending on the criterion in question. Studies of job applicants (as compared to studies of present-employees) yielded higher validity estimates when integrity was used to predict job performance, but studies of present-employees (as compared to studies of job applicants) yielded higher estimates of validity when integrity was used to predict counter-productive behavior.

The issue of incumbent and applicant differences is further complicated by the possibility that incumbents and applicants would adopt a different frame of reference when responding to personality test items (Schmit & Ryan, 1993). The self-presentational goals of incumbents participating on a voluntary basis are likely to differ from the self-presentational goals of job applicants (McAdams, 1992). Schmit and Ryan (1993) contend that incumbent and applicant differences might be better understood by considering the ***person-in-situation*** schemas that are enacted during test-taking (Cantor, Mischel, & Schwartz, 1982). Applicants wish to convey competence relative to other applicants, and therefore might operate according to an ***ideal-***

---

evidence that response distortion does not influence the validity of personality scale scores. Seemingly this is a reference to Figure 1 from the 1998b study. Unfortunately there is insufficient description of the data contributing to that figure. For that reason, only the data presented in 1998a are reviewed here.

employee frame-or-reference. Incumbents may enact a ***stranger-description*** frame-of-reference, where they communicate basic information as they would during an initial meeting with a stranger (Schmit & Ryan, 1993, p. 967). These divergent frames-of-reference can influence not only the predictor-criterion correlations (criterion-related validities), but also the correlations among predictor scale scores (Schmit & Ryan, 1993; Schmit, Ryan, Stierwalt, & Powell, 1995; Van Iddekinge, Raymark, Eidson, & Putka, 2003; for an opposing view, see Smith et al., 2001).

This is not to say that divergence in frames-of-reference between incumbents and applicants must have a negative effect on the criterion-related validity of personality scale scores. Hauenstein (1998) and Kroger (1974) suggested that criterion validity could be ***enhanced*** when those who successfully enact a particular role in responding to a test in a motivated condition also perform well on the job. J. Hogan and R. Hogan (1998; R. Hogan & J. Hogan, 1992) submit that even if people do attempt to respond in a desirable manner in selection situations, there are individual differences in how successful people are at presenting a favorable image, and these are important individual differences related to social skill. Thus, motivated responding could be a source of bias that is related to job performance. As an example, consider the Need for Affiliation component of McClelland's Leadership Motive Pattern (McClelland & Boyatzis, 1982). McClelland and Boyatzis (1982) found that the personality pattern of successful managers at AT&T included a *low* Need for Affiliation. Imagine a particular individual who happens to be dispositionally low in the Need for Affiliation, but who would not be successful as a manager. If this individual adopted a predominantly honest role when responding to the test, presenting his or her low Need for Affiliation, the consequent would be that his or her performance would be over-predicted on the basis of his or her Need for Affiliation score. Now suppose that this individual had instead responded with a motivation to present himself or herself as a successful manager, but had incorrectly chosen to enact the role of a manager who is high on the Need for Affiliation. In this case the hypothetical poor performing manager is motivated to adopt a specific role, and by doing so, communicates to the test interpreter that he or she does not understand the behaviors and characteristics reflective of a successful manager. This person's profile, then, becomes a more accurate predictor of their job performance when they are motivated to respond in a more favorable manner.

Divergence in frames-of-reference adopted by incumbents and applicants as a source of bias in correlations among personality predictors draws attention to a more important issue,

though. Specifically, comparisons of bivariate validity coefficients between present-employee and job applicant based validation studies might not present a complete picture of the comparability of these two different types of samples. Because the correlations both among personality scales as well as between personality scales and the criterion can differ by sample type, a comprehensive comparison of incumbent and applicant samples must also examine regression coefficients associated with each predictor across the two types of samples.

There is evidence that samples differing in motivation levels will yield diverse prediction equations. In the study by Schmit and Ryan (1992), they found in a sample of individuals who were motivated to present themselves favorably (simulated applicants), cognitive ability tests were strongly ($r = 0.38$) related to success (GPA) and personality tests were weakly related to success ($r = 0.15$). However, in a sample of individuals who were less motivated to present themselves favorably (as is assumed to be the case with present-employees), both cognitive ability tests ($r = 0.31$) and personality tests ($r = 0.52$) were correlated with success. If the prediction equation derived from the less motivated sample of individuals had been utilized to predict performance among the motivated sample of individuals, the cross-validation would likely have been quite poor.

Hauenstein (1998) also provided evidence concerning the potential problems associated with applying prediction equations across populations that differ in terms of their motivation to present themselves favorably. Utilizing a sample of college students who had completed the CPI, he found the equations for predicting GPA differed as a function of the motivation of his study participants. Three conditions were included: (a) students who were motivated to present themselves in a maximally socially desirable manner; (b) students who were motivated to present themselves as an excellent student; and (c) students who were asked to present themselves honestly. To estimate the potential loss in utility when a prediction equation is applied across populations that differ in motivation, he first estimated the utility of using a prediction equation derived from students motivated to present themselves as ideal college students. Assuming a base rate of .50 and a selection ratio of .20, he simulated which of those students would have been "selected" on the basis of the "ideal college student" prediction equation. He found that 67% of those who would have been selected had GPAs equal to or higher then the GPA established as a cutoff for successful performance. When he utilized the prediction equation derived from the honest respondents to predict performance in the ideal college student sample,

again assuming a base rate of .50 and a selection ratio of .20, he found that only 55% of those who would have been selected had GPAs equal to or higher then the pre-determined cutoff for success.

A final study that illustrates the potential drawback to applying present-employee results to job applicants is a study by Stokes, Hogan, and Snell (1993). They used regression analysis to empirically key a biodata instrument. This was done separately in a sample of present-employees as well as in a sample of job applicants. When they compared the empirically derived keys from the two samples, there were *no* overlapping items in the two resulting keys. In addition, when the present-employee based item key was applied to job applicant responses, the validity for predicting the criterion (tenure) was only .08. Finally, they found that when option-keying (as opposed to item-keying) was used, there were 59 response options that were related to tenure in both the job applicant and the present-employee samples. However, 23 of these 59 options were keyed to tenure in the *opposite* direction in the two different samples.

If the results of the Stokes et al. (1993), Hauenstein (1998), and Schmit and Ryan (1992) studies are indicative of a similar process operating in other settings, the implications for the use of present-employees in validation studies involving personality tests are nontrivial. The regression equation that optimizes predicted job performance among present-employees might bear little resemblance to the regression equation that optimizes the prediction of job performance among job applicants. If a present-employee based prediction equation fails to generalize to a sample of future job applicants, estimates of utility based on present-employee studies will overestimate the actual utility gain when personality tests are used to hire employees.

The point of this discussion is to emphasize that there are reasons to suspect that present-employee based studies are not interchangeable with studies of job applicants, and that efforts to evaluate the interchangeability of data sampled from these two distinct populations must move beyond simple comparisons of bivariate validity coefficients. Efforts to compare present-employee and job applicant studies should focus on the prediction equations derived from these two types of samples. If differences in sample type are related to differences in prediction equations and differences in predicted performance, they will also yield differences in applicant rank-orders. Ultimately, differing rank orders can lead to differing levels of the actual ***utility*** gained from the use of personality inventories in selection.

The preceding discussion is not intended to be an argument that estimates of validity

coefficients are not important. If the purpose of an investigation is to estimate the operational validity of a personality trait as a predictor of performance, a bivariate validity coefficient based on a sample of job applicants is an appropriate index. However, the purpose of the current investigation is *not* only to estimate the operational validity of personality traits in the prediction of job performance. The purpose of the current investigation is to estimate the comparability of present-employee and job applicant samples as estimates of the utility of personality inventories in personnel selection. To address this issue, it is necessary to take a more expansive view that includes not only validity coefficients, but also regression coefficients, prediction equations, and utility. The next chapter introduces a study designed to explicitly test the comparability of present-employee validation studies with job applicant validation studies in the context of personality tests. The study tests the following hypotheses:

> Hypothesis 1: Present-employee and job applicant based validation studies will yield different estimates of the bivariate criterion-related validity of personality tests.
>
> Hypothesis 2: Present-employee based validation studies will overestimate the incumbent-applicant cross-validation validity of personality trait measures as predictors of job performance when used in job applicant settings.
>
> Hypothesis 3: Present-employee based validation studies will overestimate the financial utility of implementing personality trait measures as predictors of job performance in job applicant settings.

Summary

Research on the use of personality inventories in personnel selection suggests Conscientiousness, Emotional Stability, integrity, and customer service orientation are valid predictors of job performance across settings. However, much of this research has not examined sample type as a potential moderator of the validity of personality test scores. Even if there were evidence that sample type did not moderate the validity of personality test scores, it would not be prudent to assume such a finding reflects immaterial differences between present-employee and job applicant based studies. To get a more informative estimate of the influence of sample type on validation study results, it is necessary to examine the influence of sample type on prediction equations and utility. The present study examines the influence of sample type on validation study results.

First, a meta-analysis of the validity of personality as a predictor of job performance is conducted, where studies are sub-grouped according to sample type. In addition to estimating the relationships between personality traits and job performance, the inter-correlations among personality traits are also estimated. The results of this meta-analytic investigation yield two population parameter estimate correlation matrices (one based on present-employee studies, the other based on job applicant studies). On the basis of the population parameter estimates from this meta-analysis, cases of hypothetical present-employees and job applicants are simulated. Utilizing the population of present-employee data, a regression equation is estimated, which is then cross-validated on the population of job applicant data. This provides an estimate of the incumbent-applicant cross-validation R when present-employee derived equations are applied to future job applicants.[5] This value is then compared to the multiple R that is obtained when the incumbent and the job applicant meta-analytic correlation matrices are analyzed with multiple regression analysis.

Next, the Brogden-Cronbach-Gleser utility formula is used to estimate the utility gain from using personality inventories in personnel selection. In order to compare the results from present-employee based studies with job applicant studies, two utility estimates are computed. The first is based on the utility of present-employee studies, and makes use of the R estimated from present-employee studies. The second is based on the application of the present-employee derived prediction equation to job applicants, and makes use of the incumbent-applicant cross-validation R. The incumbent-applicant cross-validation R is the correlation between job performance scores for the simulated job applicant observations with predicted performance of those job applicants (when predicted performance is based on the prediction equation derived from the present-employee data).

[5] Traditionally, the term cross-validation refers to the application of a regression equation derived from one sample of data to another sample of data drawn from the same population. As the current study implicitly views incumbents and applicants as two distinct populations , the term incumbent-applicant cross-validation R is used to refer to the application of the incumbent-based prediction equation to the applicant sample data.

Chapter Three: Research Methodology

A series of meta-analyses was conducted in order to derive meta-analytic correlation matrices among personality predictor constructs and a job performance criterion construct. One strength of using meta-analysis to construct the correlation matrices is that it is not necessary that any single study include measures of all the constructs under investigation (Viswesvaran & Ones, 1995). In the current situation, many studies report only criterion-related validity coefficients. Other studies report correlations among as few as two predictor constructs, but not correlations with any outcome variables. Reporting correlations among personality scale scores without reporting criterion-related validities was common in studies that compared the factor structure of personality measures in diverse groups (Collins & Gleaves, 1998; Ellingson, Smith, & Sackett, 2001; Hofer, Horn, & Eber, 1997; Smith & Ellingson, 2002).

Identification of Studies and Inclusion Criteria

The meta-analyses completed here were limited to research conducted in workplace settings. There have been a number of studies that have compared measurement properties of personality inventories between job applicants and non-applicants (Birkeland, Manson, Kisamore, Brannick, & Liu, 2003; Schmit & Ryan, 1993; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). As the focus of the present research was on comparing applicants with incumbents, studies were included only if participants completed the personality inventory in conjunction with their current or potential occupation. Undergraduate student populations were included only if the study was conducted in a career counseling or placement office (e.g., Judge & Cable, 1997), and graduate student populations were included only if the study was conducted in a practicuum setting related to their course of study.

With regard to the self-report personality measures, the taxonomy presented by Hough and Ones (2001) served as the foundation. As indicated in Chapter One, forced-choice measures were included in the criterion-related validity meta-analyses, but only single-stimulus measures were used to estimate the correlations between personality constructs. As such, studies using the Edwards Personal Preference Schedule, for example, were excluded from the inter-correlation meta-analyses.

A computerized database search was conducted using the PsycLit, National Technical Information Service, and ERIC databases in September, 2001. A keyword search of the terms

"*personality or temperament or dispositions*" and "*job performance or occupational success*" resulted in 1,433 matches. Studies that were not available in English were eliminated. A review of the abstracts of the remaining studies led to the elimination of a number of studies that were clearly non-empirical, or, were conducted in a laboratory setting. An effort was made to obtain all promising studies, though many dissertations and a few technical reports were not accessible.

Studies were also identified by hand-searching the 1991 through 2002 volumes of *Journal of Applied Psychology*, *Personnel Psychology*, *Human Performance*, *Journal of Business and Psychology*, *Journal of Occupational and Organizational Psychology*, *Academy of Management Journal*, *Journal of Management*, *Journal of Organizational Behavior*, *Journal of Vocational Behavior*, and *Educational and Psychological Measurement.* A less inclusive search was conducted of *Leadership Quarterly* (1995 – 2002), *Administrative Science Quarterly* (1991 – 1997), *Organizational Behavior and Human Decision Processes* (1991 – 1992; 2000 – 2001), and *International Journal of Selection and Assessment* (1998 – 2001). Those journals searched less inclusively were either not available in the University libraries accessible to the author or were not yielding any promising studies and were therefore abandoned. A manual search of all studies published in the Validity Information Exchange of *Personnel Psychology* was also conducted. Next, programs from the 1996 through 2002 Annual Conferences of the Society for Industrial and Organizational Psychology were searched to identify additional studies to include in the current review. Authors of potentially promising studies were contacted in an effort to obtain copies of their papers. Finally, a number of test publishers, applied researchers, assessment specialists, and consultants were contacted in order to locate unpublished technical reports and unpublished data collected in conjunction with selection and validation projects completed by those firms.

Of the studies reviewed, many failed to report complete information. In some cases, statistically significant results were reported whereas non-significant results were not presented. In such cases, efforts were made to locate authors through a search of the SIOP and APA member email directories as well as the internet search engine "google.com". Of the hundreds of authors contacted, a number responded by sending output and/or raw data. Some authors failed to reply to repeated requests, many responded they did not have the information sought after, and one author refused to release the requested correlations while the manuscript was under review. After an effort had been made to obtain additional data for each study, any usable information

for that study was coded. If only significant correlations were reported and the author could not provide additional data, the available correlations were included in the meta-analysis. In some cases it was unclear if the data presented in a manuscript overlapped with data that had previously been presented in another manuscript. Authors were contacted for clarification on this matter, and, if the sample was in fact the same as the sample in another study, the sample was included only once. If the author did not provide a definitive response, the author made the final decision. Ultimately, 429 samples from 317 studies contributed a correlation to at least one cell in the meta-analytic matrix. The 317 studies contributing one or more samples to one or more meta-analyses are marked in the reference list by an asterisk.

Coding of Personality Constructs and Study Characteristics

Data points in the meta-analysis were coded as having been drawn from a job applicant setting or a present-employee setting. In many cases, the determination of applicant versus present-employee was straightforward. In other, cases, the distinction was not so apparent. For example, for many years, Federal Aviation Administration Air Traffic Controllers were provisionally accepted into the Nonradar Screen training program on the basis of a battery of cognitive tests (Quartetti, Kieckhaefer, & Houston, 2001). Prior to commencing the Nonradar Screen, they went through a medical examination where they completed the 16PF. Not everyone who was qualified on the basis of the cognitive test was actually accepted into the Nonradar Screen. As the medical and psychological exam appeared to be a selection hurdle, the Oakes, Ferris, Martocchio, Buckley, and Broach (2001) study was classified as a study of job applicants. In contrast, Schroeder, Broach, and Young (1993) administered the NEO-PI to Air Traffic Control Specialists ***during*** the Nonradar Screen; this study was categorized as present-employees. Although it was originally planned that studies conducted during orientation/training would be coded as applicant studies, I decided against this because when administered during training, the personality test did not appear to be a barrier to employment or continued employment. If present-employees completed a personality inventory while participating in an assessment center that was to be used for promotion purposes, such samples were categorized as applicants. The reason for this was that the motivational context of a selection-oriented assessment center was thought to be similar to that of a selection context in terms of the desire to present oneself favorably.

In order to estimate the reliability of the coding system for applicants and incumbents, the coding of studies in the current study was compared with the determination made by Tett et al. (1991). Tett et al. (1991) were primarily interested in the chronological nature of the study (predictive versus concurrent), but their coding of studies nonetheless serves as a comparative index for the coding of the current studies. There were 34 overlapping studies between the current sample and those of Tett et al. The percent agreement was 79%. For nearly all of the seven studies wherein there was disagreement between coding in this investigation and the Tett et al. coding, the study had been conducted in a training setting. In the current analysis, the following were coded as studies of incumbents despite their being coded as recruits in the Tett et al. study: Dicken (1969); Lafer (1989); Pugh (1985); and Whisman (1978). Moreover, the following studies were coded as applicants despite their being coded as incumbents by Tett et al. (1991): Arvey, Mussio, and Payne (1972); Burke and Hall (1986); and Hiatt and Hargrave (1988).

Hough and Ones (2001) present a comprehensive taxonomy of personality scales that encompasses the big five, facets of the big five, and compound personality factors. Compound personality factors are defined operationally; compound personality factors emerge when existing personality traits (that may or may not be related to each other) are combined in order to predict an external variable, such as occupational success (Hough & Schneider, 1996). Examples of compound personality factors in personnel selection research are integrity, customer service potential, and stress tolerance (Ones & Viswesvaran, 2001).

In the current study, a seven-category coding system was used to classify personality inventory scales. Inventories and scales that are categorized as indicators of global- or facet-level big five constructs in the Hough and Ones taxonomy were included. In addition, personality scales grouped in the Hough and Ones (2001) optimism and ambition compound factor categories were included. There is some disagreement regarding the classification of personality scales according to the big five. For example, many studies utilized the California Psychological Inventory Dominance scale. Hough and Ones (2001) classify CPI Dominance as an indicator of Ambition (R. Hogan & J. Hogan, 1992), while Barrick and Mount (1991) classified CPI Dominance as indicative of Extraversion. R. Hogan and J. Hogan (1992) take the position that Ambition and Extraversion are two conceptually distinct components of surgency. Hough and Ones suggest that Ambition is a compound factor based upon Extraversion and

Conscientiousness, and is indicated by the CPI Dominance scale and the HPI Ambition scale, among others. Based on the arguments of R. Hogan and J. Hogan (1992) and Hough and Ones (2001), the current study went beyond the big five to include compound factors. The specific compound personality factors chosen were selected for two reasons. First, on the basis of a cursory review of the initial studies obtained, the Ambition and Optimism compound personality factors appeared to be two compound factors represented by a sufficient number of studies. Second, previous research efforts have detailed the psychometric properties of other compound personality factors such as integrity, customer service, and self-destructive autonomy (Frei & McDaniel, 1998; Ones et al., 1993; Ones and Viswesvaran, 2001).

Users' manuals, reviews of personality inventories (e.g., Buros' *Mental Measurements Yearbook*; *Tests in Print*), and personality inventory item lists were examined to identify the response format of each personality inventory. The following inventories were designated as forced-choice measures: Norman's *Descriptive Adjective Inventory, Edwards Personal Preference Schedule,* Ghiselli's *Self Description Inventory, Gordon Personal Profile & Inventory, Jenkins Activity Survey,* and the *Occupational Personality Questionnaire*. Each of these inventories utilizes a forced-choice response format that requires the test-taker to endorse the alternative that is most descriptive of him or her. All other inventories were classified as single-stimulus measures.

Occupational category was also coded as a potential moderator. The Standard Occupational Classification system was used to classify samples. However, there were too few studies in the applicant condition to allow for a meaningful hierarchical moderator analysis by occupation, so this information was not utilized.

Meta-analytic Method and Computation of Correlation Coefficients

For each unique sample in a study, Pearson bivariate correlations between indicators of different personality constructs and/or Pearson bivariate correlations between indicators of each personality construct and an indicator of job performance were recorded. There were a number of studies using a dichotomous performance criterion such as passing or failing a training program, staying in or leaving the organization, or being categorized as successful or unsuccessful. In most of these studies, t-values, F-values, or means and standard deviations for the two levels of the criterion were used to compute point-biserial correlations between the

personality scales and performance. These point-biserial correlations were corrected for the attenuating effect of dichotomization.

Some studies reported a point-biserial correlation between personality scale scores and a dichotomous turnover criterion but did not report the exact number of individuals leaving versus staying. Without this information, it is not possible to correct the point-biserial correlations for dichotomization. One possibility would be to drop these studies. This was not a desirable alternative as the discarding of otherwise relevant studies would only serve to limit the size of the total meta-analysis sample and increase second-order sampling error (Hunter & Schmidt, 1990). Alternatively, the studies could be included without making corrections for dichotomization. This was also not desirable as the inclusion of point-biserial correlations in meta-analysis leads to an underestimate of sampling error variance (Schmidt, Law, Hunter, Rothstein, Pearlman, & McDaniel, 1993). In studies reporting point-biserial correlations with turnover without reporting the percentage of individuals leaving the organization, the percentage of leavers was assumed to be 22% (the median turnover rate from McEvoy & Cascio's (1987) meta-analysis).

There was also one study wherein both the outcome variable and the predictor variable were dichotomized. In this case, the association between personality and performance was represented by a chi-square value. This chi-square value was transformed to a phi correlation and that in turn was transformed into a Pearson correlation.

In the meta-analyses of predictor constructs, there were five studies (Bernstein, Schoenfeld, & Costello, 1982; Birenbaum & Montag, 1986; Gellatly, Paunonen, Meyer, Jackson, & Goffin, 1991; Hofer et al., 1997; Salomon, 2000) that reported results of factor analyses by presenting the pattern (loading) matrix or the factor correlation matrix and structure matrix. These matrices were used to reproduce the correlation matrix, and the reproduced correlations were included in the meta-analysis. These reproduced correlations contain residual variance because the loading matrix included loadings on extracted factors (and the number of factors extracted was less than the number of indicators).

In some studies, a single indicator of each construct was utilized (e.g., studies using the NEO-FFI report one correlation between each of the big five constructs). In these cases, the reported bivariate correlation was included in the meta-analysis. While most studies reported correlations uncorrected for the attenuating effects of measurement error, there were one or two

studies reporting disattenuated correlations. In order to be consistent with the other correlations to be included in an artifact distribution meta-analysis, the reported correlations were attenuated on the basis of the reliability reported in those papers.

There were many instances where multiple indicators of each personality construct were included in a single study; in these instances, a linear composite correlation was computed following the procedures outlined in Hunter and Schmidt (1990, pp. 457-463). With some frequency, correlations of multiple indicators of the same personality construct with a performance criterion were available, while the correlation ***between*** the indicators was not available. This would preclude the possibility of computing a composite score correlation. One approach that could be taken in such a scenario would be to average the correlations. This is not an advisable approach as use of the average correlation leads to a downwardly biased estimate of the correlation between constructs (Hunter & Schmidt, 1990). The approach taken in the current study was similar to that taken by Hurtz and Donovan (2000). The normative correlation between indicator scales (as given in the inventory manual) was imputed in order to compute a composite correlation. If the test publisher's manual was not available or did not contain correlations among the personality scales, the weighted average observed correlation from studies that did report the needed correlation was used as the imputed correlation.

Similarly, if more than one criterion measure was included, composite correlations were computed. In a number of cases, there were multiple criteria but inter-correlations between the criteria were not provided. As was done in the computation of predictor composite correlations, values were imputed from other sources when a given study did not report correlations between criterion measures. Obviously, there are no technical manuals reporting correlations between job performance criterion measures. There are, however, a number of meta-analyses that have estimated population correlations between performance criteria. For example, Rich, Bommer, MacKenzie, Podsakoff, and Johnson (1999) conducted a meta-analysis of the relationship between subjective ratings of job performance and objective indices of sales performance. In the current meta-analysis, the average overall observed correlation from Rich et al. (1999) was used to compute composite correlations in studies that reported personality-sales and personality-rating correlations without reporting the correlation between objective sales and subjective rated performance. Similarly, McEvoy and Cascio (1987) computed the meta-analytic correlation between turnover and performance. The resulting value from their study was used to compute

composite correlations in studies reporting personality-turnover and personality-rating correlations without reporting turnover-rating correlations. Additional meta-analytic studies reporting correlations among criterion constructs that were used to compute composite correlations in the current meta-analysis were Koslowsky, Sagie, Krausz, and Singer (1997; lateness with performance, turnover, and absenteeism); Conway (1999; ratings of interpersonal facilitation, job dedication, technical/administrative performance, leadership, and overall performance); and Bycio (1992; absenteeism and subjective and objective indices of performance).

The Hunter and Schmidt (1990) artifact distribution meta-analysis methods were used in the current study to estimate the criterion-related validities of the personality inventories. The weighted average correlation was used in the estimation of sampling error variance with the noninteractive sampling error formula (see Hunter & Schmidt, 1990, pp. 177 – 182; Hunter & Schmidt, 1994). Criterion-related validities were corrected for measurement error in the criterion only. In the meta-analyses of inter-correlations between personality constructs, a "bare-bones" meta-analysis was conducted, correcting only for sampling error. It was decided that corrections for range restriction would not be appropriate in the current investigation. Range restriction corrections are appropriate when exogenous factors artificially restrict the variability of a measure (e.g., a validation study where selection decisions were based on the predictor scores). In the current study, incumbents and applicants are conceptualized as adopting distinct self-presentation processes when completing personality inventories. As such, the causes of scale variability are due to conceptually different processes between incumbents and applicants. Finally, corrections were not made for unreliability in the personality measures as personnel decisions must be made on the basis of observed, albeit fallible, test scores. All analyses were conducted using cell formulas in Microsoft Excel.

Methods for testing Moderator Effects

The hypotheses to be tested in this study are, in essence, hypotheses regarding sample type as a moderator of the correlations among personality traits as well as the criterion-related validities of personality traits. If sample type does moderate any of the applicable correlations, it is critical that such moderation is detected (e.g., Type II error is controlled). At the same time, if there is no moderating effect, it is desirable that corresponding false conclusions are not drawn.

The process of detecting moderators, then, is a critical step in the current research.

It can be said that there are generally two steps involved in the search for moderating effects in meta-analysis. These steps are slightly different, depending on the meta-analytic approach (e.g., Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1991) taken. However, the two steps generally encompass (1) identifying if there are likely to be any moderators; and (2) formally testing potential moderators. In the Hunter and Schmidt (1990) approach, the first step is conducted by calculating the percentage of the variance in observed effect sizes that can be attributed to sampling error and statistical artifacts. If sampling error and statistical artifacts can account for 75% or more of the observed variance, they argue that there are unlikely to be any substantive moderators (Hunter & Schmidt, 1990, p. 68; Schmidt, Hunter, & Pearlman, 1980, p. 173). While this approach to detecting the presence of moderators is well established, the second step (formally testing proposed moderators) is less definite. For example, on page 112 of their meta-analysis text, they state:

> A moderator variable will show itself in two ways: (1) the average correlation will vary from subset to subset, and (2) the corrected variance will *average* lower in the subsets than for the data as a whole (emphasis in original).

Many authors appear to use this approach to testing moderators. Hauenstein, McGonigle, and Flinder (2002; p. 46) explicitly state that they used this approach. Other authors (Ones et al., 1993; Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994) appear to be using this approach to identifying moderators, while not explicitly stating so.

An alternative method to testing proposed moderators presented by Hunter and Schmidt (1990) entails comparing the distributions of the effect sizes for the subgroups using a test of statistical significance (pp. 437 – 438; p. 447). This approach has been used by Brown (1996), Riketta (2002), and Russell, Settoon, McGrath, Blanton, Kidwell, Lohrke, Scifres, and Danforth (1994).

Alternatives to the Hunter and Schmidt procedures exist as well. Hedges and Olkin (1985; p. 153) present their Q statistic, which is a test of the homogeneity of observed effect sizes and is based on the chi-square distribution. A statistically significant Q value indicates that the observed effect sizes are sufficiently heterogeneous so as to suggest moderators are present. Proposed moderators are then compared using the Q_B statistic (Hedges & Olkin, 1985, p. 154), which is a between groups comparison of the distributions of observed effect sizes. Aguinis and

Pierce (1998) present an extension of the Hedges and Olkin procedures that compare the distributions of corrected (as opposed to observed) correlations. The Hedges and Olkin (1985) and extensions thereof have been utilized by Stajkovic and Luthans (2003), Webber and Donahue (2001), and Donovan and Radosevich (1998).[6]

A number of studies have compared the tests for homogeneity and moderating effects in terms of Type I (falsely concluding that a moderator is present when in fact it is not) and Type II error (incorrectly concluding that there is no moderator present, when in fact, there is). There are a number of important findings from this research. First, Osburn, Callender, Greener, and Ashworth (1983) found that the power of meta-analysis to detect small to moderate true variance among effect sizes is low when the number of participants per study was below 100. Second, Sackett, Harris, and Orr (1986) found that ***small*** moderating effects are unlikely to be detected, regardless of N and k, and, ***moderate*** differences are unlikely to be detected if N and k are small. Aguinis, Sturman, and Pierce (2002) confirmed these findings, concluding that "Type II error rates are in many conditions quite large" (p. 21). It is also worth pointing out that in the Aguinis et al. (2002) study, small moderating effects were not detected using the tests of the homogeneity of effect sizes, nor were they detected by the more pointed test of potential moderator effects. As such, there is opportunity for a Type II error when a researcher presented with meta-analytic data meeting the homogeneity test chooses not to conduct a moderator test. Yet, there is also opportunity for a Type II error when a researcher chooses to conduct a moderator test, despite evidence of homogeneous effect sizes. Stated more succinctly, in the presence of a small moderating effect, the power of the homogeneity tests are poor, and, the power of the moderating effect tests are also poor.

In addition to the general finding that power to detect moderators is often low, another finding that previous research has converged on is that the Hunter and Schmidt techniques generally perform as well or better than the Q statistics with regard to controlling both Type I and Type II errors (Aguinis et al., 2002; Osburn et al., 1983; Sackett et al., 1986). Because the Hunter and Schmidt procedures are generally the most accurate, their procedures for testing

---

[6] Additionally, some authors recommend the use of credibility intervals (Whitener, 1990) or contrast coefficients (Rosenthal & Dimatteo, 2001) to detect moderators. As these procedures have not been extensively utilized and evaluated in industrial and organizational psychology literature, they are not considered here. Also overlooked here are procedures that test continuous

moderators will be used here. More precisely, the percentage of the observed variance in the overall analyses will be computed. If this percentage is equal to or greater than 75%, it will be concluded that there are no substantive moderators, and the overall estimate of the correlation will be imputed as the population estimate for both incumbents as well as applicants. If the 75% rule is not met, the distributions of the observed correlations will be compared using the following independent samples t-test:

$$t = \frac{\left| r_1 - r_2 \right|}{\sqrt{\frac{Var(r_1)}{k_1} + \frac{Var(r_2)}{k_2}}} \qquad (1)$$

In this equation, r_1 is the sample size weighted average correlation in the first subgroup, r_2 is the sample size weighted average correlation in the second subgroup, $Var(r_1)$ is the observed variance among effect sizes in the first subgroup, $Var(r_2)$ is the observed variance among effect sizes in the second subgroup, k_1 is the number of studies in the first subgroup, k_2 is the number of studies in the second subgroup, and t is evaluated against the critical t-value based on the degrees of freedom determined by the number of studies in the two subgroups being compared.[7] In the current case, the critical value for a two-tailed test (as directional hypotheses were not proffered) with a nominal alpha of 0.10 will be used. If the observed t-value is less than the critical value, it will be concluded that sample type is not a substantive moderator of the applicable correlation, and the overall estimate of the correlation will be imputed as the population estimate for both incumbents as well as applicants.

Given the consistent finding of low power to detect small to moderate moderating effects, it is quite possible that the above tests of moderation will lack power to detect a moderating effect of sample type, if it is present. As such, two sets of simulation analyses will be conducted. The first set of simulation analyses will use the aforementioned rules for identifying moderators and, when evidence of moderation is not obtained, the overall correlation values will be imputed in the incumbent as well as the applicant matrices. The second set of simulations will use the

---

(as opposed to categorical) moderators.

[7] The denominator term presented in the Aguinis et al. (2002) paper is simply $\frac{Var(r_1)}{k_1} + \frac{Var(r_2)}{k_2}$. I have assumed that they inadvertently omitted the square root symbol from the denominator expression.

subgroup correlations for each cell of the matrix, regardless of the evidence for homogeneity of effect sizes or evidence for sample type as a moderator.

Before continuing, a comment regarding small moderating effects is in order. As noted above, power to detect small moderating effects is low in most all meta-analytic conditions (Aguinis et al., 2002; Sackett et al., 1986). Some researchers might contend that detection of small moderating effects is unimportant, both theoretically and practically. Sackett et al. (1986; p. 310) addressed this issue, pointing out that small validity differences can lead to large utility differences under certain selection ratios. For this reason, in the test of hypothesis three, a variety of selection ratios will be examined in order to reveal potential practical effects of potential moderating effects of sample type.

Artifact Distributions

In order to correct observed correlations for measurement error in the performance criteria, criterion reliability artifact distributions were drawn from previous research. Viswesvaran, Ones, and Schmidt (1996) found that the average single-rater reliability of overall job performance ratings across 40 reliability estimates encompassing 14,650 ratees was 0.52 (SD = 0.095). In the current meta-analyses based only on studies using ratings criteria, this artifact distribution was used. Ones et al. (1993) constructed artifact distributions based on previous efforts by Rothstein (1990) and Hunter et al. (1990). Specifically, Ones et al. (1993) combined the mean reliability estimate for production records from the Hunter et al. (1990) study with the mean reliability estimate from Rothstein (1990), weighting each value according to the relative frequency of production records and ratings as performance criteria in the Ones et al. (1993) sample of validation studies. The result was a mean reliability estimate of 0.54 (SD = 0.09). This distribution was used in the current meta-analysis for analyses involving all criteria. The means and standard deviations of the observed reliabilities and the square roots of the reliabilities are reported in Table 1.

Table 1. Criterion Reliability Artifact Distributions

	Mean of Reliability Estimates	Standard Deviation of Reliability Estimates	Mean of the Square Root of Reliability Estimates	Standard Deviation of the Square Root of Reliability Estimates
All criteria	0.54	0.09	0.73	0.05
Ratings criteria	0.52	0.10	0.72	0.06

Utility Analyses

Although cross-validation of regression analyses can provide a statistical index of the accuracy with which present-employee validation studies estimate the validity of job-applicant validation studies, it will be useful to present the results of these analyses using a more practical index, such as the dollar value gain from using personality tests in personnel selection.

To do this, the Brogden-Cronbach-Gleser (BCG) utility formula will be used to estimate the utility gain from using personality inventories in personnel selection. Inserting the multiple R resulting from incumbent studies in the BCG formula (below) will yield the utility estimate anticipated on the basis of a validation study conducted using present employees. Inserting the cross-validation correlation from the generalization of the present-employee regression equation to job-applicant data into the BCG formula will yield the actual utility gain from the use of the incumbent-derived prediction equation. The Brogden-Cronbach-Gleser utility formula is:

$$\Delta U = N_S * T * r_{xy} * SD_y * \lambda/\phi - (N_S * C/\phi) \quad (2)$$

Where:

ΔU = Total utility gain in dollars.

N_S = Number of applicants selected.

T = Expected tenure of selected group, in years.

r_{xy} = R derived above.

SD_y = Standard deviation of job performance, expressed in dollars.

λ = Ordinate of the normal curve at the cut point of the predictor.

ϕ = Selection ratio.

C = Cost of testing a single applicant.

For the purposes of this analysis, selection ratios ranging from 0.10 to 0.90 will be examined (in 0.10 increments). The Schmidt and Hunter (1983) estimation of SD_y as 40% of annual salary was used for these analyses. The median salary for management occupations in the United States, as estimated by the Department of Labor, was used as the estimate of annual salary (United States Department of Labor, 2002).[8] The median annual salary for all management occupations was $70,800 per year and 40% of this value is $28,320. Cost of testing (C) was estimated by taking the average of each test publishers' charges for computer administration and scoring of 100 administrations of the NEO-FFI, the CPI, and the 16PF. Psychological Assessment Resources, Inc. (publisher of the NEO-FFI) offers unlimited computer administration and scoring of the NEO-FFI. The advertised price for this service was divided by 100 to obtain a per administration cost of the NEO-FFI. The Institute for Personality and Ability Testing (publisher of the 16PF) offers their computer administration software free if the user purchases a certain dollar value in Interpretive Reports. The per administration charge for Interpretive Reports when ordered in quantities of 100 was taken as the estimate of each 16PF administration. Similarly, Consulting Psychologists Press (publisher of the CPI) offers online administration and scoring of the CPI with their per administration charge. Across these three sources, the average cost per administration of each personality inventory, assuming 100 administrations, was $9.63. This does not include the cost to the management or human resource professional that must situate and coordinate the test-taker. For the current analysis, it was assumed that an HR professional must dedicate 10 minutes of her or his time to situating each test taker and reviewing each test-taker's interpretive report. This ten minute time estimate is one-half the estimate that Mabon (1998) suggested as a per applicant time commitment for administering, scoring, and interpreting personality inventories (when administered in a group format). Again turning to the Department of Labor salary and wage data, the average hourly wage for an HR assistant (SOC code 43-4161) is $14.17 per hour. Dividing this value by six (assuming the processing of 10 test-takers per hour) results in an additional per administration cost of $2.36, for a total per administration estimate of $11.99. For purposes of simplicity, this value will be rounded up to $12.00. Finally, in the current illustration, tenure will be held constant at one year.

---

[8] Managerial occupations were chosen as they were the modal standard occupational

For each selection ratio investigated, there will be two estimates of dollar value gain from the use of personality inventories in selection. The two utility estimates will then be compared in an effort to provide the magnitude of the potential over-estimation of utility resulting from the use of present-employee samples in validation studies. Support for Hypothesis three will be determined by the difference between the utility estimate based on the incumbent regression equation and the cross-validation index. The results of testing hypotheses two and three are likely to be largely redundant. They are tested independently for two reasons. First, it will be useful to present the results in a more practically meaningful format (dollar value utility gain) to highlight the potential extent to which we are "fooling ourselves" with present-employee validation studies. Second, it is possible that the utility overestimate will differ across selection ratios. Thus, it may be that incumbent prediction models overestimate actual utility gain, but the degree of overestimation is only practically meaningful at certain selection ratios.

---

classification in the studies included in the current meta-analysis.

Chapter Four: Analyses and Results

The first set of meta-analyses reported concern the bivariate correlations between each personality predictor construct and the performance criterion measures (criterion-related validities) and serve as a test of the first Hypothesis. The meta-analytic results for the criterion-related validities of the seven personality predictor constructs are presented in Table 2. In addition, the results of the sample type, scale type, and sample type by scale type moderator analyses are also presented (subgroup analyses were conducted only if there were at least three studies in each subgroup). The first column indicates the pair of variables detailed on each row of the table. This is followed by the number of studies (k) and the total sample size across those k studies. The next two columns present the weighted average correlation and the variance among observed correlations. Next are the average sample size across studies, the sampling error variance, variance attributable to variation in statistical artifacts (measurement error in the criterion), and the percentage of the observed variance that is attributable to sampling error and variation in statistical artifacts. Following these are the columns presenting the corrected variance (variance among operational validities), the operational validity estimates (corrected for measurement error in the criterion), and the standard deviation of the operational validities. The next two columns present one-tailed 90% credibility intervals for the operational validity (90% credibility intervals are derived using the critical t-value for degrees of freedom equal to the number of studies minus one). The final column presents the t-test comparing the distribution of correlations between the line on which the t-test appears and the ensuing line. Bolded t-values indicate a statistically significant moderating effect.

Based on the overall meta-analytic estimates of the seven personality constructs with all indicators of occupational performance, the strongest predictor of performance was Ambition, with an operational validity $\rho = 0.14$ ($SD\rho = 0.14$). The next strongest predictors were Conscientiousness ($\rho = 0.12$, $SD\rho = 0.14$) and Optimism ($\rho = 0.11$, $SD\rho = 0.12$). No other predictors were related to performance stronger than an absolute value of $\rho = 0.08$. In addition, sampling error and variance due to variation in the reliability of criterion measures never accounted for more than 50% of the variance in observed validity estimates. As a result, the $SD\rho$ values were quite large, the credibility intervals spanned a large range, and no predictor exhibited generalizable validity.

Table 2. Meta-analysis results: Criterion-related validities of personality constructs and all performance criteria.

	k	N	$\bar{r}$	σ^2_{OBS}	$\bar{N}$	σ^2_{SE}	σ^2_{ART}
Overall: Neuroticism-Performance	219	51791	-0.06	0.0177	236	0.0042	0.0000
Neuroticism: Incumbents	169	34052	-0.08	0.0220	201	0.0049	0.0000
Neuroticism: Applicants	50	17739	-0.02	0.0066	355	0.0028	0.0000
Neuroticism: Single-stimulus	175	42220	-0.07	0.0191	241	0.0041	0.0000
Neuroticism: Forced-choice	44	9370	-0.02	0.0092	213	0.0047	0.0000
Neuroticism: Incumbents-Single-stimulus	140	25726	-0.11	0.0243	184	0.0053	0.0001
Neuroticism: Applicants-Single-stimulus	35	16494	-0.01	0.0054	471	0.0021	0.0000
Neuroticism: Incumbents-Forced-choice	30	8395	-0.01	0.0072	280	0.0036	0.0000
Neuroticism: Applicants-Forced-choice	14	975	-0.10	0.0188	70	0.0143	0.0000
Overall: Extraversion-Performance	263	68797	0.04	0.0126	262	0.0038	0.0000
Extraversion: Incumbents	220	52047	0.05	0.0147	237	0.0042	0.0000
Extraversion: Applicants	43	16750	0.03	0.0056	390	0.0026	0.0000
Extraversion: Single-stimulus	226	59391	0.04	0.0128	263	0.0038	0.0000
Extraversion: Forced-choice	48	10063	0.07	0.0131	210	0.0047	0.0000
Extraversion: Incumbents-Single-stimulus	186	43389	0.05	0.0159	233	0.0043	0.0000

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	σ^2_{ART}
Extraversion: Applicants-Single-stimulus	40	16002	0.02	0.0038	400	0.0025	0.0000
Extraversion: Incumbents-Forced-choice	44	9442	0.06	0.0129	215	0.0046	0.0000
Extraversion: Applicants-Forced-choice	4	621	0.15	0.0087	155	0.0062	0.0001
Overall: Openness-Performance	108	17686	0.03	0.0123	164	0.0061	0.0000
Openness: Incumbents	93	14479	0.03	0.0115	156	0.0065	0.0000
Openness: Applicants	15	3207	0.05	0.0156	214	0.0047	0.0000
Openness: Single-stimulus	86	15540	0.02	0.0101	181	0.0056	0.0000
Openness: Forced-choice	21	1876	0.07	0.0224	89	0.0112	0.0000
Openness: Incumbents-Single-stimulus	77	13399	0.03	0.0106	174	0.0058	0.0000
Openness: Applicants-Single-stimulus	9	2141	0.01	0.0072	238	0.0042	0.0000
Openness: Incumbents-Forced-choice	16	1080	0.06	0.0212	68	0.0149	0.0000
Openness: Applicants-Forced-choice	5	796	0.08	0.0240	159	0.0062	0.0000
Overall: Agreeableness-Performance	113	27473	0.05	0.0125	243	0.0041	0.0000
Agreeableness: Incumbents	99	24403	0.05	0.0115	246	0.0041	0.0000
Agreeableness: Applicants	14	3070	0.03	0.0203	219	0.0046	0.0000
Agreeableness: Single-stimulus	94	19614	0.08	0.0126	209	0.0048	0.0000
Agreeableness: Forced-choice	18	7589	-0.03	0.0040	422	0.0024	0.0000

	k	N	$\bar{r}$	σ^2_{OBS}	$\bar{N}$	σ^2_{SE}	σ^2_{ART}
Agreeableness: Incumbents-Single-stimulus	84	17397	0.08	0.0112	207	0.0048	0.0000
Agreeableness: Applicants-Single-stimulus	10	2217	0.06	0.0236	222	0.0045	0.0000
Agreeableness: Incumbents-Forced-choice	15	7006	-0.03	0.0039	467	0.0021	0.0000
Agreeableness: Applicants-Forced-choice	3	583	-0.05	0.0039	194	0.0051	0.0000
Overall: Conscientiousness-Performance	266	69148	0.09	0.0147	260	0.0038	0.0000
Conscientiousness: Incumbents	220	53992	0.10	0.0155	245	0.0040	0.0000
Conscientiousness: Applicants	46	15156	0.03	0.0080	329	0.0030	0.0000
Conscientiousness: Single-stimulus	201	57559	0.10	0.0125	286	0.0034	0.0000
Conscientiousness: Forced-choice	70	12046	0.04	0.0219	172	0.0058	0.0000
Conscientiousness: Incumbents-Single-stimulus	172	43861	0.12	0.0130	255	0.0038	0.0001
Conscientiousness: Applicants-Single-stimulus	29	13698	0.03	0.0055	472	0.0021	0.0000
Conscientiousness: Incumbents-Forced-choice	52	10715	0.04	0.0204	206	0.0049	0.0000
Conscientiousness: Applicants-Forced-choice	18	1331	0.04	0.0336	74	0.0137	0.0000
Overall: Optimism-Performance	80	24973	0.08	0.0111	312	0.0032	0.0000
Optimism: Incumbents	63	12951	0.12	0.0152	206	0.0047	0.0001
Optimism: Applicants	17	12022	0.04	0.0036	707	0.0014	0.0000
Optimism: Single-stimulus	77	23826	0.08	0.0115	309	0.0032	0.0000
Optimism: Forced-choice							

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	σ^2_{ART}
Optimism: Incumbents-Single-stimulus	62	12514	0.12	0.0157	202	0.0048	0.0001
Optimism: Applicants-Single-stimulus	15	11312	0.04	0.0034	754	0.0013	0.0000
Optimism: Incumbents-Forced-choice							
Optimism: Applicants-Forced-choice							
Overall: Ambition-Performance	69	14413	0.10	0.0157	209	0.0047	0.0000
Ambition: Incumbents	59	12123	0.11	0.0174	205	0.0048	0.0001
Ambition: Applicants	10	2290	0.06	0.0047	229	0.0044	0.0000
Ambition: Single-stimulus	57	11846	0.10	0.0146	208	0.0047	0.0000
Ambition: Forced-choice	11	2297	0.13	0.02214	209	0.0047	0.0001
Ambition: Incumbents-Single-stimulus	51	10479	0.10	0.0156	205	0.0048	0.0000
Ambition: Applicants-Single-stimulus	6	1367	0.04	0.0029	228	0.0044	0.0000
Ambition: Incumbents-Forced-choice	8	1644	0.15	0.0276	206	0.0047	0.0001
Ambition: Applicants-Forced-choice	3	653	0.06	0.0028	218	0.0046	0.0000

Note: k = number of studies; N = total sample size; $\bar{r}$ = weighted average observed correlation; σ^2_{OBS} = variance in observed correlations; $\overline{N}$ = average study sample size; σ^2_{SE} = variance attributable to sampling error; σ^2_{ART} = variance attributable to variation in statistical artifacts; % σ^2_{OBS} due to SE and Artifacts = percentage of observed variance attributable to sampling error and variation in statistical artifacts; σ^2 = variance in operational validities; ρ_v = operational validity estimate; $SD_{\rho v}$ = standard deviation of operational validity estimate; 90% CV_{LOWER} = Lower limit of 90% credibility interval; 90% CV_{UPPER} = Upper limit of 90% credibility interval; Moderator t-test = t-test of potential moderating effect. Each t-test represents a comparison of the distribution of validity coefficients between the line on which the t-test appears and the ensuing line; t-values in bold reflect statistically significant differences.

	% σ^2_{OBS} due to SE and Artifacts	σ^2	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Neuroticism-Performance	23.94%	0.0252	-0.08	0.16	-0.29	0.12	
Neuroticism: Incumbents	22.47%	0.0321	-0.11	0.18	-0.35	0.12	**3.98**
Neuroticism: Applicants	42.90%	0.0071	-0.03	0.08	-0.14	0.08	
Neuroticism: Single-stimulus	21.71%	0.0280	-0.10	0.17	-0.31	0.12	**3.15**
Neuroticism: Forced-choice	51.22%	0.0084	-0.02	0.09	-0.14	0.10	
Neuroticism: Incumbents-Single-stimulus	22.24%	0.0354	-0.15	0.19	-0.39	0.09	**5.32**
Neuroticism: Applicants-Single-stimulus	39.57%	0.0061	-0.02	0.08	-0.12	0.08	
Neuroticism: Incumbents-Forced-choice	50.05%	0.0067	-0.01	0.08	-0.11	0.10	**2.37**
Neuroticism: Applicants-Forced-choice	76.02%	0.0085	-0.14	0.09	-0.26	-0.01	
Overall: Extraversion-Performance	30.37%	0.0165	0.06	0.13	-0.10	0.23	
Extraversion: Incumbents	28.74%	0.0197	0.07	0.14	-0.11	0.25	**1.75**
Extraversion: Applicants	46.22%	0.0056	0.04	0.07	-0.06	0.13	
Extraversion: Single-stimulus	29.70%	0.0169	0.06	0.13	-0.11	0.22	1.40
Extraversion: Forced-choice	36.38%	0.0157	0.09	0.13	-0.07	0.25	
Extraversion: Incumbents-Single-stimulus	27.10%	0.0217	0.07	0.15	-0.12	0.26	**2.55**

	% σ^2_{OBS} due to SE and Artifacts	σ^2	ρ_v	$SD\rho_v$	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Extraversion: Applicants-Single-stimulus	65.34%	0.0025	0.02	0.05	-0.04	0.09	
Extraversion: Incumbents-Forced-choice	36.25%	0.0154	0.08	0.12	-0.08	0.25	**1.88**
Extraversion: Applicants-Forced-choice	72.36%	0.0045	0.21	0.07	0.10	0.32	
Overall: Openness-Performance	49.86%	0.0116	0.04	0.11	-0.09	0.18	
Openness: Incumbents	56.21%	0.0094	0.04	0.10	-0.09	0.16	0.63
Openness: Applicants	29.99%	0.0205	0.07	0.14	-0.12	0.26	
Openness: Single-stimulus	54.84%	0.0086	0.03	0.09	-0.09	0.15	1.34
Openness: Forced-choice	50.11%	0.0210	0.10	0.14	-0.10	0.29	
Openness: Incumbents-Single-stimulus	54.51%	0.0090	0.04	0.10	-0.09	0.16	0.43
Openness: Applicants-Single-stimulus	59.01%	0.0055	0.02	0.07	-0.09	0.12	
Openness: Incumbents-Forced-choice	70.45%	0.0118	0.09	0.11	-0.06	0.23	0.18
Openness: Applicants-Forced-choice	26.19%	0.0332	0.11	0.18	-0.17	0.39	
Overall: Agreeableness-Performance	32.96%	0.0157	0.06	0.13	-0.10	0.22	
Agreeableness: Incumbents	35.40%	0.0139	0.06	0.12	-0.09	0.22	0.49
Agreeableness: Applicants	22.54%	0.0295	0.04	0.17	-0.19	0.27	
Agreeableness: Single-stimulus	37.94%	0.0147	0.10	0.12	-0.05	0.26	**5.73**
Agreeableness: Forced-choice	59.98%	0.0030	-0.04	0.05	-0.12	0.03	

	% σ^2_{OBS} due to SE and Artifacts	σ^2	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Agreeableness: Incumbents-Single-stimulus	43.15%	0.0119	0.11	0.11	-0.03	0.25	0.38
Agreeableness: Applicants-Single-stimulus	19.12%	0.0358	0.08	0.19	-0.18	0.34	
Agreeableness: Incumbents-Forced-choice	54.65%	0.0033	-0.04	0.06	-0.12	0.04	0.62
Agreeableness: Applicants-Forced-choice	100.00%	0.0000	-0.07	0.00	-0.07	-0.07	
Overall: Conscientiousness-Performance	26.14%	0.0204	0.12	0.14	-0.07	0.30	
Conscientiousness: Incumbents	26.19%	0.0215	0.14	0.15	-0.05	0.33	**4.50**
Conscientiousness: Applicants	38.15%	0.0093	0.04	0.10	-0.08	0.17	
Conscientiousness: Single-stimulus	27.86%	0.0169	0.13	0.13	-0.04	0.30	**2.85**
Conscientiousness: Forced-choice	26.65%	0.0301	0.06	0.17	-0.17	0.28	
Conscientiousness: Incumbents-Single-stimulus	30.00%	0.0171	0.16	0.13	-0.01	0.33	**5.20**
Conscientiousness: Applicants-Single-stimulus	38.54%	0.0063	0.04	0.08	-0.06	0.15	
Conscientiousness: Incumbents-Forced-choice	23.83%	0.0292	0.06	0.17	-0.16	0.28	0.11
Conscientiousness: Applicants-Forced-choice	40.69%	0.0374	0.05	0.19	-0.21	0.31	
Overall: Optimism-Performance	28.91%	0.0148	0.11	0.12	-0.04	0.27	
Optimism: Incumbents	31.79%	0.0194	0.16	0.14	-0.02	0.34	**3.63**
Optimism: Applicants	39.41%	0.0041	0.06	0.06	-0.03	0.14	
Optimism: Single-stimulus	28.10%	0.0155	0.11	0.12	-0.05	0.27	
Optimism: Forced-choice							

	% σ^2_{OBS} due to SE and Artifacts	σ^2	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Optimism: Incumbents-Single-stimulus	31.29%	0.0202	0.16	0.14	-0.02	0.35	**3.70**
Optimism: Applicants-Single-stimulus	38.96%	0.0039	0.05	0.06	-0.03	0.14	
Optimism: Incumbents-Forced-choice							
Optimism: Applicants-Forced-choice							
Overall: Ambition-Performance	30.30%	0.0205	0.14	0.14	-0.05	0.33	
Ambition: Incumbents	27.72%	0.0236	0.15	0.15	-0.05	0.35	**1.74**
Ambition: Applicants	93.69%	0.0006	0.08	0.02	0.05	0.12	
Ambition: Single-stimulus	32.84%	0.0184	0.13	0.14	-0.05	0.31	0.68
Ambition: Forced-choice	21.37%	0.0327	0.18	0.18	-0.07	0.42	
Ambition: Incumbents-Single-stimulus	31.02%	0.0202	0.14	0.14	-0.04	0.33	**2.40**
Ambition: Applicants-Single-stimulus	100.00%	0.0000	0.05	0.00	0.05	0.05	
Ambition: Incumbents-Forced-choice	17.28%	0.0429	0.21	0.21	-0.08	0.50	1.33
Ambition: Applicants-Forced-choice	100.00%	0.0000	0.09	0.00	0.09	0.09	

In comparison to previous meta-analyses of the criterion-related validity of the Big Five, the results were generally similar to Barrick and Mount (1991), Hurtz and Donovan (2000), and Salgado (1997). Table 3 presents the weighted average (observed) validity estimates for each of the big five personality factors from the current as well as these three earlier investigations. In every study, the weighted average validity of Openness to Experience is less than 0.05. Every study has found Conscientiousness to be the strongest predictor of performance among the big five constructs, ranging from a low of 0.09 in the current study to a high of 0.14 in Hurtz and Donovan (2000). The meta-analytic observed validity estimates for Extraversion have been consistent across studies, with a low in the current study (estimated observed validity $\bar{r}$ = 0.04) and a high in the Barrick and Mount (1991) study (estimated $\bar{r}$ = 0.08).

Table 3. Comparison of Overall Observed Validities from Four Meta-Analyses

Personality Construct	Current Study $\bar{r}$	Barrick and Mount (1991) $\bar{r}$	Salgado (1997) $\bar{r}$	Hurtz and Donovan (2000) $\bar{r}$
Neuroticism	-0.06	-0.05	-0.09	-0.09
Extraversion	0.04	0.08	0.05	0.06
Openness	0.03	0.03	0.04	0.04
Agreeableness	0.05	0.04	0.01	0.07
Conscientiousness	0.09	0.13	0.10	0.14

Note: $\bar{r}$ = Weighted average observed correlation. Emotional Stability validity estimates from Barrick and Mount (1991), Salgado (1997), and Hurtz and Donovan (2000) have been reflected here and reported as Neuroticism.

Three of the four meta-analyses found Neuroticism to be the second strongest predictor of job performance, with meta-analytic observed validities of $\bar{r}$ = -0.06 (the current study), $\bar{r}$ = -0.09 (Hurtz & Donovan, 2000; Salgado, 1997). The widest range across the four meta-analyses discussed here involves the validity of Agreeableness. Hurtz and Donovan (2000) found the observed validity of Agreeableness measures to be 0.07; this value is seven times larger than the corresponding estimate from Salgado (1997), and is almost two times larger than the corresponding estimate from Barrick and Mount (1991). Considering the differences in inclusion criteria and coding systems across studies, and further taking into account the range of the standard deviations of the meta-analytic observed validities ***within*** each meta-analysis, (e.g.,

current study range: 0.11 to 0.13; Hurtz & Donovan range: 0.09 to 0.13), the differences in the mean observed validities *across* meta-analyses seem quite small.

The most notable discrepancy between the current analyses and previous efforts is that in Barrick and Mount (1991), Salgado (1997), and Hurtz and Donovan (2000), Conscientiousness was found to exhibit generalizable validity across settings. In the current study, such evidence was not observed. The likely explanation for this difference from previous research lies in a number of small differences between this study and previous efforts. First, the current study was less restrictive in terms of exclusion criteria. Hurtz and Donovan (2000) included only personality inventories explicitly designed to measure the big five. Salgado (1997) included only studies conducted in the European Community. The current study included all inventories in the Hough and Ones (2001) taxonomy, and included studies regardless of geographic location. Note that the magnitude of the variance of observed validity estimates is nearly always larger in the current study than in previous studies (four of five comparisons against Hurtz and Donovan, 2000; four of five comparisons against Salgado, 1997). Second, the average sample size per study was larger in the current meta-analysis than in these previous studies. As a result, less variance is attributable to sampling error in the current meta-analytic findings.

Potential moderators of the criterion-related validity estimates were examined next. First, sample type, scale type, and a hierarchical analysis involving sample type by scale type was conducted including measures of Neuroticism. Both sample type and scale type were identified as moderators of the validity of Neuroticism measures according to a statistically significant t-value comparing the subgroup distributions of observed validity estimates. The operational validity of Neuroticism measures was stronger in incumbent (ρ_v = -0.11, SDρ_v = 0.18) as opposed to applicant samples (ρ_v = -0.03, SDρ_v = 0.08). And, the operational validity of single-stimulus measures (ρ_v = -0.10, SDρ_v = 0.17) was stronger than that of forced-choice measures (ρ_v = -0.02, SDρ_v = 0.09). However, the hierarchical moderator analysis results reveal that Neuroticism criterion-related validity estimates were jointly influenced by sample type and scale type. Single-stimulus measures were related to performance in incumbent (ρ_v = -0.15, SDρ_v = 0.19), but not applicant (ρ_v = -0.02, SDρ_v = 0.08) samples. Yet the opposite was true for forced-choice measures: forced-choice measures exhibited criterion-related validity in applicant (ρ_v = -0.14, SDρ_v = 0.09), but not incumbent (ρ_v = -0.01, SDρ_v = 0.08) samples. Finally, only the criterion-related validity of forced-choice measures in applicant samples yielded generalizable

validity (upper credibility limit = -0.01) with no apparent further moderators (76% of observed variance attributable to sampling error and variability in criterion measurement error).

The subgroup analyses for Extraversion revealed that incumbent validity estimates (ρ_v = 0.07, SDρ_v = 0.14) were slightly larger than applicant validity estimates (ρ_v = 0.04, SDρ_v = 0.07). The difference between forced-choice and single-stimulus measures was statistically insignificant. Single-stimulus measures were only weakly related to performance, and the magnitude of this relationship was slightly stronger in incumbent (ρ_v = 0.07, SDρ_v = 0.15) samples (applicant ρ_v = 0.02, SDρ_v = 0.05). Forced-choice measures were more strongly related to performance in applicant (ρ_v = 0.21, SDρ_v = 0.07) as opposed to incumbent samples (ρ_v = 0.08, SDρ_v = 0.12). Finally, only the criterion-related validity of forced-choice measures in applicant samples yielded generalizable validity (lower credibility limit = 0.10).

Meta-analyses of the subgroup distribution of Openness validity estimates indicate that there was a small and statistically insignificant difference between the incumbent and applicant validity estimates. Forced-choice measures were more strongly related to performance (ρ_v = 0.10, SDρ_v = 0.14) than were single-stimulus measures (ρ_v = 0.03, SDρ_v = 0.09), but this difference was not statistically significant. Within types of measures, there were only small, statistically insignificant differences between incumbents and applicants. No subgroup distribution exhibited generalizable validity for Openness measures.

Meta-analyses of the subgroup distribution of Agreeableness validity estimates indicate that there was a small and statistically insignificant difference between the incumbent and applicant validity estimates. Single-stimulus measures were more strongly related to performance (ρ_v = 0.10, SDρ_v = 0.12) than were forced-choice measures (ρ_v = -0.04, SDρ_v = 0.05). Within the specific types of measures (forced-choice and single-stimulus), the differences between incumbents and applicants were small and not statistically significantly different. The distribution of validity estimates for forced-choice measures in applicant samples suggested a generalizable operational validity estimate of ρ = -0.07 (SDρ_v = 0.00).

Conscientiousness was more strongly related to performance in incumbent (ρ_v = 0.14 SDρ_v = 0.15) as opposed to applicant samples (ρ_v = 0.04, SDρ_v = 0.10), and single-stimulus measures (ρ_v = 0.13 SDρ_v = 0.13) were stronger predictors than were forced-choice measures (ρ_v = 0.06, SDρ_v = 0.17). Furthermore, within single-stimulus measures, incumbent samples

exhibited higher validity estimates ($\rho_v = 0.16$, $SD\rho_v = 0.13$) than did applicant samples ($\rho_v = 0.04$, $SD\rho_v = 0.08$). Within forced-choice instruments, there was a small and statistically insignificant difference between the incumbent and applicant validity estimates, with neither estimate being practically meaningful. There was no generalizable validity evidence within any subgroup for Conscientiousness measures.

The validity of Optimism measures was also more strongly related to performance in incumbent ($\rho_v = 0.16$ $SD\rho_v = 0.14$) as opposed to applicant samples ($\rho_v = 0.06$, $SD\rho_v = 0.06$). Single-stimulus and forced-choice measures of Optimism were not compared due to an insufficient number of studies utilizing forced-choice measures of Optimism. Within single-stimulus measures, incumbent samples exhibited higher criterion-related validity estimates ($\rho_v = 0.16$ $SD\rho_v = 0.14$) than did applicant samples ($\rho_v = 0.05$, $SD\rho_v = 0.06$). No subgroup demonstrated generalizable validity evidence for Optimism measures.

Finally, Ambition measures were more strongly related to performance in incumbent samples ($\rho_v = 0.\ 15$, $SD\rho_v = 0.15$; applicant $\rho_v = 0.\ 08$, $SD\rho_v = 0.02$). The difference between forced-choice and single-stimulus measures was not statistically significant, but suggested that forced-choice measures are more strongly related to performance. Among single-stimulus measures, incumbent samples indicated stronger validity estimates ($\rho_v = 0.\ 14$, $SD\rho_v = 0.14$) than did applicant samples ($\rho_v = 0.\ 05$, $SD\rho_v = 0.00$). The $SD\rho_v = 0.00$ indicates that the validity of Ambition measures in applicant samples generalizes, and is estimated as being $\rho_v = 0.\ 05$ in all settings. Among forced-choice measures, the difference between incumbents and samples was not statistically significant, but suggested that incumbent estimates were higher than applicant estimates. Moreover, there was generalizable evidence of validity for forced-choice measures of Ambition in applicant samples ($\rho_v = 0.\ 09$, $SD\rho_v = 0.00$).

Summarizing the results, there were a total of thirteen tests (seven constructs by two scale types, excepting forced-choice measures of Optimism) of sample type as a moderator of criterion-related validity estimates. Seven of these thirteen tests revealed a statistically significant moderating effect of sample type. In five of the seven cases, the validity estimate was higher in incumbent samples, while in two cases the applicant estimate exceeded the incumbent estimate. It is interesting to note that the five cases of heightened incumbent validity estimates occurred with single-stimulus measures, whereas the two cases of enhanced applicant validities were

realized using forced-choice scales.

A problematic concern is that the differences that are seemingly due to sample type might actually be due to a possible confound between sample type and *performance criterion.* One possibility is that applicant studies might be more likely to utilize a criterion that is less a function of dispositional characteristics than a function of situational characteristics. For instance, applicant studies might be more likely than present-employee studies to use performance during training as a criterion variable. This would allow the researcher to collect criterion data without allowing a substantial time period to elapse after the personality data was gathered. Helmreich and his colleagues (Helmreich et al., 1988) provided evidence that personality is less likely to influence performance early in one's tenure with an organization because there exists for most individuals, a "Honeymoon Effect", during which they put forth maximal effort. This is a situational, as opposed to a dispositional determinant of performance. This raises the possibility personality would not predict training performance so well as it predicts performance on the job. The alternative is also possible, though. Barrick and Mount (1991) found that Extraversion and Openness to Experience were better predictors of training (as opposed to on the job) performance.

In order to control for the possibility that differences between applicants and incumbents could be due to differences in the criteria used in those studies, studies using criteria other than a subjective rating criteria (peer, supervisor, subordinate, or client ratings, rankings, or dichotomous "effectiveness" classifications) were eliminated. Results of this analysis are presented in Table 4, including the hierarchical moderator breakdown for all conditions with at least three studies.

When only studies using a ratings criterion are included, there are some differences from the analysis of studies using all criteria. First, focusing on the validity estimates for all samples (regardless of sample type or scale type) using a ratings criterion, Optimism was the strongest predictor of performance, with an operational validity $\rho = 0.15$ ($SD\rho = 0.10$). The next strongest predictors were Conscientiousness ($\rho = 0.11$, $SD\rho = 0.14$) and Ambition ($\rho = 0.11$, $SD\rho = 0.15$). The remaining predictors had overall validity estimates with absolute values less than 0.10. In addition, sampling error and variance due to variation in the reliability of criterion measures never accounted for more than 55% of the variance in observed validity estimates. As a result, the $SD\rho$ values were quite large, the credibility intervals spanned a large range, and only

Optimism exhibited generalizable validity (lower credibility limit = 0.02).

The moderator analyses for measures of Neuroticism revealed that there was a small statistically insignificant difference between incumbent and applicant samples. Single-stimulus measures were more strongly related to performance (ρ = -0.11, SDρ = 0.15) than were forced-choice measures (ρ = -0.02, SDρ = 0.06). Within single-stimulus measures, incumbent samples yielded higher validity estimates (ρ = -0.12, SDρ = 0.16) than applicant samples (ρ = -0.05, SDρ = 0.10), while the converse was true for forced-choice measures: incumbent estimates were lower (ρ = -0.01, SDρ = 0.06) than applicant estimates (ρ = -0.09, SDρ = 0.00) for forced-choice measures. The pattern of results for Neuroticism again suggests that single-stimulus measures yield higher levels of criterion-related validity in incumbent samples (as compared to applicant samples), while applicant validity estimates are higher than incumbent estimates for forced-choice measures.

Neither sample type nor scale type was identified as a moderator of the validity of Extraversion measures. Within scale type, however, the apparent interaction between sample type and scale type again emerged. Single-stimulus measures yielded higher criterion-related validity estimates in incumbent samples (ρ = 0.08, SDρ = 0.14) as compared to applicant samples (ρ = 0.01, SDρ = 0.06). Forced-choice measures had a low criterion-related validity estimate in incumbent samples (ρ = 0.05, SDρ = 0.06), with a more useful estimate derived from applicant samples (ρ = 0.21, SDρ = 0.07). It should be noted that the applicant estimate for forced-choice measures is based on only four studies, and one study (Saville, Sik, Nyfield, Hackston, & MacIver, 1996) accounts for 70% of the total N. The joint influence of sample type and scale type on the validity estimates for Neuroticism and Extraversion are presented in Figure 1 (panels a and b). Note that in Figure 1 the validities for Neuroticism have been reflected for clarity of presentation, such that positive correlations would indicate that emotionally stable individuals exhibit higher-quality performance.

Table 4. Meta-analysis results: Criterion-related validities of personality constructs with performance ratings criteria.

	k	N	$\bar{r}$	σ^2_{OBS}	$\bar{N}$	σ^2_{SE}	σ^2_{ART}
Overall: Neuroticism-Performance	174	32161	-0.06	0.0149	185	0.00540	0.0000
Neuroticism: Incumbents	135	26528	-0.06	0.0158	197	0.00507	0.0000
Neuroticism: Applicants	39	5633	-0.05	0.0106	144	0.0069	0.0000
Neuroticism: Single-stimulus	133	23199	-0.08	0.0172	174	0.0057	0.0000
Neuroticism: Forced-choice	41	8761	-0.02	0.0064	214	0.0047	0.0000
Neuroticism: Incumbents-Single-stimulus	108	18730	-0.09	0.0183	173	0.0057	0.0001
Neuroticism: Applicants-Single-stimulus	25	4469	-0.04	0.0109	179	0.0056	0.0000
Neuroticism: Incumbents-Forced-choice	28	7867	-0.01	0.0057	281	0.0036	0.0000
Neuroticism: Applicants-Forced-choice	13	894	-0.06	0.0099	69	0.0146	0.0000
Overall: Extraversion-Performance	214	38513	0.05	0.0134	180	0.0056	0.0000
Extraversion: Incumbents	182	33855	0.05	0.0133	186	0.0054	0.0000
Extraversion: Applicants	32	4658	0.04	0.0141	146	0.0069	0.0000
Extraversion: Single-stimulus	180	29726	0.05	0.0150	165	0.0061	0.0000
Extraversion: Forced-choice	37	8784	0.04	0.0073	237	0.0042	0.0000
Extraversion: Incumbents-Single-stimulus	151	25816	0.06	0.0156	171	0.0058	0.0000

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	σ^2_{ART}
Extraversion: Applicants-Single-stimulus	29	3910	0.01	0.0091	135	0.0075	0.0000
Extraversion: Incumbents-Forced-choice	33	8163	0.04	0.0062	247	0.0040	0.0000
Extraversion: Applicants-Forced-choice	4	621	0.15	0.0088	155	0.0062	0.0002
Overall: Openness-Performance	87	14553	0.04	0.0111	167	0.0060	0.0000
Openness: Incumbents	76	12137	0.03	0.0101	160	0.0063	0.0000
Openness: Applicants	11	2416	0.06	0.0156	220	0.0045	0.0000
Openness: Single-stimulus	74	12889	0.02	0.0092	174	0.0058	0.0000
Openness: Forced-choice	12	1394	0.12	0.0131	116	0.0084	0.0001
Openness: Incumbents-Single-stimulus	67	11326	0.03	0.0094	169	0.0059	0.0000
Openness: Applicants-Single-stimulus	7	1563	0.00	0.0071	223	0.0045	0.0000
Openness: Incumbents-Forced-choice	9	811	0.10	0.0156	90	0.0110	0.0001
Openness: Applicants-Forced-choice	3	583	0.15	0.0081	194	0.0049	0.0002
Overall: Agreeableness-Performance	94	24565	0.05	0.0114	261	0.0038	0.0000
Agreeableness: Incumbents	83	22149	0.06	0.0100	267	0.0037	0.0000
Agreeableness: Applicants	11	2416	0.03	0.0241	220	0.0046	0.0000
Agreeableness: Single-stimulus	83	17107	0.08	0.0130	206	0.0048	0.0000
Agreeableness: Forced-choice	10	7188	0.00	0.0027	719	0.0014	0.0000

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	σ^2_{ART}
Agreeableness: Incumbents-Single-stimulus	76	15544	0.08	0.0113	205	0.0048	0.0000
Agreeableness: Applicants-Single-stimulus	7	1563	0.08	0.0297	223	0.0044	0.0000
Agreeableness: Incumbents-Forced-choice	7	6605	0.00	0.0024	944	0.0011	0.0000
Agreeableness: Applicants-Forced-choice	3	583	-0.05	0.0039	194	0.0051	0.0000
Overall: Conscientiousness-Performance	217	41631	0.08	0.0155	192	0.0052	0.0000
Conscientiousness: Incumbents	180	37320	0.08	0.0155	207	0.0048	0.0000
Conscientiousness: Applicants	37	4311	0.09	0.0159	117	0.0085	0.0001
Conscientiousness: Single-stimulus	166	31356	0.10	0.0146	189	0.0052	0.0001
Conscientiousness: Forced-choice	55	10341	0.04	0.0161	188	0.0053	0.0000
Conscientiousness: Incumbents-Single-stimulus	143	28209	0.09	0.0154	197	0.0050	0.0001
Conscientiousness: Applicants-Single-stimulus	23	3147	0.12	0.0074	137	0.0072	0.0001
Conscientiousness: Incumbents-Forced-choice	40	9304	0.05	0.0142	233	0.0043	0.0000
Conscientiousness: Applicants-Forced-choice	15	1037	0.00	0.0313	69	0.0147	0.0000
Overall: Optimism-Performance	63	10194	0.11	0.0112	162	0.0061	0.0001
Optimism: Incumbents	51	8295	0.10	0.0115	163	0.0061	0.0001
Optimism: Applicants	12	1899	0.13	0.0093	158	0.0062	0.0001
Optimism: Single-stimulus	60	9047	0.11	0.0124	151	0.0065	0.0001
Optimism: Forced-choice							

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	σ^2_{ART}
Optimism: Incumbents-Single-stimulus	50	7858	0.10	0.0121	157	0.0063	0.0001
Optimism: Applicants-Single-stimulus	10	1189	0.14	0.0124	119	0.0081	0.0001
Optimism: Incumbents-Forced-choice							
Optimism: Applicants-Forced-choice							
Overall: Ambition-Performance	60	10681	0.08	0.0168	178	0.0056	0.0000
Ambition: Incumbents	53	8897	0.08	0.0191	168	0.0059	0.0000
Ambition: Applicants	7	1784	0.05	0.0045	255	0.0039	0.0000
Ambition: Single-stimulus	51	8445	0.06	0.0142	166	0.0060	0.0000
Ambition: Forced-choice	8	1966	0.14	0.0240	246	0.0039	0.0001
Ambition: Incumbents-Single-stimulus	46	7371	0.07	0.0157	160	0.0062	0.0000
Ambition: Applicants-Single-stimulus	5	1074	0.01	0.0011	215	0.0047	0.0000
Ambition: Incumbents-Forced-choice	7	1526	0.15	0.0297	218	0.0044	0.0002
Ambition: Applicants-Forced-choice							

Note: k = number of studies; N = total sample size; $\bar{r}$ = weighted average observed correlation; σ^2_{OBS} = variance in observed correlations; $\overline{N}$ = average study sample size; σ^2_{SE} = variance attributable to sampling error; σ^2_{ART} = variance attributable to variation in statistical artifacts; % σ^2_{OBS} due to SE and Artifacts = percentage of observed variance attributable to sampling error and variation in statistical artifacts; σ^2 = variance in operational validities; ρ_v = operational validity estimate; $SD_{\rho v}$ = standard deviation of operational validity estimate; 90% CV_{LOWER} = Lower limit of 90% credibility interval; 90% CV_{UPPER} = Upper limit of 90% credibility interval; Moderator t-test = t-test of potential moderating effect. Each t-test represents a comparison of the distribution of validity coefficients between the line on which the t-test appears and the ensuing line; t-values in bold reflect statistically significant differences.

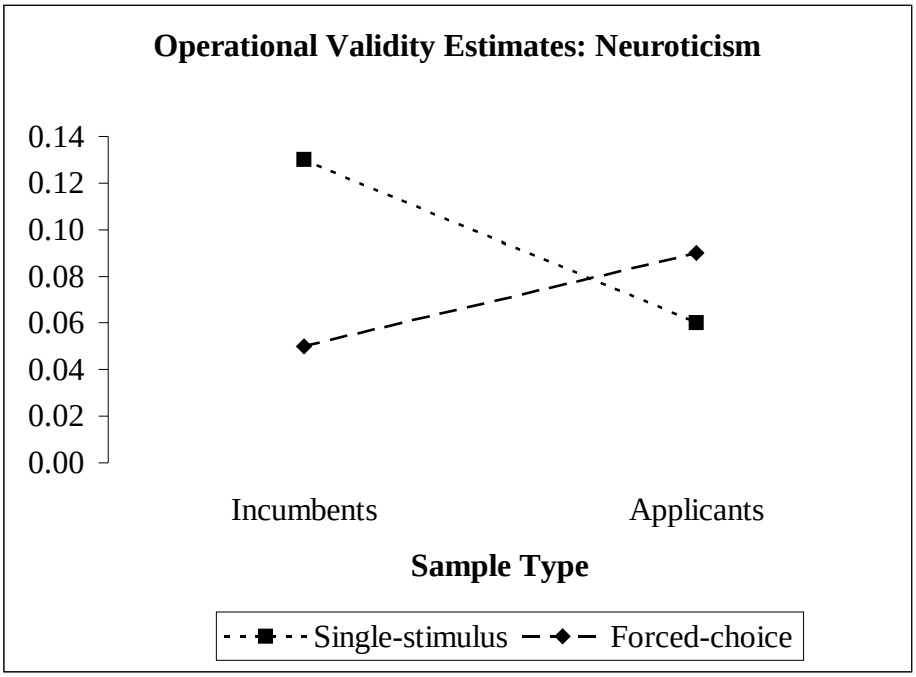
	% σ^2_{OBS} due to SE and Artifacts	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Neuroticism-Performance	36.37%	-0.09	0.14	-0.26	0.09	
Neuroticism: Incumbents	32.34%	-0.09	0.14	-0.27	0.10	0.90
Neuroticism: Applicants	65.57%	-0.06	0.08	-0.17	0.04	
Neuroticism: Single-stimulus	33.28%	-0.11	0.15	-0.30	0.08	**3.68**
Neuroticism: Forced-choice	73.74%	-0.02	0.06	-0.10	0.05	
Neuroticism: Incumbents-Single-stimulus	31.49%	-0.12	0.16	-0.32	0.08	**1.96**
Neuroticism: Applicants-Single-stimulus	51.45%	-0.05	0.10	-0.19	0.08	
Neuroticism: Incumbents-Forced-choice	62.87%	-0.01	0.06	-0.10	0.07	**1.73**
Neuroticism: Applicants-Forced-choice	100.00%	-0.09	0.00	-0.09	-0.09	
Overall: Extraversion-Performance	41.56%	0.07	0.12	-0.09	0.23	
Extraversion: Incumbents	40.50%	0.07	0.12	-0.09	0.23	0.37
Extraversion: Applicants	49.04%	0.06	0.12	-0.09	0.21	
Extraversion: Single-stimulus	40.43%	0.07	0.13	-0.10	0.24	0.31
Extraversion: Forced-choice	57.82%	0.06	0.08	-0.04	0.16	
Extraversion: Incumbents-Single-stimulus	37.52%	0.08	0.14	-0.10	0.26	**2.31**

	% σ^2_{OBS} due to SE and Artifacts	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Extraversion: Applicants-Single-stimulus	82.14%	0.01	0.06	-0.06	0.09	
Extraversion: Incumbents-Forced-choice	65.15%	0.05	0.06	-0.03	0.14	**2.40**
Extraversion: Applicants-Forced-choice	72.21%	0.21	0.07	0.10	0.33	
Overall: Openness-Performance	53.90%	0.05	0.10	-0.08	0.18	
Openness: Incumbents	62.39%	0.04	0.09	-0.07	0.15	0.80
Openness: Applicants	29.23%	0.09	0.15	-0.11	0.29	
Openness: Single-stimulus	62.76%	0.03	0.08	-0.07	0.14	**2.72**
Openness: Forced-choice	64.95%	0.16	0.09	0.03	0.29	
Openness: Incumbents-Single-stimulus	63.45%	0.04	0.08	-0.07	0.14	0.89
Openness: Applicants-Single-stimulus	63.26%	-0.01	0.07	-0.11	0.10	
Openness: Incumbents-Forced-choice	70.88%	0.13	0.09	0.00	0.26	0.78
Openness: Applicants-Forced-choice	62.90%	0.21	0.08	0.06	0.35	
Overall: Agreeableness-Performance	33.58%	0.08	0.12	-0.08	0.23	
Agreeableness: Incumbents	37.64%	0.08	0.11	-0.06	0.22	0.50
Agreeableness: Applicants	18.94%	0.05	0.19	-0.22	0.31	
Agreeableness: Single-stimulus	37.37%	0.11	0.13	-0.05	0.27	**4.06**
Agreeableness: Forced-choice	50.94%	0.00	0.05	-0.08	0.07	

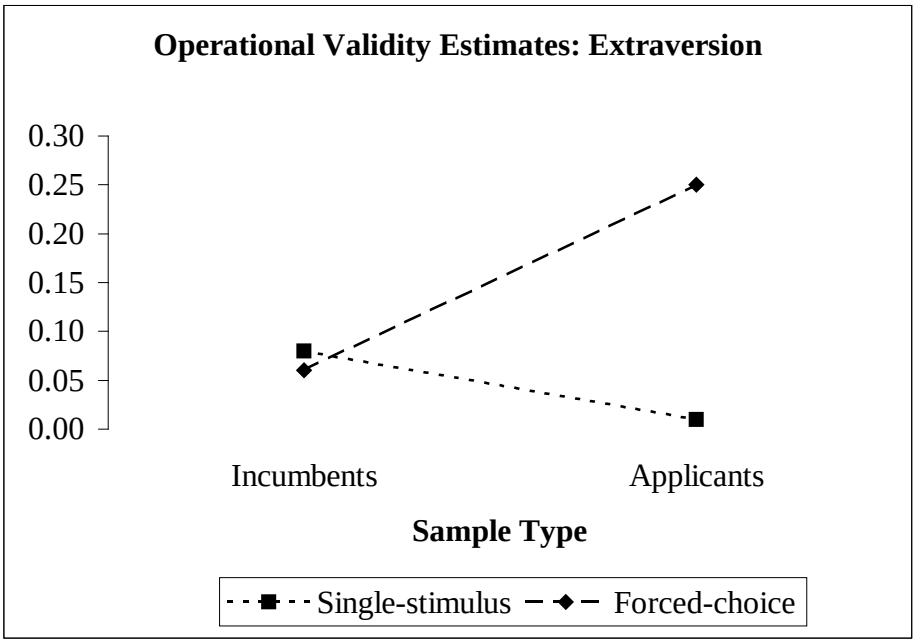
	% σ^2_{OBS} due to SE and Artifacts	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Agreeableness: Incumbents-Single-stimulus	43.23%	0.11	0.11	-0.03	0.26	0.02
Agreeableness: Applicants-Single-stimulus	15.11%	0.11	0.22	-0.21	0.43	
Agreeableness: Incumbents-Forced-choice	44.42%	0.00	0.05	-0.07	0.07	1.37
Agreeableness: Applicants-Forced-choice	100.00%	-0.08	0.00	-0.08	-0.08	
Overall: Conscientiousness-Performance	33.57%	0.11	0.14	-0.07	0.30	
Conscientiousness: Incumbents	31.15%	0.11	0.14	-0.07	0.30	0.19
Conscientiousness: Applicants	54.09%	0.12	0.12	-0.04	0.27	
Conscientiousness: Single-stimulus	36.11%	0.13	0.13	-0.04	0.31	**2.70**
Conscientiousness: Forced-choice	33.23%	0.06	0.14	-0.13	0.25	
Conscientiousness: Incumbents-Single-stimulus	32.93%	0.13	0.14	-0.05	0.31	1.27
Conscientiousness: Applicants-Single-stimulus	98.27%	0.17	0.02	0.14	0.19	
Conscientiousness: Incumbents-Forced-choice	30.39%	0.06	0.14	-0.12	0.24	0.85
Conscientiousness: Applicants-Forced-choice	46.89%	0.01	0.18	-0.23	0.25	
Overall: Optimism-Performance	54.99%	0.15	0.10	0.02	0.27	
Optimism: Incumbents	53.37%	0.14	0.10	0.01	0.27	0.85
Optimism: Applicants	67.15%	0.18	0.08	0.07	0.28	
Optimism: Single-stimulus	53.37%	0.15	0.11	0.01	0.28	
Optimism: Forced-choice						

	% σ^2_{OBS} due to SE and Artifacts	ρ_v	SDρ_v	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Optimism: Incumbents-Single-stimulus	52.35%	0.14	0.11	0.00	0.28	1.15
Optimism: Applicants-Single-stimulus	66.92%	0.20	0.09	0.08	0.32	
Optimism: Incumbents-Forced-choice						
Optimism: Applicants-Forced-choice						
Overall: Ambition-Performance	33.55%	0.11	0.15	-0.08	0.30	
Ambition: Incumbents	31.25%	0.12	0.16	-0.09	0.32	0.95
Ambition: Applicants	87.60%	0.07	0.03	0.03	0.12	
Ambition: Single-stimulus	42.58%	0.09	0.13	-0.08	0.25	1.28
Ambition: Forced-choice	16.91%	0.19	0.20	-0.09	0.47	
Ambition: Incumbents-Single-stimulus	39.84%	0.10	0.13	-0.08	0.27	**2.46**
Ambition: Applicants-Single-stimulus	100.00%	0.02	0.00	0.02	0.02	
Ambition: Incumbents-Forced-choice	15.33%	0.21	0.22	-0.11	0.53	
Ambition: Applicants-Forced-choice						

Figure 1. Operational validity of Neuroticism and Extraversion as a Function of Sample Type and Scale Type.



Panel A: Neuroticism.



Panel B: Extraversion.

Sample type was not identified as a moderator of validity estimates for Openness measures (neither across scale types nor within scale types). Forced-choice measures were identified as stronger predictors of performance (ρ = 0.16, SDρ = 0.09) than were single-stimulus measures (ρ = 0.03, SDρ = 0.08). Although not significantly different, the applicant validity estimate (ρ = 0.21, SDρ = 0.08) for forced-choice measures was higher than the corresponding incumbent estimate (ρ = 0.13, SDρ = 0.09). Again, though, the applicant estimate was based almost solely on the Saville et al. (1996) study.

As with Openness, sample type was not identified as a moderator of validity estimates for Agreeableness measures (neither across scale types nor within scale types). Single-stimulus measures were identified as stronger predictors of performance (ρ = 0.11, SDρ = 0.13) than were forced-choice measures (ρ = 0.00, SDρ = 0.05).

Similarly, sample type was not identified as a moderator of the validity estimates for Conscientiousness measures. Single-stimulus measures were found to be stronger predictors of ratings criteria (ρ = 0.13, SDρ = 0.13) than were forced-choice instruments (ρ = 0.06, SDρ = 0.14). Also noteworthy is the applicant operational validity estimate within single-stimulus measures (ρ = 0.17, SDρ = 0.02) exhibited generalizable validity (98% of the observed variance attributable to sampling error and variation in criterion measurement error). This estimate is based on 23 studies with a total sample size of 3,147.

There were very few studies that utilized forced-choice measures of Optimism as predictors of a ratings criterion. As such, only tests of sample type as a moderator of Optimism validity estimates across all measures and within single-stimulus measures were conducted. Both indicated that sample type was not likely a moderator of the validity of Optimism measures, as the differences between incumbents and applicants were small and not statistically significant. It is again worth noting that the meta-analysis suggests that there is evidence that the validity of Optimism measures in applicant studies of single-stimulus measures generalizes across settings, with a lower credibility limit of 0.08.

Finally, there was some evidence of Ambition measures being more strongly related to performance in incumbent as opposed to applicant samples, but the difference was not statistically significant. Similarly, forced-choice measures yielded higher validity estimates than single-stimulus measures revealed, but again this was not a statistically significant difference. Within single-stimulus measures, incumbent validity estimates (ρ = 0.10, SDρ = 0.13) were

higher than applicant estimates ($\rho = 0.02$, $SD\rho = 0.00$).

When only studies using a ratings criterion are included, twelve tests of sample type as a moderator were conducted, and five were significantly different. Of the five that were significantly different, three indicated stronger validity estimates in incumbent samples and two indicated stronger validity estimates in applicant samples. Again, the stronger applicant estimates occurred for forced-choice measures (Neuroticism and Extraversion) while the stronger incumbent estimates occurred for single-stimulus measures (Neuroticism, Extraversion, and Ambition).

The two noteworthy differences between the analyses based on all criteria versus the analyses based on only ratings criteria are that in the analyses of all criteria, the criterion-related validity of ***single-stimulus*** measures of Conscientiousness and Optimism was stronger in incumbent as compared to applicant samples. When only ratings criteria are included, the moderating effect of sample type disappears for these two predictors. Upon further inspection, it can be seen that the meta-analytic results are strongly influenced by a large-scale study that included an indicator of Conscientiousness and Optimism that failed to predict two non-rating criteria (training performance composite and attainment of full performance level) in a sample of applicants (Oakes et al., 2001). When this study is eliminated from the test of sample type as a moderator of the validity of single-stimulus measures as predictors of all criteria, there is not a significant difference according to sample type for Conscientiousness or Optimism.

The findings from the statistical significance tests of moderation provide mixed support for Hypothesis One. Although the statistical significance tests provide one way to test Hypothesis One, an alternative test of moderation entails examining subgroup operational validities and standard deviations of those validities. More specifically, if the operational validities differ between the subgroups, and the average subgroup SD_ρ is smaller than the overall SD_ρ, than the grouping variable is designated as a moderator of the population parameter estimate. The operational validity estimates and the subgroup SD_ρ values were examined for the studies utilizing a rating criterion.

Considering subgroup differences in operational validity estimates and the average SD_ρ values across subgroups, Neuroticism, when measured with single-stimulus measures, was more strongly related to performance in incumbent as opposed to applicant samples. Similarly, when measured with forced-choice scales, the operational validity of Neuroticism differs across sample

type and the average SD_ρ within subgroups is smaller than the overall SD_ρ.

As with the results from the statistical significance tests, sample type was identified as a moderator of both single-stimulus and forced-choice measures of Extraversion. The operational validity estimates differ across subgroups and the average SD_ρ within subgroups is smaller than the overall SD_ρ.

Based on the examination of subgroup validity estimates and average SD_ρ values within subgroups, sample type was found to moderate the validity of both single-stimulus and forced-choice measures of Openness. Single-stimulus measures of Openness were weakly related to performance in incumbent ($\rho = 0.04$, $SD_\rho = 0.08$) as well as applicant samples ($\rho = -0.01$, $SD_\rho = 0.07$). Forced-choice measures of Openness were related to performance in incumbent ($\rho = 0.13$, $SD_\rho = 0.09$) as well as applicant samples ($\rho = 0.21$, $SD_\rho = 0.08$).

The incumbent operational validity estimate for Agreeableness ($\rho = 0.11$, $SD_\rho = 0.11$) was equal to that of applicants ($\rho = 0.11$, $SD_\rho = 0.22$) when Agreeableness was measured with single-stimulus inventories. This does not pass the test of moderation, as the subgroup validity estimates did not differ, and the average subgroup SD_ρ was larger than the overall SD_ρ. When forced-choice measures were used, the operational validity of Agreeableness was moderated by sample type: the incumbent operational validity $\rho = 0.00$ ($SD_\rho = 0.05$), and the applicant operational validity $\rho = -0.08$ ($SD_\rho = 0.00$).

Single-stimulus measures of conscientiousness were found to be moderated by sample type. When Conscientiousness was measured with single-stimulus inventories, the operational validity was stronger in applicant ($\rho = 0.17$, $SD_\rho = 0.02$) as opposed to incumbent ($\rho = 0.13$, $SD_\rho = 0.14$) samples. When Conscientiousness was operationally defined with forced-choice measures, sample type was not identified as a moderator of validity because the average of the subgroup SD_ρ values was larger than the SD_ρ value for all forced-choice measures of Conscientiousness.

The criterion-related validity of single-stimulus measures of Optimism was also moderated by sample type. Specifically, the operational validity of Optimism was slightly stronger in applicant ($\rho = 0.20$, $SD_\rho = 0.09$) as opposed to incumbent ($\rho = 0.14$, $SD_\rho = 0.11$) samples. Due to insufficient extant validation of forced-choice measures of Optimism, it was not possible to examine potential subgroup differences.

Finally, the criterion-related validity of single-stimulus measures of Ambition was moderated by sample type. The operational validity of Ambition was stronger in incumbent (ρ = 0.10, SD_ρ = 0.13) as opposed to applicant (ρ = 0.02, SD_ρ = 0.00) samples. As with Optimism, it was not possible to examine subgroup differences in the validity of forced-choice measures of Ambition.

The two different methods of testing sample type as a moderator of the criterion-related validity estimates of personality measures reached the same conclusion on seven of 12 tests. On five tests of moderation different conclusions were reached. The five comparisons that arrived at different conclusions depending on the method used to test for moderating effects were: single-stimulus and forced-choice measures of Openness; forced-choice measures of Agreeableness; single-stimulus measures of Conscientiousness; and, single-stimulus measures of Optimism. The reason the two methods led to different conclusions generally seemed to be that the magnitude of the differences was small and the number of studies was small. Despite the fact that examination of the operational validity estimates and average within group SD_ρ values would lead to the conclusion that sample type moderates the validity of Openness measures (both single-stimulus and forced-choice), such a conclusion should be tempered by the fact that the validity estimates for single-stimulus measures were practically zero in both types of samples (incumbents and applicants). And, the estimates for forced-choice measures, while indicating that Openness is useful as a predictor of performance ratings, were based on relatively few studies and total sample sizes (as indicated above, the Openness meta-analytic validity estimate for applicants is based primarily on the Saville et al., 1996 study).

The validity of forced-choice measures of Agreeableness was found to be moderated by sample type when relying on subgroup validity estimates and average subgroup SD_ρ values. This difference was also not very meaningful, as the absolute value of the operational validity estimate was less than 0.10 in each subgroup. Next, the validity of single-stimulus measures of Conscientiousness was found to be moderated by sample type when relying on subgroup validity estimates and average subgroup SD_ρ values. This difference was small (incumbent ρ = 0.13, SD_ρ = 0.14; applicant ρ = 0.17, SD_ρ = 0.02). Finally, the validity of single-stimulus measures of Optimism was found to be moderated by sample type when relying on subgroup validity estimates and average subgroup SD_ρ values. This difference was also relatively small (incumbent ρ = 0.14, SD_ρ = 0.11; applicant ρ = 0.20, SD_ρ = 0.09).

As noted above, tests of potential moderators in meta-analysis generally lack power to detect small moderating effects. The differences between incumbent and applicant samples for single-stimulus measures of Conscientiousness and Optimism were small, and that is one reason that the moderating effect of sample type was not revealed by the t-test comparing the subgroup validity distributions. Based on the differences between the validity estimates and the average subgroup SD_ρ values, sample type is revealed as a moderator. From a practical standpoint, it is worth noting that the incumbent validity estimates for single-stimulus measures of Conscientiousness and Optimism were *lower* than the corresponding applicant validity estimates. To the extent that sample type moderates the validity of single-stimulus measures of Conscientiousness and Optimism, it appears that higher degrees of criterion-related validity will be found in applicant settings.

Meta-analyses of Correlations Among Personality Constructs

Next, analyses of the correlations among personality constructs were undertaken. Because there were very few studies reporting correlations between personality constructs that utilized forced-choice measures, and because forced-choice measures can lead to biased estimates of the correlations among personality constructs (Baron, 1996), only single-stimulus measures were included in these analyses.

Results of the meta-analyses of correlations between personality constructs are presented in Table 5. The first column indicates the pair of variables detailed on each row of the table. This is followed by the number of studies (k) and the total sample size across those k studies. The next two columns present the weighted average correlation and the variance among observed correlations. Next is the average sample size across studies, the sampling error variance, and the percentage of the observed variance that is attributable to sampling error. Following these are the columns presenting the corrected variance and the standard deviation of the weighted average correlations. The next two columns present one-tailed 90% credibility intervals for the operational validity (90% credibility intervals are derived using the critical t-value for degrees of freedom equal to the number of studies minus one). The final column presents the t-test comparing the distribution of correlations in incumbent samples with those from applicant samples. Bolded t-values indicate a statistically significant moderating effect.

As with the initial criterion-related validity estimates, it is worth noting that even after

correcting for sampling error, there is a great deal of variability in many of the correlations. The most dramatic example of this among the overall correlations is the correlation between Extraversion and Conscientiousness. The correlation between these two constructs is estimated as +0.19, and the standard deviation of the corrected correlation (SDρ) is 0.23. The resulting one-tailed 90% credibility interval ranges from -0.11 to 0.48. While many of the correlations between constructs exhibit generalizable correlations (that is, credibility intervals do not include zero), none of the twenty-one meta-analyses of overall correlations result in 75% or more of the variance in observed correlations being attributable to sampling error (the maximum percentage of observed variance that could be attributed to sampling error was 40%). Hence, all correlations appear to be moderated by some substantive factors. First, sample type was examined as a moderator of the meta-analytic correlations (subject to the requirement that there must be at least three studies from each sample type contributing correlations to the moderator analysis). These results also appear in Table 5.

Based on a t-test comparing the distributions of observed correlations, there is evidence of sample type acting as a moderator of the inter-correlations among personality constructs for eight of the 21 personality predictor pairs: Neuroticism-Openness; Neuroticism-Conscientiousness; Extraversion-Openness; Extraversion-Optimism; Openness-Conscientiousness; Openness-Optimism; Agreeableness-Optimism; and Conscientiousness-Ambition. Of the eight correlations that were identified as being moderated by sample type, the correlations were stronger in applicant samples in five instances.

Table 5. Meta-analysis Results for Correlations Between Predictors

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% $\sigma^2_{OBS:}$ SE
Overall: Neuroticism-Extraversion	139	68313	-0.28	0.0300	491	0.0017	5.80%
Incumbents: Neuroticism-Extraversion	104	21473	-0.28	0.0260	206	0.0042	15.97%
Applicants: Neuroticism-Extraversion	35	46840	-0.28	0.0318	1338	0.0006	2.01%
Overall: Neuroticism-Openness	89	28841	-0.15	0.0133	324	0.0030	22.22%
Incumbents: Neuroticism-Openness	75	14521	-0.13	0.0190	194	0.0050	26.45%
Applicants: Neuroticism-Openness	14	14320	-0.18	0.0063	1023	0.0009	14.48%
Overall: Neuroticism-Agreeableness	92	30470	-0.30	0.0257	331	0.0025	9.78%
Incumbents: Neuroticism-Agreeableness	77	15560	-0.27	0.0234	202	0.0043	18.29%
Applicants: Neuroticism-Agreeableness	15	14910	-0.33	0.0262	994	0.0008	3.05%
Overall: Neuroticism-Conscientiousness	118	62139	-0.46	0.0260	527	0.0012	4.53%
Incumbents: Neuroticism-Conscientiousness	92	18277	-0.37	0.0294	199	0.0038	12.86%
Applicants: Neuroticism-Conscientiousness	26	43862	-0.50	0.0193	1687	0.0003	1.72%
Overall: Neuroticism-Optimism	33	33357	-0.47	0.0133	1011	0.0006	4.55%
Incumbents: Neuroticism-Optimism	21	3574	-0.47	0.0429	170	0.0036	8.29%
Applicants: Neuroticism-Optimism	12	29783	-0.46	0.0098	2482	0.0002	2.53%
Overall: Neuroticism-Ambition	14	2401	-0.23	0.0287	172	0.0053	18.33%
Incumbents: Neuroticism-Ambition							
Applicants: Neuroticism-Ambition							

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% $\sigma^2_{OBS:}$ SE
Overall: Extraversion-Openness	94	31582	0.34	0.0249	336	0.0023	9.35%
Incumbents: Extraversion-Openness	77	14839	0.30	0.0233	193	0.0043	18.69%
Applicants: Extraversion-Openness	17	16743	0.38	0.0228	985	0.0007	3.25%
Overall: Extraversion-Agreeableness	97	32599	0.21	0.0144	336	0.0027	18.85%
Incumbents: Extraversion-Agreeableness	79	15266	0.19	0.0218	193	0.0048	22.15%
Applicants: Extraversion-Agreeableness	18	17333	0.23	0.0071	963	0.0009	13.11%
Overall: Extraversion-Conscientiousness	156	79788	0.19	0.0531	511	0.0018	3.44%
Incumbents: Extraversion-Conscientiousness	120	26099	0.19	0.0370	217	0.0043	11.57%
Applicants: Extraversion-Conscientiousness	36	53689	0.18	0.0609	1491	0.0006	1.03%
Overall: Extraversion-Optimism	55	47875	0.50	0.0159	870	0.0006	4.05%
Incumbents: Extraversion-Optimism	34	8528	0.55	0.0178	251	0.0019	10.88%
Applicants: Extraversion-Optimism	21	39347	0.49	0.0149	1874	0.0003	2.07%
Overall: Extraversion-Ambition	29	14118	0.43	0.0398	487	0.0014	3.43%
Incumbents: Extraversion-Ambition	22	6055	0.46	0.0202	275	0.0023	11.33%
Applicants: Extraversion-Ambition	7	8063	0.41	0.0537	1152	0.0006	1.12%
Overall: Openness-Agreeableness	91	30968	0.16	0.0136	340	0.0028	20.69%
Incumbents: Openness-Agreeableness	75	14488	0.16	0.0196	193	0.0049	25.07%
Applicants: Openness-Agreeableness	16	16480	0.15	0.0081	1030	0.0009	11.52%

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% $\sigma^2_{OBS:}$ SE
Overall: Openness-Conscientiousness	99	32774	0.12	0.0288	331	0.0029	10.22%
Incumbents: Openness-Conscientiousness	83	16294	0.08	0.0415	196	0.0051	12.19%
Applicants: Openness-Conscientiousness	16	16480	0.16	0.0131	1030	0.0009	7.05%
Overall: Openness-Optimism	18	5358	0.18	0.0225	298	0.0032	14.01%
Incumbents: Openness-Optimism	15	2828	0.30	0.0118	189	0.0044	37.71%
Applicants: Openness-Optimism	3	2530	0.05	0.0033	843	0.0012	36.03%
Overall: Openness-Ambition	12	1518	0.23	0.0180	127	0.0071	39.78%
Incumbents: Openness-Ambition							
Applicants: Openness-Ambition							
Overall: Agreeableness-Conscientiousness	103	34065	0.32	0.0240	331	0.0024	10.10%
Incumbents: Agreeableness-Conscientiousness	87	17585	0.30	0.0177	202	0.0041	23.20%
Applicants: Agreeableness-Conscientiousness	16	16480	0.35	0.0299	1030	0.0008	2.52%
Overall: Agreeableness-Optimism	18	5793	0.14	0.0287	322	0.0030	10.45%
Incumbents: Agreeableness-Optimism	15	3263	0.24	0.0224	218	0.0041	18.35%
Applicants: Agreeableness-Optimism	3	2530	0.01	0.0082	843	0.0012	14.41%
Overall: Agreeableness-Ambition	12	1518	0.15	0.0435	127	0.0076	17.44%
Incumbents: Agreeableness-Ambition							
Applicants: Agreeableness-Ambition							

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% $\sigma^2_{OBS:}$ SE
Overall: Conscientiousness-Optimism	56	48482	0.27	0.0349	866	0.0010	2.86%
Incumbents: Conscientiousness-Optimism	35	9135	0.28	0.0346	261	0.0033	9.42%
Applicants: Conscientiousness-Optimism	21	39347	0.26	0.0349	1874	0.0005	1.33%
Overall: Conscientiousness-Ambition	28	15707	0.34	0.0099	561	0.0014	14.11%
Incumbents: Conscientiousness-Ambition	22	6055	0.29	0.0173	275	0.0031	17.65%
Applicants: Conscientiousness-Ambition	6	9652	0.38	0.0024	1609	0.0005	19.03%
Overall: Optimism-Ambition	18	12324	0.54	0.0097	685	0.0007	7.53%
Incumbents: Optimism-Ambition	13	5144	0.52	0.0158	396	0.0013	8.54%
Applicants: Optimism-Ambition	5	7180	0.56	0.0047	1436	0.0003	6.98%

Note: k = number of studies; N = total sample size; $\bar{r}$ = weighted average observed correlation; σ^2_{OBS} = variance in observed correlations; $\overline{N}$ = average study sample size; σ^2_{SE} = variance attributable to sampling error; % $\sigma^2_{OBS:}$ SE = percentage of observed variance attributable to sampling error; σ^2 = variance in corrected correlations; SD_ρ = standard deviation of corrected correlations; 90% CV_{LOWER} = Lower 90% credibility interval for corrected correlation; 90% CV_{UPPER} = Upper 90% credibility interval for corrected correlation; Moderator t-test = t-test of sample type as a moderator of observed correlations. Each t-test represents a comparison of the distribution of correlations between the line on which the t-test appears and the ensuing line. Subgroup analyses were not conducted for the following correlations due to an insufficient number (less than three) of applicant studies: Neuroticism-Ambition; Openness-Ambition; and Agreeableness-Ambition.

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Neuroticism-Extraversion	0.0282	0.17	-0.49	-0.06	
Incumbents: Neuroticism-Extraversion	0.0219	0.15	-0.47	-0.09	0.04
Applicants: Neuroticism-Extraversion	0.0311	0.18	-0.51	-0.05	
Overall: Neuroticism-Openness	0.0103	0.10	-0.29	-0.02	
Incumbents: Neuroticism-Openness	0.0139	0.12	-0.28	0.02	**1.82**
Applicants: Neuroticism-Openness	0.0054	0.07	-0.28	-0.08	
Overall: Neuroticism-Agreeableness	0.0231	0.15	-0.50	-0.10	
Incumbents: Neuroticism-Agreeableness	0.0191	0.14	-0.45	-0.09	1.32
Applicants: Neuroticism-Agreeableness	0.0254	0.16	-0.54	-0.12	
Overall: Neuroticism-Conscientiousness	0.0248	0.16	-0.67	-0.26	
Incumbents: Neuroticism-Conscientiousness	0.0256	0.16	-0.57	-0.16	**4.11**
Applicants: Neuroticism-Conscientiousness	0.0189	0.14	-0.68	-0.32	
Overall: Neuroticism-Optimism	0.0127	0.11	-0.61	-0.32	
Incumbents: Neuroticism-Optimism	0.0393	0.20	-0.74	-0.21	0.16
Applicants: Neuroticism-Optimism	0.0095	0.10	-0.60	-0.33	
Overall: Neuroticism-Ambition	0.0235	0.15	-0.44	-0.02	
Incumbents: Neuroticism-Ambition					
Applicants: Neuroticism-Ambition					

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Extraversion-Openness	0.0226	0.15	0.15	0.54	
Incumbents: Extraversion-Openness	0.0189	0.14	0.12	0.47	**2.16**
Applicants: Extraversion-Openness	0.0220	0.15	0.18	0.58	
Overall: Extraversion-Agreeableness	0.0117	0.11	0.07	0.35	
Incumbents: Extraversion-Agreeableness	0.0170	0.13	0.02	0.36	1.63
Applicants: Extraversion-Agreeableness	0.0062	0.08	0.13	0.34	
Overall: Extraversion-Conscientiousness	0.0513	0.23	-0.11	0.48	
Incumbents: Extraversion-Conscientiousness	0.0327	0.18	-0.04	0.43	0.26
Applicants: Extraversion-Conscientiousness	0.0602	0.25	-0.14	0.50	
Overall: Extraversion-Optimism	0.0153	0.12	0.34	0.66	
Incumbents: Extraversion-Optimism	0.0159	0.13	0.39	0.72	**1.72**
Applicants: Extraversion-Optimism	0.0146	0.12	0.33	0.65	
Overall: Extraversion-Ambition	0.0384	0.20	0.17	0.69	
Incumbents: Extraversion-Ambition	0.0179	0.13	0.28	0.63	0.48
Applicants: Extraversion-Ambition	0.0531	0.23	0.08	0.74	
Overall: Openness-Agreeableness	0.0108	0.10	0.02	0.29	
Incumbents: Openness-Agreeableness	0.0147	0.12	0.01	0.32	0.64
Applicants: Openness-Agreeableness	0.0071	0.08	0.03	0.26	

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Openness-Conscientiousness	0.0259	0.16	-0.09	0.32	
Incumbents: Openness-Conscientiousness	0.0364	0.19	-0.17	0.32	**2.21**
Applicants: Openness-Conscientiousness	0.0122	0.11	0.01	0.30	
Overall: Openness-Optimism	0.0194	0.14	0.00	0.37	
Incumbents: Openness-Optimism	0.0073	0.09	0.18	0.41	**5.61**
Applicants: Openness-Optimism	0.0021	0.05	-0.03	0.14	
Overall: Openness-Ambition	0.0108	0.10	0.09	0.37	
Incumbents: Openness-Ambition					
Applicants: Openness-Ambition					
Overall: Agreeableness-Conscientiousness	0.0216	0.15	0.13	0.51	
Incumbents: Agreeableness-Conscientiousness	0.0136	0.12	0.15	0.45	0.95
Applicants: Agreeableness-Conscientiousness	0.0291	0.17	0.12	0.58	
Overall: Agreeableness-Optimism	0.0257	0.16	-0.07	0.35	
Incumbents: Agreeableness-Optimism	0.0183	0.14	0.06	0.42	**3.46**
Applicants: Agreeableness-Optimism	0.0070	0.08	-0.15	0.17	
Overall: Agreeableness-Ambition	0.0359	0.19	-0.10	0.41	
Incumbents: Agreeableness-Ambition					
Applicants: Agreeableness-Ambition					

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Conscientiousness-Optimism	0.0339	0.18	0.03	0.51	
Incumbents: Conscientiousness-Optimism	0.0314	0.18	0.05	0.51	0.36
Applicants: Conscientiousness-Optimism	0.0344	0.19	0.02	0.51	
Overall: Conscientiousness-Ambition	0.0085	0.09	0.22	0.46	
Incumbents: Conscientiousness-Ambition	0.0143	0.12	0.13	0.45	**2.46**
Applicants: Conscientiousness-Ambition	0.0020	0.04	0.31	0.44	
Overall: Optimism-Ambition	0.0089	0.09	0.42	0.67	
Incumbents: Optimism-Ambition	0.0144	0.12	0.36	0.68	0.80
Applicants: Optimism-Ambition	0.0044	0.07	0.46	0.66	

Relying on absolute differences between subgroup correlations and average subgroup SD_ρ values (as opposed to t-tests), the same conclusion is reached in 13 instances whereas a different conclusion about sample type as a moderator would be reached in five cases.[9] The correlation between Extraversion and Optimism was found to be moderated by sample type when the distributions were compared with a t-test; however, the subgroup SD_ρ values averaged larger than the overall SD_ρ value. And, the magnitude of the difference between the subgroup meta-analytic correlations was relatively small (incumbent $\rho = 0.55$ versus applicant $\rho = 0.49$). The other four instances leading to different conclusions about sample type as a moderator involved moderating effects that were not identified by the t-test but were identified when weighted average correlations and average subgroup SD_ρ values were examined. The correlations involved were Extraversion-Agreeableness; Extraversion-Conscientiousness; Extraversion-Ambition; and Agreeableness-Conscientiousness. In each case, the magnitude of the moderating effect was small, with the absolute difference in the subgroup correlations ranging from a low of 0.01 to a high of 0.05.

Within the subgroups, many of the standard deviations of the range-corrected correlations remain quite large (e.g., the standard deviation of the corrected correlation between Neuroticism and Conscientiousness is 0.16 in the incumbent subgroup and 0.14 in the applicant subgroup). In a related manner, the 90% credibility intervals overlap substantially between the two subgroups. The lower limits for the Openness-Conscientiousness correlation in the incumbent and applicant subgroups are –0.17 and +0.01 with corresponding upper limits of +0.32 and +0.30.

Again, though, a more serious concern is that the differences that are seemingly due to sample type might actually be due to a possible confound between sample type and the specific inventory utilized. This would happen if two conditions were met. The first is that correlations between personality constructs would have to be differentially related as a function of the specific inventory used to measure those constructs. This is a strong possibility, as Optimism and Conscientiousness are negatively correlated if operationally indicated by PRF Need for Play and Need for Achievement (see Jackson, 1999), but are positively correlated when measured by other inventories. The second condition would be that particular inventories are disproportionately

[9] Three distributions were not tested for moderation by sample type due to an insufficient number of applicant studies.

represented in one or the other of the two sample types. Inspection of the studies included in the current analyses indicates that there was in fact disproportionate representation of inventories by sample type. The MMPI and the MMPI-2 appeared in 19 studies in the current analyses and 15 of those (79%) were applicant samples. On the other hand, the NEO-PI, the NEO-PI-R, and the NEO-FFI were used in a total of 60 studies: seven (12%) were studies of job applicants. As a result, the potential confound between sample type and inventory presents an alternative explanation for existing differences between correlations derived from incumbents and applicants.

In an effort to eliminate differential representation of personality inventories as a confound, an additional set of analyses was conducted. These analyses isolated the modal personality inventories used to measure each pair of personality constructs. Thus, only studies using the NEO-FFI were included in the meta-analysis of the following pairs of variables: Neuroticism-Extraversion; Neuroticism-Openness; Neuroticism-Agreeableness; Neuroticism-Conscientiousness; Extraversion-Openness; Extraversion-Agreeableness; Extraversion-Conscientiousness; Openness-Agreeableness; Openness-Conscientiousness; Agreeableness-Conscientiousness. Studies using the NEO-PI or the NEO-PI-R were included in the meta-analyses of Openness-Optimism and Agreeableness-Optimism. The meta-analyses for Extraversion-Optimism, Conscientiousness-Optimism, Extraversion-Ambition, Conscientiousness-Ambition, and Optimism-Ambition were based only on studies using the CPI. Studies using the HPI were used in the meta-analyses of Neuroticism-Ambition, Openness-Ambition, and Agreeableness-Ambition. Studies using the 16PF were used to estimate the meta-analytic correlation between Neuroticism and Optimism. Obviously, this severely reduced the sample sizes included in the meta-analysis, but this was necessary in order to rule out differences in personality tests as a confound. The results of this analysis are presented in Table 6.

Table 6. Meta-analysis Results for Correlations Between Predictors: Including only modal personality inventory in each predictor pair.

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% σ^2_{OBS}: SE
Overall: Neuroticism-Extraversion	25	10882	-0.38	0.0132	435	0.0017	12.82%
Incumbents: Neuroticism-Extraversion	20	6365	-0.38	0.0051	318	0.0023	45.60%
Applicants: Neuroticism-Extraversion	5	4517	-0.38	0.0246	903	0.0008	3.31%
Overall: Neuroticism-Openness	21	7348	-0.10	0.0098	350	0.0028	28.63%
Incumbents: Neuroticism-Openness	16	2831	-0.06	0.0185	177	0.0056	30.50%
Applicants: Neuroticism-Openness	5	4517	-0.13	0.0025	903	0.0011	43.69%
Overall: Neuroticism-Agreeableness	23	7952	-0.26	0.0071	346	0.0025	35.76%
Incumbents: Neuroticism-Agreeableness	18	3435	-0.28	0.0088	191	0.0044	50.79%
Applicants: Neuroticism-Agreeableness	5	4517	-0.24	0.0049	903	0.0010	19.93%
Overall: Neuroticism-Conscientiousness	23	7576	-0.37	0.0033	329	0.0023	67.57%
Incumbents: Neuroticism-Conscientiousness	18	3059	-0.38	0.0048	170	0.0043	89.71%
Applicants: Neuroticism-Conscientiousness	5	4517	-0.37	0.0023	903	0.0008	36.60%
Overall: Neuroticism-Optimism	18	29819	-0.46	0.0102	1657	0.0004	3.64%
Incumbents: Neuroticism-Optimism	7	406	-0.37	0.0253	58	0.0130	51.25%
Applicants: Neuroticism-Optimism	11	29413	-0.47	0.0099	2674	0.0002	2.32%

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% σ^2_{OBS}: SE
Overall: Neuroticism-Ambition	10	1143	-0.38	0.0152	114	0.0065	42.81%
Incumbents: Neuroticism-Ambition							
Applicants: Neuroticism-Ambition							
Overall: Extraversion-Openness	21	7348	0.19	0.0049	350	0.0027	54.16%
Incumbents: Extraversion-Openness	16	2831	0.15	0.0064	177	0.0054	85.38%
Applicants: Extraversion-Openness	5	4517	0.22	0.0017	903	0.0010	58.14%
Overall: Extraversion-Agreeableness	24	8148	0.23	0.0107	340	0.0026	24.67%
Incumbents: Extraversion-Agreeableness	19	3631	0.24	0.0097	191	0.0047	48.09%
Applicants: Extraversion-Agreeableness	5	4517	0.22	0.0114	903	0.0010	8.79%
Overall: Extraversion-Conscientiousness	25	8059	0.32	0.0032	322	0.0025	79.02%
Incumbents: Extraversion-Conscientiousness	20	3542	0.32	0.0058	177	0.0046	78.77%
Applicants: Extraversion-Conscientiousness	5	4517	0.32	0.0011	903	0.0009	80.74%
Overall: Extraversion-Optimism	15	11717	0.65	0.0019	781	0.0004	22.96%
Incumbents: Extraversion-Optimism	10	4537	0.64	0.0016	454	0.0008	46.75%
Applicants: Extraversion-Optimism	5	7180	0.65	0.0020	1436	0.0002	11.44%
Overall: Extraversion-Ambition	15	11717	0.49	0.0127	781	0.0007	5.83%
Incumbents: Extraversion-Ambition	10	4537	0.51	0.0153	454	0.0012	7.98%
Applicants: Extraversion-Ambition	5	7180	0.48	0.0108	1436	0.0004	3.85%

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% σ^2_{OBS}: SE
Overall: Openness-Agreeableness	21	7348	0.08	0.0040	350	0.0028	69.92%
Incumbents: Openness-Agreeableness	16	2831	0.06	0.0091	177	0.0056	62.25%
Applicants: Openness-Agreeableness	5	4517	0.09	0.0005	903	0.0011	100.00%
Overall: Openness-Conscientiousness	22	7497	-0.01	0.0067	341	0.0029	43.87%
Incumbents: Openness-Conscientiousness	17	2980	-0.04	0.0119	175	0.0057	48.07%
Applicants: Openness-Conscientiousness	5	4517	0.02	0.0018	903	0.0011	61.42%
Overall: Openness-Optimism	6	1506	0.33	0.0018	251	0.0032	100.00%
Incumbents: Openness-Optimism							
Applicants: Openness-Optimism							
Overall: Openness-Ambition	10	1143	0.26	0.0154	114	0.0077	49.92%
Incumbents: Openness-Ambition							
Applicants: Openness-Ambition							
Overall: Agreeableness-Conscientiousness	24	7978	0.20	0.0129	332	0.0028	21.55%
Incumbents: Agreeableness-Conscientiousness	19	3461	0.23	0.0128	182	0.0050	38.81%
Applicants: Agreeableness-Conscientiousness	5	4517	0.17	0.0118	903	0.0010	8.84%
Overall: Agreeableness-Optimism	6	1506	0.35	0.0092	251	0.0031	33.37%
Incumbents: Agreeableness-Optimism							
Applicants: Agreeableness-Optimism							

	k	N	$\bar{r}$	σ^2_{OBS}	$\overline{N}$	σ^2_{SE}	% σ^2_{OBS}: SE
Overall: Agreeableness-Ambition	10	1143	0.23	0.0181	114	0.0079	43.55%
Incumbents: Agreeableness-Ambition							
Applicants: Agreeableness-Ambition							
Overall: Conscientiousness-Optimism	15	11717	0.32	0.0102	781	0.0010	10.15%
Incumbents: Conscientiousness-Optimism	10	4537	0.42	0.0063	454	0.0015	23.86%
Applicants: Conscientiousness-Optimism	5	7180	0.26	0.0030	1436	0.0006	20.47%
Overall: Conscientiousness-Ambition	15	11717	0.38	0.0046	781	0.0009	20.77%
Incumbents: Conscientiousness-Ambition	10	4537	0.35	0.0074	454	0.0017	23.08%
Applicants: Conscientiousness-Ambition	5	7180	0.39	0.0018	1436	0.0005	27.13%
Overall: Optimism-Ambition	15	11717	0.56	0.0041	781	0.0006	14.70%
Incumbents: Optimism-Ambition	10	4537	0.56	0.0032	454	0.0011	32.88%
Applicants: Optimism-Ambition	5	7180	0.56	0.0047	1436	0.0003	6.98%

Note: k = number of studies; N = total sample size; $\bar{r}$ = weighted average observed correlation; σ^2_{OBS} = variance in observed correlations; $\overline{N}$ = average study sample size; σ^2_{SE} = variance attributable to sampling error; % $\sigma^2_{OBS:}$ SE = percentage of observed variance attributable to sampling error; σ^2 = variance in corrected correlations; SD_ρ = standard deviation of corrected correlations; 90% CV_{LOWER} = Lower 90% credibility interval for corrected correlation; 90% CV_{UPPER} = Upper 90% credibility interval for corrected correlation; Moderator t-test = t-test of sample type as a moderator of observed correlations. Each t-test represents a comparison of the distribution of correlations between the line on which the t-test appears and the ensuing line. Subgroup analyses were not conducted for the following correlations due to an insufficient number (less than three) of applicant studies: Neuroticism-Ambition; Openness-Optimism; Openness-Ambition; Agreeableness-Optimism; and Agreeableness-Ambition.

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Neuroticism-Extraversion	0.0115	0.11	-0.52	-0.24	
Incumbents: Neuroticism-Extraversion	0.0028	0.05	-0.45	-0.31	0.01
Applicants: Neuroticism-Extraversion	0.0238	0.15	-0.61	-0.14	
Overall: Neuroticism-Openness	0.0070	0.08	-0.21	0.01	
Incumbents: Neuroticism-Openness	0.0129	0.11	-0.21	0.10	**1.73**
Applicants: Neuroticism-Openness	0.0014	0.04	-0.18	-0.07	
Overall: Neuroticism-Agreeableness	0.0045	0.07	-0.35	-0.17	
Incumbents: Neuroticism-Agreeableness	0.0043	0.07	-0.37	-0.20	1.14
Applicants: Neuroticism-Agreeableness	0.0039	0.06	-0.34	-0.14	
Overall: Neuroticism-Conscientiousness	0.0011	0.03	-0.42	-0.33	
Incumbents: Neuroticism-Conscientiousness	0.0005	0.02	-0.41	-0.35	0.49
Applicants: Neuroticism-Conscientiousness	0.0014	0.04	-0.43	-0.31	
Overall: Neuroticism-Optimism	0.0099	-0.46	0.15	-1.08	
Incumbents: Neuroticism-Optimism	0.0123	-0.37	0.16	-0.91	0.11
Applicants: Neuroticism-Optimism	0.0097	0.10	-0.60	-0.33	

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Neuroticism-Ambition	0.0087	0.09	-0.50	-0.25	
Incumbents: Neuroticism-Ambition					
Applicants: Neuroticism-Ambition					
Overall: Extraversion-Openness	0.0022	0.05	0.13	0.26	
Incumbents: Extraversion-Openness	0.0009	0.03	0.11	0.19	**2.81**
Applicants: Extraversion-Openness	0.0007	0.03	0.18	0.26	
Overall: Extraversion-Agreeableness	0.0081	0.09	0.11	0.35	
Incumbents: Extraversion-Agreeableness	0.0050	0.07	0.15	0.34	0.37
Applicants: Extraversion-Agreeableness	0.0104	0.10	0.07	0.38	
Overall: Extraversion-Conscientiousness	0.0007	0.03	0.28	0.35	
Incumbents: Extraversion-Conscientiousness	0.0012	0.04	0.28	0.37	
Applicants: Extraversion-Conscientiousness	0.0002	0.01	0.29	0.34	
Overall: Extraversion-Optimism	0.0015	0.04	0.60	0.70	
Incumbents: Extraversion-Optimism	0.0009	0.03	0.60	0.68	0.27
Applicants: Extraversion-Optimism	0.0018	0.04	0.58	0.71	
Overall: Extraversion-Ambition	0.0120	0.11	0.34	0.64	
Incumbents: Extraversion-Ambition	0.0141	0.12	0.34	0.67	0.46
Applicants: Extraversion-Ambition	0.0104	0.10	0.32	0.63	

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Openness-Agreeableness	0.0012	0.03	0.04	0.13	
Incumbents: Openness-Agreeableness	0.0034	0.06	-0.02	0.14	1.24
Applicants: Openness-Agreeableness	0.0000	0.00	0.09	0.09	
Overall: Openness-Conscientiousness	0.0038	0.06	-0.09	0.08	
Incumbents: Openness-Conscientiousness	0.0062	0.08	-0.15	0.06	**1.88**
Applicants: Openness-Conscientiousness	0.0007	0.03	-0.02	0.06	
Overall: Openness-Optimism	0.0000	0.00	0.33	0.33	
Incumbents: Openness-Optimism					
Applicants: Openness-Optimism					
Overall: Openness-Ambition	0.0077	0.09	0.13	0.38	
Incumbents: Openness-Ambition					
Applicants: Openness-Ambition					
Overall: Agreeableness-Conscientiousness	0.0102	0.10	0.06	0.33	
Incumbents: Agreeableness-Conscientiousness	0.0078	0.09	0.11	0.34	0.98
Applicants: Agreeableness-Conscientiousness	0.0108	0.10	0.01	0.33	
Overall: Agreeableness-Optimism	0.0061	0.08	0.24	0.47	
Incumbents: Agreeableness-Optimism					
Applicants: Agreeableness-Optimism					

Sample Type and Predictor Construct Pair	σ^2	SD_ρ	90% CV_{LOWER}	90% CV_{UPPER}	Moderator t-test
Overall: Agreeableness-Ambition	0.0102	0.10	0.09	0.37	
Incumbents: Agreeableness-Ambition					
Applicants: Agreeableness-Ambition					
Overall: Conscientiousness-Optimism	0.0091	0.10	0.19	0.45	
Incumbents: Conscientiousness-Optimism	0.0048	0.07	0.32	0.51	**4.51**
Applicants: Conscientiousness-Optimism	0.0024	0.05	0.18	0.33	
Overall: Conscientiousness-Ambition	0.0036	0.06	0.29	0.46	
Incumbents: Conscientiousness-Ambition	0.0057	0.08	0.24	0.45	1.46
Applicants: Conscientiousness-Ambition	0.0013	0.04	0.34	0.45	
Overall: Optimism-Ambition	0.0035	0.06	0.48	0.64	
Incumbents: Optimism-Ambition	0.0021	0.05	0.49	0.62	0.04
Applicants: Optimism-Ambition	0.0044	0.07	0.46	0.66	

Isolating one personality inventory for each pair of personality constructs tended to reduce the standard deviation of the corrected correlations. In the overall analyses that included all personality inventories (Table 5), the average standard deviation of the corrected correlations across 21 personality construct pairs was 0.14. When only the modal personality inventories were included (Table 6), the average standard deviation of the corrected correlations across 21 personality construct pairs was 0.04. Sample type was tested as a moderator of all correlations between personality constructs that contained at least three studies in each subgroup, subject to the constraint that sampling error did not account for 75% of the observed variance in the overall correlations. Of the sixteen moderator tests conducted, there was evidence for sample type acting as a moderator of four of these. The corrected correlation between Neuroticism and Openness was stronger among applicants (ρ = -0.13, SD_ρ = 0.04) than among incumbents (ρ = -0.06, SD_ρ = 0.11); the corrected correlation between Extraversion and Openness was stronger among applicants (ρ = 0.22, SD_ρ = 0.03) than among incumbents (ρ = 0.15, SD_ρ = 0.03); the corrected correlation between Openness and Conscientiousness was small and negative in incumbent samples (ρ = -0.04, SD_ρ = 0.08), whereas the same correlation was small and positive in applicant samples (ρ = 0.02, SD_ρ = 0.03). Finally, the correlation between Conscientiousness and Optimism was stronger (ρ = 0.42, SD_ρ = 0.07) in the incumbent sample as compared to the applicant sample (ρ = 0.26, SD_ρ = 0.05).

Relying on subgroup correlations and average within group SD_ρ values to test sample type as a moderator would identify four additional correlations as being moderated by sample type. Specifically, the following correlations would also be identified as being moderated by sample type: Neuroticism-Agreeableness, Extraversion-Agreeableness, Extraversion-Optimism, and Agreeableness-Conscientiousness. These moderating effects were small, with the magnitude of the difference between groups ranging from 0.01 to 0.06.

To summarize the results to this point, it is evident that the overall estimates of the operational validities are generally consistent with previous meta-analytic investigations (Table 3). The operational validities for personality inventories as predictors of performance rating criteria suggest that there are some differences in the criterion-related validities estimated in incumbent and applicant studies (Table 4). Ten of 14 validity distributions demonstrated some evidence of moderation by sample (either a statistically significant t-test or a difference between the subgroup operational validity estimates accompanied by a lower average SD_ρ value between

subgroups than in the overall analysis). In some of those cases, the correlations are stronger in the incumbent studies (Single-stimulus measures of Neuroticism, Extraversion, Openness, and Ambition), whereas in other cases, the relationship is stronger in applicant studies (Forced-choice measures of Neuroticism, Extraversion, Openness, and Agreeableness, as well as single-stimulus measures of Conscientiousness and Optimism). Finally, there is some evidence of sample type moderating the correlations for half the pairs of personality constructs. Again this evidence was a statistically significant t-test comparing the distribution of observed correlations and/or different weighted average correlations between subgroups accompanied by an average of the subgroup SD_ρ values that was lower than the overall SD_ρ value. Some of the differences between the incumbent and applicant correlations were small. The results from the simulation study based on the meta-analytic correlation matrices provides a more clear indication of the practical implications of the moderating effect of sample type.

Simulation Study

In order to test Hypotheses Two and Three, the meta-analytic correlation matrices were used as input for two sets of simulation analyses. In both sets of simulation analyses, separate correlation matrices were constructed based on applicant and incumbent parameter estimates (criterion-related validities and inter-correlations among personality constructs). The criterion-related validities were taken from analysis of only studies that utilized a performance rating criterion, whereas the inter-correlations were taken from the analyses of studies that utilized the modal personality inventory to measure each pair of constructs. As the correlations among personality constructs were based on single-stimulus personality measures, the validities utilized in these analyses were the criterion-related validity estimates for single-stimulus measures.

In the first set of simulations, a strict decision criterion was put in place for designating sample type as a moderator of the correlation. Specifically, two conditions were necessary for identifying sample type as a moderator of the population parameter. First, less than 75% of the variance in observed correlations overall could be explained by statistical artifacts. Second, the t-test result (reported in Tables 4 and 6) comparing the distribution of observed correlations across sample type was required to be statistically significant. Unless both conditions were met, the correlation was designated as not being moderated by sample type. If sample type was not identified as a moderator, the overall population correlation was imputed in the correlation

matrices for both incumbents and applicants. Of the 28 cells in the correlation matrix (seven validity estimates and 21 correlations among predictor constructs), seven cells passed the strict evidence test for sample type as a moderator. Three of these were criterion-related validity estimates (Neuroticism, Extraversion, and Ambition) and four were inter-correlations among personality constructs (Neuroticism-Openness; Extraversion-Openness; Openness-Conscientiousness; and Conscientiousness-Optimism).

In constructing the correlation matrices that would be the input for the simulated data, correlations corrected for measurement error in the criterion were entered as the criterion-related validity estimates. Weighted average observed correlations were entered as the correlations between personality constructs. The correlation matrices used in the first set of simulation analyses are presented in Table 7 below.

The second set of simulation analyses did not place any constraints on the identification of sample type as a moderator. For every cell in the correlation matrix, the subgroup correlations were entered, regardless of the evidence supporting moderation. Thus, the incumbent correlation matrix was constructed by including all the incumbent subgroup correlations, and, the applicant matrix was constructed by including all the applicant subgroup correlations. The only exceptions to this rule were in the cases where an insufficient number of applicant studies existed. In these cases, the overall estimate was imputed as both the incumbent and the applicant estimate. Table 8 presents the correlation matrices used in the second set of simulation analyses.

Table 7. Meta-analytic Correlation Matrices: Strict evidence of moderation

	N	E	O	A	C	Opt	Amb	Ratings
Neuroticism		-0.38	**-0.13**	-0.26	-0.37	-0.46	-0.38	**-0.05**
Extraversion	-0.38		**0.22**	0.23	0.32	0.65	0.49	**0.01**
Openness	**-0.06**	**0.15**		0.08	**0.02**	0.33	0.26	0.03
Agreeableness	-0.26	0.23	0.08		0.20	0.35	0.23	0.11
Conscientious	-0.37	0.32	**-0.04**	0.20		**0.26**	0.38	0.13
Optimism	-0.46	0.65	0.33	0.35	**0.42**		0.56	0.15
Ambition	-0.38	0.49	0.26	0.23	0.38	0.56		**0.02**
Ratings	**-0.12**	**0.08**	0.03	0.11	0.13	0.15	**0.10**	

Note: Incumbent Correlations below diagonal; applicant correlations above diagonal. Values in bold were identified as being moderated by sample type. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; Opt = Optimism; Amb = Ambition.

Table 8. Meta-analytic Correlation Matrices: All subgroups correlations used regardless of evidence of moderation

	N	E	O	A	C	Opt	Amb	Ratings
Neuroticism		-0.38	-0.13	-0.24	-0.37	-0.47	-0.38	-0.05
Extraversion	-0.38		0.22	0.22	0.32	0.65	0.48	0.01
Openness	-0.06	0.15		0.09	0.02	0.33	0.26	-0.01
Agreeableness	-0.28	0.24	0.06		0.17	0.35	0.23	0.11
Conscientious	-0.38	0.32	-0.04	0.23		0.26	0.39	0.17
Optimism	-0.37	0.64	0.33	0.35	0.42		0.56	0.20
Ambition	-0.38	0.51	0.26	0.23	0.35	0.56		0.02
Ratings	-0.12	0.08	0.04	0.11	0.13	0.14	0.10	

Note: Incumbent Correlations below diagonal; applicant correlations above diagonal. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; Opt = Optimism; Amb = Ambition.

Simulation Study Results: Strict evidence of moderation

Hypothesis two posited that regression equations derived from studies of job incumbents would overestimate the predictive validity of personality inventories when implemented in applicant settings. In order to test this hypothesis, it is necessary to derive a regression equation from data based on job incumbents, and apply it to data from job applicants. The job incumbent and job applicant data in the current analyses were simulated on the basis of the meta-analytic correlation matrices in Table 7.

Howell (2003) has documented the procedures for generating data as if they were drawn from a population with a designated correlation matrix. In the present case, 10,000 hypothetical participants were generated. The generated data are scores on vectors representing each of the seven personality variables and the performance ratings criterion and were generated so as to be normally distributed with a mean of zero and a standard deviation of unity for each variable. In the next step, these random normally distributed values are factor analyzed, eight factors are extracted, and factor scores are saved for each participant. Subsequently, these factor scores are post-multiplied by the Cholesky decomposition of the desired correlation matrix. The result of this are normally distributed factor scores on each variable with correlations between factors as

dictated by the meta-analytic estimates of the correlations.[10] The data generation phase is conducted separately for both job incumbent and job applicant data. The SPSS command syntax for the generation of simulated data based on the incumbent parameter estimates from Table 7 is presented in Appendix A.

Prediction Model Using Incumbent Meta-Analytic Correlations: Strict moderation evidence

Using the simulated data, a regression equation was identified that combined the personality constructs in order to predict the performance ratings criterion for the simulated incumbents. This mirrors the situation in which a personnel psychologist has gathered data on job incumbents, and is identifying a desirable way to weight and combine scores on those predictors to predict performance for future job applicants. First, the outcome variable representing the ratings criteria was regressed on the seven personality predictor variables. In the seven-predictor case, the multiple R-value was 0.183 and the standard error of the estimate (root mean square residual) was 0.983. The absolute values of the standardized regression coefficients for four of the seven predictors were less than 0.05, and only one (Optimism) exceeded 0.10.

Inclusion of all personality predictors did not seem necessary or beneficial, so alternative models with reduced numbers of predictors were examined. First, Extraversion, Openness, and Agreeableness were eliminated from the prediction equation. This resulted in a regression equation with a multiple R-value equal to 0.172 with a standard error of the estimate equal to 0.985. Next, Ambition was eliminated. There was no change in model fit from the four-predictor model. Neuroticism was eliminated next, and the resulting two-predictor model (Conscientiousness and Optimism) had a multiple R equal to 0.167 with a standard error of the estimate was equal to 0.986. This equation was selected as the final equation to interpret. Inclusion of any additional predictors beyond Conscientiousness and Optimism did not seem to be warranted. The maximum gain in explanatory power (ΔR) by adding any predictor above Conscientiousness and Optimism was 0.01 (Agreeableness). The standardized regression coefficients (as well as the zero-order correlation with the performance criterion) associated with

[10] As a check on the accuracy of this procedure and transcriptions completed during this procedure, I generated bivariate correlation matrices from the simulated data. In all cases, the correlation matrices computed on the basis of the simulated data matched the meta-analytic correlations precisely.

each predictor in the final two-predictor model are presented in Table 9.

Table 9. Regression Coefficients Associated with each Predictor in the Final Regression Model: Incumbent data, strict moderation evidence

Predictor Construct	Meta-Analytic Zero-order Correlation with Performance Ratings	Standardized Regression Coefficient Associated with Predictor
Conscientiousness	0.13	0.08
Optimism	0.15	0.12

Using Incumbent Model to Predict Performance of Applicants: Strict moderation evidence

The regression weights appearing in Table 9 (e.g., optimal weights) were then applied to the corresponding personality scores for job applicants so as to predict job performance. This is similar to the situation wherein job applicants have provided responses to personality test scales, and those scores are combined and weighted to predict future performance using a prediction equation developed on the basis of job incumbent data. A common technique used to assess the quality of the prediction model is to correlate these predicted job performance scores with actual performance scores obtained on the job at a later time. This cross-validation process can be simulated in the present data by correlating the predicted job performance scores of the applicant sample based on the incumbent prediction model with the actual performance scores generated from the applicant meta-analytic correlation matrix. Hypothesis two predicted that the cross-validation correlation would be lower than the multiple correlation coefficient for the incumbent regression model, thereby indicating that the use of the incumbent model adversely affects the utility of the selection battery. The cross-validation coefficient was 0.177, which is 6% ***larger*** than the multiple R (0.167) value obtained in the incumbent data.

The cross-validation coefficient is usually smaller than the multiple R simply due to sampling error. Sampling error is not an issue in our simulation given that there are 20,000 total simulated individuals. Instead, the expectation was that the degradation of the cross validation coefficient would be indicative of the problem of using an incumbent-derived equation to predict the job performance of applicants. The final prediction model chosen on the basis of the incumbent population parameter estimates included only Conscientiousness and Optimism as

predictors of performance. The operational validity of neither Conscientiousness nor Optimism was found to be moderated by sample type. The correlation between Conscientiousness and Optimism was found to be moderated by sample type, though. The results suggest that the correlation between Conscientiousness and Optimism is stronger in the incumbent data. As a result, less unique variance in performance is explained by Conscientiousness and Optimism in the incumbent data. In the applicant data, there is less overlap between Conscientiousness and Optimism, and more unique variance in performance is accounted for. Examining the results from a hierarchical regression analysis that includes only Conscientiousness and Optimism makes this point very clear. Based on either the incumbent or the applicant data, when performance is regressed on Conscientiousness, the resulting R-value is 0.13. In the incumbent data, the incremental variance accounted for by Optimism, beyond that which is accounted for by Conscientiousness, is $\Delta R = 0.037$. In the applicant data, the incremental variance accounted for by Optimism is $\Delta R = 0.047$. For all intents and purposes, this is a very small difference. Nevertheless, the findings are in the opposite direction than had been hypothesized in Hypothesis Two. Rather than overestimating the operational validity of a multiple predictor regression equation applied to applicant data, incumbent based equations may be an underestimate of the functional validity.

Based on this initial evidence, there is no support for Hypothesis Two. Recall, though, that this is based on the strict evidentiary standards for moderation. It is possible that when all subgroup correlations are used in the data simulation phase, the conclusions drawn would be very different. To investigate this possibility, the simulation analyses were repeated using all subgroup parameter estimates (Table 8).

Prediction Model Using Incumbent Meta-Analytic Correlations: All subgroup correlations

As with the “strict evidence of moderation” simulation conducted above, the simulated data was utilized to estimate a regression equation combining the personality constructs in order to predict the performance ratings criterion for the simulated incumbents. The outcome variable representing the ratings criterion was regressed on the seven personality predictor variables. In the seven-predictor case, the multiple R-value was 0.179 and the standard error of the estimate (root mean square residual) was 0.984. The absolute values of the standardized regression coefficients for three of the seven predictors were less than 0.05, while none exceeded 0.10.

Extraversion, Openness, and Ambition were eliminated from the subsequent model, due to the very small regression coefficients associated with these predictors. The four-predictor (Neuroticism, Agreeableness, Conscientiousness, and Optimism) model was examined, and the resulting multiple R-value was equal to 0.177. Again, no predictor had an associated regression coefficient with an absolute value greater than 0.10. The regression coefficients associated with Agreeableness and Neuroticism were only 0.05, so Agreeableness and Neuroticism were eliminated next. A two-predictor equation that included only Conscientiousness and Optimism was examined and was selected as the final model, with a multiple R-value equal to 0.160. The parsimony of this model was deemed to outweigh the small gain in predictive value that would be gained by including Neuroticism, Agreeableness, or both Neuroticism and Agreeableness. The standardized regression coefficients associated with each predictor in the final two-predictor model are presented in Table 10.

Table 10. Regression Coefficients Associated with each Predictor in the Final Regression Model: Incumbent subgroup correlations

Predictor Construct	Meta-Analytic Zero-order Correlation with Performance Ratings	Standardized Regression Coefficient Associated with Predictor
Conscientiousness	0.13	0.09
Optimism	0.14	0.10

Using Incumbent Model to Predict Performance of Applicants: All subgroup correlations

The regression weights appearing in Table 11 (e.g., optimal weights) were then applied to the corresponding personality scores for job applicants so as to predict job performance. As noted above, this analogizes the situation wherein job applicants have provided responses to personality test scales, and those scores have been combined and weighted to predict future performance using a prediction equation developed on the basis of job incumbent data. To assess the quality of the prediction model, simulated job applicants' predicted job performance scores based on the incumbent prediction model were correlated with the actual performance scores generated from the applicant meta-analytic correlation matrix. The cross-validation coefficient

was 0.234, which is 46% *larger* than the R (0.160) value obtained in the incumbent data.

Once again, the reason that the cross-validation coefficient is larger than the developmental equation R is that the data are known to be drawn from different populations (as opposed to representing two samples drawn from a single population), and, the parameter estimates of interest differ across those populations. First, the operational validity estimates for the predictors captured in the incumbent analysis (Conscientiousness and Optimism) are higher in the applicant population. As shown in Table 9, the operational validity estimates of Conscientiousness and Optimism were 0.13 and 0.14 in the incumbent sample and were 0.17 and 0.20 in the applicant sample. In addition, the correlation between Conscientiousness and Optimism was lower in the applicant data as compared to the incumbent data. These two factors combined assured that more unique variance in performance would be accounted for in the applicant data.

The evidence from the two cross-validation analyses (e.g., the cross-validation based on the strict moderation evidence and that based on full subgroup correlation matrices) does not support Hypothesis Two. In the strict moderation evidence example, the incumbent multiple R was a slight underestimate of the cross-validation coefficient when the incumbent based equation was applied to simulated applicant personality scores. In the full subgroup correlations analysis, the incumbent derived equation R was a substantial underestimate of the cross-validation index.

Prediction Model Using Applicant Meta-Analytic Correlations

The primary purpose of this study was, in regards to personality measures, to assess the interchangeability of regression weights derived from incumbent samples versus regression weights derived from applicant samples. In retrospect, Hypothesis 2 and its reliance on the cross-validation coefficient is not a complete test of the argument that sample type moderates the validity/utility of personality predictors. The cross validation approach does not address the issue of whether or not personality tests are more or less predictive when based on applicant samples versus incumbent samples. In part, this question was addressed via comparison of the bivariate validity coefficients. However, it is possible that results based on regression analyses would differ from those based on bivariate estimates alone. To test this more complete notion of interchangeability, I compared the prediction model derived from the applicant meta-analytic correlations to those derived from the incumbent samples. This was done using the applicant

correlations from the meta-analytic matrix requiring a significant t-test to conclude that sample type moderates the correlations (see Table 7). In addition, this was repeated using all applicant subgroup correlations in Table 8.

As with the simulated incumbent data, the seven-predictor model was examined first. For the simulations based on the "strict evidence of moderation" correlations, the seven-predictor model yielded a multiple R equal to 0.250 (standard error of the estimate = 0.969). Openness ($\beta = 0.00$), and Agreeableness ($\beta = 0.06$) did not appear to add meaningful variance to the other predictors, and were eliminated from the next model. In addition, there was some evidence of multicollinearity involving Extraversion and Optimism. The evidence of multicollinearity was based on large variance proportions associated with the largest condition indices. Although none of the condition indices were "large" according to the rules of thumb presented by Pedhazur (1997; p. 305), it was noteworthy that both Optimism and Extraversion did have large variance proportions associated with the largest condition index. As Optimism was related to performance whereas Extraversion was not, Extraversion was eliminated from subsequent analyses. The four-predictor model including Neuroticism, Conscientiousness, Optimism, and Ambition yielded a multiple R equal to 0.212 (standard error of the estimate = 0.977). In addition, the high correlation between Optimism and Ambition ($\rho = 0.56$) and the finding that Optimism appeared to suppress irrelevant variance in Ambition appeared problematic. Specifically, the operational validity of Ambition was $\rho = +0.02$, whereas the regression coefficient associated with Ambition was $\beta = -0.13$. Further, Optimism and Ambition had variance proportions greater than 0.50 associated with the largest condition index. Removing Ambition decreased the multiple R-value to R = 0.184, but this result seemed more tenable than the results that included Ambition. Finally, there was a similar concern in connection with Neuroticism. That is, the operational validity of Neuroticism was $\rho = -0.05$, whereas the regression coefficient associated with Neuroticism was $\beta = +0.06$. The correlation between Neuroticism and Optimism was $\rho = -0.46$, and Neuroticism and Optimism both has variance proportions greater than 0.50 associated with the largest condition index. Omitting Neuroticism from the final model resulted in a two-predictor model consisting of Conscientiousness and Optimism, with a multiple R = 0.177. The meta-analytic correlations between each of these personality constructs and job performance, as well as the standardized regression coefficient associated with each, are presented in Table 11.

Table 11. Regression Coefficients Associated with each Predictor in the Final Regression Model: Applicant data

Predictor Construct	Meta-Analytic Zero-order Correlation with Performance Ratings	Standardized Regression Coefficient Associated with Predictor
Conscientiousness	0.13	0.10
Optimism	0.15	0.13

The results are effectively the same as those reported above when the incumbent-derived prediction equation was applied to the applicant data. In comparison to the incumbent based prediction equation (see Table 9), the same predictors are included, and, again, the magnitude of the multiple R is slightly larger (0.177 in the applicant data, 0.167 in the incumbent data). There is a slight difference in the magnitude of the regression coefficients associated with each predictor. The reader is reminded that the only relevant difference between the incumbent correlations and the applicant correlations in this strict evidence analysis is in the correlation between Conscientiousness and Optimism. In the incumbent data, these predictors were more strongly related, and as a result, including only Conscientiousness and Optimism in the incumbent prediction model accounted for less unique variance in performance than when these two predictors were included in the applicant model.

Finally, a prediction model based on all applicant subgroup parameter estimates (Table 8) was derived. The seven-predictor model was examined first, yielding a multiple R equal to 0.341 (standard error of the estimate = 0.940). Once again, the results were somewhat suspect. First, the high correlation between Extraversion and Optimism appeared to cause multicollinearity in the data, as each of these predictors had variance proportions greater than 0.40 associated with the largest condition index. Removing Extraversion and examining the six-predictor model revealed a similar state of affairs involving Optimism and Ambition (variance proportions greater than 0.50 associated with the largest condition index). Eliminating Ambition revealed that in the five-predictor model (Neuroticism, Agreeableness, Openness, Conscientiousness, and Optimism), Neuroticism and Optimism shared large variance proportions with the largest condition index. As a result, Neuroticism was eliminated, and this appeared to resolve problems of multicollinearity in the data.

The four-predictor model (Agreeableness, Openness, Conscientiousness, and Optimism) had a multiple R = 0.247, and a standard error of the estimate = 0.969. Agreeableness had a weak regression coefficient associated with it, and was removed from the model. In turn, Openness did not appear to add much explanatory power above and beyond the parsimonious two-predictor model that contained only Conscientiousness and Optimism. The $\Delta R = 0.011$ when Openness was added to Conscientiousness and Optimism. The prediction equation that included only Conscientiousness and Optimism had a multiple R equal to 0.234 and a standard error of the estimate equal to 0.972. The zero-order correlations and the regression coefficients associated with each predictor are presented in Table 12.

Table 12. Regression Coefficients Associated with each Predictor in the Final Regression Model: Applicant subgroup correlations

Predictor Construct	Meta-Analytic Zero-order Correlation with Performance Ratings	Standardized Regression Coefficient Associated with Predictor
Conscientiousness	0.17	0.13
Optimism	0.20	0.17

In comparison to the incumbent prediction model (Table 12), the applicant prediction model includes identical predictors. Moreover, slightly more variance in job performance is accounted for in the applicant data (R = 0.234) than in the incumbent data (R = 0.160). Again, this is due in part to the finding that in the population of applicant data, Conscientiousness and Optimism are each more strongly related to performance, while being less strongly related to each other.

Summary of Results: Comparison of prediction models

The direct comparison of regression models suggests that while sample type does act as a moderator of regression models for personality predictors, the results are not as had been anticipated. When data were simulated on the basis of incumbent-derived population parameter estimates, and a prediction equation relating personality predictors to occupational performance was estimated on the basis of that data, the resulting R was smaller than what would be expected based on data simulated from applicant population parameter estimates. This underestimation of

the applicant validity held true in two different cases. When a statistically significant t-test was a prerequisite for designating sample type as a moderator of any population parameter estimate in the correlation matrix, regression analyses and cross-validation of those regression results revealed that incumbent-based data underestimated applicant validity. Similarly, when all subgroup population parameter estimates were imputed in the correlation matrix (regardless of the statistical significance test for moderation by sample type), regression analyses and cross-validation of those regression results revealed that incumbent-based data underestimated applicant validity.

Utility Analyses

Based on the results pertaining to Hypothesis Two, it is known that Hypothesis Three is not supported. Incumbent regression equations do not appear to overestimate applicant validity, and therefore, they will not overestimate utility. Nevertheless, the degree of underestimation will be examined by applying the results from the above regression analyses in a Brogden-Cronbach-Gleser utility model of the financial utility gain. Two sets of utility analyses were conducted. First, the results of the cross-validation estimates based on the strict evidence of moderation data were used. This data included the incumbent multiple R-value of 0.167 and the applicant cross-validation estimate of 0.177. Second, the results of the subgroup correlations were used. These values included the incumbent multiple R-value equal to 0.160 and the applicant cross-validation index (R = 0.234).

Selection ratios ranging from 10% to 90% were examined. The number of assumed applicants tested was maintained at 100. In addition, SDy and cost per applicant are held constant. Finally, tenure is held constant at one year. Results are presented in Table 13.

The magnitude of the underestimation of the financial gain is nearly equal to the underestimation of the R-value, regardless of the selection ratio.

Next, the results from the subgroup correlations cross-validation analyses were investigated. The incumbent multiple R-value was equal to 0.160 and the applicant cross-validation index R = 0.234. As with the utility analyses for the strict moderation data presented in Table 13, selection ratios ranging from 10% to 90% were examined. Once again, the number of applicants tested was maintained at 100, SDy was set equal to $28,320, and cost per applicant was set at $12.00. Tenure was held constant at one year. The results of this analysis are presented

in Table 14.

Table 13. Utility Estimates Derived from Strict Evidence of Moderation Analyses

Φ	λ	ΔU: Incumbent Estimate	ΔU: Actual	% Underestimation
0.10	0.18	$81,801	$86,771	5.73%
0.20	0.28	$131,206	$139,135	5.70%
0.30	0.35	$163,239	$173,086	5.69%
0.40	0.39	$181,518	$192,460	5.68%
0.50	0.40	$187,477	$198,775	5.68%
0.60	0.39	$181,518	$192,460	5.68%
0.70	0.35	$163,239	$173,086	5.69%
0.80	0.28	$131,206	$139,135	5.70%
0.90	0.18	$81,801	$86,771	5.73%

Note: Φ = Selection Ratio; λ = Normal curve ordinate at Selection Ratio; ΔU: Incumbent Estimate is the estimated dollar value gain based on the incumbent estimated R = 0.167; ΔU: Actual is the estimated dollar value gain based on the cross-validation coefficient when the incumbent prediction equation is applied to applicant personality scores, and cross-validated against actual (simulated) applicant performance scores (R = 0.177). % Underestimation is the magnitude of the incumbent utility underestimation of the applicant utility estimate. Number of applicants is fixed at 100; SD_y is fixed at $28,320; per applicant testing cost is fixed at $12.

Once again, the magnitude of the underestimation of the financial gain is nearly equal to the underestimation of the R-value, regardless of the selection ratio. The results are more dramatic than in the strict evidence case, and suggest that incumbent based prediction equations can substantially underestimate the actual utility of personality inventories. Under the conditions investigated here, a selection ratio of 50% would result in an estimated economic utility gain that was $75,697 less than the actual gain. As was discussed above, the underestimation is due largely to the fact that sample type moderates the operational validity of Conscientiousness and Optimism, such that these personality attributes are more strongly related to performance in applicant samples. And, there is less apparent overlap in the measurement of Conscientiousness and Optimism in applicant as opposed to incumbent samples.

Table 14. Utility Estimates Derived from Subgroup Correlations

Φ	λ	ΔU: Incumbent Estimate	ΔU: Actual	% Underestimation
0.10	0.18	$78,322	$115,101	31.95%
0.20	0.28	$125,656	$184,327	31.83%
0.30	0.35	$156,346	$229,212	31.79%
0.40	0.39	$173,860	$254,825	31.77%
0.50	0.40	$179,569	$263,174	31.77%
0.60	0.39	$173,860	$254,825	31.77%
0.70	0.35	$156,346	$229,212	31.79%
0.80	0.28	$125,656	$184,327	31.83%
0.90	0.18	$78,322	$115,101	31.95%

Note: Φ = Selection Ratio; λ = Normal curve ordinate at Selection Ratio; ΔU: Incumbent Estimate is the estimated dollar value gain based on the incumbent estimated R = 0.160; ΔU: Actual is the estimated dollar value gain based on the cross-validation coefficient when the incumbent prediction equation is applied to applicant personality scores, and cross-validated against actual (simulated) applicant performance scores (R = 0.234). % Underestimation is the magnitude of the incumbent utility underestimation of the applicant utility estimate. Number of applicants is fixed at 100; SD_y is fixed at $28,320; per applicant testing cost is fixed at $12.

As with Hypothesis Two, there is no support for Hypothesis Three. The findings from both the strict evidence of moderation analyses and the analyses of all subgroup correlations suggest that incumbent derived equations will underestimate the actual utility gain observed in practice (when tests are used to select among applicants).

Summary of Results

The results indicate that there is mixed support for Hypothesis One: some of the bivariate validity estimates from incumbent studies differ from those estimated on the basis of job applicant studies. Hypotheses Two and Three were not supported: incumbent derived equations do not appear to overestimate the overall validity (multiple R) or utility of personality tests in applicant settings. Instead, incumbent studies appear to underestimate the validity and utility of personality tests when used in personnel selection.

Chapter Five: Discussion

The discussion of the results from the current investigation is organized to present first a resolution of the Hypotheses. Next, some limitations of the current study are brought to the reader's attention, and, to the extent possible, these limitations are addressed. Next, a general discussion of the implications of the results for present employee and job applicant validation studies is presented. This is followed by a discussion of some noteworthy operational validity estimates discovered in the present investigation, again with an eye toward implications for the use of personality tests in personnel selection. Finally, some avenues for future research are introduced.

Resolution of Hypothesis One

Hypothesis one posited that criterion-related validity estimates would differ as a function of the sample type (job-incumbent versus job-applicant) utilized in the validation studies. Resolution of this hypothesis relies primarily on the meta-analysis of studies that used performance ratings as the criterion (Table 4). Although the overall analysis would contain more studies and a larger total sample size, it was decided that controlling the potential confound between sample type and criterion type was worth the omission of those studies that did not include a ratings criterion.

Based on the statistical significance tests of differences according to sample type, five of the 12 distributions of observed criterion-related validities were found to be moderated by test-taking status (incumbent versus applicant). Specifically, the criterion-related validities of single-stimulus measures of Neuroticism, Extraversion, and Ambition differed by sample type, while the criterion-related validities of forced-choice measures of Neuroticism and Extraversion varied by sample. Note that the incumbent estimate of the operational validity of single-stimulus measures of Extraversion is eight times greater than the corresponding applicant estimated operational validity. And, the incumbent validity estimate for single-stimulus measures of Ambition is five times larger as the corresponding applicant estimate. However, because the validity of single-stimulus measures of Extraversion and Ambition are so low (operational validity estimates equal to or less than $\rho = 0.10$), their overestimation of the applicant operational validity is scantly worth concern. With regard to forced-choice measures of Extraversion, the operational validity estimate is based on only four studies with a total sample size of 621. As

such, it would be imprudent to place too much faith in this estimate.

According to the statistical significance test of moderation of validity, with the further constraints that the differences: a) would likely be considered practically meaningful; and, b) were based on total sample sizes of at least 1,000 individuals in each subgroup, Hypothesis One is supported for one of the fourteen (seven predictor constructs by two scale types) possible between group comparisons (the criterion-related validity of single-stimulus measures of Neuroticism). The incumbent and applicant operational validity estimates for single-stimulus measures of Neuroticism were $\rho = -0.12$ ($SD_\rho = 0.16$) and $\rho = -0.05$ ($SD_\rho = 0.10$), respectively. A difference of this magnitude would be considered small according to commonly referenced interpretations of effect sizes (e.g., Cohen, 1992; p. 157). Of the other thirteen validity distributions, subgroup analyses were not conducted for two of these (due to an insufficient number of studies), seven were found not to be moderated by sample type, and four were moderated by sample type but do not exhibit practically meaningful differences or are based on too few studies to draw concrete conclusions.

Considering subgroup operational validities and SD_ρ values, in addition to significance tests of sample type as a moderator of the criterion-related validity, Hypothesis One is further supported for single-stimulus measures of Openness, Conscientiousness and Optimism, and forced-choice measures of Openness. Single-stimulus measures of Conscientiousness and Optimism were each more strongly related to performance in applicant (as opposed to incumbent) samples. In both cases, the differences were quite small: Conscientiousness incumbent and applicant operational validities $\rho = 0.13$ ($SD_\rho = 0.14$) and $\rho = 0.17$ ($SD_\rho = 0.02$), respectively; Optimism incumbent and applicant operational validities $\rho = 0.14$ ($SD_\rho = 0.11$) and $\rho = 0.20$ ($SD_\rho = 0.09$), respectively.

To be sure, there was *some* evidence of sample type as a moderator for 10 of the 14 criterion-related validities examined here. The only four validity estimates with *no* documented evidence of moderation according to sample type were forced-choice measures of Conscientiousness, Ambition and Optimism (moderation tests were not able to be conducted), and single-stimulus measures of Agreeableness. All other validities presented evidence of moderation via the t-test, the inspection of subgroup validities and SD_ρ values, or both of these conditions. Of the 10 moderated validities, though, five were so small in both subgroups (absolute values of the operational validity estimates less than or equal to 0.10) that they would

not warrant concern (these included single-stimulus measures of Extraversion, Openness, and Ambition, and forced-choice measures of Neuroticism and Agreeableness). Of the remaining five, two were based on too few studies ($k < 5$) and participants ($N < 625$) in the applicant subgroup to justify firm conviction in the results. Of the remaining three, single-stimulus measures of Neuroticism were more strongly related to performance in incumbent samples, while single-stimulus measures of Conscientiousness and Optimism were more strongly related to performance in applicant samples.

In short, Hypothesis One is supported as sample type demonstrates evidence of moderating 10 of 14 possible validity distributions. However, the direction of the moderating effect varies, with some validity estimates being stronger in incumbent samples, and other validity estimates being stronger in applicant samples. And, the magnitude of the both the operational validity estimates as well as the between-sample type differences in the operational validity estimates were generally small. As was mentioned earlier, though, small statistical differences can be practically meaningful. If a hiring organization calculates an economic utility estimate based on an assumed validity of $\rho = -0.12$ (incumbent operational validity for single-stimulus measures of Neuroticism), when the actual operational validity of the measure when used with job applicants is $\rho = -0.05$, this organization will have overestimated utility by approximately 140%. From this perspective, small validity differences would likely be practically important differences.

Resolution of Hypotheses Two and Three

Hypotheses two and three posited that present-employee validation studies would overestimate the cross-validation coefficient and utility for personality measures when an incumbent-based prediction equation was applied to applicant data. These hypotheses were not supported. Based on the strict requirement of statistically significant differences between sample type in the estimates of criterion-related validities and correlations between predictor constructs, the selected prediction equation based on the incumbent data is smaller (by approximately 6%) than the cross-validation R when applied to applicant data. In turn, the estimated utility gain from the use of personality tests was also approximately 6% lower based on the incumbent data. When all subgroup estimates of validities and predictor inter-correlations were used, the incumbent prediction equation is smaller than the cross-validation R by approximately 30%. As a

consequence, the utility gain was also 30% lower than what would be expected when personality tests are used in applicant settings.

Hypotheses two and three were based largely on the assumption that the pattern of correlations between personality constructs would diverge between incumbent and applicant samples. This has been a question of some interest lately (Smith et al., 2001; Ones & Viswesvaran, 1998, p. 252; Weekley, Ployhart, & Harold, 2003). Despite the lack of support for Hypotheses Two and Three, it is instructive to examine the pattern of inter-correlations among personality traits, and sample type as a moderator of those correlations.

Focusing on the inter-correlations among constructs when measured by the modal scales used in studies that reported a correlation for a given pair of constructs (Table 7), the answer to this question seems to be that sample type generally has a very small moderating effect on the correlations between personality constructs. Based on the strict evidence requirement of a statistically significant t-test, four of the 16 correlations that could be tested for moderation were found to be moderated by sample type. Aside from the tests of statistical significance, the potential moderating effect of sample type was examined by inspecting the subgroup corrected correlations and standard deviations of the corrected correlations. Again, if the corrected correlations differed and the averaged subgroup SD_ρ was less than the overall SD_ρ, it was concluded that sample type was a moderator of the correlations between personality traits. Using this guideline, sample type was identified as a moderator in eight of 16 instances. The magnitude of the differences was very small (an absolute difference of 0.05 or less) in three of the eight cases, and ranged from 0.06 to 0.10 in four cases. There is only one correlation between personality traits that appears to be moderated by sample type to an appreciable degree (difference greater than 0.10) and is based on meta-analytic samples of at least 1,000 participants in each subgroup. This is the correlation between Conscientiousness and Optimism (stronger relationship in the incumbent group; see Table 6).[11]

[11] I also dis-aggregated the construct level correlation between Conscientiousness and Optimism as measured by the CPI into two scale level correlations. This analysis seems to weaken the case for sample type as a meaningful moderator of the personality trait inter-correlations. The weighted mean correlation between Achievement via Conformance (Conscientiousness) and Well-being (Optimism) was r = 0.47 for incumbents and r = 0.44 for applicants. The weighted mean correlation between Achievement via Conformance and Self-acceptance (Optimism) was r = 0.19 for incumbents and r = 0.20 for applicants. In the meta-analytic results, the construct level

Overall, the evidence suggests that sample type does not moderate personality trait inter-correlations to any meaningful degree. On the other hand, it should be pointed out that the inventory used to operationally define personality traits (e.g., Neuroticism, Extraversion, Ambition) *does* influence the resulting correlations between constructs. For example, the sample weighted observed correlation between Openness to Experience and Conscientiousness in the current base of studies is, alternatively, r = 0.34 (Goldberg's big five markers), r = 0.01 (Hogan Personality Inventory), and r = -0.02 (NEO-FFI). Similarly, the sample weighted observed correlation between Extraversion and Conscientiousness is, alternatively, r = 0.32 (NEO-FFI), r = 0.21 (California Psychological Inventory), and r = 0.00 (16PF).

In addition to the belief that inter-correlations among personality traits would differ by sample type (which they ostensibly do not), it was assumed that those differences would matter in the multivariate prediction equation. Not only do the inter-correlations generally not differ by much, but even if they did, they would not matter in the general case. This is because most predictors are not related to performance, and therefore are not included in the prediction equation. Initially it seemed as though the correlation between Conscientiousness and Optimism would present cause for concern. These two traits are related to performance, and it appeared that the correlation between Conscientiousness and Optimism was moderated by sample type. However, as noted in Footnote 11, this was due to the fact that the Ellingson et al. (2001) did not include Well-being as an operational measure of Optimism. As such, there is no consequential evidence of inflated overlap between trait measures in applicant settings, and therefore, there is no evidence that applicant personality profiles will account for diminished unique variance in occupational performance.

Limitations

There are a number of limitations from this study that should be addressed before attempting to draw firm conclusions regarding present-employee and job-applicant validation

---

correlations differ by sample type because of the operational definition of Optimism in the Ellingson et al. (2001) study. Specifically, Ellingson et al. (2001) used the CPI Well-being scale to separate their sample into high and low socially-desirable responding. As such, the estimate of the correlation between Conscientiousness and Optimism from the two large samples in that study were based solely on the correlation between Achievement via Conformance and Self-Acceptance. This led to a downwardly biased estimate of the Conscientiousness-Optimism

samples. First, a number of possible confounds exist that have not been controlled. For example, there is the possibility of differential publication bias according to sample type, such that authors might be less likely to publish studies that have failed to find support for personality measures as predictors of performance in applicant settings. This could happen if the host organization did not wish to publish the fact that an employment tool they had used was not related to performance. The result of such differential suppression of negative results would be upwardly biased estimates of the operational validity in applicant settings. In an attempt to alleviate this concern, an effort was made to obtain unpublished doctoral dissertations, conference presentations, and raw data from researchers and testing specialists. This certainly would not, in and of itself, guarantee that null or unimpressive results would be equally likely to surface, regardless of sample type. Still, based on the findings that in some cases the incumbent validity estimates exceeded the applicant validity estimates, while in other cases the applicant validity estimates were higher, it does not appear that poor results from applicant studies have been universally suppressed at a differential rate than those from incumbent studies.

There are other confounds that may exist, though. Some of these are speculative, while others are known to be present in the existing data. For example, applicant validation studies have historically been viewed as more scientifically rigorous than incumbent validation studies (Guion, 1998). If a researcher or organization is willing to expend the additional time, effort, and money to conduct an applicant-based validation study, it might also be true that they would devote more time and effort into: a) conducting a job analysis; b) linking the job requirements to personal dispositions that would likely be related to success; c) identifying and considering alternative predictor measures; and d) developing a reliable criterion measurement system. If any or all of these were true, the likely result would be more favorable results in applicant studies. Again, though, this does not appear to be a problem that influenced the results in a universal manner, as evidenced by the fact that applicant validity estimates were not uniformly stronger than incumbent validity estimates.

It is possible to determine the number of additional studies averaging a zero correlation that would be needed to decrease the meta-analytic estimate to a specified value. This number is known as the Failsafe N (Hunter & Schmidt, 1990). Two correlations of particular interest are

---

correlation in the applicant sample.

the applicant validity estimates for single-stimulus measures of Conscientiousness and Optimism. In order to eliminate potential concern over differential suppression of null results according to sample type, a failsafe N analysis was conducted. This was done by computing the number of studies averaging a correlation of zero between Conscientiousness (Optimism) and rated job performance that would be needed to lower the meta-analytic observed validity estimate for *applicants* to equal the meta-analytic observed validity estimate for incumbents. The number of applicant studies of Conscientiousness averaging null results that would be needed to lower the applicant estimate to the incumbent estimate is seven, while four applicant studies averaging null results for Optimism would bring the applicant Optimism validity estimate in line with the incumbent Optimism validity estimate.

Often failsafe N analyses are conducted to demonstrate that an improbable number of studies averaging null results would have to exist before there would be concern that the meta-analytic results were unduly influenced by biased availability of studies. For example, in the Ones et al. (1996) meta-analysis of the relationships between social desirability and the big five, they found that a total sample size of 388,244 cases (1,261 studies) averaging null results would need to exist for the true correlation between social desirability and Emotional Stability to be lowered from $\rho = 0.37$ to $\rho = 0.10$. It is reasonable to conclude in their case that the required studies with null results simply would not exist. The same conclusion can not be reached in the current analysis: one would be hard-pressed to make the claim that there are not four studies of Optimism (and seven studies of Conscientiousness) that have been conducted in applicant settings that resulted in an average zero correlation with rated job performance. As such, it should be borne in mind the observed moderating effects of sample type on Conscientiousness and Optimism validity estimates of single-stimulus measures uncovered in this investigation could be overturned by a handful of studies.

More confidence can be placed in the moderating effect of sample type as a moderator of the criterion-related validity of single-stimulus measures of Neuroticism. Specifically, 14 applicant studies averaging $\bar{r} = -0.17$ (the correlation at the 80^{th} percentile of obtained applicant studies) would need to be uncovered for the applicant validity estimate to match the incumbent validity estimate of single-stimulus measures of Neuroticism.

An additional confound is that of the specific personality inventory chosen. While this issue was addressed in part by conducting a hierarchical moderator analysis that crossed scale

type (single-stimulus versus forced-choice) with sample type, the possibility remains that within scale type, there might be widespread utilization of some measures in applicant settings, while in incumbent settings other inventories might be more prevalent. Indeed, the example given earlier remains a relevant case in point. The MMPI (a single-stimulus measure) is popular in applicant (but not incumbent) settings, while the NEO-PI-R (also a single-stimulus measure) is widespread in incumbent (but not applicant) settings. The potential for one of these two measures being more strongly related to occupational performance is a confound left uncontrolled in the current analyses.

A final confound raised here is occupation. It is possible that some occupations are more likely to be included in applicant studies while others may be more commonly studied as incumbents. A case in point is protective service occupations (law enforcement, security guards, and firefighters). Samples of protective service employees comprised 31% of the applicant validation studies for Neuroticism. Protective service occupations made up only 6% of the incumbent validation studies for Neuroticism. If criterion-related validity were related to occupational representation, this would also be a source of bias in the current results.

These criticisms could be countered if the SD_ρ values within each subgroup were zero or near zero. If there were no true variance in the subgroup validities, then scale type or occupation as confounding sources of variance would be moot criticisms. While the SD_ρ values are greater than zero in most subgroup conditions, the SD_ρ value is near zero in one critical subgroup: single-stimulus measures of Conscientiousness in the Applicant condition. So, while the presence of unknown substantive moderators could yield an incumbent validity estimate for single-stimulus measures of Conscientiousness as low as $\rho = -0.05$ or as high as $\rho = 0.31$ (90% confidence limits), the applicant validity would be anticipated to range from $\rho = 0.14$ to $\rho = 0.19$. This reveals that there may be cases when the incumbent-based study would overestimate the applicant validity of single-stimulus measures of Conscientiousness. These cases would be in the minority, though.

Aside from these (as well as other, unmentioned) confounds, a further limitation of this study is that the criterion was ratings of overall job performance. This was selected in an effort to control for criterion as a confound, and because it was the most commonly utilized criterion. Personality measures might be better suited as predictors of specific components of performance, though (Borman & Motowidlo, 1997; J. Hogan & Holland, 2003). The current study is not able

to address whether or not incumbent based studies would overestimate applicant criterion-related validities when predictor and criterion measures are conceptually aligned. Barrick, Stewart, and Piotrowski (2002) argued that status striving would act a mediating variable linking personality to performance. One possibility is that Ambition would predict status striving. If so, the question remains as to whether or not incumbent validation studies provide an accurate representation of the applicant criterion-related validity for conceptually aligned predictors and criteria (such as Ambition and status attained in the organization).

Finally, because not all data were reported in each study, a number of liberties were taken with some of the studies included in these meta-analyses. Some of the correlations among personality constructs were based on reproduced correlations. And, some of the composite score correlations were based on intra-composite correlations that were imputed from other studies. In order to assure that such correlations did not have an undue influence on the results, observed correlations were examined for outliers. None of the studies that were the subject of these permissive decisions was identified as outliers in their distributions.

Present-employee and Job-applicant Samples

One of the foremost implications of the results of this study is that samples of job incumbents seem to provide a reasonable proxy for job applicants in validation studies of personality tests. When differences in the validity and trait inter-correlations were observed, they were generally small. Confidence in the generalizability of the findings from the trait inter-correlation estimates is bolstered by the fact that the correlations reported in Table 6 represent five different personality inventories (16PF, CPI, HPI, NEO-FFI, and NEO-PI-R).

The small and sparse differences between samples on criterion-related validity estimates, combined with the small differences between samples on trait inter-correlations combine to reveal that incumbent based prediction equations do not overestimate the cross-validation coefficient or utility when incumbent equations are applied to applicant data.

It seems pertinent to offer a potential explanation for some findings that appear to be conflicting. Incumbents and applicants have been shown to exhibit mean-level differences in personality attributes (Birkeland et al. 2003; Heron, 1956; Hough, 1998b; Robie et al., 2001; Rosse et al. 1998). However, the inter-correlations among personality traits and the higher-order factor structure do not differ by sample type (current study; Smith et al., 2001). Nor do the

criterion-related validities differ by sample type (current study). It seems peculiar that the means would be markedly influenced by testing circumstances, yet, the correlations with other attributes and external criteria would be unaffected. The most commonly advanced explanation for why increased socially desirable responding would not lead to a degradation in the validity (criterion-related or construct-oriented) of personality measures is that offered by Hogan (1983). As outlined in Chapter Two of this report, Hogan's theory suggests that personality test responses are a form of social communication where the test-taker presents an identity, informing the test-interpreter how he or she would like to be regarded. Furthermore, test-takers are thought to claim an identity that they would sustain on the job. Individuals who are able to adopt an appropriate identity (or role) during the test-taking process might also adopt a successful role on the job. Although the current study does not offer any process oriented data that can confirm or refute this explanation, it remains the explanation offered by most researchers that study applicant personality profiles (Ones & Viswesvaran, 1998; Ruch & Ruch, 1980; Weekley et al., 2003).

An additional potential explanation is that the relationships between personality constructs and occupational performance are so weak that any between group (incumbent versus applicant) differences in roles adopted or test-taking strategies can have very marginal influences on criterion-related validities. This explication is unlikely, based on the results of the correlations between personality trait measures. The correlations between personality trait constructs (Table 6) range from strong and negative (Neuroticism with Extraversion) to near zero (Openness with Conscientiousness) to strong and positive (Extraversion with Optimism). Across the range of magnitudes of relationships, sample type was generally not found to moderate the correlations between personality constructs.

Operational Validity of Personality in Applicant Settings

The current study also has implications for the use of personality as a predictor of job performance. Specifically, it is noted that at the outset of this study, criticisms were levied against existing meta-analyses of personality inventories as predictors of occupational performance on the grounds that test-taking status is rarely, if ever, considered as an important variable to be taken into account. The current study found that the operational validity of single-stimulus measures of Conscientiousness as predictors of performance ratings in applicant settings is $\rho = 0.17$, with nearly all variability in operational validities attributable to sampling

error and statistical artifacts. This estimate is based on 23 studies with a total sample size of 3,147. Although the total sample size is far smaller than those in prior meta-analyses that include primarily incumbent-based studies (e.g., J. Hogan & Holland, 2003), this is an important finding as it demonstrates that Conscientiousness is related to performance not only in incumbent settings, but in applicant settings as well. A failsafe N analysis indicates that 15 studies (total additional N = 2,055) averaging null results would be needed to bring the operational validity estimate for applicant studies of single-stimulus measures of Conscientiousness to $\rho = 0.10$.

Similarly, the operational validity of single-stimulus measures of Optimism as predictors of performance ratings in applicant settings is $\rho = 0.20$. This estimate is based on 10 studies with a total sample size of 1,189, and a failsafe N analysis indicates that 10 studies (total additional N = 1,190) averaging null results would be needed to bring the operational validity estimate for applicant studies of single-stimulus measures of Optimism to $\rho = 0.10$. This finding also highlights the possibility that while the big five provides a convenient organizing taxonomy for personality research, compound personality attributes may be more likely to demonstrate generalizable criterion-related validity across occupational settings. Specifically, most of the big five attributes have been found not to demonstrate generalizable criterion-related validity with overall performance. Only Conscientiousness and compound personality attributes such as integrity (Ones et al., 1993), customer service (Frei & McDaniel, 1998), and optimism (this study) appear to predict job performance across settings.

Future research

This study suggests a number of avenues for future research. First, extending the current study to examine criteria other than ratings of overall performance is in order. There are two aspects of this that need to be addressed by such research. One aspect is the alignment of predictor measures with theoretically relevant criteria. Stewart (1999) showed that different facets of Conscientiousness are related to job performance at different stages of acclimation to a job. J. Hogan and Holland (2003) mapped performance criteria (mostly rating criteria) onto the characteristics assessed by the HPI and found that the strongest criterion-related validity estimate for each predictor was with the conceptually aligned criterion. Demonstrating that results from studies of incumbents generalize to applicant settings when predictors and criteria are more strongly linked would be an important practical contribution.

A second aspect of the criterion problem that would need to be addressed is the issue of the reliability of criterion. That is, while J. Hogan and Holland (2003) aligned predictors and criteria, the criteria they used were primarily rating criteria. Despite the fact that they were ratings of more specific domains of job performance (as opposed to ratings of overall performance), the reliability of the criteria were still likely to be quite low. Viswesvaran et al. (1996) reviewed the reliability of ratings of various dimensions of job performance and found that no dimensions were rated with an average reliability greater than $r_{yy} = 0.52$. As such, it would seem prudent to examine predictors that are conceptually aligned with outcome measures that are measured more reliably than rating criteria (such as promotional progress, productivity and sales, accidents, and turnover).

Second, the usefulness of forced-choice measures as predictors of occupational performance should be reconsidered. The operational validity of forced-choice measures has not been sufficiently examined in previous research, and for that reason, a number of hierarchical subgroup meta-analyses involving forced-choice measures were not conducted here. Forced-choice measures do demonstrate some promise though. The most striking example of the potential benefit of using forced-choice measures comes from an examination of the operational validities of forced-choice measures of Ambition. Across eight studies and 1,966 participants (incumbents and applicants), the operational validity of forced-choice measures of Ambition (predicting ratings criteria) was $\rho = 0.19$. There was, however, substantial variability in the operational validity estimate ($SD_\rho = 0.20$). Identifying factors related to the success of forced-choice measures in predicting performance would seem to be a practically useful endeavor. One possibility is that some forced-choice measures of Ambition are more useful than others. Alternatively, the merits of forced-choice measures could be a function of the nature of the sample being investigated.

Another avenue for research that has been raised by the current analyses is the possibility that sample type would interact with scale type to influence criterion-related validities. There were some predictor constructs (Neuroticism and Extraversion) in the hierarchical subgroup analysis that suggested single-stimulus measures would experience a degradation of validity in applicant samples, while forced-choice measures would exhibit stronger validity in applicant samples (as compared to incumbent samples). Continued investigation of this issue should shed further light on this topic (see also Jackson et al., 2000). It was argued earlier that test-takers

might wish to self-present one or more specific characteristics when responding to a personality inventory in a selection setting. This role-adopting behavior could lead to enhanced validity of personality tests, if successful role-adoption in the test-taking scenario was related to similar role-adoption on the job, and, such role-adoption on the job were related to occupational performance. A forced-choice measure seems an ideal means to force respondents to choose a role or disposition that they wish to highlight. Perhaps in some jobs Extraversion is a more important quality than is Conscientiousness. Perhaps successful applicants for this job would disproportionately endorse the Extraversion response option over the Conscientiousness response option, and in turn, would be better able to enact the role of the Extravert on the job. Although this process by which personality might be related to performance remains speculative at this time, there is some existing research that supports this possibility.

Following the many meta-analytic reviews that have shown personality to be related to job performance, there has been more focused attention on identifying the mediating mechanisms in operation. Much of this research suggests that personality is related to performance via proximal motivational constructs (Lee, Sheldon, & Turban, 2003; Heggestad & Kanfer, 2000). One such example is that conscientious people set higher goals and remain focused on those goals. This goal striving process leads to enhanced performance (Barrick, Mount, & Strauss, 1993; Lee et al., 2003). Barrick et al. (2002) have also found support for Extraversion as a predictor of striving for status, which in turn, was predictive of sales performance. The question of why Optimism is related to performance warrants further consideration. Judge, Erez, and Bono (1998) have provided evidence of the relationship between Core Self-evaluations (Positive Self-concept) and job performance. Their treatment of Core Self-evaluations seems largely consistent with past conceptions of Optimism (Scheier & Carver, 1985). It is likely that Optimism also operates via motivational constructs; optimistic people might set higher goals and be more likely to persist in the face of initial setbacks. Optimistic people are less likely to attribute failure to internal and stable causes and are more likely to persist in the face of difficulties or setbacks (Scheier & Carver, 1985). It is reasonable to conjecture that these are the links between Optimism and performance, but empirical research supporting this process in an employment context were not uncovered by this writer.

Qualitative and quantitative reviews of personality research would benefit from further refinement of the process of sorting personality measurement scales into construct categories.

The effort of Hough and Ones is impressive, and is an important development in the furthering of personality research. At the same time, it is clear that the use of such a taxonomy is not without its shortcomings. Three observations uncovered in the current investigation highlight the limitation of "pigeon-holing" existing operational measures into construct categories in an effort to draw general conclusions about those constructs. All three observations demonstrate that correlations between constructs are moderated by the operational measures of those constructs. First, while Conscientiousness and Optimism are generally positively correlated, they are negatively correlated when operationally defined by the PRF Need for Achievement and Need for Play scales. Second, Conscientiousness and Optimism are more strongly correlated when operationally defined by the CPI Achievement via Conformance and Well-being scales than when operationally defined by the CPI Achievement via Conformance and Self-acceptance scales. Finally, the weighted mean observed correlation between Conscientiousness and Extraversion could alternatively be estimated as $r = 0.00$ (16PF), $r = 0.18$ (all scales in the Hough and Ones taxonomy), $r = 0.21$ (CPI), or $r = 0.32$ (NEO-FFI).

This raises an additional future research need: replication of a study of this nature in a single setting that controls many of the confounds mentioned in the Limitations section above. Although the current study presents what is likely to happen in the general case, as determined by existing validation study data, our confidence that incumbent samples can be used as proxies for job applicants would be strengthened by a comprehensive mixed-sample validation study conducted in a single setting. Such a study would also benefit from examination of multiple criteria that are conceptually aligned with each predictor construct. That is, as opposed to examining overall job performance, such a study would take an approach similar to that adopted by J. Hogan and Holland (2003) in their meta-analysis. These authors paired personality constructs with performance criteria constructs in an investigation of the criterion-related validity of the HPI. It was hoped that such a study could be conducted concurrently with the research presented here. Unfortunately, only one study was identified that met many of these criteria (Sinclair & Michel, 2001). That study included a large sample of applicants ($N > 500$) but a far smaller sample of incumbents (approximate $N = 100$). This discrepancy between the two samples precluded using that data here.

Finally, future research should consider the generalizability of present-employee validation studies for job interviews, biodata, and other assessment devices. Ones and

Viswesvaran (1998) have noted that personality tests are frequently lambasted for being susceptible to mis-representation by job applicants. It is the opinion of this author that self-presentational differences between incumbents and applicants are a ***potential*** concern, regardless of the assessment device. The current study provides support for incumbent samples as substitutes for applicant samples in the context of personality inventories. Be that as it may, the study by Stokes et al. (1993) suggests that incumbents are ***not*** a reasonable proxy for applicants in the validation of biodata measures. Much as the Barrett et al. (1981) analysis of concurrent and longitudinal validation study designs dealt primarily with cognitive ability tests, this study dealt solely with self-report personality inventories. Rather than assuming that the outcome of this study reflects the outcome for sample type comparisons among alternative predictors (situational judgment tests, structured interviews, assessment centers), the interchangeability of incumbents and applicants for these devices remains an empirical question. While each of these assessment techniques has demonstrated moderate to high levels of criterion-related validity, perhaps with increased understanding of applicants' frames of reference and self-presentational processes, these devices could provide still improved levels of performance forecasting.

Conclusion

The use of job incumbent samples as substitutes for job applicants in personality-test validation research has been questioned in the past (Guion, 1998; Stokes et al., 1993). Despite these and other authors' reservations about the use of present-employee samples, their use persists in practice. Despite noted limitations, the current study does not provide compelling evidence abrogating the use of incumbent samples in the validation of personality tests. Rather, the current study provides strengthened justification for utilization of present-employees when validating personality tests for personnel selection. Where there were differences between incumbents and applicants, they were generally quite small. While commonsense might dictate that the differential motivating context between job incumbents and job applicants would severely skew the results, the data presented here lead to the conclusion that context effects are minimal.

References

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

∗ Abbott, S. L. (1996). Forming and enacting job perceptions: An investigation of personality determinants and job performance outcomes. Unpublished doctoral dissertation, New York University.

Aguinis, H. & Pierce, C. A. (1998). Testing moderator variable hypotheses meta-analytically. Journal of Management, 24, 577-592.

Aguinis, H., Sturman, M. C., & Pierce, C. A. (2002, April). Comparison of three meta-analytic procedures for estimating moderating effects of categorical variables. In J. M. Cortina (Chair), in the hidden part, you will make me to know wisdom. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

∗ Alker, H. A., Straub, W. F., & Leary, J. (1973). Achieving consistency: A study of basketball officiating. Journal of Vocational Behavior, 3, 335-343.

∗ Allworth, E. & Hesketh, B. (1999). Construct-oriented biodata: Capturing change-related and contextually relevant future performance. International Journal of Selection & Assessment, 7, 97-111.

∗ Anderson, D. W. & Goffin, R. D. (2001, April). Does personality testing lead to gender bias in selecting managers? Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

∗ Ansel, E. (1968). A study of attitudes toward suicide attempters. Unpublished master's thesis, University of Florida.

∗ Antonioni, D. & Park, H. (2001). The effect of personality similarity on peer ratings of contextual work behaviors. Personnel Psychology, 54, 331-360.

∗ Arthur, W., Jr., & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. Journal of Personality, 64, 593-618.

∗ Arvey, R. D., Mussio, S. J., & Payne, G. (1972). Relationships between Minnesota Multiphasic Personality Inventory scores and job performance measures of fire fighters. Psychological Reports, 31, 199-202.

∗ Ash, P. (1960). Validity Information Exchange #13-05. Personnel Psychology, 13, 454.

* Avolio, B. & Howell, J. M. (1992). Impact of leadership behavior and leader-follower personality match on satisfaction and unit performance. In K. E. Clark (Ed.), Impact of Leadership (pp. 225-234). Greensboro, NC: Center for Creative Leadership.

* Bajaj, D. R. (1971). The relationship of certain personality traits to selected professional and social attributes of Oklahoma male county field extension personnel. Unpublished doctoral dissertation, Oklahoma State University.

* Baldwin, T. S. (1961). The relationships among personality, cognitive, and job performance variables. Unpublished doctoral dissertation, Ohio State University.

Baron, H. (1996). Strengths and limitations of ipsative measurement. Journal of Occupational and Organizational Psychology, 69, 49-56.

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. Journal of Applied Psychology, 66, 1-6.

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, 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, 715-722.

* Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. Journal of Applied Psychology, 86, 43-51.

* Barsness, R. E. (1989). Predictors of successful adaptation of North American English teachers in the Peoples Republic of China. Unpublished doctoral dissertation, California School of Professional Psychology.

* Bartol, C. R. (1991). Predictive validation of the MMPI for small-town police officers who fail. Professional Psychology: Research and Practice, 22, 127-132.

∗ Bartol, C. R., Bergen, G. T., Volckens, J. S., & Knoras, K. M. (1992). Women in small-town policing: Job performance and stress. Criminal Justice & Behavior, 19, 240-259.

∗ Bartram, D. (1992). The Personality of UK Managers: 16PF Norms for short-listed applicants. Journal of Occupational and Organizational Psychology, 65, 159-172.

∗ Bartram, D. & Dale, H. C. (1982). The Eysenck Personality Inventory as a selection test for military pilots. Journal of Occupational Psychology, 55, 287-296.

∗ Bass, B. D. (1983). The relationship of selected personality factors and successful performance in the nursing profession. Unpublished doctoral dissertation, University of San Francisco.

∗ Bass, B. M. (1957). Validity Information Exchange #10-25. Personnel Psychology, 10, 343-344.

∗ Bass, B. M., Karstendiek, B., McCullough, G., & Pruitt, R. C. (1954). Validity information exchange #7-024. Personnel Psychology, 7, 159-160

∗ 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, 125-148.

∗ Begley, T. M., Lee, C., & Czajka, J. M. (2000). The relationships of Type A behavior and optimism with job performance and blood pressure. Journal of Business and Psychology, 15, 215-227.

∗ Bell, S. & Arthur, W. (2002, April). Participant personality characteristics that influence feedback acceptance in assessment centers. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

∗ Bendig, A. W. (1955) Ability and personality characteristics of introductory psychology instructors. Journal of Educational Research, 48, 705-709.

∗ Bennett, M. (1977). Testing management theories cross-culturally. Journal of Applied Psychology, 62, 578-581.

∗ Bergman, M., Donovan, M. A., & Drasgow, F. (2001, April). Situational judgment, personality, and cognitive ability: Are we really measuring different constructs? In L. R. Taylor (Chair), Situational Judgment Tests: Assessing the assessments. Symposium presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

∗ Bernstein, I. H. (1980). Security guard MMPI profiles: Some normative data. Journal

of Personality Assessment, 44, 377-380.

∗ Bernstein, I. H., Schoenfeld, L. S., & Costello, R. M. (1982). Truncated component regression, multicollinearity, and the MMPI's use in a police officer selection setting. Multivariate Behavioral Research, 17, 99-116.

∗ Bhandari, A. & Tayal, R. (1990). Executive success in relation to personality and motivational patterns. Social Science International, 6, 28-34.

∗ Bing, M. N. & Burroughs, S. M. (2001). The predictive and interactive effects of equity sensitivity in teamwork-oriented organizations. Journal of Organizational Behavior, 22, 271-290.

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

∗ Birenbaum, M. & Montag, I. (1986). On the Location of the Sensation Seeking Construct in the Personality Domain. Multivariate Behavioral Research, 21, 357-373.

∗ Bishop, N. B., Barrett, G. V., Doverspike, D., Hall, R. J., Svyantek, D. J. (1999, May). Big five and selection: Factors impacting responses and validities. Poster presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

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

∗ Bluen, S. D., Barling, J., & Burns, W. (1990). Predicting sales performance, job satisfaction, and depression by using the Achievement Strivings and Impatience-Irritability dimensions of Type A behavior. Journal of Applied Psychology, 75, 212-216.

∗ Bohle, P. & Tilley, A. J. (1998). Early experience of shiftwork: Influences on attitudes. Journal of Occupational and Organizational Psychology, 71, 61-79.

∗ Bonnin, R. M. (1970). An assessment of relationships between certain personality variables and teacher performance in teaching assignments of higher and lower difficulty. Unpublished doctoral dissertation, University of California, Berkeley.

∗ Booth, R. F., Webster, E. G., & McNally, M. S. (1976). Schooling, occupational motivation, and personality as related to success in paramedical training. Public Health Reports, 91, 533-537.

Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual

performance: The meaning for personnel selection. Human Performance, 10, 99-109.

* Boudreau, J. W., Boswell, W. R., Judge, T. A. (2001). Effects of personality on executive career success in the United States and Europe. Journal of Vocational Behavior, 58, 53-81.

* Boudreau, R. A. (1981). Determining desirable criteria for success as a campus director at Fisher Junior College. Unpublished doctoral dissertation, Boston University.

* Bray, D. W. & Grant, D. L. (1966). The Assessment Center in the measurement of potential for business management. Psychological Monographs: General and Applied, 80 (Whole No. 625), 1-27.

* Brayfield, A. H. & Marsh, M. M. (1957). Aptitudes, interests, and personality characteristics of farmers. Journal of Applied Psychology, 41, 98-103.

* Brendle, M., Switzer, F. S., Stewart, W. H. Jr., & St. John, C. H. (2002, April). Personality and Company Culture: Important Contributions to Innovation for Small Businesses. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

Brown, S. P. (1996). A meta-analysis and review of organizational research on job involvement. Psychological Bulletin, 120, 235-255.

* Brown, S. P., Cron, W. L., & Slocum, J. W. Jr. (1998). Effects of trait competitiveness and perceived intraorganizational competition on salesperson goal setting and performance. Journal of Marketing, 62, 88-98.

* Bruce, M. M. (1956). Validity information exchange #9-31. Personnel Psychology, 9, 373-374.

* Bruce, M. M. (1957). Validity information exchange #10-3. Personnel Psychology, 10, 77-78.

* Burke, D. M. & Hall, M. (1986). Personality characteristics of volunteers in a Companion for Children program. Psychological Reports, 59(2, Pt 2), 819-825.

Bycio, P. (1992). Job performance and absenteeism: A review and meta-analysis. Human Relations, 45, 193-220.

* 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, 67-88.

Campbell, J. P. (1990). An overview of the Army selection and classification project (Project A). Personnel Psychology, 43, 231-239.

∗ Campbell, J. T., Otis, J. L., Liske, R. E., & Prien, E. P. (1962). Assessments of Higher level personnel: II. Validity of the over-all assessment process. Personnel Psychology, 15, 63-74.

Cantor, N., Mischel, W., & Schwartz, J. C. (1982). A prototype analysis of psychological situations. Cognitive Psychology, 14, 45-77.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. Psychometrika, 12, 197-220.

∗ 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, 694-704.

∗ Chan, D. (1996). Criterion and Construct Validation of an Assessment Centre. Journal of Occupational and Organizational Psychology, 69, 167-181.

∗ Chan, D. & Schmitt, N. (2002). Situational judgment and job performance. Human Performance, 15, 233-254.

∗ Chan, K. –Y. & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. Journal of Applied Psychology, 86, 481-498.

∗ Chang, Y. –L. (1996). Biodata, peer nomination, psychological testing, cognitive complexity and managerial performance: A new perspective of managerial selection and development in Taiwan. Unpublished doctoral dissertation, University of Southern Mississippi.

∗ Chay, Y. W. (1993). Social Support, individual differences and well-being: A study of small business entrepreneurs and employees. Journal of Occupational and Organizational Psychology, 66, 285-302.

∗ Christensen, G. E. (2000). Behavioral rehabilitative programming for offenders: Staff selection, operation, and administration. Unpublished doctoral dissertation, Walden University.

∗ 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, 847-860.

∗ Church, A. H. & Waclawski, J. (1998) The relationship between individual personality orientation and executive leadership behavior. Journal of Occupational and Organizational Psychology, 71, 99–125.

Churchill, G. A., Ford, N. M., Hartley, S. W., & Walker, O. C. (1985). The determinants of salesperson performance: A meta-analysis. Journal of Marketing Research, 22, 103-118.

* Chusmir, L. H.& Koberg, C. S. (1989). Perceived Work Competency and Sex Role Conflict: An empirical study. Journal of Psychology, 123, 537-546.

* Clark, W. H. (1970). The relationships of personality and performance to motivator and hygiene orientations. Unpublished doctoral dissertation, Case Western Reserve University.

* Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. Journal of Applied Psychology, 86, 410-417.

* Cohen, D. M. (1982). Relationship between personality and cognitive characteristics of raters and ratees in relation to job performance evaluation. Unpublished doctoral dissertation, New York University.

Cohen, J. (1992). A Power Primer. Psychological Bulletin, 112, 155-159.

* Colarelli, N. J. & Siegel, S. M. (1964). A method of police personnel selection. Journal of Criminal Law, Criminology, and Police Science, 55, 287-289.

* Collins, J. M. & Barucky, J. M. (1999). Big Five Personality Factors Research Using Christal's Self Description Inventory. Bryan, TX: Metrica, Inc. (NTIS No. AD-A364-039-XAB).

* Collins, J. M. & Gleaves, D. H. (1998). Race, job applicants, and the five factor model of personality: Implications for black psychology, industrial/organizational psychology, and the five-factor theory. Journal of Applied Psychology, 83, 531-544.

* Collins, W. E., Schroeder, D. J., & Nye, L. G. (1991). Relationships of anxiety scores to screening and training status of air traffic controllers. Aviation Space and Environmental Medicine, 62, 236-240.

* Connelly, M. S., Gilbert, J. A., Zaccaro, S. J., Threlfall, K. V., Marks, M. A., & Mumford, M. D. (2000). Exploring the relationship of leadership skills and knowledge to leader performance. Leadership Quarterly, 11, 65-86.

* Conte, J. M. & Jacobs, R. R. (1999, May). Temporal and personality predictors of absence and lateness. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. Journal of Applied Psychology, 84, 3-13.

* Conway, J. M. (2000). Managerial performance development constructs and

personality correlates. Human Performance, 13, 23-46.

* Cope, J. R. (1981). Personality characteristics of successful versus unsuccessful police officers. Unpublished doctoral dissertation, Florida Institute of Technology.

* Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "Big Five" personality factors in the IPI and MMPI: Predictors of police performance. Personnel Psychology, 45, 119-140.

* CPP, Inc. (undated). [California Psychological Inventory]. Unpublished raw data.

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

* Crant, J. M. & Bateman, T. S. (2000). Charismatic leadership viewed from above: The impact of proactive personality. Journal of Organizational Behavior, 21, 63-75.

* Curtis, B. G. (1984). Predicting attrition under the Trainee Discharge Program for regular Army soldiers in the military police school. Unpublished doctoral dissertation, University of San Francisco.

* Cutchin, G. C. (1998). Relationships between the big five personality factors and performance criteria for in-service high-school teachers. Unpublished doctoral dissertation, Purdue University.

* Day, D. V; Bedeian, A. G. (1991). Predicting job performance across organizations: The interaction of work orientation and psychological climate. Journal of Management, 17, 589-600.

De Fruyt, F. & Mervielde, I. (1999). RIASEC types and big five traits as predictors of employment status and nature of employment. Personnel Psychology, 52, 701-727.

* de Jong, R. D., Bouhuys, S. A., & Barnhoorn, J. C. (1999). Personality, self-efficacy, and functioning in management teams: A contribution to validation. International Journal of Selection and Assessment, 7, 46-49.

* Deary, I. J., Blenkin, H., Agius, R. M., Endler, N. S., Zealley, H., & Wood, R. (1996). Models of job-related stress and personal achievement among consultant doctors. British Journal of Psychology, 87, 3-29.

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

* Dicken, C. (1969). Predicting the success of Peace Corps community development workers. Journal of Consulting and Clinical Psychology, 33, 597-606.

Donovan, J. J. & Radosevich, D. J. (1998). The moderating role of goal commitment on the goal difficulty-performance relationship: A meta-analytic review and critical reanalysis. Journal of Applied Psychology, 83, 308-315.

* Dorner, K. R. (1991). Personality characteristics and demographic variables as predictors of job performance in female traffic officers. Unpublished doctoral dissertation, United States International University.

* Doster, J. A., Wilcox, S. E., Lambert, P. L., Rubino-Watkins, M. F., Goven, A. J., Moorefield, R., & Kofman, F. (2000) Stability and Factor Structure of the Jackson Personality Inventory-Revised. Psychological Reports, 86, 421-428.

* Dozier, L. A. (1980). Relationship of personality trait scores and selection test scores to job performance of selected agents in the Alabama Cooperative Extension Service. Unpublished doctoral dissertation, Auburn University.

* Drucker, E. H. & Schwartz, S. (1973, January). The prediction of AWOL, military skills, and leadership potential (Technical Report No. 73-1). Fort Knox, KY: Human Resources Research Organization.

* Dugan, R. D. (1961). Validity Information Exchange #14-01. Personnel Psychology, 14, 213-216.

* Dyer, E. D. (1967). Nurse performance description: Criteria, predictors, and correlates. Salt Lake City, UT: University of Utah Press. (NTIS No. HRP-0016927-6).

* Eichinger, R. W., Jr. (1975). A behavior validation of standard assessment tests and inventories for industrial employees. Unpublished doctoral dissertation, University of Minnesota.

* Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. Journal of Applied Psychology, 86, 122-133.

* Elliott, L. L. (1960). WAF Performance on the California Psychological Inventory (Technical Report WADD-TN-60-218). Lackland Air Force Base, TX: Wright Air Development Division.

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

* Fagenson, E. A. (1992). Mentoring - Who needs it? A comparison of protégés' and

nonprotégés' needs for power, achievement, affiliation, and autonomy. Journal of Vocational Behavior, 41, 48-60.

* Ferris, G. R., Youngblood, S. A., & Yates, V. L. (1985). Personality, training performance, and withdrawal: A test of the person-group fit hypothesis for organizational newcomers. Journal of Vocational Behavior, 27, 377-388.

* 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.

* Fitzpatrick, E. D. & McCarty, J. J. (1955). Validity information exchange #8-35. Personnel Psychology, 8, 501-504.

* Fletcher, C. & Baldry, C. (2000). A study of individual differences and self-awareness in the context of multi-source feedback. Journal of Occupational and Organizational Psychology, 73, 303-319.

* Fogarty, G. J., Machin, M. A., Albion, M. J., Sutherland, L. F., Lalor, G. I., & Revitt, S. (1999). Predicting occupational strain and job satisfaction: The role of stress, coping, personality, and affectivity variables. Journal of Vocational Behavior, 54, 429–452.

Ford, N. M., Walker, O. C., Churchill, G. A., & Hartley, S. W. (1987). Selecting successful salespeople: A meta-analysis of biographical and psychological selection criteria. In M. J. Houston (Ed.), Review of Marketing (pp. 90-131). Chicago: American Marketing Association.

* Frederick, J. B. (1985). Supervisors' performance ratings correlated with selected personal characteristics of attendants in a mental retardation developmental center. Unpublished doctoral dissertation, Bowling Green State University.

Frei, R. L. & McDaniel, M. A. (1998). Validity of customer service measures in personnel selection: A review of criterion and construct evidence. Human Performance, 11, 1-27.

* Funk A. P. (1997). Psychological assessment of military federal agents using the MMPI-2: A look at employment selection and performance prediction. Unpublished master's thesis, Florida State University.

* Furnham, A. (1991). Personality and occupational success: 16PF correlates of cabin crew performance. Personality and Individual Differences, 12, 87-90.

* Furnham, A. (1994). The Validity of the SHL Customer Service Questionnaire

(CSQ). International Journal of Selection and Assessment, 2, 157-165.

∗ Furnham, A. (1996). The big five versus the big four. Personality and Individual Differences, 21, 303-307.

∗ Furnham, A., Jackson, C. J., Forde, L., & Cotter, T. (2001). Correlates of the Eysenck Personality Profiler. Personality and Individual Differences, 30, 587-594.

∗ 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, 145-153.

∗ Gannon, M. J. & Nothern, J. C. (1971). A comparison of short-term and long-term part-time employees. Personnel Psychology, 24, 687-696.

∗ Garrity, M. J. & Raymark, P. H. (2002, April). A Causal Model of Accident Severity and Frequency. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

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

∗ Gellatly, I. R., Paunonen, S. V., Meyer, J. P., Jackson, D. N., & Goffin, R. D. (1991). Personality, vocational interest, and cognitive predictors of managerial job performance and satisfaction. Personality and Individual Differences, 12, 221-231.

∗ George, J. M. & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. Journal of Applied Psychology, 86, 513-524.

∗ Geraghty, M. F. (1987). The California Personality Inventory Test as a predictor of law enforcement officer job performance. Unpublished doctoral dissertation, Florida Institute of Technology.

∗ Ghiselli, E. E. (1956). Correlates of Initiative. Personnel Psychology, 9, 311-320.

∗ Ghiselli, E. E. (1969). Prediction of success of stockbrokers. Personnel Psychology, 22, 125-130.

∗ Ghiselli, E. E. (1963). The validity of management traits in relation to occupational level. Personnel Psychology, 16, 109-113.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

Ghiselli, E. E. & Barthol, R. P. (1953). The validity of personality inventories in the selection of employees. Journal of Applied Psychology, 37, 18-20.

* Giebink, J. W. & Stover, D. O. (1969). Adjustment, mental health opinions, and proficiency of child care personnel. Journal of Consulting and Clinical Psychology, 33, 532-535.

* Gniatczyk, L. A. (2001). An examination of career progress in a downsizing organization. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

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, 746-756.

* Götz, K. O. & Götz, K. (1979). Personality characteristics of successful artists. Perceptual and Motor Skills, 49, 919-924.

Gough, H. G. (1989). The California Psychological Inventory. In C. S. Newmark et al. (Ed). Major psychological assessment instruments, Vol. 2. (pp. 67-98). Needham Heights, MA: Allyn & Bacon.

* Gough, H. G., Bradley, P., & McDonald, J. S. (1991). Performance of residents in anesthesiology as related to measures of personality and interests. Psychological Reports, 68(3, Pt 1), 979-994.

* Graham, W. K. & Calendo, J. T. (1969). Personality correlates of supervisory ratings. Personnel Psychology, 22, 483-487.

Green, R. F. (1951). Does a selection situation induce testees to bias their answers on interest and temperament tests? Educational and Psychological Measurement, 11, 503-515.

* Griffin, M. A. (2001). Dispositions and work reactions: a multilevel approach. Journal of Applied Psychology, 86, 1142-1151.

* Griffith, T. L. (1991). Correlates of police and correctional officer performance. Unpublished doctoral dissertation, Florida State University.

* Grimsley, G. & Jarrett, H. F. (1973). The relation of past managerial achievement to test measures obtained in the employment situation: Methodology and results. Personnel Psychology, 26, 31-48.

* Guilford, J. S. (1952). Temperament traits of executives and supervisors measured by the Guilford Personality Inventories. Journal of Applied Psychology, 36, 228-233.

Guion, R. M. (1998). Assessment, measurement, & prediction for personnel decision. Mahwah, NJ: Lawrence Earlbaum Associates, Publishers.

Guion, R. M., & Cranny, C. J. (1982). A note on concurrent and predictive validity designs: A critical reanalysis. Journal of Applied Psychology, 67, 239-244.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-165.

∗ Haaland, S. & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. Personnel Psychology, 55, 137-163.

∗ Hakstian, A. R., Scratchley, L. S., MacLeod, A. A., Tweed, R. G., & Siddarth, S. (1997). Selection of telemarketing employees by standardized assessment procedures. Psychology and Marketing, 14, 703-726.

∗ Hankey, R. O. (1968). Personality correlates in a role of authority: The police. Unpublished doctoral dissertation, University of Southern California.

∗ Hansen, C. P. (1990, August). Personality correlates of success in insurance sales. Paper presented at the Annual Convention of the American Psychological Association, Boston, MA.

∗ Hart, P. M. (1999). Predicting employee life satisfaction: A coherent model of personality, work and nonwork experiences, and domain satisfactions. Journal of Applied Psychology, 84, 564-584.

Hauenstein, N. M. A. (1998, April). Faking personality tests: Does it really matter? In M. McDaniel (Chair), Applicant faking with non-cognitive tests: Problems and solutions. Symposium presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Hauenstein, N. M. A., McGonigle, T., & Flinder, S. W. (2001). A meta-analysis of the relationship between procedural justice and distributive justice: Implications for justice research. Employee Responsibilities and Rights Journal, 13, 39-56.

∗ Hayes, T. L., Roehm, H. A., & Castellano, J. P. (1994). Personality correlates of success in total quality manufacturing. Journal of Business and Psychology, 8, 397-411.

Hedges, L. V. & Olkin, I. (1985). Statistical Methods for Meta-Analysis. San Diego, CA: Academic Press.

Heggestad, E. D. & Kanfer, R. (2000). Individual differences in trait motivation: Development of the Motivational Trait Questionnaire. International Journal of Educational Research, 33, 751-766.

∗ Helmreich, R. L., Spence, J. T., & Pred, R. S. (1988). Making it without losing it: Type A, achievement motivation, and scientific attainment revisited. Personality and Social Psychology Bulletin, 14, 495-504.

Helmreich, R. L., Sawin, L. L., & Casrud, A. L. (1988). The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. Journal of Applied Psychology, 71, 185-188.

∗ Hense, R. L., III (2000). The Big Five and contextual performance: Expanding person-environment fit theory. Unpublished doctoral dissertation, University of South Florida.

Heron, A. (1956). The effects of real-life motivation on questionnaire response. Journal of Applied Psychology, 40, 65-68.

∗ Hiatt, D. & Hargrave, G. E. (1988). MMPI profiles of problem peace officers. Journal of Personality Assessment, 52, 722-731.

∗ 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, 472-478.

∗ Hofer, S. M., Horn, J. L., & Eber, H. W. (1997). A robust five-factor structure of the 16PF: Strong evidence from independent and confirmatory factorial invariance procedures. Personality and Individual Differences, 23, 247-269.

Hoffman, C. C. (1995). Applying range restriction corrections using published norms: Three case studies. Personnel Psychology, 48, 913-923.

∗ Hoffman, R. G. & Davis, G. L. (1995). Prospective validity study: CPI work orientation and managerial potential scales. Educational and Psychological Measurement, 55, 881-890.

∗ Hogan Assessment Systems (undated). [Hogan Personality Inventory Validation Studies]. Unpublished raw data.

Hogan, J. & Hogan, R. (1998). Theoretical frameworks for assessment. In P. R. Jeanneret & R. Silzer (Eds.) Individual Psychological Assessment: Predicting behavior in organizational settings. San Francisco: Jossey-Bass.

* Hogan, J., Hogan, R., & Gregory, S. (1992). Validation of a sales representative selection inventory. Journal of Business and Psychology, 7, 161-171.

* Hogan, J., Hogan, R., & Murtha, T. (1992). Validation of a personality measure of managerial performance. Journal of Business and Psychology, 7, 225-237.

Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. Journal of Applied Psychology, 88, 100-112.

* 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. (1971). Personality characteristics of highly rated policemen. Personnel Psychology, 24, 679-686.

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

Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of Industrial and Organizational Psychology (2nd edition, Vol. 2, pp. 873-919). Palo Alto, CA: Consulting Psychologists Press.

Hogan, R. & Hogan, J. (1992). Hogan Personality Inventory Manual (2nd edition). Tulsa, OK: Hogan Assessment Systems.

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

* Hoiberg, A. & Pugh, W. M. (1978). Predicting Navy effectiveness: Expectations, motivation, personality, aptitude, and background variables. Personnel Psychology, 31, 841-852.

* Hojat, M. & Lyons, K. (1998). Psychosocial characteristics of female students in the allied health and medical colleges: Psychometrics of the measures and personality profiles. Advances in Health Sciences Education, 3, 119-132.

Holland, J. L. (1979). The Self-Directed Search professional manual. Palo Alto, CA: Consulting Psychologists Press.

* Holmes, F. J. (1950a). Validity of tests for insurance office personnel. Personnel Psychology, 3, 57-69.

* Holmes, F. J. (1950b). Validity of tests for insurance office personnel. II. Personnel Psychology, 3, 217-220.

∗ Horváth, M. Frantík, E., & Kozená, L. (1997). Work stress and health in research and development personnel. Homeostasis, 38, 73-82.

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

Hough, L. M. (1998a). Personality at work: Issues and evidence. In M. D. Hakel (Ed.) Beyond Multiple Choice: Evaluating alternatives to traditional testing for selection. Mahwah, NJ: Lawrence Earlbaum.

Hough, L. M. (1998b). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. Human Performance, 11(2/3), 209-244.

Hough, L. M. & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. R. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), Handbook of Work Psychology (pp. 233-277). New York, NY: Sage.

Hough, L. M. & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In K. R. Murphy (Ed.), Individual Differences and Behavior in Organizations (pp. 31-88). San Francisco, CA: Jossey-Bass.

∗ Howard, A. (1986). College experiences and managerial performance [Monograph]. Journal of Applied Psychology, 71, 530-552.

Howell, D. C. (2003). Generating data with a fixed intercorrelation matrix. Internet file, retrieved from http://www.uvm.edu/~dhowell/StatPages/More_Stuff/CorrGen2.html on April 6, 2003.

∗ Howell, M. A. (1966). Personal effectiveness of physicians in a federal health organization. Journal of Applied Psychology, 50, 451-459.

∗ Hueber, J. (1954). Validity information exchange #7-089. Personnel Psychology, 7, 565-566.

Huffcutt, A. I. & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. Journal of Applied Psychology, 79, 184-190.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. Journal of Applied Psychology, 81, 459-473.

∗ Hughes, J. L. & Dodd, W. E. (1961). Validity versus stereotype: Predicting sales

performance by ipsative scoring of a personality test. Personnel Psychology, 14, 343-355.

* Hui, H., Cheng, K., & Yiqun, G. (2000). Predicting work performance with a contextual, narrow-based personality questionnaire: The Chinese experience. In M. J. Gelfand (Chair), Cross-cultural I-O Psychology: Expanding western theories of work behavior. Symposium presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

* Hundal, P. S. & Singh, S. (1978). Some correlates of progressive farm behaviour. Journal of Occupational Psychology, 51, 327-332.

Hunter, J. E. & Schmidt, F. L. (1990). Methods of Meta-Analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage.

Hunter, J. E. & Schmidt, F. L. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. Journal of Applied Psychology, 79, 171-177.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75, 28-42.

* Hunthausen, J. M.. (2000). Predictors of task and contextual performance: Frame-of-reference effects and applicant reaction effects on selection system validity. Unpublished doctoral dissertation, Portland State University.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. Journal of Applied Psychology, 85 (6), 869-879.

* Ingenohl, I. (1961). The Significance of the no-count on the Bernreuter Personality Inventory. The Journal of Social Psychology, 54, 127-140.

* Inwald, R. E. & Brockwell, A. L. (1991). Predicting the performance of government security personnel with the IPI and MMPI. Journal of Personality Assessment, 56, 522-535.

* Inwald, R. E. & Shusman, E. J. (1984). Personality and performance sex differences of law enforcement officer recruits. Journal of Police Science and Administration, 12, 339-347.

* Inwald, R. E. & Shusman, E. J. (1984). The IPI and MMPI as predictors of academy performance for police recruits. Journal of Police Science and Administration, 12, 1-11.

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

* Jackson, C. J., Furnham, A., & Miller, T. (2001). Moderating effect of ear preference

on personality in the prediction of sales performance. Laterality, 6, 133-140.

Jackson, D. N. (1999). Personality Research Form Manual (3rd edition). Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? Human Performance, 13, 371-388.

* Jacobs, R. L. (1992). Moving up the corporate ladder: A longitudinal study of motivation, personality and managerial success in women and men. Unpublished doctoral dissertation, Boston University.

* 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.

James, L. R. (1998). Measurement of personality via conditional reasoning. Organizational Research Methods, 1, 131-163.

* Jenkins, M. & Griffith, R. (2002). Using personality constructs to predict performance: Narrow or broad bandwidth. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

* Johnson, D. L. (1993). Competitiveness and performance in the workforce: Hierarchical factor analysis of managerial competitiveness, achievement motivation, and the big five personality dimensions. Unpublished doctoral dissertation, Iowa State University.

* Johnson, J. W., Schneider, R. J., & Oswald, F. L. (1997). Toward a taxonomy of managerial performance profiles. Human Performance, 10, 227-250.

* Joseph, B. A. (1977). Prediction of successful job performance for department store buyers through an evaluation of personality and background data. Unpublished doctoral dissertation, University of Texas.

* Judge, T. A. & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. Personnel Psychology, 50, 359-394.

Judge, T. A., Erez, A., & Bono, J. E. (1998). The power of being positive: The relation between positive self-concept and job performance. Human Performance, 11(2/3), 167-187.

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

* Judge, T. A. & Heller, D. (2002). The Dispositional Sources of Job Satisfaction: An

integrative test. In R. Ilies (Chair), Dispositional influences on work-related attitudes. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

* Judge, T. A., Thoresen, C. J., Pucik, V., & Welbourne, T. M. (1999). Managerial coping with organizational change: A dispositional perspective. Journal of Applied Psychology, 84, 107-122.

* Kleiman, L. S. (1978). Ability and personality factors moderating the relationships of police academy training performance with measures of selection and job performance. Unpublished doctoral dissertation, University of Tennessee.

* Kleiman, L. S. & Gordon, M. E. (1986). An examination of the relationship between police training academy performance and job performance. Journal of Police Science and Administration, 14, 293-299.

Klimoski, R. J. (1993). Predictor constructs and their measurement. In N. Schmitt, W. Borman, and Associates (Eds.) Personnel Selection in Organizations (pp. 99-134). San Francisco, CA: Jossey-Bass.

Kluger, A. N. & Tikochinsky, J. (2001). The error of accepting the "Theoretical" null hypothesis: The rise, fall, and resurrection of commonsense hypotheses in psychology. Psychological Bulletin, 127, 408-423.

* Knapp, W. M. (1970). A study of teacher personality characteristics and rated effectiveness. Unpublished doctoral dissertation, University of Southern Mississippi.

* Kolz, A. R., Cardillo, E., & Pena, S. (1998, April). Personality predictors of retail employee theft and counterproductive behavior. Poster presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

* Konovsky, M A & Organ, D W (1996). Dispositional and contextual determinants of organizational citizenship behavior. Journal of Organizational Behavior, 17, 253–266.

Koslowsky, M., Sagie, A., Krausz, M., & Singer, A. D. (1997). Correlates of employee lateness: Some theoretical considerations. Journal of Applied Psychology, 82, 79-88.

* Krautheim, M. D. (1998). The development and validation of a customer service orientation scale for university resident assistants. Unpublished doctoral dissertation, University of Tennessee.

* Kriedt, P. H. & Dawson, R. I. (1961). Response set and the prediction of clerical job

performance. Journal of Applied Psychology, 45, 175-178.

Kroger, R. O. (1974). Faking in interest measure interest measurement: A social-psychological perspective. Measurement and Evaluation Guidance, 7, 130-134.

Kroger, R. O. & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. Journal of Consulting and Clinical Psychology, 43, 48-55.

* Krueger, R. J. (1974). An investigation of personality and music teaching success. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

* Lafer, B. (1989). Predicting performance and persistence in hospice volunteers. Psychological Reports, 65, 467-472.

* Lambert, J. D. (1973). An analysis of selected personality and intelligence factors as predictors of the desirability of selected office workers to their supervisors. Unpublished doctoral dissertation, University of Michigan.

* Lamont, L. M. & Lundstrom, W. J. (1977). Identifying successful industrial salesmen by personality and personal characteristics. Journal of Marketing Research, 14, 517-529.

* LaRussa, G. W. (1981). A personality study predicting the effectiveness and satisfaction of Catholic priests in pastoral ministry. Unpublished doctoral dissertation, University of California, Berkeley.

* Lawson, T. T. (1994). The psychological profiles on the Minnesota Multiphasic Personality Inventory of persons presenting themselves for mission service. Unpublished doctoral dissertation, Mississippi State University.

* Lee, R. E. (1994). Personality characteristics of very desirable and undesirable childcare workers in a residential setting. Psychological Reports, 74, 579-584.

Lee, F. K., Sheldon, K. M., & Turban, D. B. (2003). Personality and the goal striving process: The influence of achievement goal patterns, goal level, and mental focus on performance and enjoyment. Journal of Applied Psychology, 88, 256-265.

* Lei, H. & Skinner, H. A. (1982). What difference does language make? Structural analysis of the Personality Research Form. Multivariate Behavioral Research, 17, 33-46.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971). Research design and validity assessment. Personnel Psychology, 24, 247-274.

* Lin, T. –R., Doyle, T. F., & Howard, J. M. (1990, June). The prospective employee potential inventory: A validation study with school bus drivers. Proceedings of the 1990

International Personnel Management Association Assessment Council Conference on Personnel Assessment, 14, 152-160.

* Lock, J. D. (1996). Developing an integrative model of leadership. Unpublished doctoral dissertation, University of Tulsa.

Locke, E. A. & Hulin, C. L. (1962). A review and evaluation of the validity studies of the Activity Vector Analysis. Personnel Psychology, 15, 25-42.

* Love, K. G. & De Armond, S. (2002, April). Incremental validity of personality and assessment center performance for police sergeants. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

* LoVerde, M. A. (1998). The effects of individuals' psychological needs on telecommuting's impact on job performance. Unpublished doctoral dissertation, Illinois Institute of Technology.

* Mabon, H. (1998). Utility aspects of personality and performance. Human Performance, 11(2/3), 289-304.

* Madjar, N., Oldham, G. R., & Pratt, M. G. (2002). There's no place like home? The contributions of work and nonwork creativity support to employees' creative performance. Academy of Management Journal, 45, 757-767.

* Mael, F. A., Waldman, D. A., & Mulqueen, C. (2001). From scientific work to organizational leadership: Predictors of management aspiration among technical personnel. Journal of Vocational Behavior, 59, 132-148.

* Maltby, J. & Day, L. (2000). The reliability and validity of a susceptibility to embarrassment scale among adults. Personality and Individual Differences, 29, 749-756.

* Mandel, K. (1970). The predictive validity of on-the-job performance of policemen from recruitment selection information. Unpublished doctoral dissertation, University of Utah.

* Martin, R. D. (1971). Personality correlates of life insurance underwriters. Studies in Personnel Psychology, 3, 63-67.

* 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, 764-773.

* Mass, G. (1979). Using judgment and personality measures to predict effectiveness in policework: An exploratory validation study. Unpublished doctoral dissertation, Ohio State

University.

∗ Matyas, G. S. The relationship of MMPI and biographical data to police selection and police performance. Unpublished doctoral dissertation, University of Missouri, Columbia.

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. Journal of Personality, 60, 329-361.

∗ McCarty, J. J. (1957). Validity information exchange #10-15. Personnel Psychology, 10, 204-205.

McClelland, D. C. & Boyatzis, R. E. (1982). Leadership motive pattern and long-term success in management. Journal of Applied Psychology, 67, 737-743.

∗ McClelland, D. C. & Franz, C. E. (1992). Motivational and other sources of work accomplishments in mid-life: A longitudinal study. Journal of Personality, 60, 679-707.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A. & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. Journal of Applied Psychology, 86, 730-740.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. Journal of Applied Psychology, 79(4), 599-616.

McEvoy, G. M. & Cascio, W. F. (1987). Do good or poor performers leave? A meta-analysis of the relationship between performance and turnover. Academy of Management Journal, 30, 744-762.

∗ Mcfadden, D. (2000). A personality comparison of active Idaho public school superintendents with successful public school superintendents as selected by their peers. Unpublished doctoral dissertation, University of Idaho.

∗ McMillan, F. W. (1974). Psychological variables related to effective supervision in health care agencies. Unpublished doctoral dissertation, Texas Tech University.

∗ Medcof, J. W. & Hausdorf, P. A. (1995). Instruments to measure opportunities to satisfy needs, and degree of satisfaction of needs, in the workplace. Journal of Occupational and Organizational Psychology, 68, 193-208.

∗ Medcof, J. W. & Wegener, J. G. (1992). Journal of Organizational Behavior, 13, 413-423.

∗ Melamed, T. (1995). Barriers to women's career success: Human capital, career

choices, structural determinants, or simply sex discrimination. Applied Psychology: An International Review, 44, 295-314.

* Middleman, R. J. (1989). Factors relating to counselor performance in a therapeutic wilderness camp setting. Unpublished doctoral dissertation, Texas Woman's University.

* Miller, R. L., Griffin, M. A., & Hart, P. M. (1999). Personality and organizational health: The role of conscientiousness. Work and Stress, 13, 7-19.

* Mills, C. J. & Bohannon, W. E. (1980). Personality characteristics of effective state police officers. Journal of Applied Psychology, 65, 680-684.

* Mills, M. C. (1981). The MMPI and the prediction of police job performance. Unpublished doctoral dissertation, University of Southern California, Los Angeles.

* Moberg, P. J. (1996). Social skills, personality, occupational interests, and job performance: A multivariate structural analysis. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

* Molitor, D. D. (1998). An examination of the effects of personality and job satisfaction on multiple non-workrole organizational behaviors. Unpublished doctoral dissertation, Iowa State University.

* Morgeson, F. P., Reider, M. H., & Campion, M. A. (2002). Selecting individuals in team settings: Comparing a structured interview, personality test, and ability test. In F. P. Morgeson (Chair), Selection for teams: A tale of five approaches. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, Canada.

* Morrison, K. A. (1997). Personality Correlates of the Five-Factor Model For A Sample of Business owners/managers: Associations with scores on self-monitoring, Type A behavior, locus of control, and subjective well-being. Psychological Reports, 80, 255-272.

* Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational stress: Its causes and consequences for job performance. Journal of Applied Psychology, 71, 618-629.

* 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). The big five personality dimensions: Implications for research and practice in human resources management. In K. M. Rowland & G. Ferris (Eds.) Research in Personnel and Human Resources Management (Vol. 13, pp. 153-200). Greenwich,

CT: JAI Press.

* 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, 707-721.

* 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, 299-323.

* Moyle, P. & Parkes, K. (1999). The effects of transition stress: A relocation study. Journal of Organizational Behavior, 20, 625-646.

* Muchinsky, P. M. (1993). Validation of personality constructs for the selection of insurance industry employees. Journal of Business and Psychology, 7, 475-482.

* Muecke-Gardner, L. A. (1988). Trait predictors of responses to behavioral and outcome sales incentives. Unpublished doctoral dissertation, Hofstra University.

* NCS Pearson (2000). Thurstone Temperament Schedule Information Guide. Minneapolis, MN: NCS Pearson.

* Neuman, G. A; Kickul, J. R. (1998). Organizational citizenship behaviors: Achievement orientation and personality. Journal of Business and Psychology, 13, 263-279.

* Neuman, G. A. & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. Journal of Applied Psychology, 84, 376-389.

* Nikolaou, I. & Robertson, I. T. (2001). The five-factor model of personality and work behaviour in Greece. European Journal of Work and Organizational Psychology, 10, 161-186.

* Nye, L. G., Schroeder, D. J., & Dollar, C. S. (1994). Relationships of type a behavior with biographical characteristics and training performance of air traffic controllers. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine. (NTIS No. AD-A283-813-4-XAB)

* 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, 523-548.

* O'Connor Boes, J. A. (1997). Individual differences and corruption: An investigation of the MMPI in a law enforcement setting. Unpublished doctoral dissertation, George Mason University.

∗ Oldham, G. R. & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. Academy of Management Journal, 39, 607-634.

∗ Olszewski, T. M. (1998). The positive side of faking: The validity of personality constructs as predictors of early attrition. Unpublished doctoral dissertation, George Washington University.

Ones, D. S., Mount, M. K., Barrick, M. R., & Hunter, J. E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. Personnel Psychology, 47, 147-156.

Ones, D. Z. & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. Human Performance, 11(2/3), 245-269.

Ones, D. Z. & Viswesvaran, C. (2001). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In B. W. Roberts & R. T. Hogan (Eds.), Applied Personality Psychology: The intersection of personality and I/O psychology (pp. 63-92). Washington, DC: American Psychological Association.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. Journal of Applied Psychology, 78, 679-703.

Oppler, S. H., Peterson, N. G., & Russell, T. (1992). Basic validation results for the LVI sample. In J. P. Campbell & L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning army enlisted personnel (Annual Report, 1991 Fiscal Year). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

∗ Orpen, C. (1985). The effects of need for achievement and need for independence on the relationship between perceived job attributes and managerial satisfaction and performance. International Journal of Psychology, 20, 207-219.

Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. (1983). Statistical power of tests of the situational specificity hypothesis in validity generalization studies: A cautionary note. Journal of Applied Psychology, 68, 115-122.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology (pp. 609-644). Chicago: Rand-McNally.

∗ Pandya, D. N. (1967). Personality characteristics and level of performance of male

county extension agents in Wisconsin. Unpublished doctoral dissertation, University of Wisconsin.

* Parkkola, K., Tuominen, J., & Piha, J. (1997). The life change inventory and the MMPI depression subscale in predicting failure in the compulsory conscript service. Nordic Journal of Psychiatry, 51, 371-377.

* Peacock, A. C. & O'Shea, B. (1984). Occupational therapists: Personality and job performance. American Journal of Occupational Therapy, 38, 517-521.

Pedhazur, E. J. (1997). Multiple Regression in Behavioral Research: Explanation and prediction (3rd edition). Fort Worth, TX: Harcourt Brace College Publishers.

Pervin, L. A. (1996). The Science of Personality. New York, NY: John Wiley and Sons.

* Phillips, A. S. & Bedeian. A. G. (1994). Leader-follower exchange quality: The role of personal and interpersonal attributes. Academy of Management Journal, 37, 990-1001.

* Piedmont, R. L. & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO Personality Inventory. Journal of Psychology, 128, 255-265.

* Ployhart, R. E., Lim, B. –C., & Chan, K. –Y. (2001). Exploring relations between maximum and typical performance ratings and the five factor model of personality. Personnel Psychology, 54, 809-843.

* Prien, E. P. (1970). Measuring performance criteria of bank tellers. Journal of Industrial Psychology, 5, 29-35.

* Prien, E. P. & Cassel, R. H. (1973). Predicting performance criteria of institutional aides. American Journal of Mental Deficiency, 78, 33-40.

* Puffer, S. M. (1987). Prosocial behavior, noncompliant behavior, and work performance among commission salespeople. Journal of Applied Psychology, 72, 615-621.

* Pugh, G. (1985). The California Psychological Inventory and police selection. Journal of Police Science and Administration, 13, 172-177.

* Pulakos, E. D., Borman, W. C., & Hough, L. M. (1988). Test validation for scientific understanding: Two demonstrations of an approach to studying predictor-criterion linkages. Personnel Psychology, 41, 703-716.

Quartetti, D., Kieckhaefer, W. & Houston, J. (2001). Predictor development background. In R. A. Ramos, M. C. Heil, & C. A. Manning, Documentation of Validity for the AT-SAT Computerized Test Battery (Volume I). Alexandria, VA: Human Resources Research

Organization. (NTIS No. DOT/FAA/AM-01/5)

∗ Raunikar, D. F. (1991). Identification of sales management potential as determined by personality characteristics. Unpublished doctoral dissertation, Mississippi State University.

∗ Reeves, P. S. (1996). Relationship between success in the Mississippi cooperative extension service and personality type. Unpublished doctoral dissertation, Mississippi State University.

∗ 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, 639-650.

∗ Rentsch, J. R. & Steel, R. P. (1998). Testing the durability of job characteristics as predictors of absenteeism over a six-year period. Personnel Psychology, 51, 165-190.

Rich, G. A., Bommer, W. H., MacKenzie, S. B., Podsakoff, P. M., & Johnson, J. L. (1999). Apples and Apples or Apples and Oranges? A meta-analysis of objective and subjective measures of salesperson performance. Journal of Personal Selling and Sales Management, 19, 41-52.

∗ Robbins, J. E. & King, D. C. (1961). Validity Information Exchange #14-02. Personnel Psychology, 14, 217-219.

∗ Robertson, I. T., Baron, H., Gibbons, P., MacIver, R., & Nyfield, G. (2000). Conscientiousness and managerial performance. Journal of Occupational and Organizational Psychology, 73, 171-180.

∗ Robie, C.; Ryan, A. M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. International Journal of Selection and Assessment, 7, 157-169.

Robie, C., Zickar, M. J.,& Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. Human Performance, 14, 187-207.

∗ Rogers, D. A., Strong, M. H., & Zbylut, M. (2001). Personal characteristics measures to predict safety-related work behaviors. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

∗ Roman, M. A. (1997). Exploring the upper bound of predictive validity of personality testing in selection decisions. Unpublished doctoral dissertation, Northern Illinois University.

* Ronan, W. W. (1964). Evaluation of skilled trades performance predictors. Educational and Psychological Measurement, 24, 601-608.

* Rose, R. M., Fogg, L. F., Helmreich, R. L., & McFadden, T. J. (1994). Psychological predictors of astronaut effectiveness. Aviation Space and Environmental Medicine, 65(10, Sect 1), 910-915.

Rosenthal, R. (1991). Meta-Analytic Procedures for Social Research. Newbury Park, CA: Sage.

Rosenthal, R. & Dimatteo, M. R. (2001). Meta-Analysis: Recent developments in quantitative methods for literature reviews. In S. T. Fiske (Ed.), Annual Review of Psychology (Volume 52; pp. 59-82).

* Ross, P. F. & Dunfield, N. M. (1964). Selecting salesmen for an oil company. Personnel Psychology, 17, 75-84.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. Journal of Applied Psychology, 83, 634-644.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75, 322-327.

Ruch, F. L. & Ruch, W. W. (1980). The K factor as a (validity) suppressor variable in predicting success in selling. Journal of Applied Psychology, 51, 201-204.

* Rushton, J. P., Murray, H. G., & Paunonen, S. V. (1983). Personality, research creativity, and teaching effectiveness in university professors. Scientometrics, 5, 93-116.

Russell, C. J., Settoon, R. P., McGrath, R. N., Blanton, A. E., Kidwell, R. E., Lohrke, F. T., Scifres, E. L., & Danforth, G. W. (1994). Investigator characteristics as moderators of personnel selection research: A meta-analysis. Journal of Applied Psychology, 79, 163-170.

* Russell, D. P., Oravec, J. T., & Wolf, P. P. (2000). Big five versus subscale measurement of personality for selection. In P. P. Wolf (Chair) Personality tests: Using theory to maximize predictive power. Symposium presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Russell, T., Oppler, S. H., & Peterson, N. G. (1998). Comparison of social desirability and validity on a personality/biodata measure across samples. Paper presented at the 13th Annual

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

* Ryan, A. M., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. Personnel Psychology, 51, 599-621.

* Sackett, P. R., Gruys, M. L., & Ellingson, J. E. (1998). Ability-personality interactions when predicting job performance. Journal of Applied Psychology, 83, 545-556.

Sackett, P. R., Harris, M. M., & Orr, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error. Journal of Applied Psychology, 71, 302-310.

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. & Rumbo, A. (1997). Personality and job performance in financial services managers. International Journal of Selection and Assessment, 5, 91-100.

* Salomon, L. M. (2000). The impact of personality variables on different facets of contextual performance. Unpublished doctoral dissertation, University of Houston.

* Sartain, A. Q. (1946). A study of Bernreuter Personality Inventory scores made by candidates for supervisory positions in an aircraft factory. Journal of Social Psychology, 24, 255-259.

* 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, 243-262.

* Scarpello, V. & Whitten, B. J. (1984) Multitrait-multimethod validation of personality traits possessed by industrial personnel in research and development. Educational and Psychological Measurement, 44, 395-404.

* Schaubroeck, J., Ganster, D. C., & Fox, M. L. (1992). Dispositional affect and work related stress. Journal of Applied Psychology, 77, 322-335.

* Schaubroeck, J., Ganster, D. C., & Jones, (1998). Organization and occupation influences on the attraction-selection-attrition process. Journal of Applied Psychology, 83, 869-891.

Scheier, M. F. & Carver, C. S. (1985). Optimism, coping, and health: Assessment and

implications of generalized outcome expectancies. Health Psychology, 4, 219-247.

∗ Schinka, J. A., Dye, D. A., Curtiss, G. (1997). Correspondence between five-factor and RIASEC models of personality. Journal of Personality Assessment, 68, 355-368.

Schmidt, F. L. & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. Journal of Applied Psychology, 68, 407-414.

Schmidt, F. L. & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1980). Task differences as moderators of aptitude test validity in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. Journal of Applied Psychology, 78, 3-12.

Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? Journal of Applied Psychology, 77, 629-637.

Schmit, M. J. & Ryan, A. M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. Journal of Applied Psychology, 78, 966-974.

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

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: how to sink science in five dimensions or less. Journal of Organizational Behavior, 17, 639-655.

∗ Schroeder, D. J., Broach, D., & Young, W. C. (1993). Contribution of Personality to the Prediction of Success in Initial Air Traffic Control Specialist Training. Springfield, VA: National Technical Information Service. (NTIS Report No. AD-A264-699-0-XAB)

∗ Schuerger, J. M., Kochevar, K. F., & Reinwald, J. E. (1982). Male and female corrections officers: Personality and rated performance. Psychological Reports, 51, 223-228.

∗ Schuerger, J. M., Ekeberg, S. E., & Kustis, G. A. (1994). 16 PF scores and machine

operators' performance. Perceptual and Motor Skills, 79(3, Pt 2), 1426.

* Shain, V. L. (1972). A study of teacher personality preferences in relationship to teacher satisfaction and teacher competency ratings in the open space elementary school. Unpublished doctoral dissertation, University of Kansas.

* Shaver, D. P. (1980). A descriptive study of police officers in selected towns of northwest Arkansas. Unpublished doctoral dissertation, University of Arkansas.

* Shenk, F., Watson, T. W., & Hazel, J. T. (1973). Relationship Between Personality Traits and Officer Performance and Retention Criteria. Lackland Airforce Base, Texas: United States Air Force Human Resources Laboratory. (US AFHRL Technical Report. No. TR-73-4)

* Sinclair, R. R. & Michel, R. P. (2001, April). A construct-oriented approach to modeling entry-level job performance. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

* Skinner, H. A., Jackson, D. N., & Rampton, G. M. (1976). The Personality Research Form in a Canadian context: Does language make a difference? Canadian Journal of Behavioral Sciences, 8, 156-168.

* Slimak, M. (1997). Job autonomy as a moderator of conscientiousness and extraversion as related to performance: Exploring individual- and job-level effects. Unpublished doctoral dissertation, Texas A & M University.

* Slocum, J. W. & Hand, H. H. (1971). Prediction of job success and employee satisfaction for executives and foremen. Training and Development Journal, 25, 28-36.

* Smith, D. B. & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. Journal of Applied Psychology, 87, 211–219.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. Journal of Applied Psychology, 86, 304-315.

* Smith, M. A. & Lutrick, E. C. (2001). Facet analysis of the NEO PI-R in an applicant samples. Poster presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

* Sormin, B. H. (1984). The relationships between personality and teaching effectiveness of vocational industrial teachers in selected high schools in Northeast Texas. Unpublished doctoral dissertation, East Texas State University.

* Sosik, J. J. & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance: The role of self-other agreement on transformational leadership perceptions. Group and Organization Management, 24, 367-390.

Spangler, W. D. (1992). Validity of questionnaire and TAT measure of need for achievement: Two meta-analyses. Psychological Bulletin, 112, 140-154.

* Sparks, C. P. (1951). Limitations of the Bernreuter Personality Inventory in selection of supervisors. Journal of Applied Psychology, 35, 403-406.

* Sprinkle, S. D. (1990). Predicting Southern Baptist foreign missionary attrition with the MMPI. Unpublished doctoral dissertation, University of Virginia.

Stajkovic, A. D. & Luthans, F. (2003). Behavioral management and task performance in organizations: Conceptual background, meta-analysis, and test of alternative models. Personnel Psychology, 56, 155-194.

* Stark, S., Chernyshenko, O. S., Chan, K. –Y., Lee, W. C., & Drasgow, F. (2001) Effects of the testing situation on item responding: Cause for concern. Journal of Applied Psychology, 86, 943-953.

* Steers, R. M. & Braunstein, D. N. (1976). A behaviorally-based measure of manifest needs in work settings. Journal of Vocational Behavior, 9, 251-266.

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

Stewart, G. L. (1997, August). Applicants versus incumbents: Assessing the impact of validation design on personality research. Paper presented at the Academy of Management Annual Meeting, Boston, MA.

* Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. Journal of Applied Psychology, 84, 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, 143-164.

* Stewart-Belle, S. & Lust, J. A. (1999). Career movement of female employees holding lower-level positions: An analysis of the impact of the Type A behavior pattern. Journal of Business and Psychology, 14, 187-197.

Stokes, G. S., Hogan, J. B., & Snell, A. F. (1993). Comparability of incumbent and applicant samples for the development of biodata keys: The influence of social desirability. Personnel Psychology, 46, 739-762.

Sussman, M. & Robertson, D. U. (1986). The validity of validity: An analysis of validation study designs. Journal of Applied Psychology, 71, 461-468.

∗ Talley, J. E. & Hinz, L. D. (1990). Performance prediction of public safety and law enforcement personnel: A study in race and gender differences and MMPI subscales. Springfield, IL: Charles C. Thomas, Publisher.

Tedeschi, J. T. & Melburg, V. (1984). Impression management and influence in the organization. Research in the Sociology of Organizations, 3, 31-58.

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

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

∗ Tett, R. P., Stelle, J. R., & Beauregard, R. S. (2000). Broad and narrow measures on both sides of the personality-job performance relationship. In J. C. Hogan (Chair), Specificity versus generality in personality-job performance linkages: Data speak louder than words. Symposium presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

∗ Thoms, P., Moore, K. S., & Scott, K. S. (1996). The relationship between self-efficacy for participating in self-managed work groups and the big five personality dimensions. Journal of Organizational Behavior, 17, 349-362.

∗ Thumin, F. J. (2002). Comparison of the MMPI and the MMPI-2 among job applicants. Journal of Business and Psychology, 17, 73-86.

∗ Tivendell, J. & Bourbonnais, C. (2000). Job insecurity in a sample of Canadian civil servants as a function of personality and perceived job characteristics. Psychological Reports, 87, 55-60.

∗ Tokar, D. M. & Swanson, J. L. (1995). Evaluation of the correspondence between Holland's vocational personality typology and the five-factor model of personality. Journal of Vocational Behavior, 46, 89-108.

∗ Tomini, B. A. (1996). The person-job fit: Implications of selecting police personnel

on the basis of job dimensions, aptitudes and personality traits. Unpublished doctoral dissertation, University of Windsor.

* Trites, D. K., Kurek, A., & Cobb, B. B. (1967). Personality and achievement of air traffic controllers. Aerospace Medicine, 38, 1145-1150.

* Tsaousis, I. & Nikolau, I. E. (2001) The stability of the five factor model of personality in personnel selection and assessment in Greece. International Journal of Selection and Assessment, 9, 290-301.

* Tull, K. T. (1997). The effects of faking behavior on the prediction of sales performance using the Guilford Zimmerman Temperament Survey and the NEO Five Factor Inventory. Unpublished doctoral dissertation, University of Akron.

* Turnbull, A. A., Jr. (1976). Selling and the salesman: Prediction of success and personality change. Psychological Reports, 38(3, Pt 2), 1175-1180.

* Tyagi, A. K., Gautam, S., Adya, A. K., & Bhatia, V. K. (1976). A study of the phenomenon of courage in Nagaland: Part II. Journal of Psychological Researches, 20, 36-39.

United States Department of Labor (2002). 2001 National Occupational Employment and Wage Estimates. Internet file, retrieved from http://www.bls.gov/oes/2001/oes_11Ma.htm on May 24, 2003.

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., & Putka, D. J. (2003). Applicant-incumbent differences on personality, integrity, and customer service measures. Poster presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

* Van Scotter, J. R. (1996). Evidence for the usefulness of task performance, job dedication, and interpersonal facilitation as components of overall performance. Unpublished doctoral dissertation, University of Florida.

* Vasilopoulos, N. L., Sass, M. D., Shipper, F., & Story, A. L. (2002, August). Target personality and rater agreement - implications for 360-degree feedback. Paper presented at the Academy of Management Annual Meeting, Denver, CO.

* Vickers, R. R., Jr., Hervig, L. K., & Booth, R. F. (1996). Personality and success among military enlisted personnel: An historical prospective study of U.S. Navy corpsmen. San Diego, CA: Naval Health Research Center. (US Naval Health Research Center Report No 96-15)

* Vincent, N. L. & Dugan, R. D. (1962). Validity Information Exchange #15-03.

Personnel Psychology, 15, 223-226.

Vinchur, A. J., Schreischeim, J. S., Switzer, 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. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. Personnel Psychology, 48, 865-885.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. Journal of Applied Psychology, 81, 557-574.

Wallace, S. R., Clark, W. V., & Dry, R. J. (1956). The Activity Vector Analysis as a selector of life insurance salesmen. Personnel Psychology, 9, 337-345.

∗ Walter, T. L. (1980). Prediction of sales performance and job survival of sales personnel: A regression model. Unpublished doctoral dissertation, University of Missouri, Columbia.

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

∗ Wanberg, C. R., Kanfer, R., & Banas, J. T. (2000). Predictors and outcomes of networking intensity among unemployed job seekers. Journal of Applied Psychology, 85, 491-503.

Webber, S. S. & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. Journal of Management, 27, 141-162.

∗ Weekes, E. M. (1995). The influence of personality dimensions and physical abilities on a pistol shooting task. Unpublished doctoral dissertation, University of Houston.

Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2003). Personality and situational judgment tests across applicant and incumbent settings. Poster presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

∗ Welsch, L. A. (1968). The supervisor's employee appraisal heuristic: The contribution of selected measures of employee attitude, intelligence and personality. Unpublished doctoral dissertation, University of Pittsburgh.

∗ Whisman, R. W. (1978). The relationship of selected personality traits and personal characteristics of drivers to the occupational performance of school bus drivers in Ohio.

Unpublished doctoral dissertation, Ohio State University.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. Journal of Applied Psychology, 75, 315-321.

* Wilder, B. C. (1996). The relation between executive success and executive personality in small, medium, large and very-large high technology companies. Unpublished doctoral dissertation, Pacific Graduate School of Psychology.

* Wilds, J. M. (1978). The relationship between scores on mental ability and personality tests and success on the job of sales management personnel. Unpublished doctoral dissertation, University of Pittsburgh.

* Williams, R. W. (1998). Using personality traits to predict the criterion space. Unpublished doctoral dissertation, Union Institute.

* Willock, J., Deary, I. J., McGregor, M. M., Sutherland, A., Edwards-Jones, G., Morgan, O., Dent, B., Grieve, R., Gibson, G., & Austin, E. (1999). Farmers' attitudes, objectives, behaviors, and personality traits: The Edinburgh study of decision making on farms. Journal of Vocational Behavior, 54, 5-36.

* Wilson, G. D., Tunstall, O. A., & Eysenck, H. J. (1972). Measurement of motivation in predicting industrial performance. Occupational Psychology, 46, 15-24.

* Witt, L. A. (2002). The interactive effects of extraversion and conscientiousness on performance. Journal of Management, 28, 835-851.

* Witt, L. A., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. Journal of Applied Psychology, 87, 164-169.

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

* 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.

* Wolfe, R. (1994). Experience, gender, marital status, and the 16PF Questionnaire as predictors of American teachers' effectiveness in southeast Asia schools. Unpublished doctoral dissertation, University of Oregon.

* Woodmansee, J. J. (1978). Validation of the nurturance scale of the Edwards Personal Preference Schedule. Psychological Reports, 42, 495-498.

* Woolard, C. & Brown, R. D. (1999). Moderation of personality test validity: An extension and replication of Barrick and Mount (1993). Poster presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

* Wooten, K. C., Timmerman, T. A., & Folger, R. (1999). The use of personality and the five-factor model to predict new business ventures: From outplacement to startup. Journal of Vocational Behavior, 54, 82-101.

* Wright, P. M., Kacmar, K. M., McMahan, G. C., & Deleeuw, K. (1995). P=f(M X A): Cognitive ability as a moderator of the relationship between personality and job performance. Journal of Management, 21, 1129-1139.

* Yoon, K. (1998). General mental ability and the big five personality dimensions: An investigation of the cross-cultural generalizability of their construct and criterion-related validities in Korea. Unpublished doctoral dissertation, University of Iowa.

* Young, S. A. & Parker, C. P. (1999). Predicting collective climates: Assessing the role of shared work values, needs, employee interaction and work group membership. Journal of Organizational Behavior, 20, 1199-1218.

* Zellars, K. L. & Perrewé, P. L. (2001) Affective personality and the content of emotional social support: Coping in organizations. Journal of Applied Psychology, 86, 459-467.

Appendix A

SPSS Command Syntax For The Generation Of Simulated Data Based On The Incumbent Parameter Estimates

```
***David Nicholl's commands to generate correlated data based on specified population
***parameters.


input program.
set seed 3458769.
loop #i=1 to 10000.
do repeat response=r1 to r8.
compute response=normal(1).
end repeat.
end case.
end loop.
end file.
end input program.
save outfile 'C:\Documents and Settings\kbradley\My
Documents\dissertation\analyses\revisions\simulations\incumbents_strict1.sav'.

Factor
  /variables r1 to r8
  /analysis r1 to r8
  /print correlation extraction
  /criteria Factors(8) Iterate(25)
  /extraction pc
  /rotation norotate
  /save reg(all) .
save outfile 'C:\Documents and Settings\kbradley\My
Documents\dissertation\analyses\revisions\simulations\incumbents_strict2.sav'.

set mxmemory=28000 .
execute.

Matrix.
Get X
 /File='C:\Documents and Settings\kbradley\My
Documents\dissertation\analyses\revisions\simulations\incumbents_strict2.sav'
 /Variables=fac1_1 to fac8_1.
Compute R={1.00, -0.38, -0.06, -0.26, -0.37, -0.46, -0.38, -0.12;
 -0.38, 1.00, 0.15, 0.23, 0.32, 0.65, 0.49, 0.08;
 -0.06, 0.15, 1.00, 0.08, -0.04, 0.33, 0.26, 0.03;
 -0.26, 0.23, 0.08, 1.00, 0.20, 0.35, 0.23, 0.11;
 -0.37, 0.32, -0.04, 0.20, 1.00, 0.42, 0.38, 0.13;
 -0.46, 0.65, 0.33, 0.35, 0.42, 1.00, 0.56, 0.15;
 -0.38, 0.49, 0.26, 0.23, 0.38, 0.56, 1.00, 0.10;
 -0.12, 0.08, 0.03, 0.11, 0.13, 0.15, 0.10, 1.00}.
Compute NewX=X*chol(R).
Save NewX /outfile=*/variables=nr1 to nr8.
End Matrix.
```

```
ren var (nr1 nr2 nr3 nr4 nr5 nr6 nr7 nr8=neurotic extravrt openness agreeabl conscien optimism
ambition perform).
execute.

SAVE OUTFILE='C:\Documents and Settings\kbradley\My
Documents\dissertation\analyses\revisions\simulations\incumbents_strict_final.sav'
  /COMPRESSED.
```

Curriculum Vitae

KEVIN M. BRADLEY
209 Westwood Court
Lexington, KY 40503
859.276.3834
kbradley@vt.edu

EDUCATIONAL TRAINING:

- Ph.D., Industrial and Organizational Psychology, Virginia Tech, Blacksburg, VA. December, 2003.
  Doctoral Dissertation Title: Personality Test Validation Research: Present-employee and job applicant samples

- M.S., Psychology, Rensselaer Polytechnic Institute, Troy, NY. August, 1996.
  Master's Thesis Title: The Influence of an Incentive to Provide Accurate Ratings on Performance Appraisal Accuracy.

- B.A., Psychology, University of Richmond, Richmond, VA. May, 1994
  Undergraduate research focused on impression management in the employment interview.

TEACHING EXPERIENCE

- Psychology of Personality, Virginia Tech, Blacksburg, VA (Fall, 1998 – Spring, 2000).
- Teaching Assistant: Statistics for Social and Behavioral Sciences I and II (graduate level), Rensselaer Polytechnic Institute, Troy, NY (Fall, 1995 – Spring, 1996).
- Teaching Assistant: Sensation and Perception, Learning, and Introductory Psychology, Rensselaer Polytechnic Institute, Troy, NY (Fall, 1994 – Spring, 1995).

RESEARCH INTERESTS

- Educational and Workplace Testing and Assessment
- Self-presentation processes in job applicant settings
- Personnel Validation Research

TEACHING INTERESTS

- Industrial and Organizational Psychology; Research Methods; Statistics for Social and Behavioral Sciences; Tests and Measurements

MANUSCRIPTS UNDER REVIEW:

- **Bradley, K. M.** (2003). Employment Interview Questions: Comparative analyses of respondent thought processes. Manuscript in preparation.

MANUSCRIPTS IN PREPARATION:

- **Bradley, K. M.**, O'Shea, P. G., & Hauenstein, N. M. A. (2003). Factors Related to Personality Test Response Processes and Response Endorsements. Manuscript in preparation.

CONFERENCE PRESENTATIONS:

- **Bradley, K. M.** & Hauenstein, N. M. A. (2002). Personality Test Validation Research: Present employee and job applicant samples. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, April, 2002.

- **Bradley, K. M.**, O'Shea, P. G., & Hauenstein, N. M. A. (2002). Factors Related to Personality Test Response Processes and Response Endorsements. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, April, 2002.

- **Bradley, K. M.** (2002). Employment Interview Questions: Comparative analyses of respondent thought processes. Poster presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, ON, April, 2002.

- **Bradley, K. M.** (2000). A Comparison of Situational and Behavioral Structured Interview Questions. Paper presented at the 21st Annual Industrial/Organizational Psychology & Organizational Behavior Graduate Student Conference, Knoxville, TN, March, 2000.

- Hauenstein, N. M. A., **Bradley, K. M.**, & O'Shea, P. G. (2000). Clarifying the Process: Verbal reports of honest and faked personality test responses. Poster presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA, April, 2000.

- **Bradley, K. M.**, Dorsey, D. W., Russell, D. P., & O'Connell, B. J. (1999). Task Similarities as Indicators of Occupational Skill Requirements. Poster presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, April, 1999.

- **Bradley, K. M.** (1999). The Dimensionality of Work in Diverse Jobs. Poster presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, April, 1999.

- Baughman, W. A., Russell, D. P., Dorsey, D. W., Cooke, A. E., & **Bradley, K. M.** (1999). Maximizing Information Gain for Job Classification: The utility of qualitative and indirect information. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, April, 1999.

- **Bradley, K. M.** (1998). The Prediction of Job Knowledge Acquisition Using Structured Interviews. Poster presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX, April, 1998.

- **Bradley, K. M.** (1997). The Impact of an Incentive to Provide Accurate Ratings on Performance Appraisal Accuracy. Paper presented at the 18th Annual Industrial/Organizational Psychology & Organizational Behavior Graduate Student Conference, Roanoke, VA, March, 1997.

APPLIED RESEARCH AND ADDITIONAL WORK EXPERIENCE:

- Assessment Coordinator, Department of Mathematics, Virginia Tech (July, 2000 – July, 2002):
  Oversaw educational assessment and evaluation research for department with 9,000+ person enrollments per semester. Conducted studies comparing efficacy of competing approaches to teaching Mathematics courses. Oversaw administration of surveys to assess student reactions to math classes. Advised department chair on efforts to evaluate effectiveness of instructors.

- Research Intern, American Institutes for Research, Washington, DC (Summer, 1998):
  Conducted investigative research on contrasting approaches for identifying the skill and ability requirements for diverse occupations.

- Research Intern, Development Dimensions International, Inc., Bridgeville, PA (Summer, 1997):
  Conducted investigative research on employment testing systems. Analyzed large data set from judgment test to ensure that no test items were biased against any demographic groups.

ADDITIONAL EXPERIENCE AND QUALIFICATIONS:

- Developed structured interviews as part of preliminary examination research project at Virginia Tech. Trained and supervised team of seven students in interview administration and scoring.

- Statistical and Research Methods consultant to graduate students at Virginia Tech and American University.

- Assisted Virginia Tech Social Sciences faculty in development and evaluation of web-based instructional tutorials for teaching research methods and statistics.

- Excellent database management and analytic skills; well versed in SAS and SPSS; experienced in application of Item Response Theory in psychological/educational testing; very strong computer skills.