2016

Heterogeneous Versus Homogeneous Measures: A Meta-Analysis of Predictive Efficacy

Suzanne Lee Dean
Wright State University

HETEROGENEOUS VERSUS HOMOGENEOUS MEASURES: A META-ANALYSIS OF PREDICTIVE EFFICACY

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

SUZANNE L. DEAN
M.S., Wright State University, 2008

_____

2016
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

January 3, 2016

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Suzanne L. Dean ENTITLED Heterogeneous versus Homogeneous Measures: A Meta-Analysis of Predictive Efficacy BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

_____
Corey E. Miller, Ph.D.
Dissertation Director

_____
Scott N. J. Watamaniuk, Ph.D.
Graduate Program Director

_____
Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

Committee on
Final Examination

_____
Corey E. Miller, Ph.D.

_____
David LaHuis, Ph.D.

_____
Allen Nagy, Ph.D.

_____
Debra Steele-Johnson, Ph.D.

ABSTRACT

Dean, Suzanne L. Ph.D., Industrial/Organizational Psychology Ph.D. program, Wright State University, 2016. Heterogeneous versus Homogeneous Measures: A Meta-Analysis of Predictive Efficacy.

A meta-analysis was conducted to compare the predictive validity and adverse impact of homogeneous and heterogeneous predictors on objective and subjective criteria for different sales roles. Because job performance is a dynamic and complex construct, I hypothesized that equally complex, heterogeneous predictors would have stronger correlations with objective and subjective criteria than homogeneous predictors. Forty-seven independent validation studies (N = 3,378) qualified for inclusion in this study. In general, heterogeneous predictors did not demonstrate significantly stronger correlations with the performance criteria than homogeneous predictors. Notably, heterogeneous predictors did not demonstrate adverse impact on protected classes. One noteworthy finding was that the heterogeneous new business development predictor profile demonstrated a relationship with subjective criteria that generalized across studies, which challenges some assumptions underlying Classical Test Theory.

TABLE OF CONTENTS

Page

I. INTRODUCTION……..…..….1

II. CURRENT LITERATURE……...4

Predictors of Performance…….4

Personality……..……...5

Specific predictors……..……..6

Broad versus Narrow Predictors Debate……7

Multidimensionality……..…….9

Heterogeneity of Performance……11

Over-Simplification of Job Performance Domain……14

Criteria Heterogeneity……..15

The Consistency Model……...17

Predictor Heterogeneity……...19

Research supporting heterogeneous predictors……..…….19

Research Opposing Heterogeneous Predictors……..21

Construct validity……..…….……21

Internal consistency reliability……..……..……..24

The Case for Predictor Heterogeneity……..24

Addressing the construct validity argument……..…....25

Addressing the reliability argument……..…...27

In conclusion……..……30

III. HYPOTHESES……...33

IV. METHOD……37

Participants……37

Hypothesis Testing Approaches……38

Inclusion Criteria……...39

Measures……...39

V. RESULTS……44

Sample Characteristics……44

Coding Agreement……44

Objective vs. Subjective Criteria……45

Tests of Hypotheses……46

VI. DISCUSSION AND IMPLICATIONS……..55

References……...63

LIST OF TABLES

Table Page

1. Effect Size between Criteria for Sales Roles and Heterogeneous and Homogeneous Profiles……..……..81

2. Effect Size between Subjective Criteria for Account Management Roles and Heterogeneous and Homogeneous Profiles……..82

3. Effect Size between Subjective Criteria for Account Management Roles and Heterogeneous and Homogeneous Scales.……..83

4. Effect Size between Objective Criteria for Account Management Roles and Heterogeneous and Homogeneous Profiles……..85

5. Effect Size between Objective Criteria for Account Management Roles and Heterogeneous and Homogeneous Scales.……..86

6. Effect Size between Subjective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Profiles……...88

7. Effect Size between Subjective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Scales……....89

8. Effect Size between Objective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Profiles……..91

9. Effect Size between Objective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Scales……....92

10. Adverse Impact Ratios for Heterogeneous and Homogeneous Profiles…….….94

11. Gender Adverse Impact Ratios for Heterogeneous Scales…….……95

12. Race Adverse Impact Ratios for Heterogeneous Scales…….……96

13. Gender Adverse Impact Ratios for Homogeneous Scales……..……98

14. Race Adverse Impact Ratios for Homogeneous Scales…….……...…100

I. INTRODUCTION

Job performance is undoubtedly one of the most studied and most important constructs in Industrial/Organizational psychology. However, predicting job performance is a problem that has vexed researchers for almost a century. Although there has been much research dedicated to better understanding job performance, the research continues to be mixed regarding the best model of job performance, how it should be defined, and its underlying factors (Arvey & Murphy, 1998). Despite the numerous advances that have been made, there is still much to learn. Research has cited two primary variables when predicting job performance: the predictor (or personnel selection test) and the criteria (or measure of job performance). When trying to improve prediction of job performance, researchers have typically investigated either the predictor or the criterion. However, the majority of current research in this realm has simplified the job performance predictors and criteria to such a degree that they have appeared to be homogeneous and generalizable across many contexts. Homogeneous predictors and criteria are composed of parts or elements that are of the same kind. This simplification of the job performance domain has made it increasingly difficult to accurately predict job performance.

In order to make a healthy improvement in predicting job performance, I have argued that researchers need to acknowledge that job performance and its corresponding criteria and predictors are heterogeneous. Heterogeneous

predictors and criteria are a constellation of multiple, narrow (job-specific), divergent factors. Moreover, the component parts interact and compensate for one another such that the whole is greater than the sum of its parts.

Researchers and academicians alike have tried to develop a deeper understanding of job performance in order to develop measures to predict an individual's future job performance. In fact, over the last 45 years, approximately 20% of the articles in the Journal of Applied Psychology (JAP) and about 13% of the articles in Personnel Psychology (PPsych) address the topic 'predictors of performance' (Cascio & Aguinis, 2008a). Given the extensive research conducted in this area, one would believe that I/O psychologists would be able to predict performance almost perfectly by now. However, this is far from being the case (Cascio & Aguinis, 2008b). Our strongest predictor to date, general mental ability (GMA), could only explain approximately 25% of the variance in performance after correcting for statistical and methodological artifacts (Schmidt & Hunter, 2004; Schmidt & Hunter, 1998; Hunter & Hunter, 1984).

In this dissertation, I asked the following questions: Why haven't we been able to better predict job performance to date? What literatures or areas have I/O psychologists not delved into sufficiently? Are there methods or tools that exist that can provide incremental validity in job performance? I pursued answers to these questions in this paper and arrived at the following conclusions: there have been avenues that I/O psychology has not fully explored, assumptions that should be violated, received doctrines that should be

reconsidered (Barrett,1972) so that we can get at the crux of the matter, and alternative approaches that should be resurrected into the research limelight.

II. CURRENT LITERATURE

Predictors of Performance

I/O psychologists have conducted a great deal of research on predictors of performance (e.g., Barrick, Mount, & Judge, 2001; Hunter & Hunter, 1984; Murphy & Shiarella, 1997; Schmidt & Hunter, 2004). The general consensus of the field is that general cognitive ability, or *g*, is still the predictor with the greatest predictive ability. Correlations between scores on intelligence tests and measures of job performance typically lie between $r = .30$ and $r = .50$ (Neisser, 1996). Furthermore, when cognitive ability is corrected for unreliability, the corrected correlation has been reported to be $r = .54$ (Hunter, 1983), which suggests that cognitive ability test performance accounts for approximately 29% of the variance in overall job performance. Additionally, cognitive ability also has the greatest predictive efficiency: it has a strong relationship with performance, it is relatively simple and inexpensive to administer, and it can be administered at any point in the selection system (i.e., without prior knowledge or experience). Although much data suggest that *g* is a good predictor of job performance, many employers decide not to rely on *g* and/or *g* alone. One of the foremost criticisms of *g* has been that it demonstrates adverse impact, particularly for African Americans and Hispanics (Hough, Oswald, & Ployhart, 2001). On average, African Americans have scored a full standard deviation (SD) lower on

intelligence tests than their Caucasian counterparts, and Hispanics have scored a half of a SD lower (Hough et al., 2001). These score differences can significantly impact pass rates and differential hiring, which would make it more difficult for companies to adhere to EEOC guidelines (Equal Employment Opportunity Commission, 1978).

Additionally, many researchers believe that measuring specific abilities is more useful when the goal is greater understanding as opposed to simple predictive efficiency (e.g., Alderton & Larson, 1994; Murphy, 1996). Furthermore, there are critics that have indicated that intelligence measures only correlate with task performance and have little to no relationship with contextual performance, which could be of equal importance to organizations (e.g., Borman, Hanson, & Hedge, 1997; McCloy, Campbell, & Cudeck,1994; Motowidlo & Van Scotter, 1994). These issues with the intelligence research have paved the way for research into alternate predictors (e.g., personality) that might better predict things like contextual performance without demonstrating adverse impact.

Personality. Research into personality as a predictor of job performance has experienced recently something of a renaissance (i.e., Barrick, Mount, & Judge, 2001). The general acceptance of a Five Factor Model (FFM or Big Five) of personality (Costa & McCrae, 1992) has made a major contribution to the resurgence of interest in personality as a predictor. Further, some researchers have suggested that adding personality measures in a selection context over and above cognitive tests has added incremental validity because personality predicts contextual performance (Motowidlo & Van Scotter, 1994). Contextual

performance consists of activities that contribute to the social and psychological core of the organization, as opposed to task performance, which consists of activities that contribute to the technical core (Borman & Motowidlo, 1997). Contextual performance is more likely to be voluntary in nature; whereas task activities are typically prescribed by the role. In fact, personality indices have been predictive of both Organizational Citizenship Behaviors (OCBs; Organ & Ryan, 1995) and Counterproductive Work Behaviors (Berry, Ones, & Sackett, 2007).

Specific Predictors. Research into predictors also has covered an array of predictors that assess specific abilities and job knowledge. Some of these predictors have included: situational judgment tests (SJTs), work samples, biodata, Assessment Centers (ACs), and structured interviews (e.g., McDaniel et al., 2001; Chan & Schmitt, 2002; Palumbo, 2008; Hunter & Hunter, 1984; Mount, Witt, & Barrick, 2000; Arthur, Day, McNelly, & Edens, 2003). The research on these predictors has waned in recent years. Despite evidence of the predictive validity of these predictors, issues concerning their internal construct validity remain problematic (e.g., Borman, Hanson, Hedge, 1997). There has also been another school of researchers who have focused particularly on predictor-criterion matching and finding job-specific predictors based on validation studies as opposed to broad overarching predictors. Although job-specific predictors are oftentimes the best in terms of predictive validity, they have been criticized for their expense and lack of generalizability. In other words, tailor-made, specific

predictors may lead to greater predictive validity, but oftentimes at the cost of broad application or generalizability.

Broad versus Narrow Predictors Debate

One of the intense current debates in the I/O literature is the broad versus narrow debate (Dudley et al., 2006), which has focused on whether job-specific, narrow constructs or broad, general constructs are most predictive of job performance. One of the foremost challenges underlying this debate has been that the terms 'broad' and 'narrow' have not been clearly defined (e.g., when discussing level of breadth, there has been an inherent ambiguity regarding what exactly would be considered broad or narrow). Furthermore, to date, the breadth of a measure has not necessarily been indicative of whether a construct being measured is heterogeneous or homogeneous. Despite the lack of consensus regarding the definitions of broad and narrow constructs, most researchers have agreed that broad constructs are more general and uncontextualized (i.e., they can apply across many different types of situations), whereas narrow traits are more situation-specific. The application of this debate in selection raises the question of whether broad or narrow constructs make better predictors of job performance. There have been strong proponents on both sides of this debate.

Advocates for broad predictors have demonstrated that broad factors are positively related to job performance across many different job types and performance criteria (e.g., Barrick, Mount, & Judge, 2001; Salgado, 1997). Moreover, some have touted that broad personality traits exhibited higher

predictive validity than their dimensions across a range of job performance criteria (e.g., Ones & Viswesvaran, 1996).

Critics have countered that although broad constructs often have higher criterion-related validity than most of their dimensions, broad constructs frequently have lower criterion-related validity than at least one dimension (Ashton, 1998; Paunonen, Rothstein, & Jackson, 1999). Furthermore, advocates of narrow traits have posited that narrow trait measures maximize the predictive validity of specific performance criteria (Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995).

There has been a resurgence in research investigating broad singular personality constructs despite their modest predictive validity (Morgeson et al., 2007a). In fact, researchers have presented correlations between personality and performance in the .10s as evidence that personality testing "works". Even if a validity coefficient reaches .20, which would explain 4% of the variance in performance, such validity coefficients likely would not pass the "so what" test among practitioners (Cascio & Aguinis, 2008b). Many researchers have believed that personality tests are poor predictors of job performance because they are simply too broad to predict on-the-job performance (Miller, 2009). However, replacing a singular broad construct with a singular narrow construct has not improved significantly the prediction of job performance either. The current stalemate in predictive efficacy may have arisen from the fact that both sides of the broad versus narrow debate have been focusing on pairing singular

predictors with singular criteria as opposed to looking at more multidimensional predictors and criteria.

Multidimensionality. Researchers have often considered broad constructs multidimensional constructs; however, broad constructs are more of an amalgam in which many factors are blended together to form the essence of a construct. Broad constructs such as the Big Five are actually tapping where two constructs meet (i.e., the essence of both) rather than capturing key components of both constructs. For example, conscientiousness is a combination of organization and achievement motivation; however, two people can score equally well on conscientiousness but really be very different in their work style (e.g., one could be very organized, whereas the other could be very achievement oriented). When conceptualized in this manner, broad constructs are really singular constructs. In essence, broad constructs are an amalgam or blend rather than a mosaic in which many constructs are integrated but independent.

Researchers developed the Big Five initially via lexical reduction and arrived at the essence of personality in five factors (see Digman, 1990). Although these factors have been a good representation of personality as a whole across a variety of domains, these five factors have been lacking in the nuanced differences of their component parts. A broad construct is a single construct that is meant to capture its essence and be applicable across multiple situations. Although conscientiousness might capture the essence of organization and achievement orientation, what has been most important in selection research is that it captures the aspects of the facets that are most

important for the specific job situation. For example, when conscientiousness has been used to predict job performance, the items measuring conscientiousness would have been relevant to multiple domains: work, home, or school. However, in a selection context, the hiring manager should not care whether the individual was extremely organized in his/her home life (e.g., alphabetizing their canned goods), rather, s/he should have been more concerned with how an individual's level of organization applied to the specific work context (e.g., follows a sales plan or keeps prospect lists organized and up-to-date).

In fact, research on frame of reference effects has demonstrated that simply adding the tag 'at work' to the end of personality items can enhance predictive validity with work outcomes (Bing, Whanger, Davison, & VanHook, 2004). These work-specific personality measures have led to enhanced predictive validity because they identify the specific frame-of-reference to be used when interpreting items and are more conceptually similar to work outcomes (Heggestad & Gordon, 2008; Lievens, De Corte, & Schollaert, 2008). A hiring manager should be more concerned with multiple narrow traits, or multiple unidimensional constructs connected to a single, specific situation (i.e., work). Because both sides of the broad/narrow debate have focused on unidimensional predictors, the underlying assumption has been that the performance and the criteria by which performance has been measured have been unidimensional, which is far from being the case.

Heterogeneity of Performance

Job performance is behavior and not the consequences or results of some action (Campbell, 1990). Job performance has been defined as the total expected value to the organization of the discrete behavioral episodes that an individual carries out over a standard period of time (Motowidlo, 2003). In other words, job performance is a gestalt or a complex interplay of multiple, divergent factors. Because performance is comprised of a variety of behavioral incidents, it should be measured using a variety of criteria.

Performance systems typically have multiple predictors and multiple criteria (Murphy & Shiarella, 1997). Performance in virtually any job is multidimensional (McCloy, Campbell, & Cudeck, 1994) because performance is comprised of so many discrete incidents and measured in so many ways. This means that for any specific job there are a number of substantive performance components that are distinguishable in terms of their intercorrelations and patterns of covariation with other variables (Smith, 1976). This is further complicated by research that has shown that measures of maximum performance (which are typically being assessed in research) are not highly related to measures of typical performance (Sackett, Zedeck, & Fogli, 1988; Mangos, Steele-Johnson, LaHuis, & White, 2007). Typical performance is how an individual performs on a regular basis, whereas maximum performance is how an individual performs when exerting maximum effort on a job (Sackett, Zedeck, & Fogli, 1988). Lastly, research has also demonstrated that various situational

factors moderate the relationship between predictors and performance (e.g., Tett & Burnett, 2003), further convoluting the construct of performance.

The work of Campbell and colleagues has provided some of the best evidence that job performance is multidimensional (Campbell, 1990; Campbell, McCloy, Oppler, & Sager, 1993). Campbell et al. used data from the U.S. Army Research Institute's Selection and Classification Project (Project A; Campbell, 1990) to develop a database for an improved selection and classification system for initial assignments in to US Army occupations or Military Occupational Specialties. Project A was a large-scale multiyear research program. Researchers sampled jobs from the population of entry-level positions, developed a comprehensive battery of new selection and classification predictor measures, and constructed a comprehensive array of job performance measures.

The Project A research led to a hierarchical model of job performance which consists of three major determinants of performance and an eight dimensional structure of job performance that applies broadly across a sample of jobs, which is often referred to as the "Campbell Model". Campbell postulated three direct determinants of job performance: declarative knowledge (DK), procedural knowledge and skill (PKS), and motivation (M) (Campbell, 1990). According to the Campbell Model, these three determinants of performance are the basic building blocks of performance (but because they are not actual behaviors), these three determinants are not performance itself. Further, all of the predictors discussed in this paper have an indirect effect on performance by

changing the level of DK, PKS, or M. Campbell (1990) also proposed a hierarchical multi-factor model of performance which defined eight behavioral dimensions of performance: Job-specific task proficiency, non-job-specific task proficiency, written and oral communications, demonstrating effort, maintaining personal discipline, facilitating team and peer performance, supervision, and management and administration. This eight factor model describes the top of the latent hierarchy in all jobs in the Dictionary of Occupational Titles, but these eight factors have different patterns of subgeneral factors and differential content depending upon the job (Campbell, 1990). Despite general agreement among researchers that Campbell's model is a thorough representation of the job performance domain, many researchers continue to search for and use simpler models of job performance for their own research.

There have been some researchers who have boiled performance down to a single construct, but the majority of current researchers acknowledge that there are both task and contextual performance. Task performance consists of the activities that usually appear on formal job descriptions. Contextual performance is behavior that contributes to the organizational effectiveness through its effects on the psychological, social, and organizational context of work (e.g., promoting positive affect in others). Many believe that the concept of 'overall job performance' is lost without considering contextual performance, or the individual's social and non-technical contributions to the workplace, in addition to task performance (e.g., Borman & Motowidlo, 1997).

Over-Simplification of Job Performance Domain

Researchers have been very effective at simplifying the job performance domain, such that they have bordered on over-simplification. The numerous simplistic current models of job performance are evidence of the simplification of the job performance domain. The most current, prominent models of job performance consist of few dimensions, do not address interactions between dimensions or moderator effects, ignore context-related variables, and treat performance as a static entity.

For example, Viswesvaran, Schmidt, and Ones (2005) suggested that performance can be explained largely (60% of the variance) by a single latent factor; however, this model was based strictly on supervisor ratings of job performance. Because supervisor ratings and actual performance metrics are not highly correlated (Bommer, Johnson, Rich, Podsakoff, & Mackenzie, 1995), relying on this model to understand performance would be weak at best. Further, if taken at face value, Viswesvaran et al.'s (2005) findings still suggest that 40% of the variance in job performance is unexplained. Consequently, such a simplified model has little practical applicability.

Although researchers have pushed forward the agenda of developing a broad taxonomy of job performance to enhance understanding, many researchers believe that organizations will continue to need more narrow and job-focused measures of performance for legal defensibility and employee feedback (Arvey & Murphy, 1998). Even researchers who are proponents of studying broad constructs that underlie performance have agreed that in

personnel selection, a demonstration of criterion-related validity is essential for the operational usefulness of a test to be established (Ones & Viswesvaran, 2001).

Criteria Heterogeneity

Much of the emphasis in studying the performance-prediction problem has been on developing and refining predictors, with too little attention paid to the performance criteria these measures are supposed to predict (Campbell, 1990). Researchers have paid little attention to performance criteria in part because criteria are often complex and dynamic (Steele-Johnson, Osburn, & Pieper, 2000), thereby making research on them more complicated and less generalizable. What predicts a job one year may change the next as the job evolves to meet market demands and keep up with technological innovation. Moreover, criteria may be subjective, objective, or a combination of both. This muddies the waters considerably, especially when one factors in the fact that subjective and objective criteria are not strongly correlated (Bommer et al., 1995; Harris & Schaubroeck, 1988). The fact that performance can be measured by several different criteria that have modest relationships with one another suggests that researchers and practitioners should use different predictors depending upon the criteria that are available.

A really complete ultimate criterion is multiple and complex in almost every case and is difficult to attain in practice (Smith, 1976). Although the ultimate criterion (i.e., the perfect, hypothetical measure of 'ideal' performance) (Thorndike, 1949) may be unattainable, the closest approximation to the ultimate

criterion will only be attained by capturing as much of the criterion space as possible. Further, capturing a significant amount of the criterion space can lead to better prediction.

The heterogeneity of criteria has been at the crux of the criterion problem - the solution to which has eluded both researchers and academicians. The criterion problem is that criteria are dynamic, multidimensional, situation-specific, and serve multiple functions (Austin & Villanova, 1992). Consequently, the ability to conceptualize and measure them becomes quite difficult.

One way in which researchers have attempted to circumvent the criterion problem is by simplifying the criterion domain such that the criterion of job performance could be "predicted" with a single predictor. However, research has suggested that singular constructs can be "theoretically sterile" whereas aggregates of behaviors can provide a more complete understanding of behavior in organizations (Roznowski & Hanisch, 1990). Moreover, no single criterion has encompassed all of job performance, and no single predictor has been able to explain the majority of the variance in job performance. The search for *the* elusive criterion continues despite strong evidence that it does not exist (Smith, 1976). Further, the single best predictor that I/O psychologists have put forth to date is General Mental Ability (i.e., GMA or cognitive ability) (Neisser et al., 1996), which has only been able to account for approximately 25% of the variance in performance and has been associated with high adverse impact.

Rather than trying to simplify the criterion domain, acknowledging that job performance is complex and is characterized by dynamic and multidimensional

criteria might help move forward the science of prediction. That being the case, equally complex and job specific predictors would be needed to best predict job performance according to the behavioral consistency model (Mount, Muchinsky, & Hanser, 1977). Moreover, researchers may benefit from using a combination of measurement methods and predictor constructs to represent more of the criteria space in attempt to approximate the ultimate criterion. If researchers developed a better understanding of what differentiates high and low performers, they could potentially predict as well as or better than *g* while minimizing the potential for adverse impact. The research below focused on how best to predict multidimensional criteria and posited that multidimensional predictors best predict multidimensional criteria. At the crux of this notion was the consistency model.

The Consistency Model

Over the years, many researchers have put forth the notion that matching predictors and criteria across multiple dimensions leads to better prediction (e.g., Ajzen & Fishbein, 1977; Smith, 1976; Wernimont & Campbell, 1968). Wernimont and Campbell (1968) argued that it would be much more fruitful to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance. Additionally, Ajzen and Fishbein (1977) put forth a compatibility principle, concluding that attitude-behavior connections are strongest when an attitude is matched in specificity or generality to behavior. Further, Smith (1976) articulated multiple possible sources and measures of performance variation and posited that criteria should parallel predictors in generality and immediacy. Similarly, in more recent research, frame of reference

effects have surfaced which suggests that more work-specific personality measures generally yield stronger relationships with work criteria than general personality (Bing, Whanger, Davison, & VanHook, 2004; Bowling & Burns, 2010).

Despite all of the evidence that samples of performance are the most predictive of future performance, I/O research continues to rely heavily on signs of performance (e.g., selection tests) rather than samples (Wernimont & Campbell, 1968). Researchers continue to favor the use of tests and broad constructs because of their simplicity and generalizability, which lend themselves easily to theory building. However, simplifying the performance predictor and criterion domain may be hindering researchers' ability to predict performance with greater accuracy.

The ability to accurately predict performance may be eluding I/O researchers because the approach that has been used to date has been overly simplistic. I/O researchers have often focused on measuring one dimension or construct very reliably, as opposed to looking at multiple dimensions that differentiate between high and low performers. Researchers have focused on measuring a single dimension in part because singular constructs can be measured more reliably than multiple dimensions in one test. However, if performance is multidimensional, it stands to reason that a predictor should be multidimensional in order to better represent the construct domain. Advocates of broad constructs argue that a single construct can explicate performance (e.g., cognitive ability or personality) or that performance can be reduced to a single factor (e.g., Viswesvaran et al., 2005). The current research has taken an

opposing view and suggested that performance is multidimensional and any predictor used to predict performance should also be multidimensional.

Predictor Heterogeneity

Research Supporting Heterogeneous Predictors. There has been strong criterion-related validity evidence to suggest that heterogeneous, job specific predictors such as work samples, background data, situational judgment tests, and assessment centers are useful predictors. Criterion-related validity is the statistical demonstration of a relationship between scores on an assessment and the job performance of a sample of workers. Work sample tests have been considered among the most valid predictors of job performance (Schmidt & Hunter, 2004; Hunter & Hunter, 1984). A work sample test requires an applicant to perform tasks that are similar to those that are performed on the job. Hunter and Hunter (1984) reported that the validity of work sample tests for predicting supervisory ratings was r = .54. Because of their point-to-point correspondence with the criterion (Mount, Muchinsky, & Hanser, 1977), some research has even suggested that work samples could be stronger predictors than *g* (Palumbo, 2008). Hunter and Hunter (1984) reported that for experienced workers, the validity of work sample exams was slightly higher than the validity of cognitive ability tests (r = .51). Moreover, work samples have demonstrated lower levels of standardized ethnic group differences than cognitive ability tests (Palumbo, 2008; Schmitt, Clause, & Pulakos, 1996).

Biodata measures have also been considered among the best predictors of job performance (Hunter & Hunter, 1984; Mount, Witt, & Barrick, 2000;

Mumford, Costanza, Connelly, & Johnson, 1996). Biodata is biographical data which can include questions about life and work experiences, opinions, values, beliefs, and attitudes. Biodata measures have proven to be effective predictors of college grades, yielding a validity coefficient of r = .36 (Mumford et al., 1996). Further, biodata have been predictive of objective performance indices (r = .41) and training criteria (r = .38). Research has also suggested that biodata scales were less fakeable than personality scales intended to measure the same constructs (Kilcullen, White, Mumford, & Mack, 1995).

Situational judgment tests (SJTs) are simulations requiring the respondent to exercise judgment when responding to hypothetical problem situations that occur in work settings. Many have considered SJTs heterogeneous measurement methods rather than measures of a single construct (Chan & Schmitt, 1997; McDaniel & Nguyen, 2001). SJTs have gained increasing popularity which has been driven both by the validity of the tests (McDaniel et al., 2001) and by findings of smaller mean differences among racial subgroups as compared with traditional cognitive ability tests (Motowidlo, Dunnette, & Carter, 1990; Pulakos & Schmitt, 1996). SJTs have demonstrated substantial validity for the prediction of job performance (ρ = .34) (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). Although *g* has been the largest correlate with SJTs (observed mean r = .36, $p < .01$) (McDaniel et al., 2001), black/white mean differences average .38 SDs favoring whites, as compared with the full 1 SD favoring whites of cognitive ability measures. Moreover, SJTs have typically shown incremental validity over cognitive ability tests. Lastly, some studies have

reported incremental validity of a SJT over a battery containing cognitive ability and personality (Chan & Schmitt, 2002; Clevenger et al., 2001; Weekley & Ployhart, 2006). In short, SJTs have demonstrated validity in predicting performance and smaller sub-group difference scores than measures of *g* (Ployhart & MacKenzie Jr., 2011).

Many researchers have agreed that Assessment Centers are also good predictors of job performance (Hunter & Hunter, 1984, Arthur et al., 2003). Assessment centers are typically developed to measure a candidate's management aptitude. This is determined by a variety of exercises which may include, but are not limited to, group exercises, presentations, and examinations. They have become widely used because of their predictive validity, and they typically demonstrate less adverse impact than *g* (Cascio & Aquinis, 2005, p. 372; Hoffman & Thornton, 1997). Although assessment centers have some correlation with *g* and personality, they are thought to be measuring a host of managerial competencies.

Research Opposing Heterogeneous Predictors

Construct validity. One of the foremost criticisms of situational judgment tests, assessment centers, and biodata is that they have poor construct validity. Construct validity is the degree to which a test measures what it purports to measure (Cronbach & Meehl, 1955). The underlying commonality between these predictors that contributes to their poor construct validity is that researchers have demonstrated their predictiveness empirically, but these predictors were developed independently of the nomological network and

research. These predictors have often been described in method-based terms and developed to simulate the job itself as opposed to a specific predictor construct (Roth, Bobko, McFarland, & Buster, 2008). The multidimensionality and job-specific nature of these predictors (that makes them more valid predictors) has significantly contributed to their weak construct validity and low internal consistency reliability.

Although situational judgment tests exhibit fairly strong criterion-related validities and smaller racial and sex subgroup differences than other methods, little to no construct validity evidence has been presented in most studies assessing SJT's predictive validity (Ployhart & MacKenzie Jr., 2011). Moreover, despite many researchers' attempts at creating response options that target specific constructs, there has been limited success in improving convergent validity with other measures of similar constructs (e.g., Ployhart & Ryan, 2000). Because SJTs are job-centered as opposed to construct-centered (Roth et al., 2008), they have typically had lower internal consistency reliability than constructs such as cognitive ability or personality (Ployhart & MacKenzie Jr., 2001).

Despite the established predictive validity of biodata measures, perhaps the most common criticism has been that there is limited evidence available for their content and construct validity (Katzell, 1994). Additionally, although assessment centers have emerged as one of the most popular tools for evaluating individual differences related to managerial performance, their poor

construct-related validity continues to be perceived as a weakness (Bowler & Woehr, 2009).

Assessment centers have demonstrated strong content and criterion-related validity (see Arthur & Day, 2011); however, many reviews have suggested that they have weak construct validity (Fleenor, 1996; Sackett & Harris, 1988). The assessment center's ability to display relatively satisfactory levels of content and criterion-related validity, but weak construct related validity has been referred to as the AC construct-related validity paradox (Arthur, Woehr, & Maldegan, 2000; Woehr & Arthur, 2003) because according to the Unitarian framework of validity, if a test demonstrates criterion or content validity, it should also demonstrate construct validity (Binning & Barrett, 1989).

Although researchers have developed assessment centers typically to measure specific dimensions, studies have demonstrated that exercise variance makes a greater contribution to validity than dimension-related variance. For example, Fleenor (1996) found that the average mean correlation within exercise was .22, while the mean correlation among various dimensions within each exercise was .42. This research demonstrated the impact of context variables and the importance of taking into account how performance varies across contexts. The fact that exercise effects made a greater contribution to criterion-related validity than did dimension effects suggests that exercise variance may be simply an example of a situationally-specific demonstration of performance. When exercise variance was removed, criterion-related validity had a tendency to decrease.

Internal consistency reliability. Some researchers and practitioners have vilified job-specific, heterogeneous predictors because of their low internal consistency reliability. In fact, many studies investigating these predictors have chosen not to include measures of reliability because they are historically low. Coefficient alpha, the most commonly used estimate of reliability, is an inappropriate estimate of the reliability of heterogeneous measures like SJTs (McDaniel et al., 2007). The same principle applies to work sample tests and background data scales. Researchers must design background data items to cover multiple situations, a factor that may limit the magnitude of internal consistency coefficients unless a large number of items (30 or more) are in use (Mumford & Stokes, 1992). Oftentimes, studies have opted to use alternate measures of reliability (e.g., test/retest) when dealing with job-specific heterogeneous predictors (e.g., Roth, Bobko, & McFarland, 2005).

Most I/O researchers agree that all things being equal, higher reliability leads to higher validity and lower reliability leads to lower validity (Nunnally & Bernstein, 1994). Some would say that I/O researchers have considered this notion received doctrine, or a fact that is not to be challenged (Barrett, 1972). Because classical test theory is rooted so strongly in the idea that reliability underscores validity, many researchers just try not to broach the topic when discussing heterogeneous predictors.

The Case for Predictor Heterogeneity

The two most significant arguments against predictor heterogeneity are that heterogeneous predictors have poor construct validity and poor internal

consistency reliability. The following section addresses these concerns and makes the case that despite these weaknesses, there is still merit in predictor heterogeneity.

Addressing the construct validity argument. All of the aforementioned heterogeneous predictors have demonstrated criterion-related validity (i.e., heterogeneous predictors have meaningfully and statistically significantly correlated with a particular criterion such as supervisor ratings of job performance) but have not demonstrated strong construct validity. Construct validity, or validating the measures by investigating their relationships with other measures from the research literature, has typically been more interesting to researchers as it allows one to put the relationships in a nomological network and helps researchers explain how a relationship between variables fits into previous research. Compared to the construct approach, results from criterion-related validation studies have seldom revealed how a variable fits into the research literature. Although the difficulty for researchers in establishing construct validity might be an issue, for practitioners, there may actually be greater utility in criterion-related validity over and above construct validity. Criterion-related validity has generally been more legally defensible (Van Iddekinge & Ployhart, 2008) and easier to explain to lay audiences (e.g., a company's clients) (Lissitz & Samuelsen, 2007). Construct validity is a difficult concept to explain to lay audiences, whereas criterion-related validity is fairly straight-forward to explain. As Einstein reportedly once said, "You do not really understand something unless you can explain it to your grandmother." And, construct validity, as it is

currently articulated has essentially flunked the grandmother test (Lissitz & Samuelsen, 2007). Conversely, criterion-related validity conceptually provides evidence about how well the test represents a construct and asks whether the test item differentiates between high and low performers (Carrier, Dalessio, & Brown, 1990), which is a little easier to explain to lay audiences.

From a research perspective, depending upon how you conceptualize construct validity, any demonstration of criterion-related validity can be considered one piece of the construct validity puzzle. Researchers have often explained construct validity in statistical terms by determining a single underlying factor that an assessment is measuring. However, according to the unitary view of validity, other forms of validity evidence (e.g., criterion-related validity or content validity) are essentially components of construct validity. For those who adopt a broader understanding of construct validity, criterion-related validity and construct validity are enmeshed. According to Messick (1995), the evidence and rationales supporting the trustworthiness of score meaning are what is meant by construct validity. That is, construct validity is the evidential basis for score interpretation—not just those involving so-called “theoretical constructs.” Almost any kind of information about a test can contribute to an understanding of score meaning, but the contribution becomes stronger if researchers explicitly evaluate the degree of fit of the interpretation (Cronbach, 1988; Messick, 1989). Consequently, construct validity subsumes content relevance and representativeness as well as criterion-relatedness (Messick, 1995).

Addressing the reliability argument. Heterogeneous predictors typically have had weak internal consistency reliability, which is problematic because according to classical test theory (Crocker & Algina, 1986; Lord & Novick, 1968; Novick, 1966), reliability is necessary for validity and sets the upper bounds of validity. Classical test theorists believe that there is no way to directly observe or calculate the true score, so they use a variety of methods to assess the reliability of a test (e.g., internal consistency reliability, inter-rater reliability, parallel-forms reliability, and test-retest reliability). Although many ways of assessing the reliability of a test exist, I/O researchers use and cite internal consistency reliability most often in their research (Cronbach, 1951; Hogan, Benjamin & Brezinski, 2000). Internal consistency reliability, simply put, measures how much items correlate with one another. However, internal consistency reliability is only one estimate of the reliability index. If internal consistency reliability set the upper bound for validity, this would suggest that heterogeneous predictors could not have strong validity. This statement is problematic because many heterogeneous predictors with low internal consistency exist that have demonstrated strong predictive validity (e.g., assessment centers and SJTs).

The reliability index may set the upper bound for validity; however, when the only form of reliability proffered is internal consistency reliability—it alone may not necessarily set the upper bound for validity. Internal consistency reliability cannot set the upper bound of validity because it is an imperfect estimate of the reliability index with its own sources of error variance. By its very nature, high internal consistency reliability is concurrent with more homogeneous scales or

assessments. Yet, there is much research that suggests that heterogeneous measures with relatively low internal consistency reliability can demonstrate high validity. Despite the fact that heterogeneous predictors with low internal consistency reliability have not necessarily been indicative of low validity (contrary to the notion that reliability underscores validity), there is little research in the I/O literature rectifying this seeming contradiction. However, educational testing literature has already investigated this issue.

The educational testing literature on heterogeneous measures has suggested that very few reliability analyses take the multidimensional structure of empirical data into account (Brunner & Süb, 2005). Statistically, internal consistency reliability estimation applies only to measures that are unidimensional or homogeneous (Gerbing & Anderson, 1988; Hunter & Gerbing 1982), thereby raising fundamental questions regarding the meaning and assessment of reliability for measures of multidimensional constructs (Brunner & Süb, 2005). Dimensions of multidimensional constructs are often heterogeneous (Roznowski & Hanisch, 1990). Dimensions of a multidimensional construct are necessarily heterogeneous because they represent different facets or manifestations of a construct. As dimension heterogeneity increases, correlations among the dimensions decrease, which in turn reduces the internal consistency reliability of summed dimensions scores (Brunner & Süb, 2005).

The most popular coefficient for internal consistency has been Cronbach's alpha (Cronbach, 1951; Hogan, Benjamin & Brezinski, 2000); however, reliability estimates based on internal consistency are irrelevant for measures of

heterogeneous constructs (Hanisch, Hulin, & Roznowski, 1998). In short, researchers should not use alpha to assess the reliability of heterogeneous measures. Only if the error terms are uncorrelated and the scale indicators are essentially tau equivalent will Cronbach's alpha accurately estimate the reliability of a scale (Raykov, 1998). Consequently, it is not appropriate to rely solely on benchmark values of .7 or .8 for construct reliability to decide whether scales should be used in research investigations, because this would preclude the investigation of interesting theoretical questions (Brunner & Süb, 2005).

Tests with high internal consistency are virtually always unidimensional or homogeneous, which means that they are not likely to correlate well with heterogeneous criteria. Thus, we should only expect a predictor with high internal consistency to predict unidimensional criteria. Conversely, if job performance is actually a multidimensional construct, as is proposed in this paper, it is necessary to use multidimensional predictors to predict job performance.

There is some I/O research that supports these ideas and demonstrates that internal consistency reliability does not set the upper bounds of validity when it comes to heterogeneous predictors. Data from Huffcutt and Arthur (1994) did not support the hypothesis that the superior validity of more structured interviews can be explained by their higher reliabilities (Schmidt & Zimmerman, 2004). In this study, the operational validities were r = .33 for unstructured interviews and r = .44 for structured interviews. After interview unreliability was corrected for, the true score validity was the same (r = .54) for both types of interviews (Schmidt &

Zimmerman, 2004). Perhaps, rather than viewing reliability as an internal aspect of validity (Sireci, 2007), we should adopt Campbell and Fiske's (1959) characterization of reliability and validity as ends of a continuum (i.e., reliability as monomethod-monotrait validity). Although the reliability index may set the upper bound for true validity, the notion that internal consistency reliability sets the upper bounds of validity is only true if you make the assumption that performance is a homogeneous construct.

Many consider the statement 'reliability sets the upper bound for validity' to be a received doctrine (Barrett, 1972). However, constructs must necessarily be homogeneous when asserting that internal consistency reliability sets the upper bound of validity. Although internal consistency reliability is not the only type of reliability, it is the most oft cited form of reliability in most I/O research. This conflict between depth and breadth of coverage has often been viewed as entailing a trade-off between validity and reliability (or generalizability). Notwithstanding, it might better be depicted as a trade-off between the valid description of the specifics of a complex task and the power of construct interpretation (Messick, 1995).

In conclusion. Although it is rarely explicitly stated, researchers who study predictors of job performance focus primarily on highly explanatory, theoretical predictors rather than empirical predictors that demonstrate strong statistical relationships but lack a theoretical rationale. A hundred years of research with this focus has not led us remotely close to perfect prediction, which suggests that the predictors that offer the most logical explanatory power (i.e., have the best

construct validity and internal consistency reliability) may not necessarily have the most predictive power. The predictors with the most explanatory power may have the strongest link to the theory associated with criteria in general, but they may not have the strongest link to the specific criteria in a given research or applied situation.

The majority of I/O researchers believe that empirical relationships between predictor scores and criterion measures should make theoretical sense in terms of what the predictor test measures and what the criterion embodies (Gulliksen, 1950). Some would even go so far as to suggest that the goal of research is to obtain parameter estimates that are accurate, regardless of their magnitude (Johns, 1998). However, parameter estimates that are accurate and small in magnitude really have little utility in a field that purports to provide valuable tools for practitioners as well as researchers. Many practitioners would suggest that maximizing criterion-related validity should be one of the most important goals of selection research. The academician focus on explanations concerning nomological networks, traits, and theory as opposed to a focus on maximizing criterion related validity may help explain the large researcher/practitioner divide in this field.

We have yet to find a theory that is all-encompassing for understanding job performance, which is evidenced by our inability to predict it. One of the reasons why we have yet to find a comprehensive theory that explains all of job performance is because job performance and the criteria by which it is measured are dynamic and heterogeneous, and equally heterogeneous predictors are

needed to predict it. Recent research has focused heavily on broader more homogeneous predictors for the sake of simplicity, but I suggest that more heterogeneous predictors are what are really needed to predict job performance with greater accuracy. Taking an inductive (as opposed to a deductive) approach to research might help open researchers' eyes to other predictors and lead them to develop a more comprehensive theory of job performance. Conducting more inductive research will force researchers to probe and learn more about the predictor space through observation prior to developing theories. Recent research has focused heavily on predictors with weak predictive validity because they have a strong theoretical basis (e.g., personality); however, researchers need to move away from this mentality and explore new avenues if they want to uncover how to best predict job performance. The current understanding of job performance has been hindered by a heavy focus on developing theories when the current understanding of job performance continues to be fairly narrow.

III. HYPOTHESES

After over 80 years of research, I/O psychologists have only done a modest job of predicting performance (Cascio & Aguinis, 2008). One of the reasons why we have not been able to better predict performance may be because researchers have favored singular and homogeneous predictor constructs that are easily generalizable over predictor constructs that are multiple, narrow, situation-specific, and divergent. Although the simplicity of this research focus may have led to more overarching theories of performance prediction, it likely will not lead to overarching theories that predict performance with any more accuracy than has been achieved to date.

Performance is a gestalt: its comprising factors are multiple, narrow, and heterogeneous and interact in a way such that the sum is greater than its component parts. Subsequently, choosing predictor constructs that are a constellation of multiple narrow, situation-specific, divergent constructs is a key way to capture more of the variance in performance and attempt to capture the gestalt. The following research served as an attempt to demonstrate that performance is in fact a gestalt and that a constellation of multiple, narrow, divergent component parts is a better predictor than what has been used traditionally. In short, this research has investigated whether there is greater value in using more empirically-derived, heterogeneous predictors than theoretically-derived, homogeneous predictors.

This study compared homogeneous scales that were developed to measure a single broad construct against heterogeneous scales that were developed via empirical keying and designed to measure multiple narrow constructs. Because researchers have found that objective and subjective performance criteria are not highly related (Bommer et al., 1995), the following hypotheses correlated the homogeneous and heterogeneous scales with objective and subjective job performance criteria.

The following hypotheses utilized a dataset that was largely composed of validation studies for sales jobs. The heterogeneous scales and profiles that were under investigation were developed for account management roles (i.e., Farmer roles) and new business development roles (i.e., Hunter roles). The Hunter-Farmer typology dichotomizes salespeople such that hunters acquire new business and close deals while farmers maintain and grow existing business.

Salespeople that specialize in account management are highly customer-oriented. They typically specialize in sales that involve cultivating long-term relationships and growing revenue in existing accounts. Usually, these sales people take a long-term approach that focuses on building a partnership (see Appendix A for greater detail). New business development salespeople are opportunistic closers who generally have minimal post-sale contact. Hunters specialize in generating leads, introducing new products or novel product applications to new prospects, and closing new accounts. Typically, after Hunters close an account, they quickly move on to other potential clients as

opposed to cultivating a long-term relationship (see Appendix B for greater detail).

Hypothesis 1: Heterogeneous scales that comprise an account management profile will have higher positive correlations with subjective criteria for account management roles than homogenous scales and heterogeneous scales that comprise a new business development profile.

Hypothesis 2: Heterogeneous scales that comprise an account management profile will have higher positive correlations with objective criteria for account management roles than homogenous scales and heterogeneous scales that comprise a new business development profile.

Hypothesis 3: Heterogeneous scales that comprise a new business development profile will have higher positive correlations with subjective criteria for new business development roles than homogenous scales and heterogeneous scales that comprise an account management profile.

Hypothesis 4: Heterogeneous scales that comprise a new business development profile will have higher positive correlations with objective criteria for new business development roles than homogenous scales and heterogeneous scales that comprise an account management profile.

The Federal Government has urged employers to investigate and use alternative selection procedures that are equally valid, but produce less adverse impact (Uniform Guidelines on Employee Selection Procedures, 1978), which has spurred research on predictors besides cognitive ability. One of the foremost reasons for the continued investigation of personality measures as predictors of

performance is that they do not adversely impact minorities. If an alternative selection method was presented that also did not adversely impact minorities, continued investigation into that method might be warranted. Because the empirically developed heterogeneous predictors from the previous hypotheses are not specifically tapping into cognitive ability, I believed that they would not adversely impact minorities and deserved further study.

Hypothesis 5a: Heterogeneous scales will not demonstrate adverse impact for women, African Americans or Hispanics

Hypothesis 5b: Heterogeneous scales will demonstrate less adverse impact on women, African Americans and Hispanics than homogeneous scales

IV. METHOD

Participants

The data used for this study was gathered from a proprietary archival database of incumbents and applicants from a variety of industries who completed the consulting firm's assessment. Incumbents were individuals who already had a position within a company and were likely taking the assessment for developmental purposes, whereas applicants were people applying for a position with a company. The complete archive was used to conduct the adverse impact analyses for Hypothesis 5. A subset of the archival database that had performance data and met certain inclusion criteria were used to test the first four hypotheses. The majority of participants in the complete archive were male (63.8%) and Caucasian (79.0%). Approximately 8.1% of participants were African American and 5.5% were Hispanic. The participants ranged in age from 18 to 97 (M = 35.51, SD = 10.07) and generally had a bachelor's degree or higher (57.7%). The tenure distribution for the archive ranged from 0 to 498 months, with an average tenure of approximately 24 months (M = 23.56, SD = 46.25).

The subset of the archival database that had performance data and met the inclusion criteria (i.e., validation study sample) was predominantly male (71.2%) and Caucasian (94.8%). There was a smaller proportion of participants that were African American (2.7%) and Hispanic (1.5%) in this sample. The

average participant from the validation study sample was a little older and more educated than the full archive: participants ranged in age from 21 to 78 (*M* = 40.52, *SD* = 10.14) and 66.8% had a bachelor's degree or higher. In the validation study sample, the tenure distribution ranged from 0 to 431 months, with an average tenure of approximately 76 months (*M* = 76.17, *SD* = 66.03). The differences between the validation study sample and the complete archival database likely reflect the higher proportion of incumbents versus applicants in the validation study pool. Moreover, if these differences effected the correlations reported in this study, the effect would have led to more conservative estimates of the true relationship.

Hypothesis Testing Approaches

There were a couple of approaches to testing the hypotheses. Effect sizes could be calculated using profiles (i.e., composites) of account management, new business development and homogeneous scales, or alternatively, effect sizes could be calculated looking individually at the scale level. There were pros and cons to each approach. Using profiles led to simpler interpretation and was more in line with the way in which the consulting firm analyzed its validation studies. Conversely, calculating effect sizes at the scale level provided some nuanced details regarding specific scales that one would not be able to see if a composite were used. Because there are merits to both of these approaches, I conducted the analyses looking at both the profile level and the scale level.

Inclusion Criteria

Only studies whose participants were applicants or incumbents were used (i.e., no student samples). When multiple objective or subjective criteria were provided, the primary criteria that were used in the original validation study were given priority over other criteria. In cases when the primary criteria used in the original validation study were not explicitly clear, supervisor ratings of job performance were given priority over other available measures for "subjective performance criteria", and sales dollars were given priority over other available measures for "objective performance criteria."

Additionally, any classifications that were conducted used independent coders (i.e., the primary author and a subject matter expert (SME)). For example, job descriptions and job analyses were carefully reviewed by independent coders against a list of key words and a profile role description for New Business Development and Account Management roles to classify jobs as either NBD or AM roles (see Appendices A and B). For instances where jobs were not easily classified as strictly NBD or AM roles, additional classifications included: both AM and NBD roles, neither AM nor NBD roles, non-sales role, or insufficient information to make a classification. These coders made classifications independently and then discussed any discrepancies until a consensus was reached.

Measures

The measures that this study investigated are proprietary measures developed by a Midwest consulting firm. This company specializes in the

selection of business-to-business (B2B) sales people. The original research supporting the consulting firm's sales profiles was sponsored by the U.S. Justice Department in 1973 to develop legally defensible and effectively predictive items. Their research has revealed that the characteristics of successful salespeople vary depending on the content of the job. The consulting firm believes that varied markets and industries in combination with the product/service that a salesperson is selling dictate the selling relationship. The way a salesperson sells is a function of both complexity and customer experience (Stevens & Cox, 1992). The level of complexity dictates the degree of touch that needs to be done on the part of the salesperson (e.g., staying in touch, keeping in close contact with the customer), whereas the level of a customer's experience with a product dictates the extent to which the seller has to assist the buyer in understanding and using the product/service.

Scales within the consulting firm's assessment were developed utilizing both construct-driven and criterion-driven approaches. Some scales were developed to measure a unitary hypothetical construct through traditional test development (construct-related methods), whereas others were developed by correlating job performance criteria with work-related competencies (empirical keying method). In the empirical keying method, the items that accounted for the greatest amount of variance in the performance criteria were joined parsimoniously to create a scale representing the competency. Multiple, fairly independent items were joined together to form scales that are heterogeneous, and consequentially have weak internal consistency reliability. As the

development of the consulting firm's assessment predated research on the Big Five Factor model of personality, the assessment did not have scales analogous to the Big Five factors.

This study investigated the nature and strength of the relationship between heterogeneous and homogeneous predictors and subjective and objective indices of job performance. A proprietary dataset provided this study with criterion-related validity data. This dataset included 42 local validation studies conducted using the homogeneous and heterogeneous scales outlined above. This *k* is comparable to many published meta-analyses. The availability of numerous independent samples lent itself to the application of meta-analysis methodology. Meta-analysis refers to the compilation of results from numerous independent studies to arrive at an estimate of the effect size of a relationship of the population as a whole. Examining results across multiple studies serves to eliminate various sources of error that may be due to statistical artifacts (e.g., sampling error, measurement error, and range restriction) and other methodological issues. Meta-analysis can test whether observed inconsistencies across studies are due to artifacts or substantive moderation. Furthermore, by controlling for statistical artifacts, meta-analysis has an advantage over narrative reviews in that it can uncover relationships that more closely reflect the true underlying relationships. The present study applied the meta-analysis procedure described by Hunter and Schmidt (1990).

The organizations were drawn from a number of different industries, and the majority of validated profiles were for sales roles. The overall effect size, ρ,

was estimated by calculating the weighted mean correlation across studies, thus correcting for sampling error. This corrected correlation weighted the correlation by the sample size from which it was drawn so that the overall mean correlation was more heavily weighted by larger samples than small ones. Correlations from larger samples tend to be more reliable and better estimates of the population as a whole. Additionally, 95% confidence intervals were computed to assess the accuracy of the estimate of the mean observed correlation. Confidence intervals estimate the extent that sampling error remains in the sample size-weighted mean correlation and provide a range of values that the mean observed correlations would likely take if other studies were pulled from the population and used in the meta-analysis.

The overall effect size was also corrected for range restriction and unreliability in the criterion. Range restriction corrections were applied because the data from the validation studies represent a range of test scores from an incumbent population, which is greatly reduced when compared to the range of test scores for the applicant population. The standard deviations reported in the validation studies were used to estimate the amount of range restriction present and incorporated in range restriction calculations. Corrections for measurement error or unreliability in the criterion were also conducted to account for insufficient variability in the sample. It has long been recognized that insufficient variability in a sample restricts the observed magnitude of a Pearson product moment coefficient (Mendoza & Mumford, 1987). Formulas for these calculations can be found in Hunter and Schmidt (1990).

Lastly, the lower 90% credibility value (CV) was computed also to assess whether validities were generalizable. The lower 90% CV indicates that ninety percent of the estimates of true validity lie above the given value (Whitener, 1990). If the CV is greater than zero, then one can reasonably conclude that validity generalizes (Hunter & Schmidt, 1990).

V. RESULTS

Sample Characteristics

In total, I was able to find 17 validation studies (N = 1,211) for account management roles with subjective criteria and 4 validation studies (N = 287) for account management roles with objective criteria. For new business development roles, I found 21validation studies (N=1,742) that had subjective criteria and 5 validation studies (N = 538) with objective criteria. It should be noted that unlike the majority of research studies which are carried out on a sample of subjects rather than the whole population, in this case, the data on the entire population of validation studies conducted by the consulting firm were available for study. As a result, the reported sample sizes and standard deviations were collected from the entire population and not a subset.

Coding Agreement

The jobs within the validation database were classified as either account management jobs or new business development jobs. These classifications were made by independent coders who conducted a qualitative analysis of company-provided job descriptions. Two independent coders (i.e., the author of the study and a SME) were given a list of criteria (as specified by an internal expert on the proprietary assessment; see Appendices A and B) by which to examine and categorize jobs as either account management or new business development roles. If job analysis information was not available, then the

correlation corresponding to that validation study was not used. Correlations from jobs that were deemed to be a combination of account management and new business development sales were also excluded from the study. Disagreements were resolved by discussion until a consensus was reached. Agreement had to be obtained unanimously across coders; otherwise, the correlation coefficient corresponding to a given validation was excluded from study.

Objective vs. Subjective Criteria

The hypotheses set forth in this paper looked at the relationship between homogeneous and heterogeneous predictors and objective and subjective criteria. I wanted to investigate the relationship between these predictors and different types of criteria because researchers have found that objective and subjective performance criteria are not highly related (Bommer et al., 1995). In the cases where I had both objective and subjective performance data for the account management role, the relationship between objective and subjective performance data was weak ($r(309) = .22$, $p < .01$). This finding supports my rationale for incorporating both types of performance criteria in my hypotheses. Interestingly, for the new business development role, the relationship between objective and subjective data was considerably stronger ($r(431) = .69$, $p < .01$). This may suggest that there is more of a clear correspondence between the activities that a NBD salesperson carries out with the sales achieved than an AM salesperson. In other words, the work tasks that NBD salespeople carry out may be more directly aligned with objective measures of their performance (e.g., sales

dollars) than AM salespeople. This finding may reflect the shorter life cycle of NBD sales as well as a clearer correspondence between NBD job tasks and on-the-job-performance. Account managers, on the other hand, typically have longer sales cycles that involve building trust and developing long-term relationships with clients – both of which have a more nebulous relationship with objective performance outcomes and are less easily observed.

Tests of Hypotheses

Hypotheses 1 and 2 were tested using scales that were taken from a proprietary assessment instrument designed to measure relevant work competencies. Heterogeneous scales were selected that best represented account management and new business development jobs. These scales were chosen based on their content relevance to the account management and new business development jobs. The heterogeneous scales that were selected as most representative of an account management (AM) role were called: 'Maximizes Results by Systematically Managing an Account Plan', 'Driven to Increase Sales to Existing Accounts', 'Works the System for the Customer', 'Educates Customers through Structured Training', and 'Promotes Customer Relations by Soliciting Feedback.' The heterogeneous scales that were selected as most representative of a new business development (NBD) role were called: 'Problem Solving,' 'Qualifying Prospects with Standard Probes,' 'Commits Time and Effort to Ensure Success,' and 'Closes through Logical, Incremental Steps."

As mentioned earlier, there were a couple of approaches to testing the first 4 hypotheses. Effect sizes could be calculated using profiles (i.e.,

composites) of account management, new business development and homogeneous scales or at the individual scale level. Because there are pros and cons with each of these approaches, I conducted the analyses looking at both the profile level and scale level. The overall results of the hypotheses did not significantly change depending upon whether we looked at effect sizes at the scale or profile level. Consequently, I primarily reported the results from the profile level to facilitate results interpretation. In cases where there was a sufficient sample to have reliable analyses, I reported any notable follow-up analyses that were conducted at the individual scale level.

To test Hypotheses 1 and 2, I used a validation study database of account management jobs. Effect sizes were calculated between the heterogeneous scales that comprised the account management profile, homogeneous scales, heterogeneous new business development scales, and objective and subjective performance indices for account management roles. Correlations were corrected for sampling error, range restriction, and unreliability in the criterion. Additionally, 90% credibility values were calculated and the correlations were compared.

Hypothesis 1 was not supported. The heterogeneous account management (AM) profile did not have higher positive correlations with subjective criteria for AM roles than the homogeneous profile or heterogeneous new business development (NBD) profile. The effect size between subjective criteria for AM roles and the heterogeneous AM profile ($\rho_{yy, rr} = .042$; $CV_{low} = .008$) was smaller than the effect sizes found between subjective criteria for AM roles and the homogeneous profile ($\rho_{yy, rr} = .099$; $CV_{low} = -.048$) and between subjective

criteria for AM roles and heterogeneous NBD profile ($\rho_{yy, rr}$ = .084; CV_{low} = .084) (see Table 2). Although the effect sizes were modest, based on the credibility values, the heterogeneous AM and NBD profiles both had effect sizes that would be considered generalizable or statistically significant. However, the homogeneous profile did not have a generalizable effect size.

Similarly, the follow-up analyses at the individual scale level did not support Hypothesis 1. Heterogeneous account management scales did not have higher positive correlations with subjective criteria for account management roles than homogeneous scales or heterogeneous new business development scales. The estimated true score validities between subjective criteria for AM roles and heterogeneous AM scales ranged from -.069 to.099, with an average effect size of .013. The estimated true score validities between subjective criteria for AM roles and NBD scales ranged from -.039 to .156, with an average effect size of .051. Lastly, the estimated true score validities between subjective criteria for AM roles and homogeneous scales ranged from -.136 to .118, with an average effect size of .020. Although the corresponding effect sizes differed from the profile level analyses, the overall result was the same, and Hypothesis 1 was not supported even after the follow-up analyses (see Table 3).

Hypothesis 2 was also not supported. The heterogeneous AM profile did not have higher positive correlations with objective criteria for AM roles than the homogeneous profile or heterogeneous NBD profile. The effect size between objective criteria for AM roles and the heterogeneous AM profile ($\rho_{yy, rr}$ = -.062; CV_{low} = -.062) was comparable to the effect size found between objective

criteria for AM roles and homogeneous profile ($\rho_{yy, rr}$ = -.061; CV_{low} = -.061) and lower than the effect size found between objective criteria for AM roles and heterogeneous NBD profile ($\rho_{yy, rr}$ = .005, ; CV_{low} = .005) (see Table 4, please note variability equals zero, therefore the rho and credibility value low are the same). Based on the credibility values, only the heterogeneous NBD profile had an effect size that would be considered generalizable or statistically significant; however, the effect size was quite small. These results should be interpreted with some caution because the sample of validation studies with objective criteria was too small to have reliable analyses. Consequently, details regarding the individual scale level analyses were not discussed for the hypotheses where objective criteria were a correlate (see Table 5 for scale level analyses).

To test Hypotheses 3 and 4, I used the validation study database of new business development roles. Corrected correlations were conducted between the heterogeneous scales that comprised the new business development profile, homogeneous scales, the heterogeneous scales that comprised the account management profile, and objective and subjective performance indices for new business development roles. Correlations were corrected for sampling error, range restriction, and unreliability in the criterion. Additionally, 90% credibility values were computed, and the correlations were compared.

Hypothesis 3 was supported. The heterogeneous NBD profile had higher positive correlations with subjective criteria for NBD roles than the homogeneous profile and heterogeneous AM profile. The effect size between subjective criteria for NBD roles and the heterogeneous NBD profile ($\rho_{yy, rr}$ = .102; CV_{low} = .102)

was greater than the effect sizes found between subjective criteria for NBD roles and the homogeneous profile ($\rho_{yy, rr}$ = .080; CV_{low} = .088) and between subjective criteria for NBD roles and the heterogeneous AM profile ($\rho_{yy, rr}$ = .031; CV_{low} = .031) (see Table 6). Once again, the effect sizes were modest; however, based on the credibility values, all of the profiles (i.e., the heterogeneous AM and NBD profiles and homogeneous profile) had effect sizes that would be considered generalizable or statistically significant.

The follow-up analyses at the individual scale level continued to support Hypothesis 3. Heterogeneous new business development scales had higher positive correlations with subjective criteria for new business development roles than homogeneous scales or heterogeneous account management scales. The estimated true score validities between subjective criteria for NBD roles and heterogeneous NBD scales ranged from -.007 to .103, with an average effect size of .062. The estimated true score validities between subjective criteria for NBD roles and AM scales ranged from -.022 to .086, with an average effect size of .015. Lastly, the estimated true score validities between subjective criteria for NBD roles and homogeneous scales ranged from -.080 to .133, with an average effect size of .017. The effect sizes at the individual scale level differed from the profile level analyses; however, the scale level analyses provided further support for Hypothesis 3 (see Table 7).

Hypothesis 4 was not supported. The heterogeneous NBD profile did not have higher positive correlations with objective criteria for NBD roles than the homogeneous profile or heterogeneous AM profile. The effect size between

objective criteria for NBD roles and the heterogeneous NBD profile ($\rho_{yy, rr}$ = .009; CV_{low} = .009) was slightly lower than the effect size found between objective criteria for NBD roles and the homogeneous profile ($\rho_{yy, rr}$ = .013; CV_{low} = .013). Furthermore, the effect size between objective criteria for NBD roles and the heterogeneous NBD profile was lower than the effect size found between objective criteria for NBD roles and the heterogeneous AM profile ($\rho_{yy, rr}$ = .130; CV_{low} = .130) (see Table 8). The effect sizes were small, but the credibility values indicated that all of the profiles (i.e., the heterogeneous AM and NBD profiles and homogeneous profile) had effect sizes that would be considered generalizable or statistically significant. These results should be interpreted with some caution because the sample of validation studies with objective criteria was too small to have reliable analyses. Consequently, details regarding the individual scale level analyses will not be discussed here (see Table 9 for scale level analyses).

Hypothesis 5 was tested using an archival database with over 450,000 incumbents and applicants who completed the proprietary assessment. To test Hypotheses 5a, the heterogeneous scales from Hypotheses 1 through 4 were assessed on their adverse impact according to the 4/5ths rule on women, African Americans, and Hispanics (see Table 10). The 4/5ths, or 80% rule, which was codified in the Uniform Guidelines on Employee Selection Procedures (UGESP, section 4D) was used to assess adverse impact for this study because it is one of the most commonly used practical significance measure in the EEO context. The 4/5ths rule states that the selection ratio of those individuals in protected classes

should be at least 4/5ths (80%) of the selection ratio of the majority class. 4/5ths calculations were conducted using the archival database of participants who have completed the proprietary assessment. Calculating the selection ratio simply involved dividing the mean score of the protected class by the mean score of the majority class. Additionally, to test Hypothesis 5b, the 4/5ths rule calculations were conducted on the homogeneous scales from Hypotheses 1 through 4. These 4/5ths rule calculations were compared with the 4/5ths calculations that were conducted on the heterogeneous scales. Once again, I begin by reporting results at the profile level and then follow up with scale level analyses.

Although the 4/5ths rule is one of the most commonly used practical significance measures of adverse impact, 4/5ths rule analyses can be inaccurate in some situations (e.g., see Roth, Bobko, & Switzer, 2006). Roth et al. (2006) used simulation research to identify situations where the 4/5ths rule provided erroneous conclusions and found that false-positives (situations where the 4/5th rule was violated but selection rates were essentially equal) occurred at an alarming rate. Consequently, most experts in EEO view the 4/5th rule to be a general rule of thumb that can be used in combination with other evidence. Thusly, I have opted to also include Cohen's *d* in the adverse impact tables to supplement the 4/5ths rule calculations (see Tables 10-14).

Hypothesis 5a was supported. Namely, at the profile level, the heterogeneous scales from Hypotheses 1 through 4 did not demonstrate adverse impact according to the 4/5ths rule on women, African Americans, or Hispanics

(see Table 10). The selection ratio for women was 99% of the selection ratio for men. Further, the selection ratio for African Americans was 91% of the selection ratio for Caucasians, and the selection ratio for Hispanics was 92% of the selection ratio for Caucasians.

Additional support for Hypothesis 5a was found at the scale level analyses. The selection ratios for women ranged from 81% to 128% of the selection ratios for men, with an average selection ratio of 100%. The selection ratios for African Americans ranged from 81% to 126% of the selection ratios for Caucasians, with an average selection ratio of 95%. Lastly, the selection ratios for Hispanics ranged from 84% to 109% of the selection ratios for Caucasians, with an average selection ratio of 98% (see Tables 11 and 12).

Hypothesis 5b was not supported. The heterogeneous profile did not demonstrate significantly less adverse impact on women, African Americans, or Hispanics than homogeneous scales (see Table 10). At the profile level, heterogeneous scales had selection ratios for women (99%) that were comparable to the selection ratios for women on homogeneous scales (98%). The selection ratios for African Americans and Hispanics also did not differ greatly whether they were heterogeneous or homogeneous scales. The heterogeneous profile had selection ratios for African Americans (91%) that were similar to the selection ratios for African Americans on homogeneous scales (96%). Further, the heterogeneous profile had selection ratios for Hispanics (92%) that were comparable to the selection ratios for Hispanics on the homogeneous profile (90%). In sum, the heterogeneous profile did not

demonstrate significantly less adverse impact for women, African Americans, and Hispanics than the homogeneous profile. Neither of the heterogeneous and homogeneous profiles demonstrated adverse impact on protected classes. Scale level analyses for race and gender led to similar conclusions as the profile level analyses for Hypothesis 5b (see Tables 13 and 14). The selection ratios for heterogeneous and homogeneous scales had little adverse impact on women, African Americans, and Hispanics at the scale and profile level.

VI. DISCUSSION AND IMPLICATIONS

The majority of this study's hypotheses were not supported; however, there is some evidence from this meta-analysis which suggests that heterogeneous predictors may have added value in terms of their predictive validity and lack of adverse impact. Hypotheses 1 and 2 were not supported, suggesting that the heterogeneous account management scales did not have significantly greater predictive validity than heterogeneous new business development scales or homogeneous scales on subjective and objective criteria across a variety of jobs. Hypothesis 3 was supported, which suggests that heterogeneous new business development scales had greater predictive validity than heterogeneous account management scales and homogeneous scales on subjective criteria across jobs. Although Hypothesis 3 was supported, the magnitude of the validity difference was not large enough to be deemed practically significant. Hypothesis 4 was not supported. Heterogeneous new business development scales did not yield higher positive correlations with objective criteria than heterogeneous account management scales or homogeneous scales. Lastly, there was some support for Hypothesis 5. Namely, heterogeneous scales did not demonstrate adverse impact for women or ethnic minorities.

Although the majority of our hypotheses were not supported, this study does provide some evidence that there is utility in heterogeneous predictors.

The heterogeneous predictors' validity coefficients were not higher than comparable homogenous scales by a statistically significant margin; however, there were a number of heterogeneous scales that had validity coefficients that were comparable to correlations that have been reported between personality and performance. Further, the heterogeneous scales had comparable adverse impact ratios to homogeneous scales. Although there was a general lack of support for my initial hypotheses, when interpreting the results of a meta-analysis, significance testing is not as important as it is for a single study.

In fact, Schmidt (1996) believed that hypothesis testing is misleading at best when reporting the results of a meta-analysis. In his article, Schmidt (1996) posited that significance testing slows the growth of cumulative research knowledge and should be replaced with point estimates and confidence intervals. Despite the myriad of limitations to this study, when one looks at the point estimates from the scale level data, there are a number of validity coefficients that are comparable to validity coefficients found between personality and job performance. At the scale level, there were six heterogeneous scales with effect sizes that were greater than .10, with effect sizes ranging from .10 to .16 (see Tables 3, 5, 7, & 9). Although these effect sizes are modest, they are comparable to effect sizes between personality measures and performance found in published meta-analyses (e.g., Barrick, Mount, & Judge, 2001). There were five homogeneous scales with effect sizes greater than .10, with effect sizes ranging from .10 to .13 (see Tables 3, 5, 7, & 9).

Although the majority of the hypotheses were not supported, the support found for Hypothesis 3 has some important implications. Hypothesis 3 reveals that NBD heterogeneous scales had a relationship with subjective criteria that generalized across studies. This means that the consulting firm successfully developed scales of a heterogeneous nature that predicted across different employers and industries. Classical test theorists would suggest that predictors with weak internal consistency that were not specifically designed to measure an underlying latent construct should not demonstrate predictive validity that generalizes across studies. These findings may make researchers who ascribe to the classical test model uncomfortable because at its surface it calls into question the existing research literature and established Psychometric theory. However, in reality, this finding corroborates the extensive published literature that finds that predictor measures that are not designed to maximize internal consistency (i.e., assessment centers, SJTs, etc.) can demonstrate predictive validity. Additionally, the partial support found for Hypothesis 5 suggests that heterogeneous predictors also do not demonstrate significant adverse impact on protected classes.

There were a number of limitations to this study. One of the foremost limitations of this study was the fact that the data being analyzed was archival in nature. The archival data had a number of statistical issues (e.g., insufficient job analysis data, sampling error, questionable choice of primary criterion, inability to parse applicant and incumbent data). Furthermore, because the data were not

designed for the sole purpose of testing my hypotheses, there might be some questions regarding its appropriateness.

Lack of sufficient job analysis data made coding the jobs quite challenging, which limited the number of studies that could be included in the final analyses and could raise questions regarding the studies that were included. Because of the insufficient job analysis data, approximately half of the validation studies had to be excluded. Moreover, there were a number of jobs that could not strictly be categorized as new business development or account management roles, and those validation studies were excluded as well.

In fact, there were almost as many studies where the job analyses indicated that the roles were dual hunter/farmer roles as there were job analyses indicating that the roles were separate. The difficulty in separating jobs into strict account management/new business development categories raises questions regarding the literature that suggests that there is a hunter/farmer dichotomy in sales (Stevens & Cox, 1992, Thompson, Miller, Leasher, Dean, & Tristan, 2007). Based on our review of job analyses, I would suggest that the hunter/farmer dichotomy in sales is not as clear cut as the sales literature would suggest. Rather, account management roles and new business development roles fall more along a continuum.

In the end, many of the original validation studies that were available were not used to test the hypotheses. Although this study used two independent coders to categorize jobs, if different coders were used, some of the decisions for

inclusion might have differed and had an effect on the final samples and subsequent validity coefficients.

Another limitation to this study was that the majority of studies that were included in the meta-analysis had fairly small sample sizes, which made it such that there were few representative samples and a large amount of sampling error. Although I attempted to eliminate sampling error by correcting for it by using weighted mean correlations, the sampling error may have been so great in these data that standard corrections were not sufficient to correct for the issue. Because a large proportion of available validation studies had small samples, there was very little variance, as evidenced by the number of standard deviations and credibility values that were zero as well as the significantly overlapping confidence intervals that were reported.

An additional challenge when sorting through the validation data was deciding which criterion to use as a primary objective or subjective criterion. For a number of the validation studies, multiple criteria were available (this was especially the case for subjective criteria). Generally, the criterion that was chosen was the one used in the original validation study. This decision rule was applied based on the assumption that the consulting firm chose the most appropriate criteria for the validation study; however, there is no guarantee of that being the case. For example, there were some validation studies where there were multiple years of subjective criteria, and only a given year was selected. If I chose to use the mean of all years of subjective criteria, it may have led to a different final effect size for that study. Further, in the cases where there

were objective data, sometimes the primary criterion used was sales dollars and sometimes it was quota. Again, I chose to use the primary criterion that was used in the original validation study. However, had I chosen to only use sales dollars or quota, it may have impacted the final effect size for those studies. Additionally, it was sometimes unclear which criterion variable was originally used, particularly for some of the older validation studies. As with all meta-analyses, some judgment calls had to be made for some of the studies.

The study was also limited because I was unable to parse out whether the data were from applicants or incumbents. Because the data were used for both developmental and selection purposes, sometimes there were applicant data in the same validation study as incumbent data. This was an area that was practically impossible to parse out, and it may have had an effect on the final results. Although the proprietary assessment was designed with the intent of being non-face valid to minimize faking and other threats to validity, research suggests that applicants are more motivated to do well on selection tests, which could lead to a difference in how they answer test questions (Rosse, Stecher, Miller, & Levin, 1998). Thusly, the validation study outcomes could have been affected by the test takers' differences in motivation.

By and large, the data from the validation studies represented the test scores from an incumbent population, which was likely to be considerably lower than the range of test scores for an applicant population. As a result, the standard deviations reported were very small. In fact, in some cases the standard deviations were actually zero (see Tables 1-9). Although corrections for

range restriction and criterion unreliability were conducted, they may not have been sufficient to account for the lack of variability in the sample. This lack of sample variability may have restricted the observed magnitude of the reported correlations.

An additional limitation of this study was that the "homogeneous" predictors used were proprietary in nature and not subject to the rigorous convergent and discriminant validity analyses that are conducted to assess construct validity in peer-reviewed journals. Moreover, there were no homogeneous predictors that were analogous to the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Goldman, 1980). Because the development of the proprietary assessment predated research on the Big Five Factors of personality, the assessment did not have scales that directly corresponded with the Big Five. Further research would have to be conducted on these data to establish these scales and profiles as truly homogeneous in nature.

Based on my earlier assertions regarding the relationship between reliability and homogeneity and heterogeneity, calculating alpha for the various scales used in this study would have been one method to assess their homogeneity or heterogeneity. Namely, the scales with the lowest alphas would be the most heterogeneous and those with the highest alphas would be the most homogeneous. However, conducting reliability analyses on these data presented a significant challenge because there were a number of assumptions that underlie reliability analyses that would have been violated. One of the foremost

challenges was that the scales in their raw form were ordinal as opposed to interval data. Although higher scores were indicative of higher performance on a given scale, the differences between the values was not meaningful.

Despite the challenges and limitations associated with using these data, a meta-analysis of this nature could not have been conducted on published research because the research on heterogeneous predictors is simply not available. Further, the archival data allowed access to a really large database with many years of selection data, which helps bolster arguments regarding the generalizability of the findings. Moreover, its use decreases threats to internal validity like experimenter bias.

Although this study's hypotheses were not largely supported, there is evidence to suggest that heterogeneous predictors have comparable predictive efficacy with homogeneous predictors. Moreover, heterogeneous predictors did not adversely impact protected classes. Consequently, there is utility in more heterogeneous predictors and using them as predictors could significantly broaden the job performance predictor space. Hopefully, this research will incite more research into empirical methods and resurrect predictors that have been cast aside because of their poor construct validity and internal consistency reliability. Increasing the scope of potential predictors will enable industrial/organizational psychologists to predict job performance with much greater accuracy.

References

Alderton, D. L., & Larson, G. E. (1994). Dimensions of ability: diminishing returns? In *Personnel Selection and Classification* (pp. 137-144). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Arthur, W., Jr., & Day, E. A. (2011). Assessment centers. In S. Zedeck, (Ed.) *Handbook of Industrial and Organizational Psychology: Volume 2, Selecting and developing members for the organization (pp. 205-235).* APA, Washington DC.

Arthur, W., Jr., Day, E. D., McNelly, T. L., Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*(1), 125-154.

Arthur, W., Jr., Woehr, D. J., & Maldegan, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management, 26*, 813-835.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168.

Ashton, M. C. (1998). Personality and job performance: the importance of narrow traits. *Journal of Organizational Behavior, 19*, 289-303.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E. & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific factor scales. *Journal of Research in Personality, 29*, 432-442.

Austin, J. T. & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77,* 836-874.

Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888-918.

Barrett, G. V. (1972). Research models of the future for industrial and Organizational psychology. *Personnel Psychology, 25,* 1-18.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection & Assessment, 9*, 9-30.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*(2), 410-424.

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

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*(3), 478-494.

Blickle, G., Momm., T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. Construct and Criterion-Related Validation of a Measure of Emotional Reasoning Skills: A two-study investigation. *International Journal of Selection and Assessment, 17*(1), 101-118.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of subjective and objective measures of employee performance: A metaanalysis. *Personnel Psychology, 48,* 587-605.

Borman, W. C. (1978). Exploring the upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology, 63*(2), 135-144

Borman, W. C., Hanson, M. A., Hedge, J. W. (1997). Personnel selection. *Annual Review of Psychology, 48*, 299-337.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*(2), 71-83.

Bowler, M. C., & Woehr, D. J. (2009). Assessment center construct-related validity: Stepping beyond the MTMM matrix. *Journal of Vocational Behavior, 75,* 173-182.

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

Brunner, M., & Süb, H. M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement, 65*(2), 227-240.

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

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, E. (1993). A theory of job performance. In N. Schmitt & W. C. Borman (Eds*.). Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.

Cannon, A. R., St. John, C. H. (2007). Measuring environmental complexity. *Organizational Research Methods, 10*(2), 296-321.

Carrier, M. R., Dalessio, A. T., Brown, S. H. (1990). Correspondence between estimates of content and criterion-related validity values. *Personnel Psychology, 43*, 85-100.

Cascio, W. F., & Aguinis, H. (2008a). Research on industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology, 93*(5), 1062-1081.

Cascio, W. F., & Aguinis, H. (2008b). Staffing twenty-first-century organizations. *The Academy of Management Annals, 2*(1), 133-165.

Cascio, W. F., & Aguinis, H. (2005). *Applied Psychology in Human Resource Management (6th Ed.).* Upper Saddle River, NJ: Prentice Hall.

Chan, D. & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*(3), 233-254.

Chan, D. & Schmitt, N. (1997). Video-based versus paper—and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity. *Journal of Applied Psychology, 82*(1), 143-159.

Cizek, G. J., Rosenberg, S. L. Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*(3), 397-412

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*(3), 410-417.

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325-351.

Costa, P. T., Jr. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of a test. *Psychometrika, 16*, 297-334.

Cronbach, L. J. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika, 53*(1), 63-70.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology, 41,* 417-440.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1), 40-57.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods, 4*(2), 144-192.

Equal Employment Opportunity Commission (1978). Uniform Guidelines on Employee Selection Procedures. *Federal Register, 43*, 38290-38309.

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10*(3), 319-335.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011-1027.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25*(2), 186-192.

Gulliksen, H. (1950). Intrinsic validity. *American Psychologist, 5*(10), 511-517.

Hanisch, K. A., Hulin, C. L., & Roznowski, M. (1998). The importance of individuals' repertoires of behaviors: The scientific appropriateness of studying multiple behaviors and general attitudes. *Journal of Organizational Behavior, 19*(5), 463-480.

Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*(1), 43-62.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

Heggestad, E. D., & Gordon, H. L. (2008). An argument for context-specific personality assessments. *Industrial and Organizational Psychology, 1*(3), 320-322.

Hinrichs, J. R., & Haanpera, S. (1976). Reliability of measurement in situational exercises: An assessment of the assessment center method. *Personnel Psychology, 29*, 31-40.

Hoffman, C. C., & Thornton, G. C. III (1997). Examining selection utility where competing predictors differ in adverse impact. *Personnel Psychology, 50*, 455-470.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*(4), 523-531.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1-2), 152-194.

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

Huffcutt, A. I., Conway, J. M., Roth, P. L., Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897-913.

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*(5), 459-473.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance Measurement and Theory (pp. 257-266).* Hillsdale, NJ: Erlbaum.

Hunter, J. E. & Gerbing, D. W. (1982). Machiavellian beliefs and personality: Construct invalidity of the Machiavellianism dimension. *Journal of Personality and Social Psychology, 43*(6), 1293-1305.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Newbury Park, California; London; New Delhi: SAGE Publications.

Johns, G. (1998). Aggregation or aggravation? The relative merits of a broad withdrawal construct. *Journal of Organizational Behavior, 19*, 453-462.

Jones, R. G., Born, M. Ph. (2008). Assessor constructs in use as the missing component in validation of assessment center dimensions: A critique and directions for research. *International Journal of Selection and Assessment, 16*(3), 229-238.

Katzell, R. A. (1994). Contemporary meta-trends in industrial and organizational psychology. In H. C. Triandis, M. D. Dunnette, M. D. & L. M. Hough, (Eds.) *Handbook of Industrial and Organizational Psychology, 2nd ed., (Vol.4) (pp. 1-89).* Palo Alto, CA: Consulting Psychologists Press.

Kilcullen, R. N., White, L. A., Mumford, M. D., & Mack, H. (1995). Assessing the construct validity of biodata scales. *Military Psychology, 7*(1), 17-28.

Klehe, U-C, Konig, C. J., Richter, G. M., Kleinmann, M., & Melchers, K., G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance, 21*, 107-137.

Lance, C. E., Teachout, M. S., Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology, 77*(4), 437-452.

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

Lissitz, R. W. & Samuelsen, K. (2007). Further clarification regarding validity and education. *Educational Researcher, 36*(8), 482-484.

Lord, F. M. & Novick, R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Mangos, P. M., Steele-Johnson, D., LaHuis, D., & White, E. D. (2007). A multiple-task measurement framework for assessing maximum-typical performance*. Human Performance, 20*(3), 241-258.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*(4), 493-505.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

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*(4), 730 – 740.

McDaniel, M.A. & Nguyen, N.T. (2001). Situational Judgment Tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9,* 103-113.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Messick, S. (1989). Validity. In the England American Council on Education, *Educational Measurement (3rd ed.).* New York, NY: Macmillan Publishing Co., Inc.

Miller, C. E. (2009). The trouble with personality tests. Retrieved from http://chally.com/global-research/executive-briefings/

Morgeson, F., Campion, M., Dipboye, R., Hollenbeck, J., Murphy, K., Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683-729.

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

Motowidlo, S. J. (2003). Job Performance. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.). *Handbook of Industrial and Organizational Psychology, 12,* 39-53.

Motowidlo, S. J., Dunnette, M. D., & Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(5), 640-647.

Motowidlo, S. J. & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*(4), 475-480.

Mount, M. K., Muchinsky, P. M., & Hanser, L. M. (1977). The predictive validity of a work sample: A laboratory study. *Personnel Psychology, 30*, 637-645.

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, 52*(2), 299-323.

Mumford, M. D., Costanza, D. P., Connelly, M. S., Johnson, J. F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology, 49*, 361-398.

Mumford, M. D. & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background measures. In M. D. Dunnette & L. M. Hough (Eds.). *Handbook of Industrial and Organizational Psychology, 3*(2), 61-138.

Murphy, K. R. (1996). *Individual differences and behavior in organizations*. San Francisco: Jossey-Bass Publishers.

Murphy, K. R. & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: multivariate frameworks for studying test validity. *Personnel Psychology, 50*, 823-854.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*(1), 1-18.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb III, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment, 15*(1), 19-29.

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

Ones, D. S. & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment, 9*(1/2), 31-39.

Organ, F. E., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48*(4), 775-802.

Palumbo, M. V. (2008). Cognitive ability, job knowledge, and stereotype threat: When does adverse impact result? *Dissertation Abstracts International: Section B: The Sciences and Engineering, 68*(7-B), 4877.

Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20,* 389-405.

Ployhart, R. E., & MacKenzie Jr., W. I. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.). *Handbook of Industrial and Organizational Psychology, 2*(2), 237-252.

Ployhart, R. E. & Ryan, A. M. (2000, April). *Integrating personality tests with situational judgment tests for the prediction of customer service performance.* Paper presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment, 13*(4), 304-315.

Pulakos, E. D., Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*(3), 241-258.

Raju, N. S., Pappas, S., & Williams, C. P. (1989). An empirical monte carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology, 74*(6), 901-911.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*(4), 375-385.

Rogers, W. M., Schmitt, N., & Mullins, M. E. (2002). Correction for unreliability of multifactor measures: Comparison of alpha and parallel forms approaches. *Organizational Research Methods, 5*(2), 184-199.

Rosse, J. G., Stecher, M. D., Miler, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*(4), 634-644.

Roth, P. L., Bobko, P., McFarland, L. A. (2005). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology, 61*(3), 637-661.

Roth, P. L., Bobko, P., McFarland, L. A., & Buster, M. (2008). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58,* 1009-1037.

Roth, P. L., Bobko, P., & Switzer, F. S. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology, 91*, 507-522.

Roznowski, M. & Hanisch, K. A. (1990). Building systematic heterogeneity into work attitudes and behavior measures. *Journal of Vocational Behavior, 36*, 361-375.

Sackett, P. R. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67*(4), 401-410.

Sackett, P. R. & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology, 3*(2), 214-229.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*(3), 482-486.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82*(1), 30-43.

Schippmann, J. S., Prien, E. P., & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology, 43,* 837-859.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*(2), 115-129.

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162-173.

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

Schmidt, F. L., & Zimmerman, R. D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology, 89*(3), 553-561.

Schmitt, N., Clause, C. S., Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In Cooper C. L., Robertson (Eds.), *International Review of Industrial and Organizational Psychology, 115-139.* New York: Wiley.

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.

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher, 36*(8), 477-481.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*(2), 211-219.

Smith, P. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M.D. Dunnette (Ed.) *Handbook of Industrial and Organizational Psychology,* 745-775. Chicago: Rand McNally.

Stevens, H. & Cox, J. (1992). The quadrant solution: A business novel that solves the mystery of sales success. New York, NY: Amacom.

Steele-Johnson, D., Osburn, H. G., & Pieper, K. F. (2000). A review and extension of current models of dynamic criteria. *International Journal of Selection and Assessment, 8*(3), 110-136.

Terpstra, D. E., Mohamed, A. A., & Kethley, R. B. (1999). An analysis of federal court cases involving nine selection devices. *International Journal of Selection and Assessment, 7*(1), 26-34.

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. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Tett, R. P., Jackson, D. N., Rothstein, M., Reddon, J. R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick and Hunter (1994). *Personnel Psychology*, 47(1), 157-172

Thompson, C. N., Miller, C. E., Leasher, M. K., Dean, S. L., & Tristan, E. (2007) Hunters and farmers: Increased prediction through sales role specialization. Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology, New York, April 2007.

Tracey, J. B., & Tews, M. J. (2005). Construct validity of a general training climate scale. *Organizational Research Methods, 8*(4), 353-374.

Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology, 61*(4), 871-925.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*(1), 108-131.

Walker, C. M. Ajzen, R., & Schmitt, T. (2006). Statistical versus substantive dimensionality: The effect of distributional differences on dimensionality assessment using DIMTEST. *Educational and Psychological Measurement, 66*(5), 721-738.

Weekley, J. A. & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*(5), 372-376.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75*(3), 315-321.

Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment Center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258.

Table 1

Effect Size between Criteria for Sales Roles and Heterogeneous and Homogeneous Profiles

Sales Role	Criteria	K	AM			NBD			Homogeneous		
			$\rho_{yy, rr}$	*SD*	CV_{low}	$\rho_{yy, rr}$	*SD*	CV_{low}	$\rho_{yy, rr}$	*SD*	CV_{low}
AM	Subj.	17	0.04	0.02	0.01	0.08	0.00	0.08	0.10	0.09	-0.05
AM	Obj.	4	-0.06	0.00	-0.06	0.01	0.00	0.01	-0.06	0.00	-0.06
NBD	Subj.	21	0.03	0.00	0.03	0.10	0.00	0.10	0.08	0.00	0.09
NBD	Obj.	5	0.13	0.00	0.13	0.01	0.00	0.01	0.01	0.00	0.01

Note. Hyp = Hypothesis; AM = Account Management; NBD = New Business Development; CV = credibility value; Subj. = Subjective; Obj. = Objective.

Table 2

Effect Size between Subjective Criteria for Account Management Roles and Heterogeneous and Homogeneous Profiles

Profile [a, b]	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	*SD* ρ	90% CV
AM	0.03	[-0.03, 0.09]	0.01	0.04	0.03	0.04	0.02	[0.01, 0.08]
NBD	0.07	[0.01, 0.13]	0.01	0.09	0.07	0.08	0.00	[0.08, 0.08]
Homogeneous	0.07	[0.02, 0.13]	0.02	0.10	0.08	0.10	0.09	[-0.05, 0.25]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development.

[a]n = 1,211; [b]*k* = 17

Table 3

Effect Size between Subjective Criteria for Account Management Roles and Heterogeneous and Homogeneous Scales

Scales[a]	n	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	SD ρ	90% CV
Account Management									
AM 1	1,211	0.07	[0.02, 0.13]	0.02	0.09	0.08	0.10	0.04	[0.03, 0.16]
AM 2	1,211	0.07	[0.01, 0.13]	0.01	0.09	0.07	0.09	0.00	[0.09, 0.09]
AM 3	1,211	0.00	[-0.06, 0.06]	0.01	0.00	0.00	0.00	0.00	[0.00, 0.00]
AM 4	1,211	-0.05	[-0.11, 0.00]	0.01	-0.07	-0.05	-0.07	0.01	[-0.08, -0.06]
AM 5	1,211	-0.04	[-0.09, 0.02]	0.01	-0.05	-0.04	-0.05	0.00	[-0.05, -0.05]
AM 6	1,211	0.00	[-0.05, 0.06]	0.02	0.00	0.00	0.00	0.12	[-0.19, 0.20]
New Business Development									
NBD 1	1,214	-0.03	[-0.09, 0.02]	0.02	-0.04	-0.03	-0.04	0.09	[-0.19, 0.12]
NBD 2	1,211	0.13	[0.07, 0.18]	0.02	0.16	0.12	0.16	0.10	[-0.01, 0.33]
NBD 3	1,211	0.05	[-0.01, 0.10]	0.01	0.06	0.05	0.06	0.00	[0.06, 0.06]
NBD 4	1,211	0.02	[-0.03, 0.08]	0.01	0.03	0.02	0.03	0.00	[0.03, 0.03]
Homogeneous									
Homo 1	1,214	-0.01	[-0.06, 0.05]	0.01	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
Homo 2	1,214	0.03	[-0.03, 0.09]	0.01	0.04	0.03	0.04	0.00	[0.04, 0.04]
Homo 3	1,214	0.06	[0.00, 0.12]	0.01	0.08	0.06	0.08	0.00	[0.08, 0.08]
Homo 4	1,214	0.09	[0.03, 0.15]	0.01	0.12	0.09	0.12	0.00	[0.12, 0.12]
Homo 5	1,215	0.03	[-0.03, 0.09]	0.03	0.04	0.03	0.04	0.13	[-0.17, 0.25]
Homo 6	1,213	0.01	[-0.05, 0.06]	0.02	0.01	0.01	0.01	0.04	[-0.05, 0.07]
Homo 7	1,214	0.10	[0.04, 0.15]	0.01	0.12	0.09	0.12	0.00	[0.12, 0.12]
Homo 8	1,214	0.01	[-0.04, 0.07]	0.01	0.02	0.01	0.02	0.00	[0.02, 0.02]
Homo 9	1,214	-0.01	[-0.06, 0.05]	0.02	-0.01	-0.01	-0.01	0.09	[-0.16, 0.15]
Homo 10	1,214	-0.07	[-0.13, -0.01]	0.01	-0.09	-0.07	-0.09	0.00	[-0.09, -0.09]

Homo 11	1,214	-0.10	[-0.16, -0.05]	0.01	-0.13	-0.11	-0.14	0.00	[-0.14, -0.14]
Homo 12	1,214	0.02	[-0.04, 0.07]	0.01	0.02	0.02	0.02	0.00	[0.02, 0.02]
Homo 13	1,215	0.01	[-0.05, 0.07]	0.01	0.01	0.01	0.01	0.00	[0.01, 0.01]
Homo 14	1,214	0.00	[-0.06, 0.06]	0.01	0.00	0.00	0.00	0.00	[0.00, 0.00]
Homo 15	1,214	0.02	[-0.04, 0.07]	0.02	0.02	0.02	0.02	0.09	[-0.13, 0.17]
Homo 16	1,214	0.02	[-0.04, 0.07]	0.01	0.02	0.02	0.02	0.00	[0.02, 0.02]
Homo 17	1,215	0.05	[0.00, 0.11]	0.02	0.07	0.05	0.06	0.10	[-0.1, 0.22]
Homo 18	1,215	0.06	[0.00, 0.11]	0.01	0.08	0.05	0.07	0.00	[0.07, 0.07]
Homo 19	1,214	0.02	[-0.04, 0.07]	0.02	0.02	0.02	0.02	0.05	[-0.07, 0.11]
Homo 20	1,214	-0.01	[-0.07, 0.04]	0.01	-0.02	-0.01	-0.02	0.00	[-0.02, -0.02]
Homo 21	1,215	-0.01	[-0.06, 0.05]	0.01	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
Homo 22	1,214	0.05	[-0.01, 0.11]	0.02	0.06	0.05	0.06	0.04	[-0.01, 0.13]

[a]k = 17.

Table 4

Effect Size between Objective Criteria for Account Management Roles and Heterogeneous and Homogeneous Profiles

Profile [a, b]	n	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	*SD* ρ	90% CV
AM	311	-0.04	[-0.15, 0.07]	0.01	-0.06	-0.05	-0.06	0.00	[-0.06, -0.06]
NBD	285	0.00	[-0.11, 0.12]	0.01	0.01	0.00	0.01	0.00	[0.01, 0.01]
Homo	316	-0.05	[-0.16, 0.07]	0.01	-0.06	-0.05	-0.06	0.00	[-0.06, -0.06]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development; Homo = Homogeneous.

[a] $k = 4$.

Table 5

Effect Size between Objective Criteria for Account Management Roles and Heterogeneous and Homogeneous Scales

Scales [a, b]	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	SD ρ	90% CV
Account Management								
AM 1	-0.11	[-0.22, 0.01]	0.00	-0.14	-0.11	-0.14	0.00	[-0.14, -0.14]
AM 2	-0.01	[-0.18, 0.15]	0.03	-0.02	-0.01	-0.02	0.15	[-0.26, 0.23]
AM 3	-0.03	[-0.15, 0.09]	0.01	-0.04	-0.03	-0.04	0.00	[-0.04, -0.04]
AM 4	0.05	[-0.07, 0.17]	0.02	0.07	0.05	0.06	0.07	[-0.04, 0.17]
AM 5	0.06	[-0.06, 0.18]	0.02	0.08	0.06	0.08	0.11	[-0.10, 0.26]
AM 6	-0.10	[-0.21, 0.02]	0.01	-0.13	-0.11	-0.14	0.00	[-0.14, -0.14]
New Business Development								
NBD 1	-0.09	[-0.20, 0.03]	0.01	-0.12	-0.08	-0.11	0.00	[-0.11, -0.11]
NBD 2	0.07	[-0.04, 0.19]	0.00	0.09	0.07	0.09	0.00	[0.09, 0.09]
NBD 3	-0.04	[-0.15, 0.08]	0.02	-0.05	-0.04	-0.05	0.11	[-0.23, 0.14]
NBD 4	0.02	[-0.09, 0.14]	0.01	0.03	0.02	0.03	0.00	[0.03, 0.03]
Homogeneous								
Homo 1	0.00	[-0.12, 0.11]	0.00	-0.01	0.00	-0.01	0.00	[-0.01, -0.01]
Homo 2	0.01	[-0.10, 0.13]	0.01	0.02	0.01	0.02	0.00	[0.02, 0.02]
Homo 3	0.03	[-0.09, 0.14]	0.01	0.03	0.02	0.03	0.00	[0.03, 0.03]
Homo 4	-0.02	[-0.14, 0.09]	0.01	-0.03	-0.02	-0.03	0.00	[-0.03, -0.03]
Homo 5	0.07	[-0.04, 0.19]	0.01	0.09	0.07	0.09	0.00	[0.09, 0.09]
Homo 6	0.04	[-0.07, 0.16]	0.00	0.06	0.04	0.06	0.00	[0.06, 0.06]
Homo 7	-0.05	[-0.17, 0.07]	0.01	-0.07	-0.05	-0.06	0.00	[-0.06, -0.06]
Homo 8	0.00	[-0.12, 0.11]	0.00	0.00	0.00	0.00	0.00	[0.00, 0.00]
Homo 9	-0.06	[-0.18, 0.06]	0.01	-0.08	-0.06	-0.08	0.00	[0.08, -0.08]
Homo 10	0.02	[-0.10, 0.14]	0.02	0.03	0.02	0.03	0.06	[-0.08, 0.13]

Homo 11	-0.03	[-0.14, 0.09]	0.01	-0.03	-0.03	-0.03	0.00	[-0.03, -0.03]
Homo 12	-0.04	[-0.16, 0.08]	0.00	-0.05	-0.04	-0.05	0.00	[-0.05, -0.05]
Homo 13	0.05	[-0.07, 0.16]	0.03	0.06	0.05	0.06	0.16	[-0.19, 0.32]
Homo 14	0.05	[-0.07, 0.17]	0.00	0.06	0.05	0.06	0.00	[0.06, 0.06]
Homo 15	-0.07	[-0.19, 0.04]	0.01	-0.09	-0.07	-0.09	0.00	[-0.09, -0.09]
Homo 16	0.05	[-0.06, 0.17]	0.01	0.07	0.06	0.07	0.00	[0.07, 0.07]
Homo 17	-0.01	[-0.12, 0.11]	0.00	-0.01	0.00	-0.01	0.00	[-0.01, -0.01]
Homo 18	-0.03	[-0.14, 0.09]	0.03	-0.03	-0.02	-0.03	0.15	[-0.28, 0.21]
Homo 19	-0.09	[-0.21, 0.02]	0.00	-0.12	-0.09	-0.12	0.00	[-0.12, -0.12]
Homo 20	-0.01	[-0.13, 0.10]	0.01	-0.02	-0.01	-0.02	0.00	[-0.02, -0.02]
Homo 21	-0.07	[-0.19, 0.04]	0.00	-0.10	-0.09	-0.11	0.00	[-0.11, -0.11]
Homo 22	0.01	[-0.11, 0.13]	0.00	0.01	0.01	0.01	0.00	[0.01, 0.01]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development.

[a]n = 287; [b]*k* = 4

Table 6

Effect Size between Subjective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Profiles

Profile [a, b]	n	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	*SD* ρ	90% CV
NBD	1,671	0.08	[0.03, 0.13]	0.01	0.10	0.08	0.10	0.00	[0.10, 0.10]
AM	1,671	0.02	[-0.03, 0.07]	0.01	0.03	0.02	0.03	0.00	[0.03, 0.03]
Homo	1,657	0.07	[0.02, 0.11]	0.01	0.08	0.07	0.08	0.00	[0.09, 0.09]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development; Homo = Homogeneous.

[a] $k = 21$

Table 7

Effect Size between Subjective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Scales

Scales [a]	*n*	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	SD ρ	90% CV
New Business Development									
NBD 1	1,742	-0.01	[-0.05, 0.04]	0.01	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
NBD 2	1,742	0.08	[0.03, 0.13]	0.01	0.11	0.08	0.10	0.00	[0.10, 0.10]
NBD 3	1,742	0.06	[0.02, 0.11]	0.01	0.08	0.06	0.08	0.00	[0.08, 0.08]
NBD 4	1,742	0.06	[0.01, 0.10]	0.01	0.07	0.06	0.07	0.05	[0.00, 0.15]
Account Management									
AM 1	1,742	0.07	[0.02, 0.11]	0.01	0.09	0.07	0.09	0.00	[0.09, 0.09]
AM 2	1,742	0.01	[-0.04, 0.06]	0.01	0.01	0.01	0.01	0.00	[0.01, 0.01]
AM 3	1,742	-0.02	[-0.06, 0.03]	0.01	-0.02	-0.02	-0.02	0.00	[-0.02, -0.02]
AM 4	1,742	-0.01	[-0.05, 0.04]	0.01	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
AM 5	1,742	-0.01	[-0.05, 0.04]	0.01	-0.01	-0.01	-0.01	0.05	[-0.10, 0.08]
AM 6	1,742	0.02	[-0.02, 0.07]	0.01	0.03	0.02	0.03	0.00	[0.03, 0.03]
Homogeneous									
Homo 1	1,660	-0.06	[-0.11, -0.01]	0.01	-0.08	-0.06	-0.08	0.00	[-0.08, -0.08]
Homo 2	1,658	-0.01	[-0.05, 0.04]	0.01	-0.01	-0.01	-0.01	0.05	[-0.09, 0.07]
Homo 3	1,660	0.05	[0.00, 0.10]	0.02	0.07	0.05	0.07	0.11	[-0.11, 0.25]
Homo 4	1,657	0.03	[-0.01, 0.08]	0.02	0.04	0.04	0.05	0.06	[-0.06, 0.15]
Homo 5	1,660	0.05	[0.00, 0.10]	0.01	0.07	0.05	0.06	0.00	[0.06, 0.06]
Homo 6	1,658	0.00	[-0.04, 0.05]	0.01	0.01	0.00	0.01	0.00	[0.01, 0.01]
Homo 7	1,660	0.10	[0.06, 0.15]	0.01	0.13	0.10	0.13	0.00	[0.13, 0.13]
Homo 8	1,660	0.04	[-0.01, 0.09]	0.01	0.05	0.04	0.05	0.00	[0.05, 0.05]
Homo 9	1,658	-0.02	[-0.07, 0.03]	0.01	-0.03	-0.02	-0.03	0.00	[-0.03, -0.03]
Homo 10	1,658	0.03	[-0.02, 0.08]	0.01	0.04	0.03	0.04	0.00	[0.04, 0.04]

Homo 11	1,658	-0.01	[-0.06, 0.04]	0.01	-0.01	-0.01	-0.01	0.01	[-0.02, 0.00]
Homo 12	1,658	-0.03	[-0.08, 0.02]	0.01	-0.04	-0.03	-0.04	0.00	[-0.04, -0.04]
Homo 13	1,660	0.06	[0.01, 0.10]	0.01	0.07	0.06	0.07	0.00	[0.07, 0.07]
Homo 14	1,660	0.04	[-0.01, 0.09]	0.01	0.05	0.04	0.05	0.00	[0.05, 0.05]
Homo 15	1,658	-0.01	[-0.05, 0.04]	0.01	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
Homo 16	1,658	-0.02	[-0.07, 0.03]	0.01	-0.02	-0.02	-0.02	0.00	[-0.02, -0.02]
Homo 17	1,660	0.01	[-0.04, 0.06]	0.01	0.02	0.01	0.01	0.00	[0.01, 0.01]
Homo 18	1,660	0.01	[-0.04, 0.06]	0.01	0.01	0.01	0.01	0.00	[0.01, 0.01]
Homo 19	1,661	-0.04	[-0.09, 0.01]	0.02	-0.05	-0.04	-0.05	0.06	[-0.14, 0.04]
Homo 20	1,660	0.03	[-0.02, 0.07]	0.01	0.03	0.03	0.03	0.00	[0.03, 0.03]
Homo 21	1,660	0.03	[-0.02, 0.08]	0.01	0.04	0.03	0.04	0.00	[0.04, 0.04]
Homo 22	1,659	0.00	[-0.05, 0.05]	0.01	0.00	0.00	0.00	0.00	[0.00, 0.00]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development; Homo = Homogeneous.

[a]*k* = 21

Table 8

Effect Size between Objective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Profiles

Profile [a, b]	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	*SD* ρ	90% CV
NBD	0.01	[-0.08, 0.09]	0.00	0.01	0.01	0.01	0.00	[0.01, 0.01]
AM	0.09	[0.01, 0.18]	0.01	0.12	0.10	0.13	0.00	[0.13, 0.13]
Homogeneous	0.01	[-0.07, 0.1]	0.01	0.01	0.01	0.01	0.00	[0.01, 0.01]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development.

[a]n = 538; [b]*k* = 5

Table 9

Effect Size between Objective Criteria for New Business Development Roles and Heterogeneous and Homogeneous Scales

Scales [a]	*n*	r_{obs}^{sw}	95% CI	σ^2_{obs}	ρ_{yy}	ρ_{rr}	$\rho_{yy, rr}$	SD ρ	90% CV
New Business Development									
NBD 1	538	-0.05	[-0.13, 0.04]	0.01	-0.06	-0.04	-0.05	0.00	[-0.05, -0.05]
NBD 2	539	0.03	[-0.05, 0.12]	0.00	0.04	0.03	0.04	0.00	[0.04, 0.04]
NBD 3	538	0.03	[-0.05, 0.12]	0.01	0.04	0.03	0.04	0.00	[0.04, 0.04]
NBD 4	537	0.00	[-0.08, 0.09]	0.01	0.00	0.00	0.00	0.00	[0.00, 0.00]
Account Management									
AM 1	538	0.10	[0.02, 0.18]	0.01	0.13	0.10	0.13	0.00	[0.13, 0.13]
AM 2	538	0.04	[-0.05, 0.12]	0.01	0.05	0.04	0.05	0.00	[0.05, 0.05]
AM 3	538	-0.01	[-0.09, 0.08]	0.00	-0.01	-0.01	-0.01	0.00	[-0.01, -0.01]
AM 4	538	0.07	[-0.01, 0.16]	0.00	0.09	0.08	0.10	0.00	[0.10, 0.10]
AM 5	538	-0.02	[-0.10, 0.07]	0.02	-0.02	-0.02	-0.02	0.10	[-0.19, 0.15]
AM 6	538	0.10	[0.01, 0.18]	0.01	0.13	0.10	0.12	0.00	[0.12, 0.12]
Homogeneous									
Homo 1	540	-0.10	[-0.24, 0.04]	0.03	-0.13	-0.10	-0.13	0.17	[-0.40, 0.15]
Homo 2	538	-0.08	[-0.17, 0]	0.01	-0.11	-0.08	-0.11	0.00	[-0.11, -0.11]
Homo 3	540	0.08	[-0.01, 0.16]	0.01	0.10	0.08	0.10	0.00	[0.10, 0.10]
Homo 4	538	0.06	[-0.02, 0.15]	0.01	0.08	0.07	0.08	0.00	[0.08, 0.08]
Homo 5	540	0.06	[-0.07, 0.19]	0.02	0.08	0.06	0.08	0.15	[-0.17, 0.31]
Homo 6	538	0.04	[-0.04, 0.13]	0.00	0.05	0.04	0.05	0.00	[0.05, 0.05]
Homo 7	540	0.10	[0.02, 0.19]	0.00	0.13	0.09	0.12	0.00	[0.12, 0.12]
Homo 8	540	0.00	[-0.08, 0.09]	0.01	0.00	0.00	0.00	0.00	[0.00, 0.00]
Homo 9	538	-0.03	[-0.11, 0.06]	0.01	-0.04	-0.03	-0.04	0.00	[-0.04, -0.04]
Homo 10	538	-0.06	[-0.14, 0.03]	0.01	-0.07	-0.06	-0.08	0.00	[-0.08, -0.08]

Homo 11	538	-0.07	[-0.16, 0.01]	0.01	-0.10	-0.08	-0.10	0.00	[-0.10, -0.10]
Homo 12	538	0.06	[-0.02, 0.15]	0.01	0.08	0.06	0.08	0.00	[0.08, 0.08]
Homo 13	540	-0.03	[-0.12, 0.05]	0.02	-0.04	-0.03	-0.04	0.10	[-0.21, 0.13]
Homo 14	540	-0.02	[-0.1, 0.07]	0.01	-0.03	-0.02	-0.02	0.08	[-0.15, 0.10]
Homo 15	538	0.02	[-0.07, 0.1]	0.02	0.02	0.02	0.02	0.12	[-0.17, 0.22]
Homo 16	538	0.04	[-0.04, 0.13]	0.00	0.05	0.04	0.06	0.00	[0.06, 0.06]
Homo 17	540	-0.05	[-0.14, 0.03]	0.01	-0.07	-0.05	-0.06	0.00	[-0.06, -0.06]
Homo 18	540	0.01	[-0.08, 0.09]	0.00	0.01	0.01	0.01	0.00	[0.01, 0.01]
Homo 19	540	-0.06	[-0.15, 0.02]	0.02	-0.08	-0.06	-0.08	0.10	[-0.25, 0.09]
Homo 20	540	0.02	[-0.07, 0.1]	0.01	0.02	0.02	0.02	0.00	[0.02, 0.02]
Homo 21	540	0.02	[-0.06, 0.11]	0.01	0.03	0.03	0.03	0.00	[0.03, 0.03]
Homo 22	539	0.07	[-0.01, 0.16]	0.00	0.09	0.07	0.09	0.00	[0.09, 0.09]

Note. CI = confidence interval; CV = credibility value; AM = Account Management; NBD = New Business Development.

[a]$k = 5$

Table 10

Adverse Impact Ratios for Heterogeneous and Homogeneous Profiles

Category	*n*	*M*	*SD*	AI Ratio	Cohen's *d*
			Heterogeneous		
Gender					
Male	234,877	0.845	0.362		
Female	134,289	0.834	0.372	0.987	0.029
Race					
Caucasian	277,224	0.864	0.343		
African American	28,860	0.784	0.411	0.908	0.210
Hispanic	19,330	0.793	0.405	0.918	0.189
			Homogeneous		
Gender					
Male	234,892	0.844	0.363		
Female	134,298	0.828	0.377	0.981	0.043
Race					
Caucasian	277,243	0.759	0.428		
African American	28,861	0.729	0.445	0.960	0.069
Hispanic	19,330	0.683	0.465	0.900	0.170

Note. AI = Adverse Impact.

Table 11

Gender Adverse Impact Ratios for Heterogeneous Scales

Scale	Gender	*n*	*M*	*SD*	AI Ratio	Cohen's *d*
Account Management						
AM Scale 1	Male	234,892	0.436	0.496		
	Female	134,298	0.469	0.499	1.075	-0.066
AM Scale 2	Male	234,897	0.735	0.441		
	Female	134,305	0.800	0.400	1.088	-0.153
AM Scale 3	Male	234,897	0.564	0.496		
	Female	134,305	0.536	0.499	0.950	0.057
AM Scale 4	Male	234,897	0.609	0.488		
	Female	134,305	0.584	0.493	0.960	0.050
AM Scale 5	Male	235,195	0.490	0.500		
	Female	134,447	0.628	0.483	1.282	-0.281
AM Scale 6	Male	235,243	0.642	0.479		
	Female	134,493	0.653	0.476	1.017	-0.023
New Business Development						
NBD Scale 1	Male	235,243	0.837	0.369		
	Female	134,493	0.812	0.391	0.969	0.067
NBD Scale 2	Male	236,543	0.589	0.492		
	Female	134,767	0.477	0.499	0.810	0.226
NBD Scale 3	Male	236,543	0.740	0.439		
	Female	134,767	0.688	0.463	0.931	0.114
NBD Scale 4	Male	236,543	0.672	0.469		
	Female	134,767	0.606	0.489	0.902	0.138

Table 12

Race Adverse Impact Ratios for Heterogeneous Scales

Scale	Race	*n*	*M*	*SD*	AI Ratio	Cohen's *d*
Account Management						
AM Scale 1	Caucasian	277,243	0.450	0.497		
	African American	28,861	0.450	0.498	1.002	-0.002
	Hispanic	19,330	0.454	0.498	1.009	-0.008
AM Scale 2	Caucasian	277,252	0.763	0.425		
	African American	28,862	0.733	0.442	0.962	0.068
	Hispanic	19,331	0.749	0.434	1.021	0.33
AM Scale 3	Caucasian	277,252	0.580	0.494		
	African American	28,862	0.473	0.499	0.814	0.217
	Hispanic	19,331	0.465	0.499	0.984	0.232
AM Scale 4	Caucasian	277,252	0.609	0.488		
	African American	28,862	0.552	0.497	0.907	0.115
	Hispanic	19,331	0.514	0.500	0.844	0.193
AM Scale 5	Caucasian	277,541	0.523	0.499		
	African American	28,945	0.659	0.474	1.260	-0.279
	Hispanic	19,345	0.571	0.495	1.092	-0.097
AM Scale 6	Caucasian	277,611	0.665	0.472		
	African American	28,955	0.605	0.489	0.910	0.125
	Hispanic	19,346	0.577	0.494	0.868	0.181

New Business Development						
NBD Scale 1	Caucasian	277,611	0.841	0.366		
	African American	28,955	0.806	0.396	0.958	0.093
	Hispanic	19,346	0.839	0.367	1.042	0.005
NBD Scale 2	Caucasian	278,929	0.559	0.496		
	African American	28,973	0.524	0.499	0.937	0.071
	Hispanic	19,386	0.539	0.499	1.027	0.042
NBD Scale 3	Caucasian	278,929	0.739	0.439		
	African American	28,973	0.640	0.480	0.866	0.216
	Hispanic	19,386	0.724	0.447	0.979	0.034
NBD Scale 4	Caucasian	278,929	0.672	0.470		
	African American	28,973	0.567	0.496	0.844	0.217
	Hispanic	19,386	0.602	0.490	0.896	0.146

Table 13

Gender Adverse Impact Ratios for Homogeneous Scales

Scale	Gender	*n*	*M*	*SD*	AI Ratio	Cohen's *d*
Homo 1	Male	236,926	0.661	0.473		
	Female	134,776	0.754	0.430	1.142	-0.208
Homo 2	Male	237,119	0.627	0.483		
	Female	134,848	0.654	0.476	1.043	-0.056
Homo 3	Male	237,119	0.560	0.496		
	Female	134,848	0.404	0.491	0.721	0.317
Homo 4	Male	237,119	0.523	0.499		
	Female	134,848	0.513	0.500	0.980	0.021
Homo 5	Male	237,121	0.424	0.494		
	Female	134,850	0.352	0.478	0.829	0.149
Homo 6	Male	237,121	0.640	0.480		
	Female	134,850	0.662	0.473	1.035	-0.046
Homo 7	Male	237,121	0.665	0.472		
	Female	134,850	0.602	0.489	0.906	0.130
Homo 8	Male	237,121	0.692	0.462		
	Female	134,850	0.699	0.459	1.010	-0.015
Homo 9	Male	237,121	0.634	0.482		
	Female	134,850	0.656	0.475	1.035	-0.046
Homo 10	Male	237,121	0.610	0.488		
	Female	134,850	0.659	0.474	1.079	-0.101
Homo 11	Male	237,121	0.506	0.500		
	Female	134,850	0.541	0.498	1.069	-0.070
Homo 12	Male	237,121	0.495	0.500		
	Female	134,850	0.490	0.500	0.990	0.010
Homo 13	Male	237,121	0.763	0.425		

	Female	134,850	0.750	0.433	0.983	0.031
Homo 14	Male	237,121	0.674	0.469		
	Female	134,850	0.665	0.472	0.987	0.019
Homo 15	Male	237,121	0.481	0.500		
	Female	134,850	0.448	0.497	0.932	0.065
Homo 16	Male	237,121	0.555	0.497		
	Female	134,850	0.513	0.500	0.925	0.083
Homo 17	Male	237,121	0.676	0.468		
	Female	134,850	0.681	0.466	1.008	-0.012
Homo 18	Male	237,121	0.769	0.422		
	Female	134,850	0.759	0.428	0.988	0.022
Homo 19	Male	237,121	0.487	0.500		
	Female	134,850	0.493	0.500	1.013	-0.012
Homo 20	Male	237,121	0.336	0.472		
	Female	134,850	0.397	0.489	1.181	-0.127
Homo 21	Male	237,121	0.407	0.491		
	Female	134,850	0.395	0.489	0.971	0.024
Homo 22	Male	237,121	0.447	0.497		
	Female	134,850	0.440	0.496	0.984	0.014

Table 14

Race Adverse Impact Ratios for Homogeneous Scales

Scale	Race	*n*	*M*	*SD*	AI Ratio	Cohen's *d*
Homo 1	Caucasian	279,237	0.714	0.452		
	African American	29,018	0.694	0.461	0.971	0.045
	Hispanic	19,396	0.689	0.463	0.964	0.056
Homo 2	Caucasian	279,467	0.645	0.478		
	African American	29,034	0.651	0.477	1.010	-0.013
	Hispanic	19,398	0.664	0.472	1.029	-0.039
Homo 3	Caucasian	279,467	0.518	0.500		
	African American	29,034	0.436	0.496	0.842	0.165
	Hispanic	19,398	0.462	0.499	0.892	0.113
Homo 4	Caucasian	279,467	0.514	0.500		
	African American	29,034	0.541	0.498	1.051	-0.053
	Hispanic	19,398	0.525	0.499	1.021	-0.022
Homo 5	Caucasian	279,471	0.405	0.491		
	African American	29,034	0.286	0.452	0.707	0.251
	Hispanic	19,398	0.343	0.475	0.846	0.129
Homo 6	Caucasian	279,471	0.621	0.485		
	African American	29,034	0.735	0.441	1.182	-0.245
	Hispanic	19,398	0.721	0.449	1.159	-0.212
Homo 7	Caucasian	279,471	0.666	0.472		
	African American	29,034	0.604	0.489	0.906	0.130
	Hispanic	19,398	0.621	0.485	0.933	0.094
Homo 8	Caucasian	279,471	0.711	0.453		
	African American	29,034	0.692	0.461	0.974	0.040
	Hispanic	19,398	0.711	0.453	1.000	0.000
Homo 9	Caucasian	279,471	0.660	0.474		
	African American	29,034	0.572	0.495	0.867	0.181
	Hispanic	19,398	0.614	0.487	0.931	0.095
Homo 10	Caucasian	279,471	0.630	0.483		
	African American	29,034	0.563	0.496	0.893	0.138
	Hispanic	19,398	0.688	0.463	1.091	-0.122
Homo 11	Caucasian	279,471	0.538	0.499		
	African American	29,034	0.440	0.496	0.817	0.198

	Hispanic	19,398	0.508	0.500	0.943	0.061
Homo 12	Caucasian	279,471	0.486	0.500		
	African American	29,034	0.565	0.496	1.162	-0.158
	Hispanic	19,398	0.479	0.500	0.985	0.015
Homo 13	Caucasian	279,471	0.780	0.415		
	African American	29,034	0.676	0.468	0.868	0.234
	Hispanic	19,398	0.711	0.453	0.912	0.158
Homo 14	Caucasian	279,471	0.673	0.469		
	African American	29,034	0.701	0.458	1.042	-0.061
	Hispanic	19,398	0.717	0.450	1.066	-0.096
Homo 15	Caucasian	279,471	0.516	0.500		
	African American	29,034	0.270	0.444	0.524	0.520
	Hispanic	19,398	0.290	0.454	0.561	0.475
Homo 16	Caucasian	279,471	0.516	0.500		
	African American	29,034	0.649	0.477	1.258	-0.273
	Hispanic	19,398	0.567	0.496	1.099	-0.102
Homo 17	Caucasian	279,471	0.666	0.472		
	African American	29,034	0.741	0.438	1.113	-0.165
	Hispanic	19,398	0.727	0.445	1.093	-0.134
Homo 18	Caucasian	279,471	0.765	0.424		
	African American	29,034	0.842	0.365	1.101	-0.195
	Hispanic	19,398	0.830	0.375	1.086	-0.164
Homo 19	Caucasian	279,471	0.495	0.500		
	African American	29,034	0.605	0.489	1.223	-0.223
	Hispanic	19,398	0.512	0.500	1.034	-0.033
Homo 20	Caucasian	279,471	0.349	0.477		
	African American	29,034	0.332	0.471	0.952	0.035
	Hispanic	19,398	0.305	0.460	0.875	0.093
Homo 21	Caucasian	279,471	0.431	0.495		
	African American	29,034	0.259	0.438	0.602	0.368
	Hispanic	19,398	0.264	0.441	0.612	0.358
Homo 22	Caucasian	279,471	0.449	0.497		
	African American	29,034	0.479	0.500	1.069	-0.062
	Hispanic	19,398	0.393	0.489	0.877	0.112

APPENDIX A

JOB DESCRIPTION AND JOB ANALYSIS DATA

FOR ACCOUNT MANAGEMENT ROLES

Please review the job description and job analysis data and categorize the following jobs as Hunter, Farmer, Neither, Both Hunter and Farmer, or Insufficient Information. Customer Service and Retail sales jobs are not considered true Hunter/Farmer roles.

Key words/themes for Farmer (typical sales roles: account manager/rep, customer service rep, inside sales rep)

- Existing accounts
- Focus on retention
- Builds/cultivates relationships
- Service beyond the initial sale
- Collaborate
- Team players
- Account management
- Follow-through
- Maintains customer base
- Continued contact
- Alleviating concerns
- Existing customers
- Upselling/renewals
- Helping customers
- Advising customers, trusted advisor
- Patient/patience
- Account penetration
- Previously assigned territory/accounts
- Upselling, renewals, upgrades
- Base salary
- Following-up with customers
- Customer loyalty
- Develop Relationships
- Manages details
- Customer satisfaction
- Relationship selling
- Ongoing client engagement
- Help others

PROFILE ROLE DESCRIPTION

Account Management (Farmer)

TYPE OF CONTACT:	Face-to-face sales calls (outside)
CONTACT WITH END-USER:	Direct to purchaser
SCOPE OF PRODUCTS:	Full line
TYPE OF CUSTOMER:	Big enough prospect to represent a major client for seller
TYPICAL DECISION MAKER:	Financial and/or technical buyer; cross-functional team for larger purchases

WHAT IT IS	WHAT IT IS NOT
• Almost exclusively sells to an existing book of customers with the expectation of growing share of wallet – may engage in cross-selling • Role that develops deep and broad relationships within the customer • Strong customer advocate within the seller's organization and does proactive planning • Sales cycle is ongoing with annual purchase agreements possible for the larger accounts • Strong focus on quality of personal service to the client	• Fast expansion of sales through active prospecting outside of the current customer base • Territorial sales to small and medium-size customers • Reactive and content to keep existing customers at current usage levels

This position is responsible for optimizing sales in an existing set of major accounts. This is usually accomplished by increasing the seller's share of current purchases as well as expanding sales both by line extension and penetrating additional buyers within the accounts. This type of sales is proactive in strengthening customer relations and focuses on continually growing revenue by meeting customer needs and making it easy to do business with the seller.

Some typical account management roles would include selling financial planning to high net worth clients, raw materials to mid and large manufacturers, or second-tier suppliers to major manufacturers.

Top account managers tend to be disciplined in their account planning process and set a contact schedule that reinforces their commitment to increasing customer satisfaction. They develop a network of internal support to facilitate special requests and quick problem resolution. When forced to increase the account base, they will be slow to add customers and use a referral process almost exclusively to identify possible prospects.

SELECTION MEASURES

ACCOUNT MANAGEMENT SCALE 1

Consistently meets or exceeds sales targets through personally controlling the critical aspects of the sales and delivery processes; systematically works each account plan and anticipates problems in order to work around them; is driven to win customers' attention and treats their business as an honor, never letting them feel taken for granted

HIGH SCORES

- Meets or surpasses his sales goals as a means to gain more freedom and independence
- Maintains personal control over those aspects of the sales and delivery process identified as top priorities
- Sets high expectations for himself and holds associates to the same stringent standards of dedication
- Expects to encounter barriers a fair percentage of the time and prepares to deal with setbacks or delays in order to ensure the success of his plans
- Concentrates full energy and attention on systematically accomplishing key tasks
- Places his customers on a pedestal and shows them how valued they are

LOW SCORES

- May adopt a casual or relaxed approach that fails to project personal commitment or dedication
- Is comfortable with achieving average results, rising above the bottom of the performance chart but not striving to be the best
- Adopts a more interdependent and team-oriented approach to accomplishing goals
- May depend too much on fate or 'being in the right place at the right time' rather than plan and work to make things happen
- Expects his account management plan to proceed without the need for careful monitoring and contingencies
- Can take customer relationships for granted and may not work at continually demonstrating appreciation for the business

This skill is common among top-producing Relationship oriented sales professionals. They are driven to meet or surpass their goals as a means to gaining more freedom and independence. They prefer to personally control the critical elements of their job function. They expect things to go wrong a fair percentage of the time, so they are prepared to deal with setbacks or delays in order to ensure the success of their plans. They set high expectations for themselves and hold their associates the same stringent standards of dedication.

As part of controlling the factors that influence their success, they tend to develop specific sales plans for each account so they can maintain and where possible increase the sales volume. They don't simply hope for an increase in overall sales by trying to get across the board gains from each customer with only superficial analysis of each one's potential and product requirements.

LOWER SCORES MAY INDICATE:

- A drive for excellence through a more flamboyant or charismatic approach
- A comfort with achieving average results; it is not imperative to be the best, it is important not to be at the bottom of the performance charts
- An expectation that things will go as planned without the need for careful monitoring and contingencies
- A more interdependent and team-oriented approach to accomplishing goals

ACCOUNT MANAGEMENT SCALE 2

Motivated to advance in a sales career by increasing the product diversity offered to existing accounts; concentrates on finding avenues to generate increased sales from established accounts; works to increase revenue by identifying additional products to complement what is currently sold to the existing customer base

HIGH SCORES

- Driven to generate growth through account penetration and referral and directs effort toward building sales
- Unwilling to settle for maintaining the status quo in generating revenue or increasing sales just incrementally
- Prefers the efficiency of increasing sales with existing customers to spending the time cultivating new accounts
- Takes advantage of opportunities to penetrate existing accounts via new product applications
- Builds business with existing accounts by utilizing creative means to answer their needs with his products

LOW SCORES

- May be satisfied with moderate growth in sales
- Diverts resources from his sales effort to respond to administrative tasks
- May offer what is familiar to existing customers rather than build a repertoire of products or services that can match additional needs

Those who score high in this skill are intense people who channel their drive into expanding their book of business. They are not easily satisfied with maintaining current levels of business or even modest growth. They understand that their products don't just sell themselves, and they look for creative ways to match their products and services to diverse needs. Likewise, they are willing to explore new product applications to be able to increase the repertoire of products they can offer to existing customers.

They understand that their intensity and strong values can come across as being opinionated or judgmental; so they are careful not to provoke others and prefer to keep the focus on the customer. Their goal is to grow sales either by finding additional opportunities to serve existing customers or expanding the base through referral to other contacts. They want to reach their growth goals and maximize their efficiency by selling more to fewer people rather than spend a significant amount of time trying to acquire new customers.

LOWER SCORES MAY INDICATE:

- A tendency to be comfortable with little or modest growth in sales
- A preference for a standard sales approach with little innovation and variation
- A strong value system that supports the belief that everyone needs the product and should respond to one's prepared sales proposition

ACCOUNT MANAGEMENT SCALE 3

Shows skill and ease at working in groups; works with others to achieve better results and forges close working relationships and alliances in order to get things done; is quick to cooperate versus expect others to bend to his wishes; supports joint ownership and shared recognition for results; recognizes the advantages of group participation in planning and problem solving

HIGH SCORES

- Works collaboratively in a group effort to effectively accomplish a goal or task
- Puts team and management objectives ahead of his own
- Performs own role effectively on the team and helps team members as needed
- Forges working relationships and alliances with others in order to get things done
- Willing to compromise and give others the benefit of the doubt when he disagrees
- Proactively shares information, ideas, suggestions and support
- Encourages joint ownership and shared recognition for results

LOW SCORES

- Functions as more of an individualist than as a team player
- Finds it simpler or more convenient to focus independently on his own tasks or direction
- Prefers to exercise personal control over results and outcomes
- Is more likely to emerge as a leader than an equal player
- Believes compromise for the sake of cooperation can potentially threaten the quality of the end result
- Wants to be personally recognized for individual contributions

High scores in this skill indicate a preference for working with others to achieve better results than are possible by working on one's own. These individuals are comfortable putting the group goal ahead of their own and are proactive in offering support and help where they can. They are congenial and willing to give others the benefit of the doubt when they disagree or don't understand their point of view. They are quick to cooperate versus expect others to bend to their wishes.

They are comfortable working in an environment of shared control. They do not feel a need to impose their judgment or will on the group. They are not expecting everything to be enjoyable or fun in their work life and don’t take disagreements or differences personally.

LOWER SCORES MAY INDICATE:

- A strong need for individual contribution or personal recognition
- A preference for working alone and ensuring the results through personal control
- A desire to be more of a leader than one of the pack
- A belief that compromise may lengthen the time to reach the end result and/or threaten the quality of the desired outcome

ACCOUNT MANAGEMENT SCALE 4

Helps the customer to gain maximum benefit from the product or service by committing to continuous education that provides information updates or product training; prepares more structured sessions to cover the most critical areas of learning; stays on top of information needed by customers in an effort to serve as a resource; takes responsibility for motivating customers to update their information; reinforces shared information through periodic repetition; routinely adjusts education process to address individual and group progress or understanding

HIGH SCORES

- Prepares scheduled and consistent programs to train or educate others
- Establishes measurable criteria for assessing progress in the learning process
- Demonstrates patience and a willingness to repeat or reinforce ideas and information until the audience understands
- Focuses training sessions on those competencies that will make a difference in the group's ultimate effectiveness
- Concentrates more on the results produced or change accomplished through his training than with how attractive or entertaining the training can be

LOW SCORES

- Prefers one-on-one training or a more loosely organized curriculum to the structured requirements of a scheduled class session
- Expects the people he is training to be self-motivated to learn and becomes impatient when required to repeat or reinforce information he has already covered
- Does not implement a tracking process for assessing the effectiveness of his teaching efforts or the progress of his trainees
- Enjoys working on content delivery and may be more concerned with the audience's assessment of his public speaking skills than with the subject matter
- Tries to make the training entertaining at the expense of providing only relevant information

There is an old expression that says, "there are some who do and some who teach." While it is often used in a derogatory sense to imply that people who teach can't do, the greater likelihood is that people who are competent at doing, have the ability to demonstrate to others how to do, and expand their overall potential by working through others, do have a strong offering to make. Those who demonstrate proficiency in this skill understand the need for continued reinforcement of those issues previously covered, in addition to presenting new concepts and ideas. They are comfortable seeing themselves as responsible for the group's continued learning and their motivation for the process. This skill is focused on structures sessions to provide information to a group to help them do their job more effectively.

This skill is applied by managers to clearly communicate changes and new information while integrating them into the current set of policies and procedures for their reports. These sessions are also used to reinforce and support any corporate training initiatives. This is usually done in regularly scheduled meetings that can be face-to-face or by teleconference.

Salespeople and individual contributors apply this skill to help internal or external customers take best advantage of the deliverables that the individual is responsible for providing. Those who score high will work to provide regular updates and reinforcement for past communication so that their constituents get a clear and consistent message.

LOWER SCORES MAY INDICATE:

- A tendency to rush through the training without gauging how much is being absorbed or retained by the group
- A preference for one-on-one coaching and/or more free form curriculum than formal class sessions
- A preference for a more off-the-cuff training style

ACCOUNT MANAGEMENT SCALE 5

Seeks customer feedback regularly to verify satisfaction and uncover minor issues which could escalate if left unattended; solicits suggestions for continuous improvement and demonstrates personal attention to the customer even when there is no problem; listens to feedback without judgment or defensiveness, remaining focused on results rather than personal feelings; stays logical and objective and refrains from expressing personal frustrations to the customer; ensures future business by anticipating and removing potential sources of dissatisfaction

HIGH SCORES

- Demonstrates continued personal attention to the customer, calling regularly and soliciting suggestions for improvement
- Is disciplined and consistent in initiating contact with the intent of monitoring satisfaction and detecting potential problems
- Welcomes opportunities to demonstrate his willingness to make the customer's experience even better
- Aims emotional intensity at addressing a customer's needs or concerns, not at the customer personally
- Does not allow the mood or feeling about a customer to alter the degree or quality of the follow-up

LOW SCORES

- Assumes that a customer will make him aware of problems but is otherwise satisfied
- Tends to 'let sleeping dogs lie' and is reticent to probe for problems that may be in the early stage of development
- Finds it difficult to keep emotions in check and remain logical when presented with suggestions for improvement
- Takes customer feedback or criticism as a personal jab and may inappropriately display a negative response

People who score high in this skill tend to be able to separate their feelings from their intention to maximize their ability to satisfy customer needs. They understand that the strength of their relationship with a customer depends on frequency of contact and demonstrating a keen interest in their satisfaction. They are comfortable asking for feedback and suggestions for improvement on a continuous basis to prevent small frustrations from building to become more serious issues.

They can listen to the feedback and remain focused on the result rather than personally identify with others' concerns and need to justify past actions. They stay logical and objective. They welcome opportunities to demonstrate their willingness to make the customer's experience even better. If they do experience personal frustration, they will choose an appropriate time and place to express their feelings without the customer's knowledge.

LOWER SCORES MAY INDICATE:

- A tendency to assume that the customer is satisfied if there is no overt evidence to the contrary
- A reticence to probe the customer or a preference for letting the customer dictate the direction and flow of the interaction
- A tendency to try and anticipate needs and act without checking to be sure the customer's needs are really understood
- A difficulty keeping emotions in check and remaining logical when confronted with suggestions for improvement

ACCOUNT MANAGEMENT SCALE 6

Organizes time to cover ongoing priorities and will make arrangements to cover whenever absence is unavoidable; understands that rewards are commensurate with effort and invests the time to get the job done; uses time efficiently during regularly scheduled hours so that extra hours are not inevitable, but is responsive to additional requirements or demands when necessary

HIGH SCORES

- Makes himself available after hours to respond to the needs of both internal and external customers
- Organized and efficient in how he spends his time during the day so that the need to work unscheduled hours is the exception rather than the rule
- Resists the distraction of nonwork-related issues that can interfere with his coverage of ongoing priorities
- Accepts that his strong sense of duty to those who are counting on him cannot always be met during regular working hours
- Prepares a back-up system to cover unpredictable or unpreventable overloads in his absence

LOW SCORES

- Believes his personal time is off-limits and may resist interrupting his leisure activities to deal with work demands
- Loses control of his action list and can become overwhelmed by ongoing priorities
- May not feel a need to put forth more effort, but will accept delays instead
- Allows outside distractions, commitments or time demands to consistently take priority and prevent an efficient use of his time and resources
- Fails to put into place systems or resources to cover customer needs when he is unprepared or unavailable, creating delays in response time

This is a measure of the amount of time a person is willing to devote to dealing with the demands of internal and external customers. People who score high tend to have a strong sense of duty to their constituents. Likewise, they know that unexpected issues will arise, and they try to remain available to deal with them appropriately. They are likely to have backups in place for the times they are not available so the customer can get contact satisfaction in their absence. Their goal is to prevent the escalation of a problem as a primary priority and they are not distracted by non-work issues.

LOWER SCORES MAY INDICATE:

- A higher need for short-term, personal gratification from one's activities or the tendency to use one's mood to determine the amount of time to devote to a specific outcome
- A tendency to place a higher value on non-work endeavors
- A preference for personally handling customer interactions even if that forces delays in response time
- A tendency to assume that the customer is satisfied if there is no overt evidence to the contrary
- A reticence to probe the customer or a preference for letting the customer dictate the direction and flow of the interaction
- A tendency to try and anticipate needs and act without checking to be sure the customer's needs are really understood
- A difficulty keeping emotions in check and remaining logical when confronted with suggestions for improvement

APPENDIX B

JOB DESCRIPTION AND JOB ANALYSIS DATA

FOR NEW BUSINESS DEVELOPMENT ROLES

Please review the job description and job analysis data and categorize the following jobs as Hunter, Farmer, Neither, Both Hunter and Farmer, or Insufficient Information. Customer Service and Retail sales jobs are not considered true Hunter/Farmer roles.

Key words/themes for Hunter (typical sales roles include: Account executive, Business Development Rep, Field Sales Rep)

- Lead generating
- Seek out opportunities
- Initiating contact
- Sells to new customers
- Closes/closing sales
- Persistent
- Tenacious
- Qualifies prospects
- Commission-based salary
- Builds customer base
- Networks
- Independent/individualist
- Prospecting
- Aggressive
- Competitive
- Cold-calling
- Respond to bids or quote requests
- Deal well with rejection
- Focus on growth
- New business
- Marketing
- Persuasive/ Communicates persuasively
- Sales calls/Numerous sales calls per week
- Establish new business with potential customers
- Handoff account to others for maintaining accounts

PROFILE ROLE DESCRIPTION

New Business Development (Hunter)

TYPE OF CONTACT:	Face-to-face sales calls (outside)
CONTACT WITH END-USER:	Direct to purchaser
SCOPE OF PRODUCTS:	Full line
TYPE OF CUSTOMER:	Big enough prospect to represent a major client for seller
TYPICAL DECISION MAKER:	Financial and/or technical buyer; cross-functional team for larger purchases

WHAT IT IS	WHAT IT IS NOT
• Almost exclusively sells to new customers or dormant existing relationships • Not responsible for ongoing client contact after the sale • Rain-maker for new product introduction • Sales cycle is typically 3 months or less, but can extend to 6 months for complex deals	• Builds a book of customers and then services them to maintain and increase revenue from them • Territory sales role to deal with mid and small customers • A “closer” whose only function is to seal the deal

This position is usually responsible for bringing in new business from larger accounts. Those who excel tend be comfortable uncovering opportunities and continuously promoting their benefits in the hopes of uncovering a match. The position focuses almost exclusively on the prospect conversion process, and the salesperson typically does minimal post-sale contact. This type of position is used to develop a presence in a competitive stronghold or to expand sales coverage beyond the firm’s current customer base. Most often, hunters sell products or services that are relatively new or, at least, represent a new approach to the potential customer.

The sales cycle is typically six months or less but may extend longer if service is a small percentage of the sales. An example of this approach is Office Equipment salespeople who demonstrate and sell while others service until renewal is required.

Hunters tend to be very opportunistic and look for novel ways to apply their products and services. This process needs to be carefully managed if the organization is not equipped to cost-effectively customize the offering for diverse customer applications.

Those who excel at new business development are not likely to easily move to a customer maintenance mode once they have sold a specified number of customers. They would be likely to look for a new position if required to continuously service the customers that they have closed. They will thrive in a situation where they can hand off their recently closed customer to a sales professional who is responsible for account maintenance so they are free to continue their quest for new customers.

SELECTION MEASURES

NEW BUSINESS DEVELOPMENT SCALE 1

Objectively analyzes a problem situation and takes steps to provide a solution; identifies the root of the problem before pressing for a resolution; remains engaged until a solution is reached; tries to see all sides of the problem and thus understand others' assessment of the issue or response; takes personal responsibility for identifying a resolution

HIGH SCORES

- Objectively isolates and defines problem areas clearly
- Determines the true nature of the problem rather than deal with its symptoms
- Is willing to 'think outside the box' to find a solution
- Displays sensitivity and genuine interest in understanding others' perspectives and will not ignore their concerns
- Regards any problem as a challenge to be met with eagerness and enthusiasm
- Remains engaged until a problem has been resolved
- Takes personal accountability for the result

LOW SCORES

- Can be biased and make judgmental or inappropriate assumptions without analyzing the situation objectively
- May press toward resolution without identifying the root of the problem
- Becomes wrapped up in his own views and loses sight of how others may see the problem or response
- Becomes frustrated with solving the same or similar problems over and over again
- Tends to take complaints and problems personally and feel oppressed by them
- Sees problem resolution as an inconvenience and a distraction
- May oversimplify a problem and its solution and disengage his efforts before identifying a satisfactory solution
- Resists taking ownership of the solution

This scale was originally developed to assess the problem-solving competency of people in customer service or inside sales roles. This skill focuses on the process and measures the ability to objectively analyze a customer's situation and work to provide a solution. It is not a measure of the creativity or exquisite precision of the solution. People who score high are genuinely interested in understanding the customer's perspective when problems occur. They remain focused and positive while looking for a solution. They are generally optimistic that they can work through the issues and provide an acceptable outcome. They see the situation as an opportunity to increase the strength of the relationship rather than as an inconvenience to them.

They are willing to go beyond standard processes and procedures to uncover the cause of the problem and its possible solutions. Once the solution is identified, they remain engaged until they are sure the problem has been resolved and the customer is satisfied. This may involve interfacing with internal people on the customer's behalf or finding external resources to provide the solution. All the while, personal accountability for the result is retained.

This skill can be demonstrated throughout the sales cycle. With prospective customers, those who score high are interested in providing solutions to those issues that could be seen as barriers to proceeding toward a sale. They will work for resolution so that they can ask for the order or close the deal. With existing customers, the goal shifts to retention and increasing customer satisfaction by removing obstacles to fully experiencing the expected benefits.

LOWER SCORES MAY INDICATE:

- A frustration with seemingly solving the same problem over and over again
- A press toward resolution before the root of the problem has been identified
- A tendency to take complaints and problems personally and feel oppressed by them

  A tendency to rigidly follow procedures or pass the problem on to another to close the loop with the customer

NEW BUSINESS DEVELOPMENT SCALE 2

Uses a formula or series of questions to determine the prospect's fit with the product; expects to sell to the majority of prospects since they are known to need the seller's products; reacts quickly and objectively to the answers to standard probes by disqualifying the prospect or proceeding through the selling process

HIGH SCORES

- Probes with tough questions around profitability, capacity and readiness to buy when evaluating a prospect for potential business opportunities
- Uses a formula approach to qualifying prospects that addresses typical interest points, such as price and ability to deliver
- Does not allow his time to be consumed by prospects who should be disengaged or deferred based on the quality of business they will provide
- Remains emotionally detached from the prospect's situation, allowing him to stay objective in assessing the potential for profitable business

LOW SCORES

- Uses an inconsistent approach to uncovering needs
- May have difficulty probing to obtain sufficient information to accurately assess the likelihood of a profitable match
- Exhibits too much optimism and not enough pragmatism to cut loose potential business that would not be profitably serviced
- Finds it difficult to be objective in assessing the potential of demanding or challenging contacts

This measure of qualifying prospects centers on the use of some fairly standard questions to assess the potential buyer's readiness to make a decision and willingness to pay at a rate that includes a reasonable profit for the seller. Once the determination is made those who score high allocate their efforts and resources to spend the most time with the prospects that are most ready and capable of reaching a profitable deal. Those who could be good future customers are contacted periodically to establish timing for the next steps. Low likelihood prospects are tactfully dropped from the sales funnel so they are not a distraction.
This process is dispassionate and relies on fit with the ideal prospect profile and purchase readiness rather than personal chemistry. Those who score high are consistent in determining the prospect's fit with the seller's existing capabilities and spend the bulk of their time with the best ones. They are not tempted to focus too much on volume and opportunity versus fit with the production capacity and profitability of the potential deal.

LOWER SCORES MAY INDICATE:

- Using interest in prospects' problems or compatibility with contacts as the criteria for spending time with them in the selling process
- Trying to be all things to all people or sell something to each lead
- Having difficulty recognizing the potential of contacts that are demanding or challenging to deal with

NEW BUSINESS DEVELOPMENT SCALE 3

Thrives on working; tends to achieve higher results in direct proportion to the time he is willing to commit to his work; remains focused on the goal and is not easily discouraged or distracted; uses work as an opportunity for interaction and incorporates interpersonal contacts into task accomplishment; sees work as a major source of personal satisfaction

HIGH SCORES

- Seeks to balance work and life priorities while committing time and resources to results accomplishment that surpasses what is just acceptable or expected
- Is dedicated to achieving business results, even at some personal sacrifice
- Defines himself to a great extent by his work and derives personal satisfaction from his success and accomplishments in his career
- Builds into his schedule the time required to lay the groundwork and develop the plans for achieving his work objectives
- Retains control of his objectives by preparing action steps and evaluating his progress
- Remains focused when the commitment to his work is threatened by internal or external distractions
- Enjoys finding opportunities to mix personal interaction with goal accomplishment

LOW SCORES

- May embrace a somewhat rigid separation of professional and personal commitments that limits flexibility in balancing the time and resources devoted to work and leisure
- Is more inclined to work a set schedule of hours than to commit resources as required by the objective
- Distinguishes personal time from work and career commitments and is not likely to make consistent sacrifices in his personal life to accommodate extraordinary advancements in his work objectives
- Becomes distracted by unexpected obstacles or may not effectively plan to deal with his job demands
- May not recognize that opportunities to advance his goals can be present in the personal interactions that occur in the workplace

This scale was originally developed using a sample of consultive salespeople who tended to achieve higher results in direct proportion to the time they were willing to commit to their work. People who demonstrate this skill thrive on working and place a high value not only on accomplishing the tasks specified in their job description, but devoting the additional time necessary for planning, preparation, and skill development. They are hard workers and like the social aspects of the job. However, they tend to use the social interaction as a means to accomplishing their goals rather than as an end in itself.
High-scoring individuals typically use the extra time to develop more in-depth plans to achieve their objectives. They also build personal sales tools and tracking processes that enable them to increase their sales volume and margins. They tend to believe that the basic job is accomplished between 9 and 5, and the effort required to be a top performer is spent from 5 to 9. They tend to be perfectionists and want to do the job right.

LOWER SCORES MAY INDICATE:

- A desire for balance in one’s life with sufficient personal and family time
- A preference for a 40-hour work week with stable and predictable time requirements
- Career advancement is not a strong priority

NEW BUSINESS DEVELOPMENT SCALE 4

Reinforces the purchase decision with a series of logical reasons that support the sale; moves closer to a purchase decision in a multiphase sale by setting an objective for each encounter and gaining agreement at each of the milestones in the process; provides ammunition for the buyer to justify the cost and defend the purchase within the customer organization; encourages a buying decision by demonstrating future savings and ease of transition to the proposed system; keeps the sales process from stalling by taking control and focusing the customer on the incremental steps leading to a buying decision; remains patient but focused on reaching the end result

HIGH SCORES

- Bases a logical appeal on the demonstration of future savings and ease of transition to the proposed system or benefit
- Provides the buyer with purchase justifications that may be needed to reach a buying decision or sway additional influencers
- Breaks the sales process into manageable steps, seeking continued agreement from the customer at each step before moving on
- Continually checks the customer's understanding of how their needs would be satisfied
- Reaches agreement on the concept or pilot test, with a contract coming later
- Recognizes that the close is soft, but remains focused on accomplishing that result
- Asks for the opportunity to begin the project or deliver the benefit

LOW SCORES

- Makes an emotional appeal for the buying decision, using the strength of the personal relationship that has developed as the basis for a commitment
- Emphasizes the potential loss due to inaction versus promoting the potential benefits gained by acting on the purchase decision
- Unknowingly loses customers early in the sales process because he does not continually assess their commitment throughout an extended sales cycle
- May lose patience in a long sales cycle and rush the customer toward a purchase decision when it is more appropriate to test the water and give the customer more control through trial closes

This approach to closing sales is based on a series of logical appeals that demonstrate the benefits to the customer or prospect. It is consistent with a consultive selling style that builds or tailors a system to specifically meet the individual's needs and usually requires at least a six-month sales cycle. Those who score high are consistently encouraging a buying decision by demonstrating future savings and ease of transition to the proposed system. They provide ammunition for the buyer to justify the cost and defend the purchase within the customer organization.
In the light of such a complex selling environment, those who score well tend to exercise their closing skills to progress toward the final close by gaining agreement at each of the milestones in the process. While not impatient, they are focused on the end result.

LOWER SCORES MAY INDICATE:

- A more emotional appeal for the buying decision based on the strength of the relationship that has developed
- A strong focus on potential loss due to inaction as an impetus to buy
- A reticence to probe to determine the key benefits for the prospect or customer and/or to ask for the sale