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.

The research reported in this article was partially supported by Grant SEJ-2008-3070 from the Ministry of Science and Innovation (Spain) to Jesús F. Salgado.

Correspondence concerning this article should be addressed to Jesús F. Salgado, Department of Social Psychology, University of Santiago de Compostela, Campus Vida, 15782 Santiago de Compostela. E-mail: jesus.salgado@usc.es.
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 of 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; Henschcovis et al., 2007; Hogan & Holland, 2003; Hough, 1992; Hülsheger, Anderson, & Salgado, 2009 Hurtz & Donovan, 2000; Judge & Bono, 2001; Judge, Bono, Ilies, & Gerhard, 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 $\rho$ 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.

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.
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 $r_s$ 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, $r_c$), the full-corrected correlation (true validity, $\rho$), and the standard deviation of $\rho$. 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 $\rho$ 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 $\rho$ in our comments.

We were able to create nine occupational groups, using the Dictionary 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 showed 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 overally 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 aipsative 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 require a number of decisions, 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


Psychological Corp. New York.


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


*Salgado, J.F. (1991). Validez del Preference and Perception Inventory (PAPI) en una compañía de servicios financieros [Validity of the Preferences and Perception Inventory En una compañía de servicios financieros].*
Inventory (PAPI) in a financial services company]. Unpublished technical report, Department of Social Psychology, University of Santiago de Compostela.


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


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

<table>
<thead>
<tr>
<th>Personality Dimension</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>.73</td>
<td>.09</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.75</td>
<td>.13</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>.81</td>
<td>.12</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.80</td>
<td>.08</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.72</td>
<td>.11</td>
</tr>
</tbody>
</table>
Table 2. Distributions of Performance Reliability for the Occupational Categories.

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Ipsative Mean</th>
<th>Ipsative SD</th>
<th>Quasi-Ipsative Mean</th>
<th>Quasi-Ipsative SD</th>
<th>Normative FC Mean</th>
<th>Normative FC SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clerical</td>
<td>.52</td>
<td>.05</td>
<td>.52</td>
<td>.05</td>
<td>.52</td>
<td>.05</td>
</tr>
<tr>
<td>Customer service</td>
<td>.56</td>
<td>.12</td>
<td>.60</td>
<td>.11</td>
<td>.59</td>
<td>.11</td>
</tr>
<tr>
<td>Health-Care</td>
<td>.50</td>
<td>.08</td>
<td>.50</td>
<td>.08</td>
<td>.50</td>
<td>.08</td>
</tr>
<tr>
<td>Manager</td>
<td>.57</td>
<td>.12</td>
<td>.65</td>
<td>.12</td>
<td>.59</td>
<td>.11</td>
</tr>
<tr>
<td>Military</td>
<td>.68</td>
<td>.11</td>
<td>.68</td>
<td>.11</td>
<td>.68</td>
<td>.11</td>
</tr>
<tr>
<td>Police</td>
<td>.61</td>
<td>.15</td>
<td>.61</td>
<td>.15</td>
<td>.61</td>
<td>.15</td>
</tr>
<tr>
<td>Salesmen</td>
<td>.82</td>
<td>.08</td>
<td>.60</td>
<td>.15</td>
<td>.60</td>
<td>.15</td>
</tr>
<tr>
<td>Skilled</td>
<td>.60</td>
<td>.11</td>
<td>.60</td>
<td>.11</td>
<td>.60</td>
<td>.11</td>
</tr>
<tr>
<td>Supervisor</td>
<td>.52</td>
<td>.05</td>
<td>.52</td>
<td>.05</td>
<td>.52</td>
<td>.05</td>
</tr>
</tbody>
</table>
Table 3. Distributions of Range Restriction for the Big Five Dimensions.

<table>
<thead>
<tr>
<th>Personality Dimension</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>0.87</td>
<td>0.16</td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.90</td>
<td>0.14</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>0.92</td>
<td>0.13</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.90</td>
<td>0.14</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.88</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Table 4. Results of the Meta-analyses of the Validity of Ipsative Measures of Big Five for Occupational Groups.

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

(Table 4 continued)
Table 4 continued

<table>
<thead>
<tr>
<th></th>
<th>Supervisor</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Stability</td>
<td>4</td>
<td>423</td>
<td>.01</td>
<td>.04</td>
<td>.01</td>
<td>.04</td>
<td>100</td>
</tr>
<tr>
<td>Extraversion</td>
<td>4</td>
<td>423</td>
<td>-.04</td>
<td>.08</td>
<td>-.07</td>
<td>-.07</td>
<td>.00</td>
</tr>
<tr>
<td>Openness</td>
<td>4</td>
<td>423</td>
<td>.01</td>
<td>.06</td>
<td>.02</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>4</td>
<td>423</td>
<td>-.06</td>
<td>.09</td>
<td>-.09</td>
<td>-.10</td>
<td>.00</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>4</td>
<td>423</td>
<td>.08</td>
<td>.09</td>
<td>.14</td>
<td>.16</td>
<td>.00</td>
</tr>
</tbody>
</table>

K=number of independent samples; N=total sample size; $r_w$=observed validity; SD$_r$= standard deviation of observed validity; $r_c$=operational validity. $\rho_i$=validity corrected for criterion reliability and indirect range restriction in predictor; SD$\rho$= standard deviation of $\rho$; %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.

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

(Table 5 continued)
Table 5 Continued.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salesmen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K=number of independent samples; N=total sample size; $r_w=$observed validity; SD$<em>r=$ standard deviation of observed validity; $r_c=$operational validity. $\rho_i=$validity corrected for criterion reliability and indirect range restriction in predictor; SD$</em>\rho=$ standard deviation of $\rho_i$; %VE= percentage of variance accounted for by artifactual errors; 90CV= 90% credibility value based on the operational validity ($r_c$).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>4</td>
<td>472</td>
<td>.11</td>
<td>.12</td>
<td>.18</td>
<td>.21</td>
<td>.13</td>
<td>65</td>
<td>.04</td>
</tr>
<tr>
<td>Extraversion</td>
<td>4</td>
<td>472</td>
<td>.05</td>
<td>.20</td>
<td>.07</td>
<td>.08</td>
<td>.29</td>
<td>23</td>
<td>-.28</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>2</td>
<td>236</td>
<td>.11</td>
<td>.09</td>
<td>.16</td>
<td>.17</td>
<td>.00</td>
<td>100</td>
<td>.17</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>2</td>
<td>236</td>
<td>.07</td>
<td>.14</td>
<td>.11</td>
<td>.12</td>
<td>.17</td>
<td>44</td>
<td>-.10</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>4</td>
<td>472</td>
<td>.22</td>
<td>.04</td>
<td>.35</td>
<td>.39</td>
<td>.00</td>
<td>100</td>
<td>.39</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>5</td>
<td>1350</td>
<td>.30</td>
<td>.19</td>
<td>.46</td>
<td>.51</td>
<td>.26</td>
<td>21</td>
<td>.18</td>
</tr>
<tr>
<td>Extraversion</td>
<td>4</td>
<td>1018</td>
<td>-.17</td>
<td>.13</td>
<td>-.25</td>
<td>-.28</td>
<td>.17</td>
<td>32</td>
<td>-.06</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>2</td>
<td>796</td>
<td>.28</td>
<td>.03</td>
<td>.42</td>
<td>.45</td>
<td>.00</td>
<td>100</td>
<td>.45</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>8</td>
<td>2338</td>
<td>.43</td>
<td>.09</td>
<td>.64</td>
<td>.71</td>
<td>.00</td>
<td>100</td>
<td>.71</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>3</td>
<td>171</td>
<td>.37</td>
<td>.07</td>
<td>.61</td>
<td>.68</td>
<td>.00</td>
<td>100</td>
<td>.68</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>3</td>
<td>171</td>
<td>.09</td>
<td>.10</td>
<td>.16</td>
<td>.18</td>
<td>.00</td>
<td>100</td>
<td>.18</td>
</tr>
</tbody>
</table>
Table 6. Results of the Meta-analyses of the Validity of Normative Forced-Choice Measures of Big Five for Occupational Groups.

<table>
<thead>
<tr>
<th>Personality Dimension</th>
<th>K</th>
<th>N</th>
<th>$r_w$</th>
<th>SD$_r$</th>
<th>$r_c$</th>
<th>$\rho_i$</th>
<th>SD$\rho$</th>
<th>%VE</th>
<th>90%CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managerial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extroversion</td>
<td>3</td>
<td>1937</td>
<td>-.05</td>
<td>.03</td>
<td>-.08</td>
<td>-.09</td>
<td>.00</td>
<td>100</td>
<td>-.09</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>3</td>
<td>1937</td>
<td>.06</td>
<td>.02</td>
<td>.09</td>
<td>.10</td>
<td>.00</td>
<td>100</td>
<td>.10</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>3</td>
<td>1937</td>
<td>.06</td>
<td>.01</td>
<td>.09</td>
<td>.09</td>
<td>.00</td>
<td>100</td>
<td>.09</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>6</td>
<td>2578</td>
<td>.09</td>
<td>.07</td>
<td>.15</td>
<td>.17</td>
<td>.08</td>
<td>60</td>
<td>.07</td>
</tr>
</tbody>
</table>

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