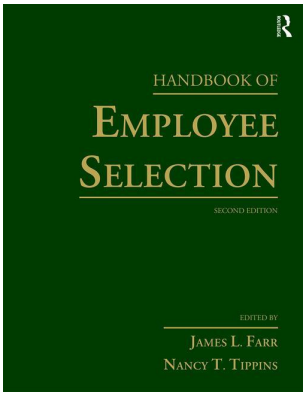
This article was downloaded by: *10.2.97.136*
On: 21 Sep 2023
Access details: *subscription number*
Publisher: *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Employee Selection

James L. Farr, Nancy T. Tippins, Walter C. Borman, David Chan, Michael D. Coovert, Rick Jacobs, P. Richard Jeanneret, Jerard F. Kehoe, Filip Lievens, S. Morton McPhail, Kevin R. Murphy, Robert E. Ployhart, Elaine D. Pulakos, Douglas H. Reynolds, Ann Marie Ryan, Neal Schmitt, Benjamin Schneider

Personality

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-13
Leaetta Hough, Stephan Dilchert
Published online on: 22 Mar 2017

How to cite :- Leaetta Hough, Stephan Dilchert. 22 Mar 2017, *Personality from:* Handbook of Employee Selection Routledge
Accessed on: 21 Sep 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-13

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://test.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

13

PERSONALITY

Its Measurement and Validity for Employee Selection

LEAETTA HOUGH AND STEPHAN DILCHERT

In our 2010 chapter, we noted that "personality variables have had a roller-coaster-like ride in employee selection" but predicted "a more stable and sanguine future as evidence continues to mount documenting the importance of personality variables as determinants of individual- and team-level performance" (Hough & Dilchert, 2010, p. 299). Indeed, this has occurred. Greater recognition of the role personality plays in individual, group, and organizational outcomes has resulted in more sophisticated thinking about personality variables and their role as determinants (predictors) of individual, team, and organizational outcomes. As a result, more personnel selection batteries include personality variables to enhance prediction of important work-related outcomes.

In this chapter, we update the issues and evidence, and describe the emerging consensus about the usefulness of personality variables in employee selection. We describe the mega-trends that have influenced the personality variables that are selected for inclusion in selection systems, how they are measured, and the outcomes they are expected to predict. We describe factors that hinder our understanding and those that help increase our knowledge of personality variables and their role in more accurately predicting work-related criteria. We address issues related to taxonomic structure, measurement methods, level of measurement, validity, and factors that threaten and enhance the validity of personality measures.

MEGA-TRENDS AND NEW TRENDS AFFECTING USE OF PERSONALITY

Several phenomena are affecting the use of personality variables in organizational settings and more specifically for personnel selection. These phenomena include:

- Rapidly changing work and social environments
- Changing demographics
- Availability of mega-data

Rapidly Changing Work and Social Environments

Intense competition demands that companies bring new and different products and services to market faster than ever. Innovation has long been of interest but is now a business necessity in

almost any industry. Creativity has become an even more important individual difference variable, and its personality determinants are of significant interest in a continually changing, highly competitive marketplace (National Research Council, 2015). Increased emphasis on understanding and assessing attributes of supervisors and managers who champion innovation and can enhance the performance of individuals and teams on creative tasks has sharpened the focus on the role of personality variables as determinants of performance.

Continuous learning is another important behavior leading to successful outcomes in rapidly changing and demanding situations. The importance of individual difference variables in learning is clear, and the role of personality is especially salient in continuous learning. With the increase in the use of technology to perform work along with rapid changes in technology, continuous learning is an important phenomenon that impacts individual and organizational success.

The speed with which change occurs in organizations and work settings has placed greater emphasis on performance variables of interest that reflect a person's performance in changing work settings. Adaptability, a criterion construct that has now been carefully explored and its components defined (see Chan, 2000; 2014; Pulakos, Arad, Donovan, & Plamondon, 2000), is an increasingly important outcome variable to organizations. Personality variables predict adaptive performance (Pulakos et al., 2002).

Work is now recognized as often accomplished in teams, both temporary and more permanent ones. This is true of knowledge work, service work, hospitality work, production work—most all work is done as a part of some sort of team or group effort. True, there are still accomplishments, innovations, and breakthroughs that might be described as single, individual efforts, but they are becoming increasingly rare (including in scientific and scholarly communities). This reality has increased the interest and focus on *group-level* variables. One new group-level variable is "Collective Intelligence" (Engel, Woolley, Jing, Chabris, & Malone, 2014; Woodley & Bell, 2011; Woolley, Aggarwal, & Malone, 2015). With its predictive validity primarily attributed to social perceptiveness, aka social awareness, there is greater interest in personality variables that affect—even determine—interpersonal behavior. Another example of a group-level personality variable is "Aggregate Personality" (Schneider & Bartram, 2015). Aggregate personality is a unit-level variable that is gaining attention through its value in predicting unit effectiveness and other important unit-level outcomes (e.g., Call, Nyberg, Ployhart, & Weekley, 2015; Ployhart, Weekley, & Baughman, 2006; Ployhart, Weekley, & Ramsey, 2009).

The social landscape is also rapidly changing and has affected measurement of personality variables. For example, research on social media profiles shows that such information can reflect an individual's personality rather than only an idealization of the self (Back et al., 2010) and will likely significantly influence how personality is assessed during the hiring process in the years to come. One can envision a selection system of the future that is without a self-report personality inventory, and instead measures personality through assessment of the individual's online behavior. The development of pertinent guidelines for cyber-vetting by organizations as well as governmental agencies (e.g., Rose et al., 2010) only underscores the relevance of this trend.

Changing Demographics

Organizations increasingly have an international workforce, and the U.S. population is increasingly diverse (U.S. Census Bureau, 2016). Life expectancy is increasing in most economically advanced countries. Many people are working longer, either due to voluntary delays of retirement (Pew Research Center, 2009) or as a result of increasing retirement ages in many countries. These demographic realities, along with Civil Rights laws in the U.S., have placed an emphasis on selection systems that are fair to all applicants and have less adverse impact on protected groups. Personality variables, especially many facet-level personality variables, typically show minimal to no mean score differences between protected and nonprotected groups. For example, older workers score, on average, higher than younger workers on Dependability, an important predictor of job performance and its components. African Americans score, on average, about the same as Whites on Dominance, an important predictor of leadership performance. Latinos

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

and African Americans score, on average, about the same as Whites on Emotional Stability and Agreeableness, both important predictors of job performance and its components (Hough, Oswald, & Ployhart, 2001). This reality provides companies with important individual difference variables to add to their predictor batteries to increase validity of their prediction equations and to compose company workforces that are more representative of their applicant pools. These benefits of personality variables have been a factor in their increased use in selection systems.

In addition to an increasingly diverse workforce, generation differences exist. Today, younger applicants and employees score significantly higher on personality variables such as self-focus/self-orientation than similarly aged cohorts of past generations (Twenge, Campbell, Hoffman, & Lance, 2010). Satisfaction with one's work, supervisor, pay, and employer is likely a more important factor in turnover than before. Increasingly, organizations are more concerned about an individual's "fit" with his or her work and with the organization. Interest variables, often considered personality variables (Holland, 1976), are increasingly a part of selection systems that are designed to benefit the individual as well as the company. The U.S. Army Research Institute, for example, is examining interest and personality variables as a way to enhance an individual's fit with work assignments that benefit both the individual and the organization, an orientation that is consistent with the principles of vocation counseling (Wolters, Heffner, & Sams, 2015; see also Wiernik, Dilchert, & Ones, 2016, for a discussion of implications).

Availability of Mega-Data

Mega- (or "big") data describes large, individual data sets as well as data sets composed of multi-organization or multi-source samples that are increasingly available to researchers and assessment system developers. Their availability has significant effects on questions and hypotheses researched, study designs employed, measurement methods used, nature and amount of data collected, analyses undertaken, and types of validation strategies employed. Macey, LoVerde, and Bartram (2016), for example, are developing leadership types using mega-data sets to cluster personality profiles into empirically homogeneous groups. This enables use of personality scales in combination (i.e., profiles) to examine relationships between personality variables and outcomes without constraining or specifying the nature of the relationships between and among the independent and dependent outcomes. The data-mining possibilities are truly significant. The phenomenon of mega-data is changing the way personality variables are and will be researched and used to select and assign people to work environments. (See Chapter 43, this volume, for additional discussion of the uses of mega-data in selection.)

All of these forces have contributed to greater use of personality variables in assessment systems for hiring and placing people in work assignments. At the same time, information technology or data-focused providers, often lacking knowledge of personality constructs and their structure, are entering the employee selection and HR market because of their ability to predict valued outcomes with indicators contained in large data sets. The generalizability (over time, as well as contexts/cultures) of mega-data findings can become a concern when measurement and prediction systems are developed without concern for constructs. If properly guided and applied, mega-data can lead to more nuanced and sophisticated research with personality variables, how they are measured, and how they are used.

STRUCTURE OF PERSONALITY VARIABLES

The taxonomic structure of personality variables is critically important to industrial-organizational (I-O) psychology, and it is nowhere more important than in employee selection research and practice. Personality constructs now play key roles in our models of individual and team performance. Researchers accumulate criterion-related validity studies to meta-analytically summarize the relationships between personality and criterion constructs. Practitioners contribute to the research base and benefit from the accumulation of knowledge generated by meta-analyses,

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

enabling us to build better prediction equations for criteria of interest. All of these activities and contributions depend on a good and generally agreed-upon taxonomic structure of personality variables.

Although criticism has waxed and waned, today the Five-Factor Model (FFM) is the most widely accepted structure of personality variables (Goldberg, 1993; Wiggins & Trapnell, 1997; for a history of the FFM, see Dilchert, Ones, Van Rooy, & Viswesvaran, 2006; Schneider & Hough, 1995). The earliest version of the FFM (emotional stability, surgency, culture, dependability, and agreeableness) dates back to Tupes and Christal's work in the 1950s and early 1960s (Tupes & Christal, 1961/1992). The specifics of the FFM have evolved somewhat over the years, and the factors are now often labeled emotional stability (or neuroticism), extraversion, openness, conscientiousness, and agreeableness (see Goldberg, 1993, for a concise summary of the FFM structure). Since Barrick and Mount (1991), most researchers followed their example of summarizing relationships between personality variables and work-related criteria according to the FFM.

Nonetheless, Hough and colleagues (Hough, 1992; Hough & Connelly, 2012; Hough & Oswald, 2000, 2005, 2008; Hough, Oswald, & Ock, 2015; Hough & Schneider, 1996; Oswald & Hough, 2008, 2011; Oswald, Hough, & Ock, 2013; Schneider & Hough, 1995; Schneider, Hough, & Dunnette, 1996) have consistently criticized the FFM, concluding it is an inadequate taxonomy of personality variables for I-O psychology to build knowledge and understand the determinants of work behavior and performance. They and others (especially Block, 1995) argued that the FFM is not comprehensive, combines variables into factors that are too heterogeneous, and is method-bound, dependent upon factor analysis. (See Hough et al., 2015, for a list of missing variables.)

Although some of these "missing" traits are included as lower-order facets in inventory-specific conceptualizations of the FFM, they are not necessarily measuring the same trait, nor are they necessarily narrow or homogenous enough to constitute personality *facets*. *Compound* traits such as integrity, managerial potential, or customer service orientation (cf. Ones & Viswesvaran, 2001) are made up of several homogeneous traits that do not necessarily covary, but all relate to a criterion of interest (Hough & Schneider, 1996). Hough and Ones (2001) have offered a working taxonomy of personality compound traits including scales available to measure them. In addition, a lack of generally accepted facet-level taxonomies for the Big Five domains and the resulting reliance on inventory-specific, lower-level trait descriptions has impeded research and practice of personality measurement relating to prediction of behaviors and performance in work settings, although empirically derived, facet-level taxonomies for Big Five domains are emerging (see, for example, Connelly, Davies, Ones, & Birkland, 2008, for agreeableness; Connelly, Ones, Davies, & Birkland, 2014; Roberts, Chernyshenko, Stark, & Goldberg, 2005, for conscientiousness).

Between Big Five factors and their facets, there are meso-level personality traits called *aspects* (DeYoung, Quilty, & Peterson, 2007). They are volatility and withdrawal (aspects of neuroticism), enthusiasm and assertiveness (aspects of extraversion), intellect and experiencing (aspects of openness), compassion and politeness (aspects of agreeableness), and industriousness and orderliness (aspects of conscientiousness). Judge, Rodell, Klinger, Simon, and Crawford (2013) meta-analyzed validities of the Big Five aspects for overall job performance, task performance, and citizenship behaviors and found general support for the DeYoung et al. (2007) approach and for the importance of facets in particular.

Since we wrote our chapter for the first edition of this book, research evidence has provided increased understanding of multiple taxonomic structures of personality variables. The HEXACO model, circumplex models, and nomological-web clustering approach are three such examples. Hough et al. (2015) provide a more in-depth description of these approaches, their limitations, and how they improve our theories, hypotheses, and prediction of work outcomes.

The HEXACO model identifies six factors (rather than five): Honesty-humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O). The HEXACO model is not simply the FFM plus Honesty-humility. Ashton, Lee, and deVries (2014) suggest that factor- and facet-level variables are substantively different in the two models (see Viswesvaran & Ones, 2016, for a contrary view).

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Circumplex models acknowledge the reality that personality variables often correlate with each other; despite the hierarchical structure envisioned for the FFM (or HEXACO model), factors correlate with other factors, and facets underlying the factors often correlate with facets in factors other than the one to which they supposedly belong. In circumplex models, two factors and their facets are considered at a time, until all 10 pairings (FFM model) are examined. In this way, the facets and factors and their inter-correlations are mapped. The disadvantage of circumplex models is that they only allow for two-dimensional space; that is, only two factors are considered at a time. Reality is more complex.

The nomological-web clustering approach is nonhierarchical and, as articulated by Hough and colleagues (Hough & Ones, 2001; Hough et al., 2015; Oswald & Hough, 2011; Oswald et al., 2013), envisions a structure of personality variables in which personality constructs (taxons) consist of personality variables that are similar in terms of (a) their relationships to each other, (b) their relationships to other variables (e.g., individual difference variables, individual and organizational outcome variables), (c) their psychobiological bases, (d) their interactions with other variables, (e) malleability over time, and (f) their patterns of relationships within demographic groups.

Although most meta-analyses have utilized the FFM to summarize the relationships among personality variables and job-related criteria, summaries of relationships at this broad level can mask relationships that emerge between narrower facets and performance constructs. Hough and colleagues (Hough, 1992, 1997, 1998; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Hough & Johnson, 2013; Hough & Oswald, 2005; Hough et al., 2015; Oswald & Hough, 2008, 2011; Oswald et al., 2013; Schneider, Hough et al., 1996) have long argued that focusing exclusively on factor-level personality traits in the prediction of heterogeneous work-related criteria can be counterproductive for a science aiming to explain the relationships between personality and work-related constructs. The predictive validity of a personality variable depends on (a) the criterion content domain being predicted (Bartram, 2005; Hough, 1992; Hogan & Holland, 2003; Ones, Dilchert, Viswesvaran, & Judge, 2007) and (b) the hierarchical match between the predictor and criterion measures (Hogan & Roberts, 1996; Ones & Viswesvaran, 1996; Schneider et al., 1996).

This is not to conclude that measurement at levels narrower than the facet level of the FFM is better. Overly narrow personality constructs can impede the growth of knowledge just as overly broad constructs can impede our science and practice (Hough, 1997; Ones, Viswesvaran, & Dilchert, 2005; Oswald & Hough, 2008, 2011; Oswald et al., 2013). Although it is appropriate to summarize the relationships between narrow constructs and various criteria, it is difficult to build a science without learning about the extent to which information and conclusions generalize at a broader construct level as well, including the Big Five and even higher-order factors (cf. Digman, 1997). Combining variables into compound variables (such as integrity and customer service orientation) that consist of multiple Big Five domains, such as conscientiousness, agreeableness, and emotional stability, can increase the predictive accuracy of personality variables (Hough & Ones, 2001; Ones et al., 2007; Ones & Viswesvaran, 2001).

Meta-analytic evidence summarizing personality-criterion relationships at various levels, including Big Five factors, facets, and compound scales, indicates that validity varies as a function of the theoretical relevance of the predictor to the criterion, which includes similarity of bandwidth (Hogan & Roberts, 1996; Hough, 1992; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Rothstein & Goffin, 2000; Schneider et al., 1996; Tett, Jackson, Rothstein, & Reddon, 1999). Personality variables that are a priori identified as theoretically relevant to a criterion correlate more highly with the criterion, and the overall predictor and criterion should be similarly heterogeneous/homogeneous.

The FFM provides an important organizing function for I-O psychology and has helped connect our science to our sister sciences. At the same time, we encourage the search for personality constructs that consist of variables with similar nomological nets to improve our understanding of personality structure. Using the nomological-web clustering model, Hough and Ones (2001) conducted a qualitative cluster analysis of such personality variable profiles and recommended that others use their model and further improve their taxonomy. Some quantitative summarizes have used this taxonomy in summarizing results across personality scales (e.g., Dudley, Orvis,

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Lebiecki, & Cortina, 2006; Foldes, Duehr, & Ones, 2008). Dudley et al. (2006) examined the criterion-related validities of four of the facets of conscientiousness defined by Hough and Ones, and found that these facets (a) have only low to moderate correlations with each other, (b) correlate differentially with the broad factor conscientiousness, and (c) depending upon the occupation and criterion, correlate higher with the criterion than does global conscientiousness. Foldes et al. (2008) used the Hough and Ones taxonomy to summarize mean score differences between Whites and different ethnic groups, finding that (a) facet-level mean score differences varied although the facets all belonged to the same domain and (b) factor-level differences varied from their facet-level mean score differences. Taken together, these summaries of very different types of information provide construct validity evidence for the Hough and Ones personality taxonomy as well as its usefulness for the science and practice of personnel selection. We urge others to report validities according to Hough and Ones' proposed structure, as well as to refine their structure to increase our understanding of the pattern of relationships between personality constructs and other constructs.

MEASUREMENT

Although personality variables are typically measured with self-report, Likert-type items and scales, other assessment methods can and are used. In this section we describe reliability, construct- and criterion-related validity evidence, and discuss practical issues such as development cost and ease of administration. In doing so, we discuss traditional, Likert-type measures and forced-choice, item response theory (IRT), and other recent innovations as well as several other methods of measuring personality, namely, biodata, interviews, situational judgment tests (SJTs), simulations, and assessment centers.

Self-Report Questionnaire Measures

Personality measurement is almost synonymous with standardized self-report questionnaires. Many other methods, some of them discussed in following sections, can also be thought of as a form of self-report. For example, the information provided in interviews and assessment centers is self-reported, despite being other-rated or recorded, and in most cases captured in a less standardized fashion. Traditional personality questionnaires elicit an individual's responses to items and use these responses (assuming Likert-type scaling) to express the individual's trait standing in comparison to a normative group. What distinguishes them from most other self-report methods is the degree of standardization they provide in eliciting test taker responses, allowing the user to reliably compare an individual's scores to those of other test takers.

Decades of research, hundreds of primary studies, and dozens of quantitative summaries have shown that such standardized, self-report tests of personality traits provide (a) reliable assessments (cf. Viswesvaran & Ones, 2000) and (b) scores that correlate at highly useful levels with valued organizational outcomes and criteria. We refer the reader to several comprehensive overviews of the validity of personality measures for predicting various valued behaviors and outcomes in organizational settings—e.g., Hough and Furnham (2003), Hough and Johnson (2013), Ones et al. (2005), and Ones et al. (2007).

Despite the strong empirical evidence for their validity (see section in this chapter titled "Validity of Personality Constructs and Factors that Affect Their Usefulness" for details), self-report measures of personality are often criticized when used in employee selection because of the possibility of intentional response distortion. Much of the research addressing the issue of response distortion has focused on standardized tests (rather than other ways of assessing personality; see below), and much of the basis of criticism of self-report measures is, as Chan (2009) suggested, likely rooted in an urban legend rather than reality. Dilchert and colleagues (Dilchert & Ones, 2011; Dilchert, Ones, Viswesvaran, & Deller, 2006) have summarized suggested palliatives and evaluated their merit to deal with intentional distortion and socially desirable responding on such measures, concluding that approaches such as score corrections or

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

exclusion of test takers on the basis of social desirability scale scores have little merit and that future improvements are more likely to come from the use of new and innovative item formats (see also McGrath, Mitchell, Kim, & Hough, 2010). Oswald and Hough (2008) and Hough and Oswald (2008) also summarized the literature, cautioning that results may differ depending on (a) item transparency (i.e., subtle vs. obvious items), (b) research setting (i.e., experimental vs. real-life applicant selection setting), and (c) research design (i.e., concurrent vs. predictive [longitudinal] design). They concluded that (a) validities for both types of scales remain essentially intact in real-life applicant selection situations using concurrent validation studies and (b) subtle-item scales also retain their validities in predictive designs. Below, we review new developments in this area and evaluate their promise for addressing concerns about response distortion in typical, Likert-type self-report personality scales.

Forced-Choice Item Response Formats

Forced-choice formats that require the respondent to choose between endorsing one statement (or characteristic) versus others force the respondent to score lower on one of the characteristics or scales. If the inventory consists of only a few traits/scales (e.g., five traits/scales), then the result is a distorted individual profile because the forced-choice nature of the measurement forces the individual to score low on some traits and high on others—a phenomenon known as ipsativity (Hicks, 1970). If the inventory consists of many scales/traits (e.g., perhaps 25), then the problem is less severe and is called quasi-ipsative. A meta-analysis of criterion-related validities of forced-choice inventories measuring personality characteristics indicates that quasi-ipsative measures might be better predictors of job performance than both forced-choice normatively scored and fully ipsative forced-choice measures (Salgado & Táuriz, 2014). A comparison of these meta-analytic results with meta-analytic results of Likert-type personality inventories (single stimulus items) found the quasi-ipsative scales correlated more highly with job performance (Salgado, Anderson, & Táuriz, 2014).

Computer Adaptive, IRT, Non-ipsative Forced Choice

One way to avoid ipsativity in forced-choice responses is to present response options that reflect different trait levels of the same construct. Rather than forcing the respondent to choose between equally attractive options loading on different traits, this approach uses item response theory (IRT) to develop more accurate measurement along the entire continuum of a given trait.

The Navy Computer Adaptive Personality Scales (NCAPS) is one example of this type of personality measurement. A computer-adaptive, forced-choice format (albeit with simultaneously presented response options loading onto the same trait) and a traditional version (non-adaptive, non-forced choice) of the NCAPS were developed (Houston, Borman, Farmer, & Bearden, 2005). Both types of scales correlated with the targeted criteria. In all but one comparison, the traditional NCAPS scales out-predicted the computer-adaptive, forced-choice scales and reached near-maximum validity with fewer items (six or seven item pairs for traditional NCAPS versus eight or nine for adaptive NCAPS). According to Underhill (2006), although the "item cutoff adaptive component of the Adaptive NCAPS version did not meet expectations" (p. viii), further research is warranted.

Another way to avoid ipsativity in forced-choice measures is a multi-unidimensional pairwise-preference model using item response theory to construct and score the items (Stark, Chernyshenko, & Drasgow, 2005). The "Tailored Adaptive Personality Assessment System" (TAPAS; Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012; Drasgow, Chernyshenko, & Stark, 2010b) is such an example. It is a computer-adaptive, forced-choice set of personality scales that yield non-ipsative (normative) measurements. Each item consists of response options that load on different traits. The U.S. Army is sufficiently impressed with the results that TAPAS is currently being used for many of its selection and placement decisions (Stark et al., 2014).

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Interestingly, though, results from a *longitudinal* validation study found that a rationally developed biodata inventory measuring personality characteristics appears to predict work-relevant criteria as well as the TAPAS scales even in high-stakes testing settings (Knapp, Owens, & Allen, 2011). More about these new strategies for measuring personality characteristics is provided in the next section on ideal point response methods.

Ideal Point Response Methods

Stark, Chernyshenko, Drasgow, and their colleagues, involved in a programmatic effort to improve current measurement of personality constructs, propose that ideal point response scales (based largely on Thurstone's, 1928, scaling method and assumptions) better fit the nature of item responding than Likert's (1932) method and assumptions. Ideal point response scales assume that people endorse items that are closer to their true trait level (i.e., an individual's ideal point) than items that are further away from their true trait level, and thus provide more precise measurement than Likert-type scales at all points on the trait continuum. Items that differentiate people at the extreme ends of the continuum are infrequently endorsed, resulting in low variances and low item-total scale correlations. Such items are retained in ideal point scaling methods but typically discarded in Likert scaling methods. With Likert-type scaling methods, desirable items have monotonically increasing item response functions, whereas items selected using ideal point response methods have bell-shaped item response functions. On the basis of item-response theory analyses, Stark, Chernyshenko, Drasgow, and colleagues conclude that ideal point response methods (a) fit monotonically increasing item response functions (although they, compared with Likert-type scales, do not require it), (b) do not negatively affect criterion-related validity of personality scales, and (c) provide more accurate measurement of high- and low-scoring individuals and thus potentially lead to better selection decisions (Chernyshenko, Stark, Drasgow, & Roberts, 2007; Stark & Chernyshenko, 2007; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & Williams, 2006).

These benefits appear to be real, but an important question is whether the complexity of the development and scoring procedures is required to attain these advantages. As Oswald and Schell (2010) state in their commentary to the Drasgow, Chernyshenko, and Stark (2010a) article describing Thurstone's approach (ideal point response method) and its advantages over Likert-style measurement: "Science prefers parsimony unless the added complexity is justified" (p. 482). Importantly, Oswald and colleagues (Oswald, 2010; Oswald, Shaw, & Farmer, 2015) successfully predicted ideal point personality scores with much simpler scoring methods.

Another scoring innovation is retrospective scoring of traditional multi-dimensional forced-choice questionnaires. Brown and Maydeu-Olivares (2011, 2013) demonstrated how to recover normative data from traditionally scored forced-choice questionnaires. Using a commercially available, traditionally scored forced-choice personality inventory, they used item response modeling (IRT methods) to re-score the data and overcome the limitations of ipsative data. This development will no doubt lead to re-analyses of significant amounts of personality data obtained using traditional forced-choice questionnaires. We envision re-analyses of many criterion-related validity studies and new meta-analyses using "recovered normative data" from traditional multi-dimensional forced-choice questionnaires that originally used non-IRT-based scoring. (See Chapter 42, this volume, for additional discussion of IRT methods related to selection.)

Intentional Distortion and Forced-Choice Item Response Formats

In the early and mid-20th century, the initial motive for developing forced-choice tests that asked respondents to choose between response options matched for level of social desirability was the desire to reduce, even eliminate, intentional distortion. In the early 21st century, much of the impetus for seeking a new personality measurement model, such as the ideal point response

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

method, is the same: When individuals are motivated or instructed accordingly, Likert-style personality scales can be easily faked, i.e., intentionally distorted in ways that make the test taker look better than they actually are (Hough et al., 1990; Viswesvaran & Ones, 1999). The concern, of course, is the possible effect on criterion-related validity in high-stakes testing situations such as in personnel selection contexts. The motivation to reduce intentional distortion continues today, even though controversy still exists about the amount of intentional distortion in Likert-type scales in real occupational (versus experimental) settings and how to overcome such distortion. Some argue that measurement strategies such as forced-choice, unidimensional (or multi-unidimensional) pairwise-preference models are needed, whereas others argue that other strategies, for example, warnings and consequences for distorting self-descriptions, are sufficient for overcoming most intentional distortion when Likert-style scales are used.

Evidence about the amount of distortion with partially ipsative, forced-choice scales indicates that less distortion occurs on partially ipsative, forced-choice scales than Likert-type scales (Jackson, Wroblewski, & Ashton, 2000; Martin, Bowen, & Hunt, 2002; Stanuch, 1997; White, Young, & Rumsey, 2001). Evidence about the amount of distortion that occurs with the new measurement strategies (ideal point response methods) indicates that multi-dimensional forced-choice inventories administered even without warnings and consequences for distortion result in less distortion than Likert-type items in high-stakes testing (Stark et al., 2014). However, other research (e.g., Heggestad, Morrison, Reeve, & McCloy, 2006) indicates that this phenomenon holds only at the group level of analysis and not at individual-level analyses (which is particularly relevant in the case of employee selection). McCloy, Heggestad, and Reeve (2005) also raised concerns that, although forced-choice measures may effectively curb faking at the item level, it can still be possible to distort one's scores on the scale level by identifying the traits hypothesized to relate to job performance and endorsing statements accordingly.

Additional research suggests that in predictive validity studies, forced-choice measures retain their validity better than Likert-scaled measures, although there is an unwanted substantial increase in the correlation between the forced-choice measure and cognitive ability (Christiansen, Burns, & Montgomery, 2005). It is possible that the apparent retained validity is actually due to the cognitive ability variance in the forced-choice measure, reducing the usefulness of personality measures as a strategy to reduce adverse impact against protected classes in hiring decisions. The evidence so far suggests that forced-choice methods are not fake-resistant in real-life settings, although the new non-ipsative forced-choice formats may produce more fake-resistant measurements than do Likert-type scales and might do so without the unwanted correlation with cognitive ability—time will tell.

Other-Reports

Organizations frequently use 360-degree feedback measures (a form of other-reports), and observers frequently rate personality characteristics of participants in simulations and assessment center exercises (see following section). Yet until recently, organizations rarely utilized other-reports of individuals' personality that are assessed with standardized personality measures. This has started to change, in part due to meta-analytic support for their predictive validity as well as evidence of their incremental validity over self-report measures. A comprehensive meta-analysis showed that unreliability-corrected consensus correlations between self and other-reports for the Big Five range from .72 to .91 (Connelly & Ones, 2010). However, when stranger ratings are excluded from analyses, convergence between self- and other-reports is higher. Although even strangers can provide valuable insight into a target individual's personality on easily observed traits, such as extraversion (Connolly, Kavanagh, & Viswesvaran, 2007), opportunity to observe behavior through increased interactions improves convergence. Largest consensus with self-reports exists for observers who have closest interpersonal intimacy with the target being rated (e.g., spouse, parents, siblings). Unreliability corrected self–family member consensus correlations range between .80 for emotional stability and .91 for agreeableness. Largest improvements in convergence due to familiarity are found for low-visibility traits (e.g., emotional stability); they are marginal for highly evaluative traits (e.g., agreeableness). Although

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

personality traits are observed and evaluated in slightly different ways by self and others, there is substantial overlap between these sources, especially once the attenuating effect of unreliability is corrected. However, less than perfect self–other convergence in personality ratings suggests that other ratings can increment criterion-related validity beyond that of self-reports, and two meta-analyses (i.e., Connelly & Ones, 2010; Oh, Wang, & Mount, 2011) indicate criterion-related validity is higher than for self-report.

In predicting academic achievement, others' ratings of extraversion and conscientiousness appear to be stronger than self-ratings. Operational validities associated with single observers are .35 and .41, respectively (Connelly & Ones, 2010). These values greatly exceed those reported for the same traits by Hough (1992; .08 and .25, respectively) and Poropat (2009; −.02 and .18, respectively), whereas single observer validities for emotional stability are similar to self-ratings (.27 reported by Connelly & Ones, 2010, and .22 reported by Hough, 1992, respectively). The availability of multiple raters offers the possibility of achieving validities of .69 for conscientiousness, .52 for extraversion, and .46 for emotional stability for predicting academic achievement (Connelly & Ones, 2010).

In predicting job performance, a single observer's description of a target's personality predicts job performance better than does a self-rating of personality (Connelly & Ones, 2010; Oh et al., 2011). As with self-ratings, the strongest validity is for conscientiousness (approximately .30s, depending on the corrections applied); validities for other Big Five dimensions are, while lower, still at useful levels. When multiple raters assess personality, validities for job performance asymptote to the .50s for conscientiousness, .30s for agreeableness, .40s for openness, and .30s for emotional stability (Connelly & Ones, 2010). For extraversion, findings are somewhat lower. Similar conclusions are reached when results from Oh et al. (2011) are considered.

Only a handful of studies have examined the predictive validity of others' ratings of personality, and further research on this topic is warranted. It would be especially valuable to examine others' ratings of personality in the prediction of major job performance dimensions of task performance, organizational citizenship behavior, and counterproductive work behavior. Further strengthening personality measurement using multiple raters is a viable strategy and promises to improve prediction. Compositing ratings across multiple raters increases reliability and compounds accuracy of raters (compared to targets themselves; Connelly & Hülsheger (2012)).

Studies investigating potential response distortion in standardized other-reports of personality are also scarce. It is safe to assume that if such measures were used to elicit information from candidates' acquaintances, the choice of rating source will influence the degree of distortion to be expected. However, it is unlikely that organizations are willing to rely on ratings obtained from spouses or friends in selecting among job applicants. We see the potential for other-ratings of personality for applications in which the source of the ratings can be standardized and verified (e.g., personality ratings made by the last two supervisors). Other tools used in employee selection already employ a similar rationale (e.g., letters of reference) but do not provide the benefit that standardized ratings of personality could provide: wide distributions of scores that could be used to select rather than identify negative indicators that allow screening out of potential candidates. We encourage researchers and practitioners to explore the potential for standardized other-ratings of personality and to conduct additional studies investigating their criterion-related validity and potential for incrementing validity.

Biodata

Biodata measures (also known as biographical data and autobiographical information) focus on previous life experiences and have a long history in I-O psychology as useful predictors of work-related criteria (see reviews by Griffeth, Hom, & Gaertner, 2000; Hough, 2010; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). The premise of their success is the old adage: past behavior is the best predictor of future behavior (consistency principle). When scale development is construct-oriented, biodata represent another method of measuring individual differences such as personality constructs (Hough & Paullin, 1994). As Tenopyr (1994) hypothesized, biodata scales developed to measure personality constructs (e.g., Big Five factors and

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

their facets) correlate appropriately with each other and with work-related criteria (cf. Kilcullen, White, Mumford, & Mack, 1995; Manley, Benavidez, & Dunn, 2007; Oviedo-Garcia, 2007; Sisco & Reilly, 2007a; Stokes & Cooper, 2001).

Although intentional distortion occurs on biodata and on traditional personality scales, the evidence on the extent of distortion compared to traditional personality scales is mixed. Some studies report less distortion on biodata scales (e.g., Kilcullen et al., 1995; Sisco & Reilly, 2007b; Stokes, Hogan, & Snell, 1993), whereas other research suggests little difference in the amount of faking on biodata versus standard personality scales (e.g., McFarland & Ryan, 2000; White et al., 2001). Evidence suggests that one fruitful approach to reduce distortion is to require respondents to elaborate on their responses to biodata items (Schmitt & Kunce, 2002; Schmitt et al., 2003). Moreover, both verifiable and subtle items (where the construct measured is less apparent) appear to retain their validity when used in real-life applicant settings (Alliger, Lilienfeld, & Mitchell, 1996; Harold, McFarland, & Weekley, 2006; White, Young, Hunter, & Rumsey, 2008).

Given the advantages of biodata, it is surprising that biodata measures are not used more frequently for employee selection (Stokes & Cooper, 2004). We expect that opportunities that the mega-trends involving social media and big data sets present will result in personality-based biodata becoming more fully utilized in future hiring processes. It is important to note that we do not argue that organizations should obtain or use (even publicly available) information from individual applicants' social media profiles in making personnel decisions; in fact, ethical and legal considerations speak against such data use in most circumstances. However, the possibilities of using data from large numbers of individuals' social media profiles to empirically identify effective biodata predictors of criteria of interest for purpose in selection tool development are truly exciting.

Interviews

Around the world, the interview is probably the most frequently used employee selection assessment method (Moscoso, 2000; Ryan, McFarland, Baron, & Page, 1999), and it is most often intended to measure personality characteristics (Huffcutt, Conway, Roth, & Stone, 2001). Huffcutt and colleagues developed a comprehensive taxonomy of possible interview constructs that interview questions might measure. The seven constructs were (1) mental ability, (2) knowledge and skills, (3) basic personality characteristics (such as the Big Five), (4) applied social skills and social competence, (5) interests and preferences, (6) organizational fit, and (7) physical attributes. They sorted 338 interview questions from 47 actual employment interviews into the seven constructs. They found that interview questions were most often intended to measure personality characteristics (35% of the questions), followed by applied social skills (28%), mental ability (16%), knowledge and skills (10%), interest and preferences (4%), physical attributes (4%), and organizational fit (3%). Sixteen percent of all questions were intended to measure conscientiousness or its facets.[1]

Huffcutt et al.'s (2001) study does not address the construct validity of interview ratings but did find that interview ratings of personality correlate well with overall job performance in various jobs. The correlations (corrected for range restriction in interview scores and measurement error in performance evaluations) with overall job performance were .33 for extraversion, .33 for conscientiousness, .51 for agreeableness, and .47 for emotional stability. Nor did the study examine the validity of a compound personality variable such as Big Five scales used in combination, which has been shown by Ones et al. (2007) to increase the validity of personality for predicting important job-relevant criteria (validities in the high .40s, high .30s, and mid .20s for predicting team performance, leadership performance, and overall job performance, respectively).

Studies that purport to investigate the construct validity of the employment interview often investigate external correlates but often leave unanswered whether or not the interview measured the construct(s) intended. The few studies that have examined the construct validity of personality scores obtained from interviews designed specifically to measure personality variables do not provide much support for the construct validity of such interview scores (e.g., Roth, Van Iddekinge, Huffcutt, Eidson, & Schmit, 2005; Van Iddekinge, Raymark, Eidson, & Attenweiler,

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

2004). Another study by Van Iddekinge and colleagues (Van Iddekinge, Raymark, & Roth, 2005) examined the construct validity of an interview for assessing the NEO Personality Inventory facets of altruism, self-discipline, and vulnerability. Interviewees described themselves using the NEO facet scales, and experienced interviewers interviewed the mock candidates, asking them questions intended to measure the three characteristics. The interviewers provided interviewer ratings of the personality constructs; they also completed the NEO facet scales to describe the candidates. The study included an honest as well as an applicant-like condition. In the honest condition, convergent validities of interviewer-based NEO ratings with the self-report NEO ratings averaged .32 (discriminant validities averaged .20); convergent validities of the interview ratings with self-report NEO ratings averaged .24 (discriminant validities averaged .16). Neither type of interviewer-based assessment of personality showed good convergent validity with the target constructs, and convergent validities were even lower in the applicant-like condition (Van Iddekinge et al., 2005), possibly because of the effect of response distortion in self-reports and interview scores in this condition.

The disappointing results for construct validity of the interview as a measure of personality characteristics can perhaps be improved with attention to four variables that moderate the accuracy of personality judgments: the judge, the target individual, the trait, and information obtained (Funder, 1995). Research suggests that (a) unstructured interviews carry more personality variance than do structured interviews (although criterion-related validity may suffer); (b) visible traits such as extroversion and agreeableness are better measured than are less visible traits; and (c) accuracy increases with more information about the target individual (Blackman & Funder, 2002).

It is also possible that interviewers' overall ratings of interviewees' personalities might provide a measure of a general personality factor or level (profile elevation) that is a useful predictor of criteria of interest. Certainly, this would not be a construct-valid measure of a particular personality variable, but it might provide a very useful level of validity for hiring purposes. The Huffcutt et al. (2001) study, as well as meta-analyses of the criterion-related validity of interviews in general (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Wiesner & Cronshaw, 1988), leave little doubt about the criterion-related validity of the interview for predicting job-relevant criteria. And, as we suggest, the interview might provide an overall assessment of personality that is useful for personnel selection.

Situational Judgment Tests

Situational Judgment Tests (SJTs) present test takers with a scenario (in written, audio, or video format) and several response options describing possible courses of action. For employee selection purposes, SJTs are most often contextualized for specific occupational domains (e.g., law enforcement or customer service) and are often designed to measure interpersonal characteristics and personality traits deemed particularly relevant (e.g., conscientiousness or extraversion). When measuring personality traits via an SJT, the development of scenarios and response options must be theory- and data-driven, and SJT scores hypothesized (based on item content) to measure a certain personality trait should relate to external measures of the same construct (Chan & Schmitt, 2005).

Choice of response instructions is critically important in measuring personality via SJTs. A major distinction in SJT response instructions is behavioral tendency versus knowledge instructions, sometimes conceptualized as "would do" versus "should do." Conceptually, SJTs administered with behavioral tendency instructions are more likely to elicit responses that resemble future behavior on the job, rather than mere knowledge of appropriate responses to a given scenario. Even though different behavioral tendency instructions produce scores that are highly correlated (Ployhart & Ehrhart, 2003), response instructions play a key role in determining whether interpersonal, personality, or cognitive characteristics are measured.

We used data from the McDaniel, Hartman, Whetzel, & Grubb (2007) meta-analysis to shed light on the constructs typically assessed using SJTs with behavioral tendency ("what would you do?") versus knowledge ("what should you do?") response instructions. Using their meta-analytic

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

true-score correlations in combination with Big Five intercorrelations, we estimated the amount of personality variance typically observed in SJTs. (We obtained the meta-analytic Big Five intercorrelations from the Ones [1993] meta-analysis[2] and attenuated them to reflect observed relationships [using meta-analytic reliability estimates from Viswesvaran & Ones, 2000].) A multiple regression of SJT scores on the Big Five indicated that at the construct level, 25% of the variance assessed by SJTs with *behavioral tendency instructions* ("what would you do?") is personality (Big Five) variance. Less than 10% of the variance is explained by the Big Five when SJTs are administered with *knowledge instructions* ("what should you do?"). This suggests that SJTs with behavioral tendency instructions are better suited to measure personality traits; that is, "would do" instructions elicit more personality-saturated responses. Meta-analyses of the criterion-related validities of SJTs using behavioral tendency ("would do") instructions indicate validity is .26 (corrected for sampling error and attenuation due to criterion unreliability) for predicting overall job performance (McDaniel, et al., 2007). Much of the SJT validity research lacks a construct-oriented approach, and because of that lack of focus on constructs, the McDaniel et al. meta-analysis was unable to examine the validities of response instructions according to criterion construct.

In an effort to understand the construct validity of SJTs, Christian, Edwards, and Bradley (2010) examined SJT inventories and classified them into six more or less homogenous groups—i.e., interpersonal skills, teamwork skills, and leadership (applied social skills), personality composites and conscientiousness (basic personality tendencies), and job knowledge. They also classified the criteria into constructs (i.e., contextual performance, task performance, and managerial performance). They then separately meta-analyzed the criterion-related validities of each SJTs content area for each criterion construct. They found highly useful levels of validity for SJTs predicting all performance criteria. In addition, they showed that SJTs designed to measure personality constructs yield validities on par with (or higher) than those of SJTs assessing knowledge and skill constructs. Clearly, construct-focused research with SJTs is beneficial, and we encourage more construct-oriented research.

As is true for all individual difference measures, reliability is an important factor when evaluating the usefulness of SJTs for employee selection purposes. SJTs are often multi-dimensional (McDaniel et al., 2007), rendering internal consistency estimates of little value (as would be the case if an internal consistency reliability estimate were to be computed across items of different scales on a traditional personality test). In these circumstances, parallel form reliability (Chan & Schmitt, 2002) and test-retest correlations (over a short time period; Schmidt & Hunter, 1996) are appropriate methods of estimating reliability. However, both types of estimates for SJTs measuring personality variables are rarely presented in the literature. We encourage scientists and practitioners alike to investigate and report on this issue to further improve our knowledge of personality measurement using situational judgment approaches.

Another important issue concerns the distribution of constructs included in the response options of each SJT item. Providing response options that load on different traits (conceptually and empirically) complicates score interpretation and makes inter-individual comparisons difficult. This is especially true in the case of personality assessment. The challenge lies in developing different response options that are all expressions of the same personality trait, albeit at different trait levels, for each SJT item/scenario. Making test takers choose between response options loading on different personality dimensions will result in ipsative or partially ipsative scores, limiting their usefulness for employee selection purposes (see earlier discussion on ipsativity).

Simulations and Assessment Centers

Assessment centers (ACs) have received much attention in the research literature, yet high development and administration costs often limit their use only to occupations in which the dollar value of performance variability is large (e.g., as selection tools for higher-level managerial positions or screening tools in high-risk jobs). This is also true for what can be considered their building blocks—single exercises or simulations that can be administered individually to assess personal characteristics.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Motowidlo, Dunnette, and Carter (1990) described a simulation as any situation that "present[s] applicants with a task stimulus that mimics an actual job situation" (p. 640). Now simulations are considered situational tests that have fidelity greater than a paper-and-pencil test (Thornton & Rupp, 2003). Construct-validity evidence for the traits underlying performance on simulations is often sparse. A systematic review of the available literature reveals that many dimensions assessed in ACs are at least conceptually related to personality dimensions (Arthur, Day, McNelly, & Edens, 2003).

Arthur et al.'s (2003) construct-based meta-analysis of the AC method has shown that personality-based AC dimensions, especially influencing others (a facet of extraversion), possess predictive validity that rivals that of cognitive-ability-based dimensions such as problem solving. A survey of AC practices among 97 organizations in western Europe and North America (Krause & Thornton, 2009) shows that personality-based, extraversion-related dimensions are among those most commonly assessed in ACs, in addition to interpersonal ones conceptually related to agreeableness (e.g., consideration of others).

In ACs, personality-relevant variance is captured using simulations and exercises such as role-plays, group discussions, or in-baskets. An early meta-analysis by Scholz and Schuler (1993) revealed an interesting pattern of findings, indicating that scores obtained in group discussion exercises mainly captured openness to experience, dominance, and self-confidence ($\rho = .46$, .34, and .39, respectively, $N = 236$–318), whereas in-basket exercises only reflected dominance ($\rho = .23$, $N = 273$). A large-scale investigation in two primary samples ($N = 3{,}748$–$4{,}770$) showed that scores on many simulations correlate only negligibly with personality characteristics, with the exception of extraversion (Ones & Dilchert, 2008).

Simulations and exercises are often tailored to a given job context to make them more realistic and face-valid. However, design features can impact the nature of the construct measured and the quality of the measurement (e.g., reliability). For example, a leaderless group discussion that is competitive (e.g., framed in a negotiation scenario) is more likely to elicit behaviors indicative of different personality traits than a discussion that is cooperative (e.g., framed in a team problem-solving context). The selection of simulations and exercises for the prediction of specific criteria should take such issues into account. Factors such as observability also affect the reliability and validity of scores. The survey by Krause and Thornton (2009) indicated that in about 50% of organizations surveyed, most (> 75%) AC exercises are specifically developed for an organization. Customization is costly. If customization elicits behavior indicative of traits that are particularly valued in a given context though, the cost is likely worthwhile.

VALIDITY OF PERSONALITY CONSTRUCTS AND FACTORS THAT AFFECT THEIR USEFULNESS

Significant evidence documents the utility of personality variables for predicting important organizational criteria. Yet there are those who sharply criticize the utility of personality variables for employee selection on the grounds of purportedly low validities. For an exchange on this issue, see Morgeson et al. (2007a, 2007b) and Murphy and Dzieweczynski (2005) for one side of the argument; and Barrick and Mount (2005), R. Hogan (2005a, 2005b), Hough and Oswald (2005), Ones et al. (2005, 2007), and Tett and Christiansen (2007) for the other side. In addition, we refer the reader to meta-analyses and reviews of the literature such as Barrick, Mount, and Judge (2001); Dudley et al. (2006); J. Hogan and Holland (2003); J. Hogan and Ones (1997); Hough and Furnham (2003); Hough and Johnson, (2013); Hough and Ones (2001); Hough and Oswald (2008); Ones et al. (2007); Ones, Viswesvaran, and Schmidt (1993); Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007); and Rothstein and Goffin (2006). These summaries indicate that personality constructs predict many important criteria, including major life outcomes. The list of criteria that are well predicted by personality variables includes, among others, the following:

- *Overall job performance*: Conscientiousness, $r_{true} = .23$ (Barrick & Mount, 1991) and $r_{operational} = .20$ (Hurtz & Donovan, 2000); integrity tests, $r_{operational} = .41$ (Ones et al., 1993); and core self-evaluations $r_{true} = .36$ (Chang, Ferris, Johnson, Rosen, & Tan, 2012)

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

- *Organizational citizenship behaviors*: Conscientiousness r_{true} = .22; agreeableness, r_{true} = .17; openness, r_{true} = .17; emotional stability, r_{true} = .15; extraversion, r_{true} = .11 [somewhat different patterns for organizational, interpersonal, and change-oriented OCB are also reported] (Chiaburu, Oh, Berry, Li, & Gardner, 2011); core self-evaluations r_{true} = .22 (Chang et al., 2012); positive affect, r_{true} = .23; negative affect, r_{true} = –.10 (Kaplan, Bradley, Luchman, & Haynes, 2009)
- *Counterproductive work behavior*: Conscientiousness, $r_{operational}$ = −.26 (Salgado, 2002), $r_{operational}$ = –.31 (Berry, Ones, & Sackett, 2007); dependability, r_{true} = −.34 (Dudley et al., 2006); emotional stability, $r_{operational}$ = –.23, agreeableness, $r_{operational}$ = –.38 (Berry et al. 2007), personality-based integrity tests, $r_{operational}$ = −.32, overt integrity tests, $r_{operational}$ = .55 (Ones et al., 1993); core self-evaluations, r_{true} = −.19 (Chang et al., 2012); negative affectivity, r_{true} = .30 (Kaplan et al., 2009). For counterproductive work behaviors rated by others: conscientiousness, $r_{operational}$ = –.19, agreeableness, $r_{operational}$ = –.22 (Berry, Carpenter, & Barratt, 2012)
- *Task performance*: core self-evaluations, r_{true} = .19 (Chang et al., 2012); positive affectivity, r_{true} = .20; negative affectivity, r_{true} = –.09 (Kaplan et al., 2009)
- *Adaptive performance at work*: emotional stability, $r_{operational}$ = .16, but .20 for managers; ambition, $r_{operational}$ = .14, but .26 for managers (Huang, Ryan, Zabel, & Palmer, 2014)
- *Managerial effectiveness*: Dominance, $r_{operational}$ = .27; energy level, $r_{operational}$ = .20; achievement orientation, $r_{operational}$ = .17 (Hough, Ones, & Viswesvaran, 1998); conscientiousness, r_{true} = .22 (Barrick & Mount, 1991)
- *Entrepreneurial performance*: Conscientiousness, r_{true} = .19; openness, r_{true} = .21; emotional stability, r_{true} = .18 (Zhao, Seibert, & Lumpkin, 2010)
- *Customer service*: Customer service scales, $r_{operational}$ = .34 (Ones & Viswesvaran, 2008)
- *Unsafe behavior*: conscientiousness, r_{true} = –.25; agreeableness, r_{true} = –.26; emotional stability, r_{true} = –.13; extraversion, r_{true} = .10 (Beus, Dhanani, & McCord, 2015)
- *Job satisfaction*: Emotional stability, r_{true} = .29; conscientiousness, r_{true} = .26; extraversion, r_{true} = .25; agreeableness, r_{true} = .17 (Judge et al., 2002); core self-evaluations, r_{true} = .44 (Chang et al., 2012); positive affectivity r_{true} = .33; negative affectivity, r_{true} = –.37 (Thoreson, Kaplan, Barsky, Warren, & de Chermont, 2003)
- *Job commitment*: core self-evaluations, r_{true} = .32, –.18, and –.27 for affective commitment, continuance commitment, and turnover intentions, respectively (Chang et al., 2012)
- *Intrinsic motivation*: core self-evaluations, r_{true} = .37 (Chang et al., 2012)
- *Goal setting—goal setting motivation*: emotional stability and agreeableness, r_{true} = .29; conscientiousness, r_{true} = .29; openness, r_{true} = .18; extraversion, r_{true} = .15 (Judge & Ilies, 2002). G*oal commitment:* core self-evaluations, r_{true} =.44 (Chang et al., 2012)
- *Life satisfaction*: core self-evaluations, r_{true} = .57 (Chang et al., 2012); emotional stability, r_{true} = –.45; extraversion, r_{true} = –.35; conscientiousness, r_{true} = .27; agreeableness, r_{true} = .19 (Steel, Schmidt, & Shultz, 2008)
- *Divorce*: Conscientiousness, $r_{observed}$ = −.13; emotional stability, $r_{observed}$ = −.17; agreeableness, $r_{observed}$ = −.18 (Roberts et al., 2007)
- *Mortality*: Conscientiousness, $r_{observed}$ = −.09; extraversion/positive emotion, $r_{observed}$ = −.07; emotional stability, $r_{observed}$ = −.05; agreeableness/lack of hostility, $r_{observed}$ = −.04 (each greater than the effects of socioeconomic status and IQ; Roberts et al., 2007)

Ones et al. (2005; 2007), as well as Hough and Ones (2001) and Hough and Johnson (2013), have summarized the meta-analytic evidence for compound personality scales in predicting work-related criteria and shown that these scales have high validity in predicting the specific criteria they were developed for, as well as for overall job performance. We readily acknowledge that it is not necessarily Big Five factors that predict valued outcomes. Indeed, we argue that (a) more specific criteria are predicted by more narrow personality traits; (b) complex criteria are predicted by theoretically appropriately matched predictors; and (c) for some of the criteria listed above, the highest predictive validities are not necessarily obtained at the factor level.

We do not want to underestimate the importance of the FFM. It has provided a structure for us to think about personality variables. Prior to its acceptance, personality and I-O psychology had little from which to generalize, the myriad of personality measures and variables numbered in the hundreds, and there were different names for the same or similar constructs or the same name for different constructs. We are not advocating a return to the "good old daze" (Hough, 1997). We applaud the interest and evidence coming from studies that examine facet-level

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

variables of the FFM. We urge more such research, especially research based on empirically derived, generalizable, facet-level personality taxonomies.

An area of increased research attention in examining personality in work contexts are maladaptive traits and measures, as well as their overlap with measures and taxonomies of adaptive personality. Often, as Dilchert, Ones, and Krueger (2014) point out, "personality constructs range between maladaptive positive and negative extremes, with the middle normal range representing typical (i.e., 'normal') traits" (p. 98).

Maladaptive personality measures have item content that is tilted toward higher negative valence. Examples include so-called dark side measures (e.g., Hogan Development Survey [HDS], R. Hogan & Hogan, 2009), measures of the "Dark Triad" of narcissism, Machiavellianism, and psychopathy (Paulhus & Williams, 2002), as well as—at the extreme end—measures of psychopathology (e.g., the MMPI, Ben-Porath & Tellegen, 2008; Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001; the Personality Inventory for DSM-5 [PID-5], Krueger, Derringer, Markon, Watson, & Skodol, 2012). Typically, such measures assess more extremes of personality variables, often describable in terms of Big Five factors, their facets, but most often compounds. For example, narcissism measures capture variance from low agreeableness and high extraversion (Moore & Ones, 2016).

I-O psychology literature on maladaptive traits is meager. A few recent meta-analyses have summarized the criterion-related validities of the Dark Triad and the dark side traits assessed by the HDS. Machiavellianism and narcissism predict counterproductive work behavior well (r_{true} = .25 and .43, respectively; O'Boyle, Forsyth, Banks, & McDaniel, 2012). It appears that the entitlement/exploitative facet of Narcissism is responsible for its predictive utility for CWB (Grijalva & Newman, 2015). For managers, a small meta-analysis (*k*s from 4 to 12) based on the Hogan Development Survey examined the validity of dark side traits for managerial performance (Gaddis & Foster, 2015). Managers who were leisurely ("indifferent to other people's requests"), skeptical (cynical and distrustful), excitable (volatile and inconsistent), and cautious (resistant to change) performed worse (operational validities ranged from –.11 to –.20). Fine-grained analyses indicated that colorful (dramatic), bold (overconfident), imaginative, mischievous (taking risks, testing limits), and skeptical managers are rated as untrustworthy by their supervisors (unreliability-corrected validities ranged from –.10 to –.29).

We urge more such research on workplace consequences of maladaptive traits and workforce-relevant nomological nets of their measures. Especially needed are studies examining how these traits and measures relate to a broader spectrum of criteria that are consequential in organizations such as negotiation tactics, conflict resolution, benefitting from HR interventions, political behavior, coaching, mentoring and derailment, among others. Also important, maladaptive behavior may not have a personality construct label that is obviously maladaptive. For example, Chan (2006) has demonstrated that the construct of proactive personality, which has almost always been taken as adaptive, can be maladaptive when high proactive personality scores are accompanied by low situational judgment effectiveness. Assessments of maladaptive traits, as well as understanding how positive traits can be negative in the presence (or lack) of other skills and abilities, can be useful in predicting interpersonal behavior as well as counterproductivity in organizations.

Incremental Validity

Personality variables can increment criterion-related validity in at least one of two ways. One way is in combination with other relevant personality variables. A second way is in combination with other individual variables such as measures of cognitive ability. Personality variables generally have low correlations with cognitive ability measures and do increment validity when jointly used (Bartram, 2005; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; White et al., 2008). When used in combination with other measures, such as the interview, biodata, and situational judgment, personality variables also increment validity (DeGroot & Kluemper, 2007; McManus & Kelly, 1999).

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Variables That Moderate Validity of Personality Constructs

Many variables affect the magnitude of the criterion-related validity that is obtained in primary and meta-analytic studies. Important factors are the type of criterion, the criterion measurement method, the relevance of the predictor for the criterion, personality measurement method (see above), research setting (experimental/laboratory vs. real-life selection), research design (concurrent vs. predictive/longitudinal), item transparency (subtle vs. obvious), and rater perspective (e.g., self vs. other). The more theoretically relevant the predictor is to the criterion, the higher the validity. The Hough and Furnham (2003) and Hough and Johnson (2013) summaries of meta-analyses according to predictor and criterion construct provide excellent examples of how predictor-criterion relevance affects the relationship between the two. In addition, validities are typically higher in concurrent validation studies compared to longitudinal validity studies (see Lievens, Ones, & Dilchert, 2009, for exceptions). Validities are also higher in "weak" situations in which people have more autonomy and control compared with "strong" situations in which people have few options. Trait activation theory (Tett & Burnett, 2003; Tett & Guterman, 2000) provides an integrated framework for understanding how the situation can explain variability in the magnitude of the relationships between personality and behavior and performance.

Nature of Predictor-Criterion Relationships

As with all employee selection measures (whether standardized tests, interviews, simulations, or ACs), their utility depends on the nature of the predictor-criterion relationship. In making top-down hiring decisions, linearity of the relationship between predictor and criterion scores is typically assumed. Pearson correlations, which are most commonly used to estimate operational validities, also assume linearity, the same assumption that is critical to traditional utility approaches.

Two plausible scenarios of nonlinearity between personality traits and criterion variables that would impact employee selection, especially with rigorous top-down selection, are: A relationship between the predictor and criterion in which an asymptote is reached after a certain level of predictor scores (e.g., beyond a certain point, all conscientious individuals keep their workplace similarly tidy, and differences in orderliness do not translate into performance differences). Additionally, a U- (or inverted U) shaped function, in which the direction of the relationship actually reverses beyond a certain level of predictor scores, is possible. In the case of an asymptotic relationship between personality and criterion scores, there is still potential utility in using personality as part of a selection process. Many organizations using minimum standards or defined cutoff scores do so because of the implicit assumption that predictor scores do not matter after a certain cutoff.

If, however, personality-performance relationships are described by an inverted U-shaped function, the detrimental effect on overall utility of a selection system could be significant. In cases where predictor-criterion relationships reverse direction, top-down hiring could result in the acceptance of applicants who display high predictor scores but actually perform worse than some lower-scoring candidates. There have been a handful of studies investigating curvilinearity of personality–job performance relationships, most focusing on conscientiousness. Classically scaled conscientiousness scale scores are linearly related to overall job performance (LaHuis, Martin, & Avis, 2005; Robie & Ryan, 1999; Walmsley, 2013; Whetzel, McDaniel, Yost, & Kim, 2010), task performance (Carter et al., 2014; Le et al., 2011), organizational citizenship behaviors, and counterproductive work behaviors (Carter et al., 2014; Le et al., 2011), as well as GPA and training performance (Cucina & Vasilopulos, 2005; Vasilopoulos, Cucina, & Hunter, 2007), although slight decrements to performance or its facets were reported for classically scored, ad hoc measures of conscientiousness (Carter et al., 2014; La Huis et al., 2005; Le et al., 2011).

Overall, for most commercially available personality inventories, curvilinearity appears not to present an impediment to predicting performance constructs (for examples, see Walmsley, 2013, for the Hogan Personality Inventory; Whetzel et al., 2010, for the Occupational Personality

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Questionnaire; Robie & Ryan, 1999, for the Personal Characteristics Inventory and NEO). The conclusions from the largest study on the topic (N > 11,000) are noteworthy:

> Any expected declines in performance at high ends of the predictor range were very small on average, and would be highly unlikely to produce scenarios in which those passing a realistic cut score would be expected to underperform those screened out due a curvilinear effect. . . . Even with slight curvilinear trends for several of the scales examined, the results suggest that curvilinearity is highly unlikely to present problems for typical uses of personality test scores in employment settings.
>
> (Walmsley, 2013, pp. ii–iii)

In general, nonlinear relationships occur when the independent or dependent variable's measures are non-normally distributed. Most personality scales used in employee selection (rather than screening) are normally distributed and thus present little concern. Nonlinearity may be more of an issue when measures are used to assess extreme ranges of the trait continuum. Benson and Campbell (2007) reported nonlinear relationships between composites of dark-side personality traits and leadership as assessed in AC dimensions and supervisory ratings. Grijalva and Newman (2015) found a mean incremental validity of .06 for nonlinearity using the Bold scale of the HDS dark-side personality measure in predicting leadership effectiveness across six samples (the authors interpreted scores on this scale to indicate narcissism). Thus, personality scales constructed to assess maladaptive ranges of personality constructs can have inverted U-shaped, nonlinear relations with performance criteria. An interesting illustration of such nonlinearity was provided by Carter et al. (2014). In their study 1, they scored a select set of conscientiousness items using the generalized graded unfolding item response theory model (see GGUM, Roberts, Donoghue, & Laughlin, 2000). For task performance and citizenship behaviors, inverted U-shaped relationships were found, whereas for CWB, a U-shaped relationship was found. In a second sample, another set of selected conscientiousness items, when scored using GGUM in especially notable nonlinear effects at scores lower than one standard deviation below the mean on conscientiousness (i.e., larger decrements to task performance and citizenship behavior, and larger increases in CWB) compared with those scoring within a standard deviation (above or below) the mean.

Classically constructed measures of maladaptive or abnormal personality designed to detect infrequently occurring psychopathological characteristics should be expected to have greater predictive value at extreme score ranges. However, most of these measures are not suitable for pre-offer employee selection and are typically employed for screening out extreme cases after a conditional job offer has been made (Dilchert et al., 2014).

Finally, although most research examining personality-criterion relationships has highlighted the predictor construct and related measurement issues, nonnormality in criterion measures can also result in nonlinearity. Future research in this area should carefully distinguish personality constructs versus their measures. Examinations in diverse samples of occupations and a broader set of criterion measures can help determine whether nonnormality and nonlinearity appreciably impact usefulness of personality measures used in employee selection. These examinations should proceed in a theory-driven manner, taking into account the distinction between test method and test content (Chan & Schmitt, 1997), the nature of the test response (Lievens, De Corte, & Westerveld, 2015), and the conceptual nature of the predictor and criterion constructs (Chan, 2005; Sackett & Lievens, 2008).

Adverse Impact

Group mean score differences on measures used in employee selection are one of the major factors determining adverse impact against protected groups, in addition to the selection ratio and score variability. Hough et al. (2001) summarized studies that examined mean score differences between Whites and various ethnic minorities, between men and women, and between older and younger people on personality traits, cognitive ability, and physical abilities. They found essentially no differences between Whites and ethnic minorities for most personality

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

variables. They also examined mean score differences between groups at the facet level of the Big Five with some unexpected findings: For some facets, mean-score differences differed from that of their respective Big Five factor (e.g., a Black–White difference of $d = -.10$ on global extraversion but a reversal, i.e., $d = .12$, on surgency/dominance, a facet of extraversion). Another meta-analysis of race and ethnic group differences on personality measures also showed modest differences between Whites and ethnic minority groups on facets of the Big Five (Foldes et al., 2008) and again established that differential patterns may exist for Big Five factors and facets (e.g., a Black–White difference of −.12 on global emotional stability measures but a reversal, i.e., .17, on self-esteem, a facet of emotional stability). Table 8 of Foldes et al. also provides a summary of scenarios based on majority/minority group selection ratios under which these observed group differences are unlikely to result in adverse impact. These two summaries highlight the usefulness of personality variables in reducing adverse impact in personnel selection systems as well as the importance of focusing on facet-level measurement.

CONCLUSIONS

We now have a better understanding of personality and its role in determining work behavior and performance. Although the FFM has provided an important framework to organize our research and systematically cumulate evidence, understanding personality and personality-criterion relationships requires more than five trait variables, including broader and narrower variables. Current research examining the taxonomic structure at the facet level of the FFM will benefit science and practice as generally accepted models emerge. Such models allow us to move beyond inventory-specific investigations of limited generalizability to cumulating results across studies and settings, thus enabling systematic investigations of moderator variables. Such models also enhance our theory building and theory testing. As our knowledge of personality-criterion relationships grows for different hierarchical levels of predictor and criterion variables, we learn how to combine predictor variables into criterion-appropriate variables that will enhance the prediction of valued outcomes in applied settings.

The prospects of better understanding the determinants of work behavior and performance are exciting. Already primary studies, meta-analyses, and second-order meta-analyses provide ample evidence that traditional self-report questionnaires of personality are among the most powerful predictors of behavior in work settings. New developments in assessment and scoring methods show promise for further improvements in measurement and prediction. Although initial optimism regarding alternate response formats (e.g., fully ipsative forced-choice scales) proved unjustified, other innovations (e.g., ideal point response methods and adaptive testing based on IRT) are promising ways to address concerns about traditional self-reports of personality on Likert-type scales. Moreover, I-O psychologists have several other assessment tools at their disposal to measure personality (e.g., biodata, interviews, other-reports, SJTs, and ACs).

In addition to improving measurement using self-report personality measures, we encourage researchers to thoroughly investigate the value of standardized other-reports in relation to occupational criteria. The few studies that have investigated their criterion-related validity suggest that other-reports may be even more valid for certain criteria than are self-report measures of personality. Other-reports can reliably capture personality variance that improves construct coverage and thus have the potential to increment criterion-related validity. More evidence for the validity of other-reports must be established and moderator variables (such as rating source) more systematically investigated before organizations will be persuaded to implement such measures more fully in employee selection.

Personality variables add significant explanatory and predictive power beyond other variables (e.g., educational credentials, cognitive ability, work experience) often assessed during employment decision making. With better understanding of the structure of personality and criterion variables and better measurement of both, personality will be more fully recognized for its very important role in affecting work behavior and performance.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

NOTES

1. The results of differential observer agreement for personality traits reviewed above can provide helpful information about which traits are best assessed with traditional employment interviews, at least with regard to issues of reliability.
2. Although some have criticized the use of these intercorrelations on the basis that they are purportedly "unrealistically low" (Morgeson et al., 2007a, p. 1035), the meta-analyses are based on data from thousands of people. Other researchers have also used these estimates to compute construct overlap between personality measures and other individual difference variables to estimate incremental validity (e.g., Judge, Heller, & Mount, 2002; Judge & Ilies, 2002; McDaniel et al., 2007).

REFERENCES

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7*, 32–39.

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

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*, 139–152.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*, 372–374.

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

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18*, 359–372.

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

Bartram, D. (2005). The Great Eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*, 1185–1203.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF: Manual for administration, scoring and interpretation*. Minneapolis, MN: University of Minnesota Press.

Benson, M. J., & Campbell, J. P. (2007). To be, or not to be, linear: An expanded representation of personality and its relationship to leadership performance. *International Journal of Selection and Assessment, 15*, 232–249.

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97*, 613–636.

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*, 410–424.

Beus, J. M., Dhanani, L. Y., & McCord, M. A. (2015). A meta-analysis of personality and workplace safety: Addressing unanswered questions. *Journal of Applied Psychology, 100*, 481–498.

Blackman, M. C., & Funder, D. C. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment, 10*, 109–116.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117*, 187–215.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*, 460–502.

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18*, 36–52.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2). Manual for administration, scoring, and interpretation* (Revised ed.). Minneapolis, MN: University of Minnesota Press.

Call, M. L., Nyberg, A. J., Ployhart, R. E., & Weekley, J. (2015). The dynamic nature of collective turnover and unit performance: The impact of time, quality, and replacements. *Academy of Management Journal, 58*, 1208–1232.

Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M-C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology, 99*, 564–586.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Chan, D. (2000). Understanding adaptation to changes in the work environment: Integrating individual difference and learning perspectives. *Research in Personnel and Human Resources Management, 18*, 1–42.

Chan, D. (2005). Current directions in personnel selection. *Current Directions in Psychological Science, 14*, 220–223.

Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology, 91*, 475–481.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 308–336). New York: Routledge.

Chan, D. (2014). *Individual adaptability to changes at work: New directions in research*. New York, NY: Routledge.

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 perceptions. *Journal of Applied Psychology, 82*, 143–159.

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

Chan, D., & Schmitt, N. (2005). Situational judgment tests. In A. Evers, O. Voskuijl, & N. Anderson (Eds.), *Handbook of selection* (pp. 219–242). Oxford, England: Blackwell.

Chang, C. H., Ferris, D. L., Johnson, R. E., Rosen, C. C., & Tan, J. A. (2012). Core self-evaluations: A review and evaluation of the literature. *Journal of Management, 38*, 81–128.

Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*, 88–106.

Chiaburu, D. S., Oh, I-S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology, 96*, 1140–1166.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83–117.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*, 267–307.

Connelly, B. S., Davies, S. E., Ones, D. S., & Birkland, A. (2008). Conscientiousness: Investigation of its facet structure through meta-analytic factor analysis. *International Journal of Psychology, 43*, 553–553.

Connelly, B. S., & Hülsheger, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of Personality, 80*, 603–631.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092–1122.

Connelly, B. S., Ones, D. S., Davies, S. E., & Birkland, A. (2014). Opening up openness: A theoretical sort following critical incidents methodology and meta-analytic investigation of the trait family measures. *Journal of Personality Assessment, 96*, 17–28.

Connelly, B. S., Davies, S. E., Ones, D. S., & Birkland, A. (2008). Conscientiousness: Investigation of its facet structure through meta-analytic factor analysis. *International Journal of Psychology, 43*, 553–553.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15*, 110–117.

Cucina, J. M., & Vasilopoulos, N. L. (2005). Nonlinear personality-performance relationships and the spurious moderating effect of traitedness. *Journal of Personality, 73*, 227–260.

DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment, 15*, 30–39.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880–896.

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

Dilchert, S., & Ones, D. S. (2011). Application of preventive strategies. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessments* (pp. 177–200). New York, NY: Oxford University Press.

Dilchert, S., Ones, D. S., & Krueger, R. F. (2014). Maladaptive personality constructs, measures, and work behaviors: Scientific background and employment practice recommendations. *Industrial and Organizational Psychology, 7*, 98–110.

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

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 48*, 209–225.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010a). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology, 3*, 465–476.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010b). *Tailored Adapative Personality Assessment System (TAPAS)*. Urbana, IL: Authors.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions (Tech. Rep. No. 1311)*. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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

Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (2014). Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PLOS One, 9*, 1–16.

Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61*, 579–616.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–670.

Gaddis, B. H., & Foster, J. L. (2015). Meta-analysis of dark side personality characteristics and critical work behaviors among leaders across the globe: Findings and implications for leadership development and executive coaching. *Applied Psychology, 64*, 25–54.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*, 463–488.

Grijalva, E., & Newman, D. A. (2015). Narcissism and counterproductive work behavior (cwb): Meta-analysis and consideration of collectivist culture, Big Five personality, and narcissism's facet structure. *Applied Psychology, 64*, 93–126.

Harold, C. M., McFarland, L. A., & Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment, 14*, 336–346.

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*, 9–24.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167–184.

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

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work. In R. Hogan & J. A. Johnson (Eds.), *Handbook of personality psychology* (pp. 849–870). San Diego, CA: Academic Press.

Hogan, J., & Roberts, B. W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior, 17*, 627–637.

Hogan, R. T. (2005a). Comments. *Human Performance, 18*, 405–407.

Hogan, R. T. (2005b). In defense of personality measurement: New wine for old whiners. *Human Performance, 18*, 331–341.

Hogan, R. T., & Hogan, J. (2009). *Hogan development survey manual*. Tulsa, OK: Hogan Assessment Systems.

Holland, J. L. (1976). Vocational preferences. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 521–570). New York, NY: Rand-McNally.

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

Hough, L. M. (1997). The millennium for personality psychology: New horizons or good old daze. *Applied Psychology: An International Review, 47*, 233–261.

Hough, L. M. (1998). Personality at work: Issues and evidence. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131–166). Mahwah, NJ: Lawrence Erlbaum.

Hough, L. M. (2010). Assessment of background and life experience: Past as prologue. In J. C. Scott & D. H. Reynolds (Eds.), *Handbook of workplace assessment: Selecting and developing organizational talent* (pp. 109–139). Hoboken, NJ: Wiley & Sons.

Hough, L. M., & Connelly, B. S. (2012). Personality measurement and use in industrial-organizational psychology. In K. F. Geisinger (Editor-in-Chief), *APA handbook on testing and assessment* and N. Kuncel (Vol.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

1 Ed.), *Test theory and testing and assessment in industrial and organizational psychology* (Vol. 1, pp. 501–531). Washington, DC: American Psychological Association.

Hough, L. M., & Dilchert, S. (2010). Personality: Its measurement and validity for personnel selection. In J. Farr & N. Tippins (Eds.), *Handbook of employee selection* (pp. 299–319). New York, NY: Routledge—Taylor & Francis Group.

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

Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of psychology. Vol. 12: Industrial and organizational psychology* (pp. 131–169). Hoboken, NJ: John Wiley & Sons.

Hough, L. M., & Johnson, J. W. (2013). Use and importance of personality variables in work settings. In I. B. Weiner (Ed.-in-Chief) & N. Schmitt & S. Highhouse (Vol. Eds.), *Handbook of psychology: Vol. 12: Industrial and organizational psychology* (pp. 211–243). New York, NY: Wiley.

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

Hough, L. M., Ones, D. S., & Viswesvaran, C. (April 1998). *Personality correlates of managerial performance constructs*. Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—Remembering the past. *Annual Review of Psychology, 51*, 631–664.

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

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

Hough, L. M., Oswald, F. L., & Ock, J. (2015). Beyond the Big Five—New directions for personality research and practice. In F. P. Morgeson (Ed.), *Annual review of organizational psychology and organizational behavior* (pp. 183–209). Palo Alto, CA: Annual Reviews.

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

Hough, L. M., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes & M. D. Mumford (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 109–145). Palo Alto, CA: CPP Books.

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

Houston, J. S., Borman, W. C., Farmer, W., & Bearden, R. M. (2005). *Development of the Enlisted Computer Adaptive Personality Scales (ENCAPS), Renamed Navy Computer Adaptive Personality Scales (NCAPS)* (Institute Report #503). Minneapolis, MN: Personnel Decisions Research Institutes.

Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology, 99*, 162–179.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and 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*, 897–913.

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

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

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*, 530–541.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*, 797–807.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the Five-Factor Model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*, 875–925.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance: A meta-analytic investigation. *Journal of Applied Psychology, 94*, 162–176.

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

Knapp, D. J., Owens, K. S., & Allen, M. T. (Eds.) (2011). *Validating future force performance measures (Army Class): In-unit performance longitudinal validation* (Tech. Rep. No. Fr-10–38). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Krause, D. E., & Thornton, G. C. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North American. *Applied Psychology: An International Review, 58*, 557–585.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*, 1879–1890.

LaHuis, D. M., Martin, N. R., & Avis, J. M. (2005). Investigating nonlinear conscientiousness—Job performance relations for clerical employees. *Human Performance, 18*, 199–212.

Le, H., Oh, I-S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology, 96*, 113–133.

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*, 1604–1627.

Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. *Journal of Applied Psychology, 94*, 1514–1535.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*, 55.

Macey, W. H., LoVerde, M. A., & Bartram, D. (2016). Evidence for a replicable leadership typology. In W. H. Macey (Chair), *Current perspectives on person-centered leadership research*. Symposium conducted at the 31st Annual Convention of the Society for Industrial and Organizational Psychology, Anaheim.

Manley, G. G., Benavidez, J., & Dunn, K. (2007). Development of a personality biodata measure to predict ethical decision making. *Journal of Managerial Psychology, 22*, 664–682.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247–256.

McCloy, R. A., Heggestad, E. D., & Reeve, C. L. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods, 8*, 222–248.

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

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

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812–821.

McGrath, R. E., Mitchell, M., Kim, B., & Hough, L. M. (2010). The validity of response bias indicators. *Psychological Bulletin, 136*, 450–470.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335–354.

McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52*, 137–148.

Moore, M., & Ones, D. S. (April 2016). *Convergent and discriminant validity of dark tetrad measures*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

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

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

Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment, 8*, 237–247.

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

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18*, 343–357.

National Research Council. (2015). *Measuring human capabilities: An agenda for basic research on the assessment of individual and group performance potential for military accession*. Committee on Measuring Human Capabilities: Performance Potential of Individuals and Collectives, Board of Behavioral, Cognitive, and Sensory

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Sciences. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
O'Boyle, E. H., Forsyth, D. R., Banks, G. R., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied Psychology, 97*, 557–579.
Oh, I-S., Wang, G., Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*, 762–773.
Ones, D. S. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.
Ones, D. S., & Dilchert, S. (February 2008). *Recent assessment center research: Dimensions, exercises, group differences, and incremental validity*. Paper presented at the annual Assessment Centre Study Group conference, Stellenbosch, South Africa.
Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995–1027.
Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609–626.
Ones, D. S., & Viswesvaran, C. (2001). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 63–92). Washington, DC: American Psychological Association.
Ones, D. S., & Viswesvaran, C. (2008). Customer service scales: Criterion-related, construct, and incremental validity evidence. In J. Deller (Ed.), *Research contributions to personality at work* (pp. 19–46). Mering, Germany: Hampp.
Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance, 18*, 389–404.
Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.
Oswald, F. L. (2010). *Practical recommendations for trait-level estimation in the Navy Computer Adaptive Personality Scales (NCAPS)*. Millington, TN: Navy Personnel Research Studies, and Technology.
Oswald, F. L., & Hough, L. M. (2008). Personality testing and I-O psychology: A productive exchange and some future directions. *Industrial and Organizational Psychology, 1*, 323–332.
Oswald, F. L., & Hough, L. M. (2011). Personality and its assessment in organizations: Theoretical and empirical developments. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol. 2. Selecting and developing members for the organization* (pp. 153–184). Washington, DC: American Psychological Association.
Oswald, F. L., Hough, L. M., & Ock, J. (2013). Theoretical and empirical structures of personality: Implications for measurement, modeling and prediction. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 11–29). New York, NY: Routledge/Taylor & Francis Group.
Oswald, F. L., & Schell, K. L. (2010). Developing and scaling personality measures: Thurstone was right—but so far, Likert was not wrong. *Industrial and Organizational Psychology, 3*, 481–484.
Oswald, F. L., Shaw, A., & Farmer, W. L. (2015). Comparing simple scoring with IRT scoring of personality measures: The Navy computer adaptive personality scales. *Applied Psychological Measurement, 39*, 144–154.
Oviedo-Garcia, M. (2007). Internal validation of a biodata extraversion scale for salespeople. *Social Behavior and Personality, 35*, 675–692.
Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*, 556–563.
Pew Research Center. (2009). *America's changing workforce: Recession turns a graying office grayer*. Washington, DC: Pew Research Center. Retrieved from http://www.pewsocialtrends.org/files/2010/10/americas-changingworkforce.pdf
Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*, 1–16.
Ployhart, R. E., Weekley, J. A., & Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal, 49*, 661–677.
Ployhart, R. E., Weekley, J. A., & Ramsey, J. (2009). The consequences of human resource stocks and flows: A longitudinal examination of unit service orientation and unit effectiveness. *Academy of Management Journal, 52*, 996–1015.
Poropat, A. E. (2009). A meta-analysis of the Five-Factor Model of personality and academic performance. *Psychological Bulletin, 135*, 322–338.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 83*, 612–624.

Pulakos, E. D., Schmitt, N., Dorsey, D. W., Arad, S., Hedge, J. W., & Borman, W. C. (2002). Predicting adaptive performance: Further tests of a model of adaptability. *Human Performance, 15*, 299–323.

Reilly, R. R., & Chao, G. R. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35*, 1–62.

Roberts, B. W., Chernyshenko, O. S., Stark, S. E., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*, 103–139.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*, 313–345.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3–32.

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

Rose, A., Timm, H., Pogson, C., Gonzales, J., Appel, E., & Kolb, N. (2010). *Developing a cybervetting strategy for law enforcement (Special Report)*. Defense Personnel Security Research Center and International Association of Chiefs of Police.

Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson, C. E., Jr., & Schmit, M. J. (2005). Personality saturation in structured interviews. *International Journal of Selection and Assessment, 4*, 261–273.

Rothstein, M. G., & Goffin, R. D. (2000). The assessment of personality constructs in industrial-organizational psychology. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 215–248). New York, NY: Kluwer Academic/Plenum.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*, 155–180.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359–391.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419–450.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*, 117–125.

Salgado, J. F., Anderson, N., & Tauriz, G. (2014). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology*, *87*, 1–38.

Salgado, J. F., & Táuriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology, 23*, 3–30.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*, 199–223.

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

Schmitt, N., & Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology, 55*, 569–587.

Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T-Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*, 979–988.

Schneider, B., & Bartram, D. (February 2015). *Aggregate personality and organizational performance*. Presentation at annual meeting of Summit Group, Livingston, TX.

Schneider, R. J., & Hough, L. M. (1995). Personality and industrial/organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 75–129). Chichester, England: 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.

Scholz, G., & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: Eine Metaanalyse. [The nomological network of the assessment center: A meta-analysis]. *Zeitschrift für Arbeits- und Organisationspsychologie, 37*, 73–85.

Sisco, H., & Reilly, R. R. (2007a). Development and validation of a biodata inventory as an alternative method to measurement of the Five Factor Model of personality. *Social Science Journal, 44*, 383–389.

Sisco, H., & Reilly, R. R. (2007b). Five Factor Biodata Inventory: Resistance to faking. *Psychological Reports, 101*, 3–17.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

Stanush, P. L. (1997). *Factors that influence the susceptibility of self-report inventories to distortion: A meta-analytic investigation*. Doctoral dissertation, College Station, TX: Texas A&M University.

Stark, S., & Chernyshenko, O. S. (2007). Adaptive testing with the multi-unidimensional pairwise preference model. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pair-wise preference items involving stimuli on different dimensions: The multi-unidimensional pair-wise-preference model. *Applied Psychological Measurement, 29*, 184–203.

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26*, 153–164.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*, 25–39.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*, 138–161.

Stokes, G. S., & Cooper, L. A. (2001). Content/construct approaches in life history form development for selection. *International Journal of Selection and Assessment, 9*, 138–151.

Stokes, G. S., & Cooper, L. A. (2004). Biodata. In J. C. Thomas (Ed.), *Comprehensive handbook of psychological assessment* (Vol. 4: Industrial and organizational assessment, pp. 243–268). Hoboken, NJ: John Wiley & Sons.

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

Tenopyr, M. L. (1994). Big Five, structural modeling, and item response theory. In G. S. Stokes, M. D. Mumford & W. A. Owens (Eds.), *Biodata handbook*: *Theory, research, and use of biographical information in selection and performance prediction* (pp. 519–533). Palo Alto, CA: Consulting Psychologists Press.

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

Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967–993.

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

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. *Human Performance, 12*, 1–29.

Thoresen, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin, 129*, 914–945.

Thornton, G. C., & Rupp, D. E. (2003). Simulations and assessment centers. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 319–344). New York, NY: Wiley & Sons.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529–554.

Tupes, E. C., & Christal, R. E. (1961/1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*, 225–251.

Twenge, J. M., Campbell, S. M., Hoffman, B. J., & Lance, C. E. (2010). Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of Management, 36*, 1117–1142.

Underhill, C. M. (2006). *Investigation of item-pair presentation and construct validity of the Navy Computer Adaptive Personality Scales (NCAPS)*. Millington, TN: Navy Personnel Research, Studies, Technology Division, Bureau of Naval Personnel.

U.S. Census Bureau (June 13 2016). Website; www.census.gov/en.html.

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., Jr., & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17*, 71–93.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536–552.

Vasilopoulos, N. L., Cucina, J. M., & Hunter, A. E. (2007). Personality and training proficiency: Issues of bandwidth-fidelity and curvilinearity. *Journal of Occupational and Organizational Psychology, 80*, 109–131.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13

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

Viswesvaran, C., & Ones, D. S. (2016). Integrity tests: A review of alternate conceptualizations and some measurement and practical issues. In U. Kumar (Ed.), *The Wiley handbook of personality assessment* (pp. 58–75). West Sussex, UK: Wiley & Sons.

Walmsley, P. T. (2013). *Investigating the presence of nonlinear personality-job performance relationships.* Doctoral dissertation, University of Minnesota. Minneapolis, MN.

Whetzel, D. L., McDaniel, M. A., Yost, A. P., & Kim, N. (2010). Linearity of personality-performance relationships: A large-scale examination. *International Journal of Selection and Assessment, 18*, 310–320.

White, L. A., Young, M. C., Hunter, A. E., & Rumsey, M. G. (2008). Lessons learned in transitioning personality measures from research to operational settings. *Industrial and Organizational Psychology, 1*, 291–295.

White, L. A., Young, M. C., & Rumsey, M. G. (2001). ABLE implementation issues and related research. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 525–558). Mahwah, NJ: Lawrence Erlbaum.

Wiernik, B. M., Dilchert, S., & Ones, D. S. (2016). Creative interests and personality: Scientific versus artistic creativity. *Zeitschrift für Arbeits- und Organisationspsychologie, 60*, 65–78.

Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61*, 275–290.

Wiggins, J. S., & Trapnell, P. D. (1997). Personality structure: The return of the Big Five. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology.* (pp. 737–765). San Diego, CA: Academic Press.

Wolters, H., Heffner, T., & Sams, M. (October 2015). *Overview and introduction of ARI's non-cognitive selection and assignment research: Enlisted personnel.* Paper presented at Briefing for Expert Classification Panel. Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Woodley, M. A., & Bell, E. (2011). Is collective intelligence (mostly) the General Factor of Personality? A comment on Woolley, Chabris, Pentland, Hashmi and Malone (2010). *Intelligence, 39*, 79–81.

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science, 24*, 420–424.

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management, 36*, 381–404.

Downloaded By: 10.2.97.136 At: 15:06 21 Sep 2023; For: 9781315690193, chapter13, 10.4324/9781315690193-13