Running Head: VALIDITY GENERALIZATION OF FC PERSONALITY
INVENTORIES

Validity Generalization of Ipsative, Quasi-ipsative, and Normative Forced-Choice

Personality Inventories across Different Occupations

Jesús F. Salgado and Gabriel Táuriz

University of Santiago de Compostela

Neil Anderson

Brunel University

Key words: forced-choice, job performance, meta-analysis, personality inventories,

quasi-ipsativity.

Abstract

A meta-analysis of three types of forced-choice personality inventories (ipsative, quasi-ipsative, and normative) across nine occupational categories is reported. Quasi-ipsative measures showed substantially larger operational validity coefficients and validity generalization than ipsative or normative measures. Results also showed that compared with the findings of previous meta-analyses, quasi-ipsative personality inventories are considerably better predictors of job performance than previously thought, but that operational validities for ipsative and normative formatted measures were notably congruent with past findings. We conclude that quasi-ipsativity matters for predicting job performance. Theoretical and practical implications of these findings for personnel selection are discussed in conclusion.

Validity Generalization of Ipsative, Quasi-ipsative, and Normative Forced-Choice Personality Inventories Across Different Occupations

Several meta-analyses conducted over the last twenty years have shown that a number personality variables, including the Big Five, integrity, self-efficacy, core self-evaluations, locus of control and positive and negative affect, among others, predict a wide range of performance outcomes, including job performance, training proficiency, counterproductive behaviors, job satisfaction, conflict, leadership, commitment, and innovative behaviors in the workplace (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Bartram, 2005; Clarke & Robertson, 2005; Dalal, 2005; Hershcovis et al., 2007; Hogan & Holland, 2003; Hough, 1992; Hülsheger, Anderson, & Salgado, 2009 Hurtz & Donovan, 2000; Judge & Bono, 2001; Judge, Bono, Ilies, & Gerhad, 2002; Judge, Bono, & Locke, 2000; Judge, Heller, & Mount, 2002; Ng, Eby, Sorensen, & Feldman, 2005; Oh, Schmidt, et al., 2011; Ones, Viswesvaran, & Schmidt, 1993; Rothmann, Meiring, & Barrick, 2002; Salgado, 1997; 1998a; 2002; 2003; Tett, Rothstein, & Jackson, 1991; Van der Walt, Meiring, Rothmann, & Barrick, 2002; Yoo & Min, 2002). Research has also demonstrated that the Five Factor Model (FFM) of personality is a robust framework for grouping the large variety of personality measures developed within the various theoretical approaches (Barrick & Mount, 1991; Hough, 1992; Hurtz & Donovan, 2000; Oh, Schmidt, et al., 2011; Rothmann, Meiring, & Barrick, 2002; Salgado, 1997; 2003; Tett et al., 1991; Yoo & Min, 2002). Across these meta-analytic efforts, conscientiousness, one of the Big Five personality dimensions, was found to be a predictor of job performance for all occupations (Barrick, Mount, & Judge, 2001) and that the compound personality variables which included conscientiousness (i.e. integrity tests, customer service scales) also showed higher

validity (Ones & Viswesvaran, 2001a, 2001b; Ones, Viswesvaran & Schmidt, 1993; 2003; Viswesvaran, Ones, & Hough, 2001; see also Hogan & Brinkmeyer, 1997).

Validity of Personality Inventories

Despite this empirical evidence, the size of the validity coefficients produced some debate among leading researchers of industrial and organizational psychology. For example, a panel of past and current editors of top I/O journals (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007a; 2007b), sustained a rather pessimistic view over the usefulness of personality measures. Morgeson et al. (2007a, 2007b) posited four criticisms: (a) the validity of personality measures is small, even when corrections for criterion reliability and range restriction are made, but they can show some incremental validity over general mental ability; (b) measures based on self-reports can be faked and this can change the rank orders of individuals and, consequently, affect the hiring decisions; (c) corrections for faking, mainly based on the scores in social desirability scales, do not seem to improve validity; (d) faking scales do not work well and fail to identify distorted responses. Counter to this view, another group of leading researchers into personality at work sustained a more optimistic view based on the fact that some of personality variables (i.e. conscientiousness, integrity) showed a validity size similar to, or even larger than, many other personnel selection procedures (i.e. assessment center, structured interviews, situational judgment tests) (Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007) They argued incremental validity over general mental ability (GMA) for predicting job performance, and that faking does not affect the criterion validity of inventories. In our view, both perspectives have a point of reason. In general, meta-analyses showed that the validity size of personality measures is not large, but simultaneously is also true that the validity size is similar or larger than other procedures and showed incremental validity. In order

to improve the criterion validity of personality inventories, some members of the panel of editors (e.g. Dipboye, Hollenbeck; see Morgeson et al. 2007) as well past and current researchers (Bartram, 2005; 2007; Brogden, 1954; Jackson, Wroblewski, & Ashton 2000; Christiansen, Burns, & Montgomery, 2005; Norman, 1964) have suggested that forced-choice personality inventories may be more useful than the traditional single-stimulus personality inventories, because the first one are more resistant to faking. Although some meta-analytic evidence (e.g. Nguyen & McDaniel, 2000) appears to support this suggestion, the amount of extant research into this supposition is not enough to be conclusive, nor have the effects of faking on the various types of forced-choice personality inventories been meta-analytically examined. This important issue is addressed in the present paper that reports a meta-analysis into the validity of different formats of forced-choice personality inventories.

Normative, Ipsative and Quasi-ipsative Personality Inventories

It is crucial to take into account that forced-choice personality inventories are not a single, discrete, or unique category of predictor measures. On the contrary, according to Hick (1970), three main types of scores can be obtained from a forced-choice inventory: *normative, ipsative, and quasi-ipsative*. The main characteristic of the first type of measures is that items representing a given bipolar scale are never paired with items representing another bipolar scale. The Myers-Briggs Type Indicator (MBTI, Myers et al., 1998) is and prototypical normative forced-choice inventory. For example, items assessing extroversion are never paired with items of emotional stability. The second type refers to those measures that totally meet Clemans' (1966) criterion of ipsativity, according to which "*any score matrix is said to be ipsative when the sum of the scores obtained over the attributes measured for each respondent is constant.*" The Occupational Personality Questionnaire (OPQ, SHL, 2006) and the Edwards Personal

Preferences Schedule (EPPS; Edwards, 1954) are two examples of ipsative inventories. The third type, quasi-ipsative scores, includes measures that do not totally meet the criterion of pure ipsativity suggested by Clemans (1966), because, for example, not all alternatives ranked by respondents are scored or the scales have different numbers of items. The Gordon Personal Profile-Inventory (GPP-I, Gordon, 1993) is an example of a quasi-ipsative inventory.

Each score type has important characteristics. For example, in the case of *normative* scoring, the scores of an individual are statistically dependent on other individuals in the population and independent of other scores of the assessed individual (e.g. scores in other attributes). Consequently, normative scoring allows the comparison of individuals or groups on each measured variable (i.e. they are inter-individual scores). In the case of *ipsative* measurement, the scores in a variable are dependent on the level of the individual in other variables which are assessed. Therefore, ipsative scores allow comparing the level of the individual across variables (i.e. they are intra-individual scores) but they are inappropriate for comparisons among individuals (Cattell & Brennan, 1994; Hicks, 1970). Additionally, ipsative scores may have some characteristics which are seen as problematic from the psychometric point of view, including negative correlations with each other measures, which limits the type of statistical analysis which is possible (Meglino & Ravlin, 1998), inflated reliability coefficients (Johnson, Wood, & Blinkhorn, 1988; Tenopyr, 1988) or less reliable scales (Bartram, 1996; Thompson et al., 1982), and inappropriate for factor analysis (Cornwell & Dunlap, 1994; Dunlap & Cornwell, 1994; Meade, 2004), On the other hand, normative scores can be transformed into ipsative scores by a simple mathematical transformation (e.g. adding a constant) but the inverse is not possible. *Quasi-ipsative* scores share some psychometric characteristics with normative and purely ipsative

scores. For example, they allow the comparison of individuals and groups but simultaneously some degree of dependence can be found among the scales of the questionnaire. Quasi-ipsative measures have not the psychometric limitations of ipsative scores (Cattell & Brennan, 1994; Hicks, 1970), and according to Hicks (1970), they may be more valid predictors than purely ipsative ones, and there may exist cases in which quasi-ipsative measures are also more valid than normative ones, because some of the advantages of the forced choice format (e.g. more resistance to faking) can be successfully exploited without the statistical limitations of ipsativity.

To the best of our knowledge, meta-analytic evidence on the validity of forced-choice personality measures is notably scant. Consequently, the main objective of this research is to extent past validity generalization research conducted mainly with normative (single-stimulus) measures and to estimate the operational validity and the true validity of the three types of FC personality inventories (i.e. normative, fully ipsative, and quasi-ipsative) across a variety of occupational categories, including a broader range of occupations. We will use the FFM as a framework for grouping the various personality inventories. Thus, this meta-analysis will be conducted on the basis of two potential moderators of the validity of personality dimensions. The score type produced by the FC personality inventories will be the first moderator. As suggested in the psychometric literature mentioned above, the three formats of FC inventories represent a remarkably different set of psychometric properties that can have important effects on the predictive validity of personality inventories. Our second moderator will be the occupational category. This is important issue as several meta-analyses found differences in the magnitude of the validity coefficients for specific jobs and occupational groups. For example Barrick and Mount (1991) examined the validity of the Big Five across occupational categories and they found that conscientiousness

showed validities ranging from .20 to .23 and 90% credibility values ranging from -.03 to .20 for professionals, police, managers, sales, and skilled workers. Salgado (1997) reported values for conscientiousness ranging from .16 to .39 and 90%CV ranging from .06 to .39 for police, managers, sales and skilled labor. Hurtz and Donovan (2000) examined the validity of conscientiousness for sales, skilled workers, customer service occupations, and managers and they reported values from .17 to .29 and 90%CV from -.19 to .17. Thus, considering the results of the meta-analyses mentioned above, it appears that there are some differences in the magnitude of the criterion-related validity of conscientiousness for different jobs and occupational groups and the same is true for the other basic personality dimensions. Therefore, beyond the minimum of validity generalization for all jobs, validity of the Big Five may be larger or smaller according to the type of occupation. This last effect was found for single-stimulus personality inventories, but it was not examined for the three types of forced-choice inventories (i.e., normative, ipsative, and quasi-ipsative inventories) and, consequently, the magnitude of the moderator effect of occupations on the validity of FC personality inventories remains unknown.

In summary, the goal of this research is to meta-analytically examine the criterion validity of the three types of forced-choice inventories to predict job performance in different jobs and occupational groups. Based on pioneering work by Hicks (1970) and Barrick and Mount (1991; Barrick, Mount, & Judge, 2001), we advance the following hypotheses:

Hypothesis 1: *Emotional Stability and Conscientiousness are valid predictors of job performance in all occupational categories.*

Hypothesis 2: *Extroversion and agreeableness are valid predictors of job performance in occupation in which interpersonal relationships are important (e.g. sales, customer service, and managerial occupations).*

Hypothesis 3: *Quasi-ipsative personality measures of conscientiousness will show larger criterion validity than normative forced-choice and purely ipsative formats.*

Method

Literature search and coding of studies

Computer-based and manual literature searches were conducted in order to identify published and unpublished studies carried out up until and including September 2011. To cover the literature on forced-choice personality measures as comprehensively as possible, and to prevent any bias in the inclusion of studies, we adopted a series of search strategies. First, we identified the most popular forced-choice and ipsative questionnaires. They included, for example, the Occupational Personality Questionnaire (OPQ), Edwards Personal Preferences Schedule (EPPS), Myers-Briggs Type Inventory (MBTI), Description en Cinq Dimensions (D5D), Survey of Interpersonal Values Inventory (SIV), and Gordon Personal Profile -Inventory (GPPI). Second, PsycInfo, Social Sciences Citation Index, and ABI/Inform databases were searched to identify studies on the relationship of forced-choice measures and organizational criteria. Several keywords were used for the computer-based literature search (e.g., ipsative, forced-choice, ipsativity, job performance), as well as the acronyms of the most popular forced-choice personality inventories. Third, electronic searches using Google were carried out systematically in order to look for articles, unpublished manuscripts, and master and doctoral dissertations not included in the most common databases. Fourth, a manual article-by-article search was carried out in a number of top-tier journals (e.g.,

Educational and Psychological Measurement, International Journal of Selection and Assessment, Journal of Applied Psychology, Journal of Occupational and Organizational Psychology, Personnel Psychology). Fifth, the reference sections of several published meta-analyses (e.g., Barrick & Mount, 1991; Bartram, 2005; 2007; Dudley et al., 2006; Hurtz & Donovan, 2000; Salgado, 1997, 2003; Tett et al., 1991) were reviewed to identify articles not covered in our computer-based search. Sixth, we contacted a number of researchers and asked for both published articles and unpublished papers on the topic in order to avoid or reduce file drawer effects and publications bias. Seventh, the technical manual of the most popular forced-choice personality questionnaires (e.g. EPPS, GPP-I, MBTI, OPQ) was examined in order to find validity coefficients. By means of these search strategies, a preliminary database of over 115 documents (i.e. articles, manuals, technical reports, unpublished papers, dissertations, and so on) was established for further inspection. There were 26 studies excluded from the total pool for various reasons: (a) some studies reported only the significant correlations, (b) a number of studies only reported multiple correlation results; (c) several of them did not report correlations or enough information to calculate the effect size (d) several studies reported findings for the same data set. As a result of these points, the present meta-analysis was conducted with 89 independent samples and a total sample size of 18,593 individuals.

The next step was to classify the scales from the inventories into the Big Five personality dimensions. This task was relatively easy because a number of studies used a Big Five measure or estimates of the Big Five (e.g. Bartram, 2007; McDaniel et al., 2004; Nye et al. 1996; Robertson et al. 2000; SHL, 2006; Warr, Bartram & Brown, 2005). With the rest of the studies, we used the following method. First, an exhaustive description of the Big Five was written and given to the coders. Then, a list and the

definition of the personality scales from each questionnaire were provided for each

coder with instructions to assign each scale to the most appropriate factor. Furthermore,

some studies reporting factor analyses of the questionnaires were also used as a basis

for the decision (e.g., Matthews et al., 1990; McCrae & Costa, 1990; Piedmont, Costa &

McCrae, 1992; SHL, 2006). Finally, we also checked the coding list used by Ones

(1993; Hough & Ones, 2001) and Salgado (2003). If the coders agreed on a dimension,

the scale was coded in that dimension. The disagreements (less than 10%) were solved

by a discussion until the coders agreed on a dimension. All the scales were assigned to a

single dimension. Two researchers served as coders, working independently to code

every study.

For each study, the following information was recorded, if available: (a) sample

characteristics, such as gender, age, education, and so forth, (b) occupation and related

information, (c) personality measures used, (d) criterion type, (e) reliability of

personality measures, (f) criterion reliability, (g) range restriction value or data for

calculating this value, (h) statistics concerning the relation between personality

measures and criterion, (i) correlation among the personality measures when more than

one was used. This complies with the APA guidelines on meta-analysis reporting

standards (APA, 2009). When a study contained conceptual replications (i.e., two or

more measures of the same construct were used in the same sample), linear composites

with unit weights for the components were formed. Linear composites provide more

construct valid estimates than the use of the average correlation. Nunnally (1978, pp.

166–168) and Hunter and Schmidt (1990, pp. 457–463; see also Hunter & Schmidt,

2004) provided a formula for the correlation of variables with composites. As

demonstrated by Warr, Bartram and Brown (2005), the average validity corrected with

Mosier's formula for composite reliability produces very accurate estimates when the appropriate inter-correlations are used.

An important difference between this meta-analysis and previous meta-analyses of the relationship between personality and job performance is that we used the type of FC measure for grouping the validity coefficients. This is especially relevant because different degrees of ipsativity can result in different validity levels (Clemans, 1966; Hicks, 1970; Radcliffe, 1963). As the number of studies allowed us to consider each one as a separate entity, using Hicks' (1970) taxonomy, we classified the personality inventories in three categories: purely ipsative, quasi-ipsative, and normative forced-choice questionnaires. In order to classify the questionnaires, each one was inspected in terms of the scoring method and the format of items. Furthermore, we used the technical manuals of the questionnaires when available, and other articles which were not relevant for this meta-analysis because they did not include validity data but did include relevant information about the questionnaire characteristics and scoring system. The initial agreement level of the coders was 95% and the disagreements were solved by a discussion until the coders agreed on a questionnaire category. We classified the questionnaires as being purely ipsative if the sum of the scores obtained over the scales measures was constant. Examples of purely ipsative questionnaires in our database are Edwards Personality Preferences Schedule (EPPS; Edwards, 1954), the Occupational Personality Questionnaire (OPQ, SHL, 2006), and the Description en Cinq Dimension (D5D, Rolland & Mogenet, 2001). We classified forced-choice questionnaires as quasi-ipsative if any of the following alternatives applied: (a) individuals only partially order the items, rather than ordering them completely; (b) scales have different numbers of items; (c) not all alternatives ranked by respondents are scored; (d) scales are scored differentially for individuals with different characteristics or involve different normative

transformations on the basis of respondent characteristics; (e) scored alternatives are differentially weighted; and (f) the questionnaire has normative sections. Examples of quasi-ipsative questionnaires in our database are the Gordon Personal Profile-Inventory (Gordon, 1993), Self-Description Inventory (Ghiselli, 1954), ESQ (Jackson, 2003), Assessment Individual Motivation (AIM; Knapp, Heggestad & Young, 2004; White, 2002). Finally, we classified a questionnaire as a forced-choice normative one if it yielded scores which posses the empirical properties of absolute measures. This is the case, for example, of the inventories in which items representing different degrees of a personality dimension are never paired with items representing another personality dimension. The Myers-Briggs Type Indicator and the Need of Achievement (Fineman, 1975) are representative examples of normative forced-choice questionnaires.

After this coding process, we examined in detail the cell and sample sizes for the resultant matrix of personality inventories by type, by occupational group categories. Thus, we examined both the number of studies (the K sample size) and the total number of individual responses in the primary studies (the N sample size) in order to verify these were large enough to perform meta-analytic procedures. In a few cases, the number of studies was small in some categories. However, as Viswesvaran, Schmidt, and Ones (2002) have persuasively argued, the smallest meta-analysis includes a larger sample size than the single studies included in it, and this sample size would be considered by many researchers larger than the samples used in published single empirical studies (see Salgado, 1998b, for an account of the average sample size of the studies published in I/O psychology).

Meta-analytic Procedure

After the studies were collated and their characteristics recorded, the following step was to apply the psychometric meta-analysis of Hunter and Schmidt (1990, 2004).

Psychometric meta-analysis estimates how much of the observed variance of findings across studies is due to artifactual errors. The artifacts considered here were sampling error, criterion and predictor reliability, and indirect range restriction in personality scores. To correct the observed validity for these last three artifacts, the most common strategy was to develop specific distributions for each of them. Some of these artifacts reduce the correlations below their operational value (e.g., criterion reliability and range restriction), and all of them produce artifactual variability in the observed validity (Carretta & Ree, 2000, 2001). In our analysis, we corrected the observed mean validity for criterion reliability and range restriction in the predictor for obtaining the operational validity which is of interest for personnel selection decisions, and we corrected the operational validity for predictor reliability in order to obtain the true validity which is of interest for modeling the theoretical relationship between personality and performance.

Predictor Reliability

The reliability of the personality dimensions was estimated from the coefficients reported in the studies included in the meta-analysis. As in previous meta-analyses, we used internal consistency coefficients as estimates of reliability. We developed an empirical distribution of Cronbach's coefficients for the Big Five. The predictor reliability estimates were used to eliminate artifactual variability in the standard deviation of ρ and for correcting the operational validity in order to obtain the theoretical value of the validity of the Big Five personality dimensions. For each personality dimension, a reliability distribution was estimated. The average reliability was .73, .75, .81, .81, and .74, for emotional stability, extraversion, openness, agreeableness, and conscientiousness, respectively. We also examined the question of whether the three types of FC scoring systems showed different levels of reliability, but

they proved to be very similar, and, therefore, we used the average reliabilities in the

meta-analytic calculations. Table 1 presents a summary of these artifact distributions. In

Table 1, we also included the reliabilities found by Ones and Viswesvaran (1999) in

their meta-analysis of the Big Five reliability. As can be seen, our estimates are very

similar to the estimates found by Ones and Viswesvaran (1999) and are also very

similar to the estimates used in previous meta-analyses (e.g. Barrick & Mount, 1991;

Hurtz & Donovan, 2000; Salgado, 1997; 2002; 2003). Consequently, the empirical

evidence suggests that when the Big Five personality dimensions are measured with

forced-choice personality inventories, the reliability is practically the same as that

obtained with normative questionnaires. This is an important finding as some

researchers had suggested that forced-choice questionnaires would show a different

reliability size than normative questionnaires, with some authors claiming that in these

cases the reliability is overestimated (inflated) and others claiming that the reliability is

underestimated (lower than it really is) (Bartram, 1996; Johnson et al., 1988; Saville &

Willson, 1991; Tenopyr, 1988).

Insert Table 1 about here

Criterion Reliability

The studies included in our data base used four types of measures of job

performance: (a) job performance ratings, (b) productivity data (e.g. sales), (c) training

proficiency, and (d) overall job performance. From the literature (e.g. Schmidt &

Zimmerman, 2004), it is well known that each type of job performance measure shows

a different degree of reliability: However, as the number of studies does not allow the

calculation of separate meta-analyses for a triple combination (FC type measure x

Occupational group x Job Performance Measure Type) and not all studies provided

information regarding the job performance reliability, we develop an empirical

distribution for the job performance criterion for each FC type measure x occupational group combination. In order to do this, we collapsed all studies in a single job performance measure and we estimated the average reliability for the total. If the study used job performance ratings, the coefficient of interest was inter-rater reliability when a meta-analysis of random effects is done (Hunter, 1986; Sackett, 2003; Schmidt and Hunter, 1996) because if this type of reliability is used in the correction for attenuation, it will correct most of the unsystematic errors in supervisor ratings (Hunter & Hirsh, 1987), although not all researchers agree with this point of view (e.g. Murphy & De Shon, 2000). We found 11 studies reporting inter-rater coefficients of job performance ratings (see Table 2). The average coefficient was .52 (SD=.05). This average coefficient is the same as that found by Viswesvaran, Ones, and Schmidt (1996; see also Salgado & Moscoso, 1996) in their meta-analysis of the inter-rater reliability of job performance ratings, and independently by Salgado et al. (2003) in European criterion-validity studies of cognitive tests. For the studies using training success, we found 2 studies reporting reliability (see Table 2). The average coefficient was .80 (SD=.09). This figure is the same as that used by Hunter (1986; see also Hunter & Hunter, 1984) and many meta-analyses of the Big Five (e.g. Barrick and Mount, 1991). For the studies using objective productivity measures, seven studies reported reliability coefficients and the average reliability was .83. This value is similar to the one found by Schmidt and Zimmerman (2004) and it is the same at that found by Schmitt et al. (1984). Pooling together the reliability coefficients and weighting for the number of studies using each criterion type, we calculated the average reliability coefficient for job performance. The average coefficient was .61 (SD=.13). Next, we estimated the criterion reliability to be used for each FC type measure x occupational group combination. In order to do this, we used the number of studies and the type of performance measures to calculate the

average reliability for within each combination. These estimates appear in Table 2.

Depending on the occupational category, reliability ranges from .52 to .82 for ipsative

personality measures, from .52 to .68 for quasi-ipsative personality measures, and it was

.59 for normative personality measures. As a whole, the average reliability estimates of

job performance were .62, .58 and .59 for the ipsative, quasi-ipsative, and normative

personality measures. Therefore, the differences in job performance reliability are

minimal and without practical effects.

<div align="center">Insert Table 2 about here</div>

Range Restriction Distributions

The distributions for range restriction were based on the following three

strategies: (a) some range restriction values were obtained from the studies that reported

both restricted and unrestricted standard deviation data, (b) another group of range

restriction values were obtained using the population values reported in the manual of

the various inventories, and (c) a third group of range restriction values was obtained

using the reported selection ratio. To use the reported selection ratio, we applied the

formula derived by Schmidt, Hunter, and Urry (1976). This triple strategy produced a

large number of range restriction estimates, and we grouped them according to the

personality dimensions. The average range restrictions (u) were .87 for emotional

stability, .90 for extroversion, .92 for openness to experience, .90 for agreeableness, and

.88 for conscientiousness. These range restriction values are very similar to the figures

used in previous meta-analysis (e.g. Barrick & Mount, 1991; Hurtz & Donovan, 2000;

Salgado, 1997, 2003) and they are in accordance with the observation by Schmidt et al.

(2008) that the range restriction of personality measures is remarkably smaller than the

range restriction found in the validity studies of cognitive ability tests. A summary of

these distributions appears in Table 3

Insert Table 3 about here

Meta-analytic Method

We used the psychometric meta-analysis methods developed by Hunter and

Schmidt (1990; 2004) and implemented in a software program by Schmidt and Le

(2004). This software includes some recent advances to correct for indirect range

restriction (IRR). According to Hunter, Schmidt and Le (2006), IRR is the most

common case of range restriction and it is present in all concurrent validity studies and

in practically all predictive validity studies conducted in personnel selection (some

studies in military selection research are the exception). Consequently, correction for

direct range restriction (DRR), rather than IRR, results in a substantial underestimation

of the operational and true validity coefficients and overestimation of the true

variability. Until very recently, all the meta-analyses conducted on the relationships

between personality variables and job performance corrected for DRR, since no formula

was available to correct for IRR in psychometric meta-analysis. Hunter and Schmidt

(2004; Hunter, Schmidt, & Le, 2006) developed a new formula.

In a series of studies, Schmidt and his colleagues have demonstrated the effects

of indirect range restriction correction on the validity of general mental ability, the

validity of specific cognitive abilities, and on the validity of the Big Five (Schmidt, Oh,

& Le, 2006; Schmidt, Shaffer, & Oh, 2008), as well the accuracy of the new procedures

via Montecarlo simulations (Le & Schmidt, 2006). The software program includes both

the new refinements and older advances and refinements, such as the use of mean r

instead of study-observed rs in the formula for sampling error variance and a new

nonlinear range-correction procedure (Law, Schmidt, & Hunter, 1994a, 1994b; Schmidt

et al., 1993).We are interested in the relationship between the Big Five and

performance, both as theoretical constructs and as operational predictors, and, therefore,

we will report both the operational validity and the true correlation. In summary, we will correct the observed validity for criterion reliability and IRR for obtaining the operational validity, and for predictor unreliability for obtaining the true correlation. The observed variance will be corrected for by artifactual errors: sampling error, criterion and predictor reliability, and IRR.

Results

We carried out three groups of meta-analyses, one for each ipsative, quasi-ipsative, and normative FC personality measures. The results of the meta-analyses of personality-occupation combinations appear in Tables 4, 5, and 6. The validity coefficients were pooled across personality dimensions and occupations. In the three tables, from left to right, the first four columns represent the number of independent coefficients (K), the total simple size (N), the observed validity weighted by the study simple size (r_w), and the standard deviation of the observed validity (SD_r). The next three columns show the observed validity corrected for the measurement error in criterion and indirect range restriction in predictor (operational validity, rc), the full-corrected correlation (true validity, ρ), and the standard deviation of ρ. Finally, the last two columns are the percentage of variance explained for by the four artifactual errors (sampling error, predictor reliability, criterion reliability, and indirect range restriction in) and the 90% credibility value (90%CV). We reported the operational validity and ρ in this table and in the next tables because they serve for different objectives. Operational validity is the coefficient to be use for predicting the criterion in applied settings (e.g. making decisions about employees). True validity is the theoretical correlation between the personality dimension and the criterion and, therefore, this coefficient is used for modeling the theoretical relationship between dependent and

independent variables. Although, we are interested in both coefficients, we will concentrate in ρ in our comments.

We were able to create nine occupational groups, using the Dictionnary of Occupational Titles (D.O.T.) (USES, 1991) as a reference. However, not all were available for the three types of FC measures, and, furthermore, the number of studies and the sample size is small in some cases (e.g., clerical and supervisor) occupations. In the following three subsections, we comment on the results separately for ipsative, quasi-ipsative and normative FC personality measures, respectively.

Meta-analysis of Personality - Occupation Combinations for Ipsative FC Measures.

Table 4 reports the results for the combination of the Big Five-occupation for the ipsative personality inventories. We found studies for five occupational groups, including customer service, managerial, police, salesmen, and supervisor. As hypothesized, conscientiousness showed to be a valid predictor for the five occupational groups, with validity coefficients ranging from .13 for managerial and police occupations to .32 for salesmen occupations. We found evidence of validity generalization for three out five occupations, as the 90% CV was positive and nonzero. Therefore, hypothesis 1 was supported regarding to conscientiousness.

Insert Table 4 about here

Extroversion showed a small but relevant validity for managerial occupations and salesmen as was hypothesized. Furthermore, in both cases, extroversion showed validity generalization. Neither emotional stability nor openness to experience showed to be relevant predictor for these five occupational groups when the Big Five are assessed with ipsative personality measures. An unexpected but very relevant result was found for agreeableness for predicting job performance of salespeople. We found a

validity coefficient of -25 and the 90%CV was -.07 which means that there is validity generalization in this case. Based in this result, less agreeable individuals will show better job performance in sales occupations.

Therefore, as a whole, Hypotheses 1 and 2 were partially supported for the ipsative measures of the Big Five personality dimensions.

Meta-analysis of Personality - Occupation Combinations for Quasi-ipsative FC Measures.

The results for the combination of the Big Five and occupations for the quasi-ipsative measures appear in Table 5. We found validity studies for eight occupational categories, including clerical, customer service, health-care, managerial, military, salesmen, skilled, and supervisor occupations.

Insert Table 5 about here

In the case of quasi-ipsative measures, conscientiousness was a valid predictor for the eight occupations, with validity coefficients ranging from .17 for managerial occupations to .64 for skilled labor occupations. All the 90%CV's were positive and very different from zero, which is evidence supporting validity generalization for predicting job performance. On average, the validity of conscientiousness was .40, which is a coefficient remarkable larger than the coefficients found in previous meta-analyses. The validity for clerical, customer service, health-care, salesmen, and skilled labor occupations were especially relevant as they were larger than .30 in all cases.

Emotional stability was a valid predictor of job performance for military, salesmen, skilled and supervisor occupations, and shoed validity generalization for these four occupations. Consequently, Hypothesis 1 was partially supported regarding to emotional stability as it did not predict job performance in the case of health-care and

managerial occupations. Extroversion was a valid predictor for managerial and skilled labor occupations and it showed validity generalization in both cases. Extroversion also showed a small but relevant validity coefficient for clerical and health-care occupations, but there was not evidence supporting validity generalization in these cases, as the 90%CV were negative. Openness predicted negatively job performance for clerical occupations, and positively for health-care, managerial, military, and salesmen occupations, i.e. for all the occupational categories with studies on the validity of openness. Furthermore, the validity size was relevant in three out four occupations, with validity coefficients ranging from .23 to -.44. Finally, agreeableness was a valid predictor of job performance for clerical, health-care, and skilled labor occupations, with coefficients ranging from .31 to .45.

Therefore, Hypothesis 1 was fully supported with regard to conscientiousness and partially supported regarding emotional stability. Hypothesis 2 was partially supported as extroversion and agreeableness predicted performance for clerical, health-care, managerial, salesmen, and skilled labor occupation. Finally, Hypothesis 3 was fully supported because, as a whole, quasi-ipsative measures of conscientiousness showed larger validity than ipsative measures.

Meta-analysis of Personality - Occupation Combinations for Normative FC Measures.

In the case of normative FC measures of the Big Five, we only found a sufficient number of validity studies for conducting the meta-analysis for managerial occupations. Table 6 presents the results for this meta-analysis.

Insert Table 6 about here

As can be seen, we found studies for extroversion, openness, agreeableness, and conscientiousness but there were not studies for emotional stability. The results showed

that the validity was small for the four personality dimensions, although for conscientiousness was a larger. In the four cases, the 90%CV allows the conclusion that there was validity generalization. Therefore, at least partially, the results supported hypothesis 1 and 2. With regard to hypothesis 3, the results are not conclusive as the validity coefficient was the same for both quasi-ipsative and normative FC inventories (=.17), although, the 90%CV was larger for the quasi-ipsative inventories (.17 vs. .07).

Discussion

This meta-analytic effort has some unique contributions that can be noted. The first contribution was to meta-analytically examine the validity of the three types of measures derived from FC personality inventories for predicting job performance in nine occupational categories. To the knowledge of the authors, this is the first time that any meta-analytic effort has been reported that examines the important question of the comparative validity of different types of forced-choice personality measures. Our results show unequivocally that the quasi-ipsative forced-choice measures of conscientiousness are more valid than any other type of questionnaire, including normative (single-stimulus), normative forced-choice and ipsative questionnaires. Their validity size is almost twice than the validity of single-stimulus questionnaires and equal or larger than the validity of assessment center, structured interview, situational judgment tests, and personality composite (i.e. integrity, COPS). Across occupations, quasi-ipsative inventories of personality predict job performance substantially better than ipsative or normative FC inventories. This held true across the five personality dimensions of the FFM. Furthermore, comparing the results of the quasi-ipsative FC inventories found in the current meta-analysis with the findings of previous meta-analyses, which were conducted largely with single stimulus (SS) personality

inventories (based on the references, over 90% were SS), quasi-ipsative are very much valid. For example, Barrick and Mount (1991), Salgado (1997), and Hurtz and Donovan (2000) found a validity of .22 for conscientiousness pooled across occupations (cumulated N= 30,066), while the validity found here was .40 pooled across occupations (N=8648). Similar results were found for extroversion and openness, and in a lower extend for emotional stability and agreeableness. Therefore, as a whole, our most important finding and our first conclusion is that quasi-ipsativity matters.

Related to this point, ipsative measures of conscientiousness showed an average validity of .16 across occupations which is smaller that the validity of SS inventories. Therefore, Hicks' hypothesis that normative inventories would show larger validity than ipsative was supported in our study, if we compare .16 against .22 although the magnitude of the difference is relatively small. This is the second contribution of this study. Our data do not allow strong conclusions regarding whether ipsative measures are better, equal, or worse predictors of job performance than normative FC measures. We cannot be conclusive about the superiority of normative FC personality inventories over single-stimulus personality inventories for predicting job performance, as we only found validity studies for managerial occupations.

A third contribution was to show that openness to experience can be a more important predictor of job performance than was thought. Previous meta-analyses generally agreed that openness to experience was a predictor of training success but it merits for predicting job performance overall and across occupations was irrelevant (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1997; 2003). The present findings suggest that, if openness to experience is assessed by quasi-ipsative personality inventories, its validity is large and important for a number of occupations. There is not a clear explanation for this differential effect of openness to experience at present. We

conjecture two possible explanations. The first explanation is based on the relationship between openness to experience and GMA. Previous research has shown that the correlation between these two variables is around .30 when openness is assessed with single-stimulus personality measures. It could be possible that quasi-ipsative measures of openness may be stronger loaded of GMA than single-stimulus measures of openness to experience and, consequently, an indirect effect of GMA explains the criterion validity. The second explanation is based on the observed relationship between openness and conscientiousness. The Big Five personality dimensions are orthogonal when the factor space is examined but when the raw scores are used for estimating the scores of the Big Five, it is frequently seen that the Big Five show moderate to large correlations among them. For example, the technical manual of the NEO-PI-R (Costa & McCrae, 1992) reports observed correlations ranging from .40 to .60. Specifically, the correlation between openness and conscientiousness was .50. At present, the correlation between openness and conscientiousness is unknown, when these two personality dimensions are assessed by quasi-ipsative inventories. But, if the correlation in this last case were similar to the one found for single-stimulus measures, then the validity of openness could be partially due to conscientiousness. Future research should be undertaken into these and other explanatory alternatives.

Study limitations and strengths

This article has some limitations which must be acknowledged. A first limitation is that the number of studies (K) is small for some combinations of personality measures occupations for some of the occupational groups covered. Therefore, some results should be interpreted with caution in these cases, because they are subjected to change as the number of studies and the sample size increase. Note also that some K

sizes were as high as 21 studies, but in addition to this the N sample sizes drawn from the original primary studies varied from just under 350 up to 4,401. Viswesvaran, Schmidt, and Ones (2002) suggested that meta-analytic procedures can be carried out on K sample sizes that are not large, and as we noted earlier in this paper, Practically, all the classical meta-analyses on the relationship between personality and performance contains a number of cases in which the number of studies is small (e.g. 2 or 3) and the total sample is even smaller than the size of our smallest sample. For example, Barrick and Mount (1991, Table 4) include a couple of cases with two correlations and 121 individuals, and the meta-analysis by Tett et al (1991, Table 5) included samples of 164 and 280 for Type A and agreeableness, respectively. Similarly, Hough (1992, Table 5) reported 11 cases with samples ranging from 116 to 329, Salgado (1997, Table 3) included two cases with n=324, and Hurtz and Donovan (2000, Table 3) conducted meta-analyses for training performance with two studies. It should also be considered that despite the small K sample size of some combinations, in fact in all cases they are larger than the sample sizes typically found in validity studies published in the top journals in industrial and organizational psychology, such as the Journal of Applied Psychology, Personnel Psychology, and Journal of Occupational and Organizational Psychology (Shen et al., 2011; Salgado, 1998b). Future primary studies are needed to address this concern and to estimate (and to reduce) the amount of second-order sampling error potentially present is our study.

A second limitation of this research is that the vast majority of the studies included in the database used personality inventories developed out the FFM. Of course, this is not a problem for grouping the validity coefficients as was demonstrated many times, but the construct validity of non-FFM personality inventories is smaller than the construct validity of inventories based on the FFM when the objective is to assess the

Big Five personality dimensions. The loss of construct validity produces to relevant

effects, one on the magnitude of the criterion validity, and another on the variability of

the validity coefficients (Hunter & Schmidt, 2004). Salgado (2003) found that there was

a remarkable difference in the magnitude of the criterion validity for FFM and non-FFM

personality inventories for predicting job performance. Therefore, our findings could be

lower estimates of the operational and true validity of the Big Five.

However, counter to these limitations, it should be noted equally that the present

meta-analysis displays several points of strength. To the knowledge of the authors, this

is the first such quantitative summary to examine the three types of FC personality

inventories across different occupational categories, and to be organized along the lines

of the FFM. Our findings have produced some notable patterns of operational validity

and suggest that personality inventories based upon a quasi-ipsative scoring procedure

generally exhibit substantially higher operational validities than those based upon a

fully ipsative or normative format. Although, as acknowledged above, the K-cell sample

sizes in some of our measure-occupation categories were not large, the total N's of the

samples from the primary studies upon which they are based, are sufficiently large to

permit meta-analytic combination.

It is important to remark too that the true validity of conscientiousness assessed

by quasi-ipsative inventories is larger than any other single psychological construct,

with the exception of cognitive abilities, and better predictor of performance than the

majority of the procedures used in personnel selection, as only work sample tests,

behavioral interviews and job knowledge test produced larger validity (see Salgado,

Ones, & Viswesvaran, 2001; Schmidt & Hunter, 1998).

Many of the strengths of this research were mentioned before, but we wish to

note that we used a very comprehensive primary study database and that this is the first

meta-analysis which differentiates the various types of forced-choice measures in combination with the personality dimensions, thus giving important clarification on the role of personality dimensions at work.

Implications for Future Research

The present findings have important implications for research and practice. A number of research questions were produced by the results. The comparison between normative forced-choice and ipsative questionnaires carried out in the present meta-analysis suggests that the latter are similarly valid, or even slightly more so, although the number of studies is relatively small for making a strong conclusion. Comparing the results of the previous meta-analyses, mainly conducted with normative single-stimulus questionnaires, with those of the normative a ipsative forced-choice questionnaires, it can not be concluded that the single-stimulus questionnaires are superior in validity, as suggested by Hicks (1970).

For example, whether quasi-ipsative personality measures based on the FFM would produce larger validity than current quasi-ipsative measures. Due to the fact that, currently, the number of personality inventories developed under the FMM is practically very small, the response to the previous question requires that new quasi-ipsative inventories are created. This could be good for both research and development points of views. A second research question is to examine the relationship between GMA and the Big Five if they are assessed with quasi-ipsative measures. Several researchers have suggested that the cognitive strategies to respond single-stimulus and forced-choice personality measures are different, being more cognitively demanding the second one. The first type of measures would require from the individual to make one

decision, agree or disagree with the item content. The second type would requires a number of decision, basically of two classes, one is agree or disagree with the content of each item and the other class of decisions is to rank the item in comparison with the rest of items included among the alternatives to choice. Therefore, based on the volume of cognitive demands, quasi-ipsative measures could be more correlated with GMA than single-stimulus measures. Recent evidence suggests some differences (e.g. Vasilopoulus et al., 2006). A third research question is about the relationships among the Big Five. This is a relevant question in order to estimate the validity of a composite of personality measures (e.g. composites of emotional stability, conscientiousness, and agreeableness) and to compare its validity with the validity of other personality compounds frequently used for selection decisions.

Implications for Personnel Selection Practice and Personality Inventory Design

Our findings also have important implications for the practice of personnel selection (e.g. Anderson, 2005) and in particular the design and formatting of item responses on personality inventories used for employee selection. Based on the validity coefficients found in this meta-analysis, two important practical implications have to be noted. First, personality measures, especially conscientiousness, should be used in hiring processes for three reasons: they are valid, they are chipper than many other selection procedures, and they do not produce discriminatory bias. The last two reasons were not discussed but the first one was systematically challenged in the last years. The empirical evidence provided by this meta-analysis, as was mentioned above, showed that quasi-ipsative personality measures are substantially more valid than either normative or fully ipsative designs, and moreover, that their validity is similar to or even more valid than other personnel selection procedures demonstrating high levels of operational validity (Schmidt & Hunter, 1996). A second implication is that quasi-

ipsative personality measures should be considered good alternatives to replace single-stimulus inventories.

Concluding comments

An unequivocal conclusion of this meta-analysis is that quasi-ipsativity in personality inventory format and design matters for predicting job performance. Quasi-ipsative measures of personality, mainly of conscientiousness, were valid predictors across all occupational categories and their operational and true validity coefficients. In general, they were twice as high as the validity of ipsative and normative personality measures. Furthermore, quasi-ipsativity appears to have positive effects on the predictive validity of other personality dimensions, such as openness, that until now showed low efficiency for predicting job performance. Future research should be devoted to examine other aspects of ipsativity not examined in this meta-analysis and new studies should be done for other occupations. Also, we recommend that new personality inventories should be developed with the quasi-ipsative technology.

References

Articles with an asterisk are included in the meta-analysis

*Adkins, C. L. & Naumann, S.E. (2001). Situational constraints on the achievement-performance relationship: a service sector study. *Journal of Organizational Behavior, 22*, 453-465.

Anderson, N. (2005). Relationships between practice and research in personnel selection: Does the left hand know what the right is doing?. In A. Evers, N. Anderson, & O. Voskuijl, (Eds.). *The Blackwell Handbook of Personnel Selection* (pp.1-24). Oxford: Blackwell.

*Antler, L., Zaretsky, H.H., & Ritter, W. (1967). The practical validity of the Gordon Personal Profile among United States and foreign medical residents. *Journal of Social Psychology, 72*, 257-263.

APA (2009). *Publication manual of the American Psychological Association*, sixth edition. Washington, DC: APA

*Azen, S.P., Snibbe, H.M., Montgomery, H.R., Fabricatore, J., & Earle, H.H. (2002). Predictors of resignation and performance of law enforcement officers. *American Journal of Community Psychology, 2*, 79-86

*Balch, D.E. (1977). *Personality trait differences between successful and non-successful police recruits at a typical police academy and veteran police officers*. Upublished Doctoral Dissertation, U.S. International University.

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. & Judge, T. (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. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology, 69*, 25-39.

*Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185-1203. DOI: 10.1037/0021-9010.90.6.1185

*Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment, 15*, 263-272.

Brogden (1954). A rationale for minimizing distortion in personality questionnaire keys. *Psychometrika, 19*, 151-148.

*Brown, A. & Bartram, D. (2009). Doing less but getting more: improving forced-choice measures with IRT. *Paper presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology*, New Orleans, April 2-4.

*Buel, W.D. & Bachner, V.M. (1961). The assessment of creativity in a research setting. *Journal of Applied Psychology, 45*, 353-358.

Carretta, T. & Ree, M.J. (2000). General and specific cognitive and psychomotor abilities in personnel selection: The prediction of training and job performance. *International Journal of Selection and Assessment, 8*, 227-236.

Carreta, T. & Ree, M.J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9*, 325-335.

Cattell, R.B. & Brennan, J. (1994). Finding personality structure when ipsative measurements are the unavoidable basis of the variables. *American Journal of Psychology, 107*, 261-274.

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

Clarke, S. & Robertson, I.T. (2005). A meta-analytic review of the Big Five personality factors and accidents involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78*, 355-36.

Clemans, W.V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14*.

*Clevenger, J., Pereira, G.M., Wiechman, D., Schmitt, N., & Schmidt-Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410-417. DOI: 10.1037/0021-9010.86.3.410

*Cocolas, G.H., Sleath, B., & Hanson-Drivers, C. (1997). Use of the Gordon Personal Profile-Inventory of pharmacists and pharmacy students. *American Journal of Pharmaceutical Education, 61*, 257-265.

*Conway, J.M. (2000). Managerial performance development constructs and personality correlates. *Human Performance, 13*, 23-46.

Cornwell, J.M. & Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Wilson (1991. *Journal of Occupational and Organizational Psychology, 67,* 89-100.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology, 90,* 1241–1255. DOI: 10.1037/0021-9010.90.6.1241

*Dodd, W.E., Wollowick, H.B., & McNamara, W.J. (1970). Task difficulty as a moderator of long-range prediction. *Journal of Applied Psychology, 34*, 265-270.

Dudley, N.M., Orvis, K.A., Liebicki, 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. DOI: 10.1037/0021-9010.91.1.40

Dunlap, W.P. & Cornwell, J.M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research, 29*, 115-126.

Edwards, A.L. (1957). *Manual for the personal preference schedule.* Psychological Corp. New York.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166. DOI: 10.1037/0021-9010.84.2.155.

*Fine, S., & Dover, S. (2005). Cognitive ability, personality, and low fidelity simulation measures in predicting training performance among customer service representatives. *Applied HRM Research*, *10*, 103-106.

*Fineman, S. (1975). The work preference questionnaire: a measure of managerial need for achievement. *Journal of Occupational Psychology, 48*, 11-32.

*Francis-Smythe, J., Tinline, G., & Allender, C. (2002). *Identifying high potential police officers and role characteristics.* Paper presentation at the Division of Occupational Psychology Conference, British Psychological Society, Bournemouth.

*Furnham, A. (1994). The Validity of the SHL Customer Service Questionnaire (CSQ). *International Journal of Selection and Assessment*, *2*, 157-165.

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

Ghiselli, E. E. (1954). The Forced-Choice Technique in Self-Description. *Personnel Psychology*, *7*, 201-208. doi:10.1111/j.1744-6570.1954.tb01593.x

*Goffin, R.D., Jan, I. & Skinner, E. (2011). Forced-choice and conventional personality assessment: each may have unique value in pre-employment testing. *Personality and Individual Differences, 5*, 840-844.

*Gordon, L.V. (1993). *Gordon Personal Profile-Inventory. Manual 1993 Revision.* Pearson-TalentLens.

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

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

*Guller, M. (2003). *Predicting performance of law enforcement personnel using the candidate and officer personnel survey and other psychological measures.* Unpublished Doctoral Dissertation. Seton Hall University.

*Helton, K.T. & Street, D.R. (1992). *The five-factor model and naval aviation candidates.* Naval Aeroespace Medical Research Laboratory, Naval Air Station, Pensacola, FL.

Hershcovis, M.S., Turner, N., Barling, J., Arnold, K.A., Dupré, K.E., Inness, M., LeBlanc, M.M., Sivanathan, N. (2007). Predicting workplace aggression: A meta-analysis. *Journal of Applied Psychology*, 92, 228-238. DOI: 10.1037/0021-9010.92.2.228

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. DOI: 10.1037/0021-9010.88.1.100

Hogan, J. & Brinkmeyer, K. (1997). Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology, 50*, 587-599. DOI: 10.1111/j.1744-6570.1997.tb00704.x

Hogan, R. & Hogan, J. (2001). Assessing leadership: a view of the dark side. *International Journal of Selection and Assessment, 9*, 40-51.

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

Hough, L.M. & Ones, D.S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work and organizational psychology. In N.R. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds). *Handbook of industrial, work, and organizational psychology. Vol 1: Personnel Psychology (pp. 233-276).* London, UK: Sage.

*Hughes, J.L. & Dood, W.E. (1961). Validity versus stereotype: predicting sales performance by ipsative scoring of a personality test. *Personnel Psychology, 15*, 343-355.

*Hughes, G. L., & Prien, E. P. (1986). An evaluation of alternate scoring methods for the mixed standard. *Personnel Psychology*, *39*, 839-847.

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology,* 94: 1128-1145.

Hunter, J.E. (1983). T*est validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (GATB).* Washington, DC: U.S. Department of Labor, Employment Service.

Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Intelligence, 29*, 340-362.

Hunter, J.E. & Hirsh, H.R. (1987). Applications of meta-analysis. *International Review of Industrial and Organizational Psychology, Vol 2*. (pp321-357). Chichester, UK: Wiley.

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

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

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

Hunter, J.E., Schmidt, F.L, & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594-612. DOI: 10.1037/0021-9010.91.3.594

Hurtz, G.M. & Donovan, J.J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869-879. DOI: 10.1037/0021-9010.85.6.869

*Illiescu, D., Ilie, A., &Aspas, D.G.(2011). Examining the criterion-related validity of the Employee Screening Questionnaire: A three-sample investigation *International Journal of selection and Assessment, 19*, 222-228.

Jackson D.N. (2002). *Employee screening questionnaire manual*. Port Huron, MI: Sigma Assessment Systems, ESQ

Jackson, D.J. & Alvin, D.F. (1980). The factor analysis of ipsative measures. *Sociological Methods & Research, 9*, 218-238.

*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.

Johnson, C.E., Wood, R., & Blinkhorn, S.F. (1988). Spriouser and spriouser: the use of ipsative personality tests. *Journal of Occupational Psychology, 61*, 153-162.

Judge, T & Bono, J.E. (2001). Relations of core self-evaluations traits –self-esteem, generalized self-efficacy, locus of control, and emotional stability- with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*, 80-92. DOI: 10.1037/0021-9010.86.1.80

Judge, T., Bono, J.E., Ilies, R. & Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780. DOI: 10.1037/0021-9010.87.6.765

Judge, T.A., Bono, J.E. & Locke, E.A. (2000). Personality and job satisfaction: The mediating role of job characteristics. *Journal of Applied Psychology, 85*, 237-249. DOI: 10.1037/0021-9010.85.2.237

Judge, T., Heller, D. & Mount, M.K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87*, 530-541. DOI: 10.1037/0021-9010.87.3.530

Knapp, D., Heggestad, E., & Young, M. (2004). *Understanding and improving the Assessment of Individual Motivation (AIM) in the Army's GED Plus Program (Study Note 2004-03).* Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

*Knauft, E.B. (1955). Test validity over seventeen-year period. *Journal of Applied Psychology, 39*, 382-383.

*Kriedt, P. H., & Dawson, R. I. (1961). Response set and the prediction of clerical job performance. *Journal of Applied Psychology*, *45*, 175-178.

*Krug, R.D. (1959). Over- and under-achievement and the Edwards Personal Preference Schedule. *Journal of Applied Psychology, 43*, 133-136.

*Kusch, R.I., Deller, J., & Albrecht, A.G. (2008). *Predicting expatriate job performance: using the normative NEO-PI-R or the ipsative OPQ32i?* Paper presented at the 29th International Congress of Psychology, Berlin, Germany, 20th -25th July.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994a). Nonlinearity of range corrections in meta-analysis: Test of an improved procedure. *Journal of Applied Psychology, 79,* 425-438.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994b). A test of two refinements in procedures for meta-analysis. *Journal of Applied Psychology, 79,* 978-986.

Le, H. & Schmidt, F.L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new analytic procedure. *Psychological Methods, 11*, 416-438.

*Lievens, F., Harris, M. M., Keer, E. V., & Bisqueret, C. (2003). Predicting cross-cultural training performance: the validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, *88*, 476-489. DOI: 10.1037/0021-9010.88.3.476

*Maher, H. (1959). Follow-up on the validity of a forced-choice study activity questionnaire in another setting. *Journal of Applied Psychology, 43*, 293-295.

Matthews, G., Stanton, N., Graham, N. C. & Brimelow, C. (1990).A factor analysis of the scales of the occupational personality questionnaire. *Personality and Individual Differences*, 11, (6), 591-596. DOI: 10.1016/0191-8869(90)90042-P

Meade, A.W. (2004) Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*, 531-552.

*McDaniel, M.A., Yost, A.P., Ludwick, M.H., Hense, R.L., & Hartman, N.S. (2004). *Incremental validity of a situational judgment test.* Paper presented at the 19th

Annual Conference of the Society for Industrial and Organizational Psychology,

Chicago, April 2004.

McCrae, R.R. & Costa, P.T. (1990). *Personality in adulthood*. New York:

Guilford Press.

Meglino, B.M. & Ravlin, E.C. (1998). Individuals values in organizations:

concepts, controversies, and research. *Journal of Management, 24*, 351-389.

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K.,

& Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel contexts.

Personnel Psychology, 60, 683-729.

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K.,

& Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations

in the use of personality tests for personnel selection. *Personnel Psychology. 60.* 1029-

1049.

Murphy, K.R. & De Shon, R. (2000). Inter-rater correlations do not estimate the

reliability of job performance ratings. *Personnel Psychology, 53*, 873-900.

Myers, I. B., McCaulley, M. H., Quenk, N., & Hammer, A. (1998). *MBTI

Handbook: A Guide to the development and use of the Myers-Briggs Type Indicator.*

Palo Alto, CA: Consulting Psychologists Press, 3rd edition. ISBN 0891061304

*Nelson, C. A. (2008). *Job type as a moderator of the relationship between

situational judgment and personality*. Unpublished doctoral dissertation. Capella

University.

*Neuman, G.A. (1991). Autonomous work group selection. *Journal of Business

and Psychology, 6*, 283-291.

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

Ng, T.W.H., Eby, L.T., Sorensen, K.L., & Feldman, D.C. (2005) predictors of objective and subjective career success: a meta-analysis. *Personnel Psychology, 58*, 367-408.

Nguyen, N.T. & McDaniel, M.A. (2000). *Faking and forced-choice scales in applicant screening: A meta-analysis.* Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA, April, 2000.

Norman, W.T. (1964). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology, 47*, 225-241.

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

*Nyfield, G., Gibbons, P. J., Baron, H., & Robertson, I. (1995). The cross-cultural validity of management assessment methods. *Surrey, UK: Saville & Holdsworth*.

Oh, I-S, Schmidt, F.L., Mount, M.K., Le, H., Guay, R.P., Takahashi, K., Chuang, A., Naito, J., Imashiro, S., Mochinushi,Y, Ranming, N., & Law, K.S. (2011). *The five-factor model of personality and performance in East Asia.* Paper presented at the Annual Conference of the Society for Industrial and Organizational psychology.

Ones, D.S. (1993). *The construct of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Ones, D.S., Dilchert, S., Viswesvaran, C., & Judge, T. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, doi/10.1111/j.1744-6570

Ones, D.S. & Viswesvaran, C. (1999). *Job-specific applicant pools and National Norms for personality scales: implications for range restriction correction in validation research*. Paper presented at the 14th Annual Conference of the SIOP, Atlanta, Georgia.

Ones, D.S. & Viswesvaran, C. (2001a). Personality at Work: criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts y R.T. Hogan (Eds), *Applied personality psychology* (pp. 63-92). Washington, DC: American Psychological Association.

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

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.

Ones, D.S., Viswesvaran, C., & Schmidt, F.L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17*, 19-38.

*Perkins, A. M., & Corr, P. J. (2005). Can worriers be winners? The association between worrying and job performance. *Personality and Individual Differences*, *38*, 25-31.

Piedmont, R.L., Costa, P.T., & McCrae, R.R. (1992). An assessment of the Edwards Personal Preference Schedule from the perspective of the Five-Factor Model. *Journal of Personality Assessment, 58*, 67-78.

Radcliffe, J. (1963). Some properties of ipsative score matrices and their relevance for some current interest tests. *Australian Journal of Psychology*, *15*(1), 1–11.

*Robertson, I., Gibbons, P., Baron, H., MacIver, R., &Nyfield, G. (1999). Understanding management performance. *British Journal of Management*, *10*, 5-12.

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

*Rolland, J.P. & Mogenet, J.L. (2001). *Système de description en cinq dimensions (D5D). Manuel réservé aux psychologues*. Paris: Les Editions du Centre de Psychologie Appliquée.

Rothman, S., Meiring, D., & Barrick, M. R. (2002). *Predicting job performance using personality measures in South Africa. In D.S. Ones (Chair) Personality at work in a cross-cultural context.* Paper presented at the 17th Annual Conference of the Society of Industrial and Organizational Psychology, Dallas, Texas.

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

*Sackett, P.R., Gruys, M.L. & Ellington, J.E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology, 83*, 545-556. DOI: 10.1037/0021-9010.83.4.545

Sackett, P.R. (2003). The status of validity generalization research: key issues in drawing inferences from cumulative research studies. In K.R. Murphy (Ed.). *Validity generalization: A critical review (pp. 91-114).* Mahaw, NJ: Lawrence Erlbaum.

*Salgado, J.F. (1991). Validez del Preference and Perception Inventory (PAPI) en una compañía de servicios financieros [*Validity of the Preferences and Perception*

Inventory (PAPI) in a financial services company]. Unpublished technical report, Department of Social Psychology, University of Santiago de Compostela.

Salgado, J.F. (1997). The five-factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*, 30-43. DOI: 10.1037/0021-9010.82.1.30

Salgado, J.F. (1998a). The Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance, 11*, 271-288.

Salgado, J.F. (1998b). Sample size in validity studies of personnel selection. *Journal of Occupational and Organizational Psychology, 71*, 161-164.

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

Salgado, J.F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76*, 323-346.

Salgado, J.F. & Moscoso, S. (1996). Meta-analysis of the interrater reliability of job performance ratings in validity studies of personnel selection. *Perceptual and Motor Skilss, 83*, 1195-1201.

Salgado, J.F., Viswesvaran, C. & Ones, D.S. (2001). Predictors used for personnel selection. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds). *Handbook of Industrial, Work, & Organizational Psychology. Vol 1. (pp. 165-199).* London, UK: Sage.

*Saville, P., Sik, G., Nyfield, G., Hackston, J., & Maclver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in

the measurement of job competencies across time and in separate organisations. *Applied Psychology*, *45*, 243-262.

Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology, 64,* 219-238.

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

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262-274.

Schmidt, F.L., Hunter, J.E. & Urry, (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*, 473-485

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology*, *78*(1), 3-12. doi:10.1037/0021-9010.78.1.3

Schmidt, F.L. & Le, H. (2004). *Software for the Hunter-Schmidt meta-analysis methods*. Iowa City: Department of Management and Organizations, University of Iowa.

Schmidt, F.L., Oh, I-S, & Le, H. (2006). Increasing the accuracy of corrections for range restriction: implications for selection procedure validities and other research practices. *Personnel Psy chology, 59*, 281-305.

Schmidt, F.L. Shaffer, J.A., & Oh, I-S (2008). Increased accuracy for range restriction corrections: implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*, 827-868.

Schmidt, F. L., & Zimmerman, R. D. (2004). A Counterintuitive Hypothesis About Employment Interview Validity and Some Supporting Evidence. *Journal of Applied Psychology*, *89*(3), 553-561. doi:10.1037/0021-9010.89.3.553

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.

*Schippmann, J.S. & Prien, E.P. (1989). An assessment of the contributions of general menatl ability and personality characteristics to management success. *Journal of Business and Psychology, 3*, 423-

*Schwab, D. P. & Packard, G.L. (1973). Response distortion on the Gordon Personal Inventory and the Gordon Personal Profile in a selection context: some implications for predicting employee tenure. *Journal of Applied Psychology, 58*, 372-374.

*Schwartz, S.L. & Gekoki, N. (1960). The supervisory inventory: A forced-choice measure of human relations attitude and technique. *Journal of Applied Psychology, 44*, 233-236.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology*, *96*(5), 1055-1064.

*SHL (2006). *OPQ32 Manual and User's Guide*. Surrey, United Kingdom: SHL.

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

*Small, R.J. & Rosenberg, L.J. (1977). Determining job performance in the industrial sales force. *Industrial Marketing Management, 6*, 99-102.

*Sommerfeld, D. (1997). *Maintenance worker selection: High validity and low adverse impact.Michigan Municipal League*. Employment Testing Consortium Project.

Tenopyr, M. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology, 73*, 749-751.

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*(4), 967–993.

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

Thompson, B., Levitov, J. E., & Miederhoff, P. A. (1982). Validity of the Rokeach value survey. *Educational and Psychological Measurement*, *42*(3), 899-905

Van der Walt, H.S., Meiring, D., Rothmann, S., & Barrick, M.R. (2002). *Meta-analysis of the relationship between personality measurements and job performance in South Africa*. Paper presented at the Conference of the South-Africa Society for Industrial and Organizational Psychology, Pretoria, June 13-14.

Viswesvaran, C., Ones, D.S., & Hough, L.M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment, 9*, 277-289.

Viswesvaran, C., Ones, D., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574. DOI: 10.1037/0021-9010.81.4.557

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (2002). The moderating influence of job performance dimensions on convergence of supervisor and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87*, 345-354

*Warr, P., Bartram, D., & Martin, T. (2005). Personality and sales performance: situational variation and interactions between traits. *International Journal of Selection and Assessment*, *13*, 87-91.

Warr, P., Bartram, D. & Brown, A. (2005). Big Five validity: aggregation method matters. *Journal of Occupational and Organizational Psychology, 78*, 377-386. DOI: 10.1348/096317905X53868

*White, L. A. (2002). A quasi-ipsative temperament measure for assessing future leaders. In *44th Annual Conference of the International Military Testing Association* (pag. 169).

*Willingham, W.W. & Ambler, R.K. (1963). The relation of the Gordon Personal Inventory to several external criteria. *Journal of Consulting Psychology, 27*, 460

*Willingham, W.W., Nelson, P., & O'Connor, W. (1958). A note on the behavioral validity of the Gordon Personal Profile. *Journal of Consulting Psychology, 22*, 378.

*Witt, L.A. & Jones, J.W. (1999). *Very particular people quit first.* Paper presented at the Annual Conference of the Society for Industrial and organizational Psychology, Atlanta, GA.

*Wollowick, H.B. & McNamara, W.J. (1969). Relationship of the components of an assessment center to management success. *Journal of Applied Psyhcology, 53*, 348-352.

Yoo, T-Y., Min, B-M. (2002). *A meta-analysis of the Big Five and performance in Korea.* Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Table 1. Distributions of Reliability of the Big Five Personality Dimensions.

Personality Dimension	Mean	SD
Emotional Stability	.73	.09
Extraversion	.75	.13
Openness to Experience	.81	.12
Agreeableness	.80	.08
Conscientiousness	.72	.11

Table 2. Distributions of Performance Reliability for the Occupational Categories.

Occupation	Ipsative		Quasi-Ipsative		Normative FC	
	Mean	SD	Mean	SD	Mean	SD
Clerical	------	------	.52	.05	------	------
Customer service	.56	.12	.60	.11	------	------
Health-Care	------	------	.50	.08	------	------
Manager	.57	.12	.65	.12	.59	.11
Military	------	------	.68	.11	-----	------
Police	.61	.15	------	------	------	------
Salesmen	.82	.08	.60	.15	-----	------
Skilled	------	-----	.60	.11	------	------
Supervisor	.52	.05	.52	.05	------	------

Table 3. Distributions of Range Restriction for the Big Five Dimensions.

Personality Dimension	Mean	SD
Emotional Stability	.87	.16
Extraversion	.90	.14
Openness to Experience	.92	.13
Agreeableness	.90	.14
Conscientiousness	.88	.17

Table 4. Results of the Meta-analyses of the Validity of Ipsative Measures of Big Five for Occupational Groups.

Personality Dimension	K	N	r_w	SD_r	r_c	ρ_i	$SD\rho$	%VE	90%CV
					Customer Service				
Emotional Stability	2	344	.04	.03	.06	.07	.00	100	.07
Extraversion	2	344	.05	.02	.08	.09	.00	100	.09
Openness to Experience	2	344	.05	.04	.07	.08	.00	100	.08
Conscientiousness	3	551	.12	.06	.19	.21	.00	100	.21
					Managerial				
Emotional Stability	21	4222	.04	.08	.07	.08	.09	68	-.03
Extraversion	20	4154	.09	.09	.14	.15	.08	68	.04
Openness to Experience	20	4154	.05	.08	.07	.08	.07	72	-.01
Agreeableness	20	4154	.00	.09	.00	.00	.09	60	-.11
Conscientiousness	21	4401	.08	.10	.13	.14	.12	54	-.01
					Police				
Emotional Stability	3	700	.04	.05	.07	.08	.00	100	.08
Extraversion	3	700	.02	.07	.03	.04	.02	96	.01
Openness to Experience	3	700	.08	.14	.11	.12	.20	22	-.13
Agreeableness	2	475	.03	.02	.04	.05	.00	100	.05
Conscientiousness	3	700	.07	.12	.12	.13	.11	58	.00
					Salesmen				
Emotional Stability	4	401	-.02	.03	-.03	-.03	.00	100	-.03
Extraversion	5	594	.08	.02	.10	.11	.00	100	.11
Openness to Experience	4	401	-.02	.15	-.03	-.03	.15	44	.15
Agreeableness	5	594	-.15	.13	-.20	-.21	.12	52	-.05
Conscientiousness	5	665	.18	.06	.24	.27	.00	100	.24

(Table 4 continued)

Table 4 continued

					Supervisor				
Emotional Stability	4	423	.01	.04	.01	.01	.10	100	.01
Extraversion	4	423	-.04	.08	-.07	-.07	.00	100	-.07
Openness	4	423	.01	.06	.02	.02	.00	100	.02
Agreeableness	4	423	-.06	.09	-.09	-.10	.00	100	-.10
Conscientiousness	4	423	.08	.09	.14	.16	.00	100	.16

K=number of independent samples; N=total sample size; r_w=observed validity; SD_r= standard deviation of observed validity; r_c=operational validity. ρ_i=validity corrected for criterion reliability and indirect range restriction in predictor; $SD\rho$= standard deviation of ρ_i; %VE= percentage of variance accounted for by artifactual errors; 90CV= 90% credibility value based on the operational validity (r_c).

Table 5. Results of the Meta-analyses of the Validity of Quasi-Ipsative Measures of Big Five for Occupational Groups.

Personality Dimension	K	N	r_w	SD_r	r_c	ρ_i	$SD\rho$	%VE	90%CV
					Clerical				
Extroversion	2	288	.14	.21	.23	.25	.33	17	-.18
Openness to Experience	2	288	-.27	.24	-.41	-.44	.34	14	-.01
Agreeableness	2	288	.25	.01	.40	.44	.00	100	.44
Conscientiousness	3	357	.16	.09	.28	.31	.00	100	.31
					Customer Service				
Conscientiousness	2	398	.29	.05	.45	.50	.00	100	.50
					Health				
Emotional Stability	4	803	.05	.20	.08	.09	.35	13	-.36
Extraversion	3	773	.10	.11	.16	.18	.16	36	-.03
Openness to Experience	3	773	.17	.04	.27	.29	.00	100	.29
Agreeableness	3	773	.17	.07	.28	.31	.00	100	.31
Conscientiousness	3	773	.21	.09	.35	.40	.05	88	.33
					Managerial				
Emotional Stability	10	1528	-.01	.10	-.02	-.02	.10	69	.10
Extraversion	11	1685	.21	.05	.31	.34	.00	100	.34
Openness to Experience	8	1085	.21	.06	.29	.32	.00	100	.32
Agreeableness	8	1085	-.04	.14	-.06	-.07	.17	38	.15
Conscientiousness	9	1132	.10	.09	.15	.17	.00	100	.17
					Military				
Emotional Stability	9	2799	.09	.08	.14	.16	.08	60	.05
Extraversion	10	3007	.07	.10	.10	.11	.12	37	-.05
Openness to Experience	6	748	.15	.05	.21	.23	.00	100	.23
Agreeableness	8	1760	.05	.12	.07	.07	.07	32	-.11
Conscientiousness	10	3007	.12	.10	.19	.21	.12	46	.05

(Table 5 continued)

Table 5 Continued.

	K	N	r_w	SD_r	ρ_i	r_c	$SD\rho$	%VE	90CV
Salesmen									
Emotional Stability	4	472	.11	.12	.18	.21	.13	65	.04
Extraversion	4	472	.05	.20	.07	.08	.29	23	-.28
Openness to Experience	2	236	.11	.09	.16	.17	.00	100	.17
Agreeableness	2	236	.07	.14	.11	.12	.17	44	-.10
Conscientiousness	4	472	.22	.04	.35	.39	.00	100	.39
Skilled									
Emotional Stability	5	1350	.30	.19	.46	.51	.26	21	.18
Extraversion	4	1018	-.17	.13	-.25	-.28	.17	32	-.06
Agreeableness	2	796	.28	.03	.42	.45	.00	100	.45
Conscientiousness	8	2338	.43	.09	.64	.71	.00	100	.71
Supervisor									
Emotional Stability	3	171	.37	.07	.61	.68	.00	100	.68
Conscientiousness	3	171	.09	.10	.16	.18	.00	100	.18

K=number of independent samples; N=total sample size; r_w=observed validity; SD_r= standard deviation of observed validity; r_c=operational validity. ρ_i=validity corrected for criterion reliability and indirect range restriction in predictor; $SD\rho$= standard deviation of ρ_i; %VE= percentage of variance accounted for by artifactual errors; 90CV= 90% credibility value based on the operational validity (r_c).

Table 6. Results of the Meta-analyses of the Validity of Normative Forced-Choice Measures of Big Five for Occupational Groups.

Personality Dimension	K	N	r_w	SD_r	r_c	ρ_i	$SD\rho$	%VE	90%CV
					Managerial				
Extroversion	3	1937	-.05	.03	-.08	-.09	.00	100	-.09
Openness to Experience	3	1937	.06	.02	.09	.10	.00	100	.10
Agreeableness	3	1937	.06	.01	.09	.09	.00	100	.09
Conscientiousness	6	2578	.09	.07	.15	.17	.08	60	.07

K=number of independent samples; N=total sample size; r_w=observed validity; SD_r= standard deviation of observed validity; r_c=operational validity. ρ_i=validity corrected for criterion reliability and indirect range restriction in predictor; $SD\rho$= standard deviation of ρ_i; %VE= percentage of variance accounted for by artifactual errors; 90CV= 90% credibility value based on the operational validity (r_c).