

Borrower: BAL
System: OCLC

Call #:

Location: online

Lending String:
COO,*COU,AZS,COF,HLS,EZC,BTL

Patron: Mitchell, Tom

Journal Title: European journal of work and organizational psychology.

Volume: 23 **Issue:** 1
Month/Year: 2014 **Pages:** 3-30

Article Author: Jesús F. Salgado* & Gabriel Táuriza

Article Title: The Five-Factor Model, forced-choice personality inventories and performance; A comprehensive meta-analysis of academic and occupational validity stud

Imprint: Hove, East Sussex, UK ; Psychology Press

ILL Number: 112575875



Shipping Address:

University of Baltimore
Langsdale Library ILL
1420 Maryland Ave
Baltimore Maryland 21201-5706 United States

ubalt.illiad.oclc.org
Fax: (410)837-4330
Email: ill@ubalt.edu

Billing Info:

Charge 12 IFM
Maxcost: 25.00IFM

NOTES:

Supplied From:

University of Colorado Anschutz Medical Campus
Health Sciences Library
OCLC Symbol: COU
DOCLINE Symbol: COUCOL

Contact Info:

Health Sciences Library - ILL/DocDel
Phone: 303-724-2111
Fax: 303-724-2150
Email: copydocs@ucdenver.edu
Odyssey: 206.107.42.56

Business Hours:

Monday-Friday, 8am to 5pm

Mailing Address:

Interlibrary Loan/Doc Del
Mail Stop A003, Health Sciences Library
12950 E. Montview Blvd., Room 2106C
Aurora, CO 80045

Health Sciences Library - Anschutz Medical Campus ILL Office



ILLiad TN: 126610

The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies

Jesús F. Salgado and Gabriel Táuriz

Department of Social Psychology, University of Santiago de Compostela, Santiago de Compostela, Spain

This article reports a comprehensive meta-analysis of the criterion-oriented validity of the Big Five personality dimensions assessed with forced-choice (FC) inventories. Six criteria (i.e., performance ratings, training proficiency, productivity, grade-point average, global occupational performance, and global academic performance) and three types of FC scores (i.e., normative, quasi-ipsative, and ipsative) served for grouping the validity coefficients. Globally, the results showed that the Big Five assessed with FC measures have similar or slightly higher validity than the Big Five assessed with single-stimulus (SS) personality inventories. Quasi-ipsative measures of conscientiousness ($K = 44$, $N = 8794$, $\rho = .40$) are found to be better predictors of job performance than normative and ipsative measures. FC inventories also showed similar reliability coefficients to SS inventories. Implications of the findings for theory and practice in academic and personnel decisions are discussed, and future research is suggested.

Keywords: Five-Factor model; Forced-choice; Meta-analysis; Performance; Personality.

Since at least the beginning of the 1990s, a large number of meta-analyses have found that personality measures predict academic and job performance, training proficiency, counterproductive behaviours, accidents, productivity data, salary, promotions and progress, grade point average, and other relevant educational and organizational criteria (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Bartram, 2005; Clarke & Robertson, 2005; Hogan & Holland, 2003; Hough, 1992; Hurtz & Donovan, 2000; Mount & Barrick, 1995; Ng, Eby, Sorensen, & Feldman, 2005; O'Connor & Paunonen, 2007; Ones, Viswesvaran, & Schmidt, 1993; Poropat, 2009; Salgado, 1997, 1998a, 2000, 2002, 2003; Tett, Rothstein, & Jackson, 1991; Trapmann, Hell, Hirn, & Schuler, 2007). Consequently, they can be useful for personnel selection and academic decisions.

Among personality measures, the Big Five personality dimensions (i.e., emotional stability,

extraversion, openness to experience, agreeableness, and conscientiousness) have received more attention than any other personality construct. The meta-analytic evidence mentioned above has demonstrated that conscientiousness and emotional stability generalized validity across samples, criteria, and occupations, and that the other three personality dimensions were valid predictors for specific criteria and specific occupations. For example, openness to experience predicted training proficiency, and extraversion and agreeableness predicted performance in occupations characterized by a high level of interpersonal relationships. This was found not only in American meta-analytic integrations (e.g., Barrick & Mount, 1991; Hogan & Holland, 2003; Hough, 1992; Hurtz & Donovan, 2000; Tett et al., 1991), but also in European meta-analyses (Salgado, 1997, 1998a), South-African meta-analyses (Rothmann, Meining, Van der Walt, & Barrick, 2002; Van der Walt,

Correspondence should be addressed to Jesús F. Salgado, Department of Social Psychology, University of Santiago de Compostela, Campus Vida, 15782 Santiago de Compostela, Spain. E-mail: jesus.salgado@usc.es

We thank Neil Anderson, Dave Bartram, Lew Goldberg, Robert Hogan, Tim Judge, Kevin Murphy, Frank Schmidt, and two anonymous reviewers for their comments on a previous version of this manuscript. The research reported in this article was partially supported by Grant SEJ-2008-3070 from the Ministry of Science and Innovation (Spain) and Grant PSI2011-27947 from the Ministry of Economy and Competitiveness to Jesús F. Salgado.

Meiring, Rothmann, & Barrick, 2002), Korean meta-analyses (Yoo & Min, 2002), and meta-analyses of the studies conducted in several Asian countries (Oh et al., 2011).

Additionally, Ones and her colleagues showed that criterion occupational personality scales (COPS), such as integrity tests, managerial potential, drug abuse, stress tolerance, and service orientation scales, were also valid predictors of job performance and that they generalized validity across samples (Ones & Viswesvaran, 2001a; Ones, Viswesvaran, & Schmidt, 1993, 2003; Viswesvaran, Ones, & Hough, 2001; see also Hogan & Brinkmeyer, 1997). The series of meta-analyses conducted by Judge and colleagues (Judge & Bono, 2001; Judge, Bono, Ilies, & Gerdhardt, 2002; Judge, Bono, & Locke, 2000; Judge, Heller, & Mount, 2002) showed that core-self evaluations predicted performance, leadership, and job satisfaction. Ng and colleagues' meta-analysis showed the validity generalization of locus of control for predicting job performance (Ng, Sorensen, & Eby, 2006). Meta-analyses by Salgado and Moscoso (2000) and Judge, Jackson, Shaw, Scott, and Rich (2007) showed that self-efficacy was a predictor of various criteria, including job performance, satisfaction, and absenteeism. More recently, meta-analyses by Dalal (2005), Hershcovis et al. (2007), and Kaplan, Bradley, Luschman, and Haynes (2010) showed that positive and negative affect predicted performance and deviant behaviours at work. It was also demonstrated that the facets of the Big Five were predictors of job performance (e.g., Dudley, Orvis, Liebicki, & Cortina, 2006; Hurtz & Donovan 2000; Salgado, 2004). Finally, there is also evidence that "dark side" personality measures predict task and contextual performance and counterproductive behaviours (Hogan & Hogan, 2001; Rolland & De Fruyt, 2003; Moscoso & Salgado, 2004). The validity coefficients ranged from .20 to .45 when the personality measures showed evidence of validity generalization.

Despite the meta-analytic evidence mentioned, there still remains no unanimous agreement that personality measures are relevant for making personnel decisions. For example, Murphy and Dzieweczynski (2005) made three general criticisms: (1) theories linking personality constructs and job performance are often vague and unconvincing, (2) little is known about how to match personality dimensions and occupations, and (3) the most valid personality-related measures have involved measures of poorly defined constructs, such as integrity.

More recently, a panel of past and current editors of the top-tier journals in industrial, work, and organizational psychology discussed the evidence on various aspects of the validity of personality inventories for personnel selection and concluded with a rather pessimistic view (Morgeson et al., 2007a,

2007b). Their main points were that: (1) the validity of personality measures is small, even when corrections for criterion reliability and range restriction are made; (2) they can show some incremental validity over general mental ability; (3) measures based on self-reports can be faked and this can change the rank orders of individuals and, consequently, affect hiring decisions; (4) corrections for faking, mainly based on the scores in social desirability scales, do not seem to improve validity; and (5) the faking scales do not work well for identifying distorted responses. The criticisms by Murphy and Dzieweczynski (2005) and Morgeson et al. (2007a) were answered by Barrick (2005), Hogan (2005a, 2005b), Hough and Oswald (2005), Ones, Viswesvaran, and Dilchert (2005), Ones, Dilchert, Viswesvaran, and Judge (2007), Tett and Christiansen (2007), among others, who provided a large amount of evidence supporting personality measures at work. With regards to faking, Hogan, Barrett, and Hogan (2007) tested a large sample of job applicants twice, 6 months apart. They found that only 5.25% improved their scores on the second occasion. Equally important, the Hogan, Barrett, and Hogan paper shows that the same number of applicants reduce their scores by trying to fake. The two groups cancel each other out, suggesting that faking may be a random process.

Overall, these findings suggest that faking on personality measures in the context of personnel selection may be less important than it was previously thought. Furthermore, personality measures remain very popular in the US, and even more so in other countries such as the European Union (see Tett, Christiansen, Robie, & Simonet, 2011; Zibarras & Woods, 2010).

The members of the panel also made two interesting recommendations: (1) that research should be done to look for alternatives to typical self-report inventories, and among these alternatives forced-choice (FC) inventories and conditional reasoning tests were suggested; and (2) that the criterion domain should be expanded and other measures should be used in addition to job performance ratings. In this article, we will centre on the relationship between the FC personality inventories and four types of work and educational criteria: job performance ratings, training proficiency, objective performance (e.g., sales), and educational success (i.e., grade point average).

FORCED-CHOICE MEASURES IN PERSONALITY ASSESSMENT AND PERSONNEL AND EDUCATIONAL DECISIONS

According to Travers (1951), it was Paul Horst and Robert Wherry who, independently of each other,

developed the seminal ideas about FC measures. Horst applied the concept to the development of personality inventories and Wherry to the problem of rating army officer performance. In this sense, FC is simply a specific format of rating and assessment procedures. Typically, the FC method gives the individual (e.g., the applicant, the rater) a number of words or phrases, along with instructions to select the ones he or she most, or in some cases least, likes when it is applied to the evaluated person. The number of words or phrases may be, for instance, pairs, triads, or tetrads, which are paired in terms of an index of preference and discrimination (e.g., social desirability). Thus, the FC formats can be distinguished from SS formats (e.g., Likert, yes/no, true/false) in that a choice must be made among the alternatives rather than rating each statement as it occurs in the single-stimulus formats.

Since the 1950s, a number of popular personality inventories have used FC modalities, including the Edwards Personal Preferences Schedule (EPPS; Edwards, 1973), the Gordon Personal Profile-Inventory (GPP-I; Gordon, 1993), Myers-Briggs Type Indicator (MBTI; Myers, McCaulley, Quenk, & Hammer, 1998), and the family of Occupational Personality Questionnaires (OPQ; SHL, 2006). FC inventories have not only been used in personality assessment, but also for assessing learning styles (Kolb, 1985), vocational interests (Kuder, 1975), social power bases (French & Raven, 1959), team conflict (Thomas & Killman, 1974), and team roles (Belbin, 1981). FC formats were also used for assessing job performance (e.g., Bartram, 2005; King Hunter & Schmidt, 1980). FC formats have been used in personnel and student selection mainly because they reduce or eliminate uniform biases such as acquiescence responding and faking, and can reduce "halo" effects (Bartram, 2005, 2007; Cheung & Chan, 2002).

A relevant characteristic of the FC formats is that they can result in several different scoring methods with specific statistical and psychometric particularities. In his seminal review of the FC personality measures, Hicks (1970) drew up a classification that remains a classic to this day. The basis of Hicks' taxonomy is the difference between normative and ipsative scores, as suggested by Cattell (1944) and Clemans (1966). According to Clemans, "any score matrix is said to be ipsative when the sum of the scores obtained over the attributes measured for each respondent is constant" (p. 4). Following this definition, Hicks (1970) suggested that three different types of FC measures can be distinguished: (1) purely ipsative measures, (2) quasi-ipsative (or partially) ipsative measures, and (3) normative FC measures. The first type refers to those measures that totally meet Clemans' criterion of ipsativity. The second type includes measures that do not totally meet the

criterion of pure ipsativity suggested by Clemans, because, for example, not all alternatives ranked by respondents are scored or the scales have different numbers of items. The main characteristic of the third type of measures is that items representing a given bipolar scale are never paired with items representing another bipolar scale. For example, items assessing extraversion are never paired with items of conscientiousness. It is important to point out that the SS format produces normative scores only, in contrast with the FC format, which can produce three types of scores.

As Cattell (1948), Clemans (1966), and Hicks (1970) have shown, each score type (i.e., normative, quasi-ipsative, and purely ipsative) has important characteristics. 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). This kind of score allows the comparison of individuals or groups on each measured variable (i.e., they are interindividual scores). In the case of purely 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 the comparison of the level of the individual across variables (i.e., they are intraindividual scores). Normative scores can be transformed into ipsative scores by a simple mathematical transformation (e.g., by subtracting each individual's average scale score from each scale) but the reverse is not possible. Quasi-ipsative scores share some psychometric characteristics with both normative and purely ipsative scores. For example, they allow the comparison of individuals and groups, which is a characteristic of normative scores (Bartram, 1996; Cattell & Brennan, 1994; Clemans, 1966; Heggstad, Morrison, Reeve, & McCloy, 2006; Horn, 1971). Simultaneously, some degree of dependence can be generally found among the quasi-ipsative scales of the questionnaire, which is a characteristic of ipsative scores (Clemans, 1966; Gordon, 1993). However, as Horn (1971) pointed out, quasi-ipsative scoring does not always introduce algebraic dependence.

In Hicks' (1970) view, ipsativity can be quantified in terms of the deviation of the conditions of ipsativity. According to the mathematical examination carried out by Clemans (1966) and Radcliffe (1963), ipsative scores show five properties: (1) Once the scores obtained in ipsative form are converted to deviation scores by columns, the sums of the columns or rows of an ipsative covariance matrix must be zero; (2) if the ipsative variances are equal, the sums of columns or rows of the ipsative intercorrelation matrices are equal to zero; (3) when the ipsative variances are equal, the average intercorrelation value will be limited by $-1/(m-1)$ where m is the number of scales; (4) the sum of the covariance terms

obtained between a specific criterion and a set of ipsative scores is zero; and (5) when ipsative variances are equal, the sum of the ipsative validity coefficients is zero. Therefore, a simple comparison between the empirical means and standard deviations of the summed validities, and the mean intercorrelation and the predicted validities, can serve to quantify ipsativity. The more deviation there is from the predicted values, the less pure ipsativity is found in the data. Hicks demonstrated that this procedure works.

There are also different strategies to produce a quasi-ipsative score or successfully reduce ipsativity (see Hicks, 1970). For example, the Gordon Personal Profile–Inventory (GPP-I; Gordon, 1993) and the IPIP-MFC (Heggestad et al., 2006) avoid full ipsativity by introducing some changes in the way FC inventories are scored. The GPP-I and the IPIP-MFC, instead of having a fixed sum, allow a range of possible values within each tetrad. This allows people to be high on all the scales, low on all, or have an intermediate level in all. Consequently, the method does not result in a fixed sum of scale scores, and normative arrangements of the individuals are possible. This scoring system is quasi-ipsative in the sense that *very high* scores in all scales are not possible. However, from the empirical point of view, this limitation is not very important because the number of people in the population who would score very highly in, for example, emotional stability, conscientiousness, openness, agreeableness, and extraversion, all at the same time, is minimal.

A number of researchers have suggested using FC measures in personnel selection because they are valid predictors of job performance and resistant to faking (Bartram, 2005, 2007; Brogden, 1954; Christiansen, Burns, & Montgomery, 2005; Jackson, Wroblewski, & Ashton, 2000; Norman, 1964). Nevertheless, if the FC inventories produce ipsative scores, they have some characteristics that some researchers consider problematic from the psychometric point of view. Among the limitations, it has been mentioned that: (1) Ipsative measures produce scores that negatively correlate with each other, which limits the type of statistical analysis which is possible (Meglino & Ravlin, 1998); (2) they can affect the size of reliability coefficients (Bartram, 1996; Horn, 1971; Johnson, Wood, & Blinkhorn, 1988; Tenopyr, 1988; Thompson et al., 1982); (3) they are not appropriate for factor analysis (Cornwell & Dunlap, 1994; Dunlap & Cornwell, 1994; Meade, 2004); (4) they are inappropriate for comparisons among individuals (Cattell & Brennan, 1994; Hicks, 1970); (5) the average correlation between m scales of a purely ipsative questionnaire is bound below by $-1/(m-1)$ and above by $(m-4)/m$, whereas for nonipsative questionnaires the average correlation ranges between $-1/(m-1)$ and 1 (Gleser, 1972); (6) they appear to

correlate with cognitive ability when individuals respond as job applicants (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006); (7) the ipsative and quasi-ipsative measures may produce gender differences in some cases and, consequently, equal opportunities may also be negatively affected (Anderson & Sleaf, 2004); and (8) according to Ones et al. (2007), they do not improve the criterion-related validity of personality measures and, therefore, are useless for most purposes in organizational decision making.

Despite these criticisms, FC inventories continue to be used in personnel selection. For example, Tett, Christiansen, Robie, and Simonet (2011) found that 30% of companies used FC inventories, although the percentage of ipsative, quasi-ipsative, and normative measures is unknown. Furthermore, several researchers have questioned the psychometric criticisms mentioned before. For example, Baron (1996) suggested that ipsative measures are appropriate for factor analysis if the number of items or scales is large (> 30) because the intercorrelations become practically zero, although this result may also be substantially artifactual (Cattell & Brennan, 1994; Dunlap & Cornwell, 1994; Meade, 2004). Bartram (1996) found that ipsative measures could show similar or even larger internal consistency coefficients than SS measures when the appropriate formula is used. Furthermore, confirmatory factor analysis (CFA) can be done using Jackson and Alwin's (1980) procedure as modified by Chan and Bentler (1993), and ipsative scores can also be transformed into a normal distribution using the formula suggested by Hayes (1967; see also Chapman, Blackburn, Austin, & Hutcheson, 1983; Feather, 1973), allowing comparisons of individuals. More recently, developments in IRT technology have produced methods for analyzing multidimensional forced-choice items (Brown & Maydeu-Olivares, 2011; Chernyshenko et al., 2007; Heggestad et al., 2006; Maydeu-Olivares & Brown, 2010; McCloy, Heggestad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & Williams, 2006). Even though the number of studies is too small at present to be conclusive about the usefulness of IRT approaches in personnel and academic selection, recent findings show that the predictive validity of FC inventories is similar or larger than the predictive validity of their SS counterparts (Brown & Bartram, 2009).

With regard to quasi-ipsative measures, Hicks (1970) hypothesized that validity increases as an inverse function of ipsativity. Therefore, quasi-ipsative measures would be more valid predictors than purely ipsative ones. Furthermore, he suggested that there may exist cases in which quasi-ipsative measures are also more valid than SS ones, because some of the advantages of the FC format (e.g., more

resistance to faking) can be successfully exploited without the statistical limitations of ipsativity. Moreover, Cattell and Brennan (1994) demonstrated that factor analysis is not affected if quasi-ipsative data are used.

An important implication of Hicks' (1970) review has to do with the conditions that ipsative measures should fulfil to be used in psychology and, specifically, in personnel selection and academic decisions. Hicks suggested that the value of ipsative measures depends on three conditions being fulfilled: (1) that SS measures are affected by faking, (2) that faking diminishes the validity of SS measures, and (3) that ipsative measures control faking better than other SS controls of faking, and simultaneously increase the validity. With regard to the third condition, research has shown that FC measures are more resistant to faking than their SS counterparts (see Nguyen & McDaniel, 2000). However, it has not been conclusively demonstrated that quasi-ipsative and purely ipsative measures are equal or more valid predictors of job and academic performance than SS inventories. Some researchers have reported high validities for FC measures. For example, Brogden (1954) found validity coefficients of .33 and .42, and Villanova, Bernardin, Johnson, and Dahmus (1994) found similar coefficients for an FC job compatibility questionnaire. However, according to Ones et al. (2007), the precise reasons for the higher correlations remain unclear.

It is surprising that despite the years which have passed since Hicks' (1970) review and the large number of meta-analyses of the criterion-oriented validity of personality measures, to date no meta-analysis has been conducted to examine the validity of FC personality inventories for predicting job and academic performance. For example, there is no meta-

analytic research examining whether the different types of scores (e.g., normative, purely ipsative, and quasi-ipsative) yielded by FC inventories have similar or different validity, or whether the validity of FC measures is moderated by variables such as criterion type. Also, previous meta-analyses did not distinguish between SS and FC personality measures. In this sense, based on the reference list of the previously published meta-analyses, it can confidently be affirmed that those meta-analyses were mainly conducted with data obtained with SS inventories. For example, an inspection of Table 1 in the meta-analysis by Tett et al. (1991) shows that more than 85% of studies used SS inventories. Similar or even larger figures were found in the meta-analyses conducted by Hurtz and Donovan (2000), O'Connor and Paunonen (2007), Poropat (2009), Salgado (1997, 2003), and Trapmann et al. (2007) on the validity of the Big Five personality dimensions. Barrick and Mount (1991) and Hough (1992) did not include the list of studies in their meta-analysis, but it is likely that they shared the majority of the studies with those of Tett et al.

AIMS OF THE STUDY

In summary, this research has five goals. The main objective of this article is to conduct a meta-analytic study of the validity of the Big Five personality dimensions as assessed with FC personality measures for predicting work and educational criteria (e.g., job performance, productivity, GPA). The second goal is to ascertain whether the type of score (i.e., purely ipsative, quasi-ipsative, and normative) moderates the validity of FC questionnaires. The third objective is to compare the results with the previous meta-analytic findings (e.g., Barrick & Mount, 1991; see

TABLE 1
Distributions of reliability of the Big Five personality dimensions assessed with forced-choice questionnaires

	<i>Emotional stability</i>	<i>Extraversion</i>	<i>Openness to experience</i>	<i>Agreeableness</i>	<i>Conscientiousness</i>
Internal consistency					
Mean	.73	.75	.81	.80	.72
SD	.09	.13	.12	.08	.12
Square root of reliabilities	.85	.87	.90	.90	.86
K	10	6	4	8	11
Range	.52-.81	.47-.84	.60-.88	.70-.89	.53-.92
Ones & Viswesvaran (1999)	.78	.78	.73	.75	.78
Square root of reliabilities	.88	.88	.85	.86	.88
Test-retest					
Mean	.76	.75	.75	.71	.77
SD	.13	.06	.04	.14	.11
Square root of reliabilities	.87	.87	.87	.84	.88
K	13	13	8	4	15
Range	.50-.92	.65-.86	.67-.81	.53-.85	.61-.96
Ones & Viswesvaran (1999)	.77	.76	.71	.69	.72
Square root of reliabilities	.87	.87	.84	.83	.84

also Schmidt, Shaffer, & Oh, 2008), conducted mainly with SS personality measures. The fourth objective is to examine if the criterion type moderates the validity of FC measures. The fifth goal is to examine the reliability of the Big Five when assessed with FC inventories. In order to achieve these objectives, we decided to use the Five-Factor Model as the framework for classifying the validity coefficients of the FC personality measures, both for theoretical and practical reasons. From the theoretical point of view, the classification of FC measures into the Big Five will allow us to compare our results with the findings of previous meta-analyses, mostly conducted with SS inventories, and respond to the question posed by Hicks (1970) about the smaller, equal, or larger validity of ipsative versus normative data (based on SS inventories). From the practical point of view, the Big Five is a useful taxonomy which is unrivalled at present, although it may not be exhaustive (Hogan, 2005a).

This research refers to the same variables used in previous meta-analyses of personality inventories (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000; Salgado, 1997) and, therefore, some of the seminal hypotheses stated by Barrick and Mount (1991) are appropriate in the present case. Consequently, we advance the following hypotheses:

Hypothesis 1: Emotional stability and conscientiousness are valid predictors for all academic and job performance criteria and they generalize validity across samples, criteria, and FC personality measures.

Hypothesis 2: Openness to experience is a valid predictor of training proficiency and it generalizes validity across samples and FC personality measures.

On the other hand, Hicks' (1970) analysis suggested that the criterion-oriented validity is dependent on the score type (e.g., normative scores would be more valid than ipsative scores) and that validity is also dependent on the assessment conditions (e.g., personnel selection vs. laboratory conditions). According to Hicks, quasi-ipsative measures may be more valid than SS ones in personnel selection conditions, because some of the advantages of the FC format (e.g., more resistance to faking) can be successfully exploited without the statistical limitations of ipsativity. Consequently, we also state a hypothesis concerning the type of scores (i.e., purely ipsative, quasi-ipsative) derived from FC inventories:

Hypothesis 3: Quasi-ipsative personality inventories will show larger criterion-oriented validity

than purely ipsative inventories for predicting job performance.

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 FC personality measures as exhaustively 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 FC inventories. They included, for example, the Occupational Personality Questionnaire (OPQ), Edwards Personal Preferences Schedule (EPPS), Myers-Briggs Type Indicator (MBTI), Description en Cinq Dimensions (D5D), Survey of Interpersonal Values Inventory (SIV), and the Gordon Personal Profile-Inventory (GPP-I). Second, PsycInfo, Social Sciences Citation Index, and ABI/Inform databases were searched to identify studies on the relationship of FC 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 FC 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 manuals of the most popular FC personality questionnaires (e.g., EPPS, GPP-I, MBTI, OPQ) were examined in order to find validity coefficients. By means of these search strategies, a preliminary database of over 180 documents (i.e., articles, manuals, technical reports, unpublished papers, dissertations, and so on) was established for further inspection. There were 58 studies excluded from the total pool for various reasons: (1) Some studies reported only the significant correlations, (2) a number of studies only reported

multiple correlation results, (3) several of them did not report correlations or enough information to calculate the effect size, and (d) several studies reported findings for the same data set. As a result of these points, the meta-analysis was conducted with 122 independent samples. With regard to these studies, we are not able to ascertain how much overlap, if any, exists between our meta-analysis and Barrick and Mount's (1991) meta-analysis because these researchers did not report the list of studies included in their meta-analysis. The same is true for Hough's (1992) meta-analysis. With regard to Tett et al.'s (1991) and Salgado's (1997) meta-analyses, we can say that the degree of overlapping is practically irrelevant. We share six studies out of 122 with Tett et al.'s meta-analysis and we share three studies with Salgado's meta-analysis.

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; Nyfield et al., 1995; Robertson, Baron, Gibbons, MacIver, & Nyfield, 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 (based on the definitions of the Big Five given by Barrick & Mount, 1991; Hough, 1992; McCrae & Costa, 1990; Hough & Ones, 2001; and Salgado, 1997; among other sources). A list and the definition of the personality scales from each inventory were then provided for each coder with instructions to assign each scale to the most appropriate factor. Furthermore, some studies reporting factor analyses of the inventories were also used as a basis for the decision (e.g., Matthews, Stanton, Graham, & Brimelow, 1990; McCrae & Costa, 1990; Piedmont, Costa, & McCrae, 1992; SHL, 2006) because these factor analyses were informative about the Five-Factor structure of some FC inventories (e.g., OPQ, EPPS). Finally, we also checked the coding lists used by Hough and Ones (2001), Ones (1993), and Salgado (2003) in order to compare our classification and the classification of these researchers when the same personality measures were used. 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. Appendix A includes a complete list of the scales that were assigned to each Big Five personality dimension.

For each study, the following information was recorded, if available: (1) sample characteristics, such as gender, age, education, and so forth; (2)

occupation and related information; (3) personality measures used; (4) criterion type; (5) reliability of personality measures; (6) criterion reliability; (7) range restriction value or data for calculating this value; (8) statistics concerning the relation between personality measures and the criterion; and (9) correlation among the personality measures when more than one was used. This complies with the American Psychological Association 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 estimates that are more construct-valid 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 Mosier's formula for the correlation of variables with composites. As demonstrated by Warr, Bartram, & Brown (2005), the average validity, corrected with Mosier's formula for composite reliability, produces very accurate estimates when the appropriate intercorrelations are used.

An important difference between this meta-analysis and previous meta-analyses of the relationship between personality and job and academic performance is that we used the type of FC measure for grouping the validity coefficients. This difference is especially relevant because different degrees of ipsativity can result in different validity levels (Clemans, 1966; Hicks, 1970; Radcliffe, 1963). Because the number of studies allowed us to consider each one as a separate entity, we classified the personality inventories in three categories: purely ipsative, quasi-ipsative, and normative FC questionnaires. In order to classify the inventories, each one was inspected in terms of the scoring method and the format of items. Furthermore, we used the technical manuals of the inventories when available, and other articles that included relevant information about the inventory 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 was constant. Examples of purely ipsative inventories in our database are the Edwards Personality Preferences Schedule (EPPS; Edwards, 1957, 1973), the Occupational Personality Questionnaire 32i (OPQ32i; SHL, 2006), the Occupational Personality Questionnaire CM4.2 (OPQ CM4.2; Saville, Holdsworth, Nyfield, Cramp, & Mabey, 1984), and the Description en Cinq Dimensions (D5D; Rolland & Mogenet, 2001). Based on Hicks' (1970) criteria of ipsativity, we classified inventories as quasi-ipsative if any of the following alternatives applied: (1)

Individuals only partially order the items, rather than ordering them completely; (2) scales have different numbers of items; (3) not all alternatives ranked by respondents are scored; (4) scales are scored differentially for individuals with different characteristics or involve different normative transformations on the basis of respondent characteristics; (5) scored alternatives are differentially weighted; and (6) the inventory has normative sections. Examples of quasi-ipsative inventories in our database are the Gordon Personal Profile-Inventory (Gordon, 1993), the IPIP-MFC (Heggestad et al., 2006), the Self-Description Inventory (Ghiselli, 1954), the ESQ (Jackson, 2002), and the Assessment Individual Motivation (AIM; Knapp, Heggestad, & Young, 2004; White, 2002). Finally, we classified inventories as normative if it yielded scores that possess the empirical properties of absolute measures. This is the case of the inventories in which items representing different degrees of a personality dimension are never paired with items representing another personality dimension. The MBTI and the "Need of Achievement" questionnaire (Fineman, 1975) are representative examples of normative FC questionnaires. Following the standards of the American Psychological Association for reporting meta-analytic research (APA Publications and Communications Board Working Group, 2008) and the recent suggestion by Sackett and Schmitt (2012) on data reporting in meta-analysis, we have included an appendix with the totality of studies and validity coefficients we have used in our meta-analysis (see Appendix B).

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 in order to obtain the operational validity (which is of interest for personnel selection and academic decisions), and we corrected the operational validity for predictor reliability in order to obtain the true validity (which is of interest for modelling the theoretical relationship between personality and performance).

Artifact distributions

Predictor reliability. The reliability of the personality dimensions was estimated from the coefficients reported in the studies included in our 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 whether the three types of FC scoring systems showed different levels of reliability, but they proved to be very similar. Therefore, we used the average reliabilities in the meta-analytic calculations. Table 1 presents a summary of these artifact distributions. In Table 1, we have also included the reliabilities found by Ones and Viswesvaran (1999) in their meta-analysis of Big Five reliability. As can be seen, our estimates are very similar to the estimates found by Ones and Viswesvaran 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).

Although we used the internal consistency coefficients in our meta-analysis, this decision may be controversial for some researchers. For example, Heggestad et al. (2006) pointed out that Cronbach's alpha is not an appropriate coefficient for estimating the reliability of FC inventories because the responses to items are not independent and, therefore, the observed item covariances are not accurate estimates of the true item covariances. They suggested that test-retest coefficients are the appropriate estimates. For this reason, we also developed an empirical distribution of test-retest coefficients for the Big Five, using the coefficients reported in the studies included in our database. The average test-retest reliability was .76, .75, .75, .71, and .77, for emotional stability, extraversion, openness, agreeableness, and conscientiousness, respectively. These estimates are very similar to the test-retest estimates reported by Viswesvaran and Ones (1999) and are also similar to the average internal consistency estimates found here. Thus, the use of Cronbach's alpha or test-retest coefficients has no differential effect in the current meta-analysis.

In summary, the empirical evidence suggests that when the Big Five personality dimensions are measured with FC personality inventories, the reliability is practically the same as that obtained

with SS inventories. This is an important finding as some researchers had suggested that FC inventories would show a different level of reliability than SS questionnaires.

Criterion reliability. In this research, six types of criteria were used: (1) supervisory ratings of job performance, (2) productivity data (e.g., sales), (3) training proficiency, (4) grade point average, (5) global academic performance, and (6) global job performance. This choice was made for two reasons: (1) Previous meta-analyses of personality measures used some of these types of criteria, and one of our objectives was to provide a comparison with those meta-analyses; consequently, it was important to retain the same criteria; (2) other criteria, such as tenure, turnover, or accidents were used in a very small number of studies, or were totally absent from our database; therefore, we would not be able to carry out meta-analyses in these cases.

Not all studies provided information regarding the criteria reliability and, consequently, we developed empirical distributions for the six criteria. Fortunately, a number of studies provided reliability coefficients for estimating criterion reliability. For job performance ratings, the coefficient of interest is interrater reliability when a meta-analysis of random effects is carried out (Hunter, 1986; Sackett, 2003; Schmidt & Hunter, 1996). This is 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 & Hirsch, 1987), although not all researchers agree with this point of view (e.g., Murphy & De Shon, 2000). We found 11 studies reporting interrater 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 interrater reliability of job performance ratings, and independently by Salgado et al. (2003) in European criterion-oriented validity studies of cognitive tests. In the case of training proficiency, we found two 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 & Mount, 1991). The average reliability for the objective productivity measures was .83, based on seven studies. This value is similar to the one found by Schmidt and Zimmerman (2004) and it is the same as that found by Schmidt, Gooding, Noe, & Kirsch (1984). We found six coefficients of GPA reliability, which produced a mean reliability of .80 ($SD = .10$) and we found two coefficients of class attendance reliability. Pooling together the reliability coefficients and weighting for the number of studies using each criterion type, we calculated the average reliability coefficient for global job performance and for global academic performance, so we had two additional composite criteria. The respective average coefficients were .61 ($SD = .13$) and .81 ($SD = .09$). These last two distributions were used for the meta-analyses carried out with the pooled sample of validity studies. Out of six criterion reliability estimates (four criterion categories plus two overall composites), four were .80 or slightly higher, one was .61 and another .52. Considered globally, this artifact produced an underestimation of the observed validity and error variance as well.

Range restriction distributions

The distributions for range restriction were based on the following three strategies: (1) Some range restriction values were obtained from the studies that reported both restricted and unrestricted standard deviation data, (2) another group of range restriction values were obtained using the population values reported in the manual of the various inventories, and (3) 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 extraversion, .92 for openness to experience, .90 for agreeableness, and .88 for conscientiousness. These u values were used when no more specific ones were available. We also found $u = .93$ ($SD = .10$) and $u = .81$ ($SD = .19$) for the ipsative and quasi-ipsative measures combined with job performance ratings; $u = .82$ for ipsative measures combined with training proficiency, and $u = .86$ ($SD = .17$) when the criterion was GPA. 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 smaller than the range restriction found

TABLE 2
Distributions of criteria reliability

Criterion	Mean	SD
Job performance ratings	.52	.05
Training proficiency	.80	.09
Productivity data	.83	.07
Grade point average	.80	.10
Occupational performance (average)	.61	.13
Academic performance (average)	.81	.09

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 in validity studies. 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 would be the exception). Consequently, correction for direct range restriction (DRR), rather than IRR, results in an underestimation of the operational and true validity coefficients, and in an 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 et al., 2006) developed a new formula for this purpose.

In a series of studies, Schmidt and his colleagues have demonstrated the effects of the IRR correction on the validity of general mental ability (GMA), on the validity of specific cognitive abilities, and on the validity of the Big Five (Schmidt, Oh, & Le, 2006; Schmidt et al., 2008), as well as the accuracy of the new procedures via Montecarlo simulations (Le & Schmidt, 2006). Globally, Schmidt et al. (2008) found that IRR correction improved the operational validity by .01. 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. 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 to obtain the operational validity, and for predictor unreliability to obtain the true correlation. The observed variance will be corrected for by four artifactual errors: sampling error, criterion and predictor reliability, and IRR.

RESULTS

Meta-analysis of the personality factors pooled across criteria and forced-choice measures

Table 4 shows the meta-analytic results of the five personality dimensions for predicting academic and job performance. The validity coefficients were pooled across criteria and FC measures. Therefore, this is the most general level of analysis of our study. From left to right, the first four columns represent the number of independent coefficients (K), the total sample size (N), the observed validity weighted by the study sample 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_o), the full-corrected correlation (true validity, ρ), and the standard deviation of ρ . Finally, the last two columns are the percentage of variance explained by the four artifactual errors (sampling error, predictor reliability, criterion reliability, and indirect range restriction in predictor scores) 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 different objectives. Operational validity is the coefficient to be used for predicting the criterion in applied settings (e.g., making decisions about employees or students). True validity is the theoretical correlation between the personality dimension and the criterion. Therefore, this coefficient is used for modelling the theoretical relationship between dependent and independent variables. Although we are interested in both coefficients, we will concentrate on ρ in our comments.

As can be seen, only conscientiousness showed to be a generalizable predictor of occupational performance, with a ρ value of .24 and 90%CV positive, which means that there is validity generalization for this personality dimension. This figure is very similar to the one found by Barrick and Mount (1991) and Salgado (1997) for the combination of criteria, although these two meta-analyses used DRR

TABLE 3
Distributions of range restriction

Personality dimension	Mean	SD
Emotional stability	.87	.16
Extraversion	.90	.14
Openness to experience	.92	.13
Agreeableness	.90	.14
Conscientiousness	.88	.17
For all personality dimensions		
Job performance ratings—ipsative	.93	.10
Training—ipsative	.82	.17
Job performance ratings—quasi-ipsative	.81	.19
GPA—all	.86	.17

TABLE 4
Results of the meta-analyses of the Big Five validity for occupational and academic performance

Personality dimension	<i>K</i>	<i>N</i>	<i>r_w</i>	<i>SD_r</i>	<i>r_c</i>	<i>ρ_i</i>	<i>SD_ρ</i>	% <i>VE</i>	90% <i>CV</i>
Occupational performance									
Emotional stability	82	16436	.06	.13	.10	.11	.19	30	-.12
Extraversion	80	17692	.06	.13	.09	.10	.18	29	-.11
Openness to experience	63	13539	.09	.12	.13	.14	.15	33	-.05
Agreeableness	65	14740	.04	.13	.05	.06	.17	28	-.15
Conscientiousness	96	20307	.14	.13	.21	.24	.18	36	.01
Academic performance									
Emotional stability	14	3916	.03	.07	.05	.05	.06	73	-.01
Extraversion	20	6884	-.02	.11	-.02	-.03	.14	25	.14
Openness to experience	16	6299	.15	.19	.19	.21	.24	11	-.09
Agreeableness	17	6560	-.06	.14	-.08	-.09	.20	13	.15
Conscientiousness	25	6314	.12	.10	.17	.19	.10	52	.05

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

correction instead of the IRR correction used here. In order to confirm whether the difference in the type of range restriction correction represents a large change in the operational and true validity, we also conducted the analysis with DRR correction. We found a slightly smaller validity for DRR correction, but the amount is so small that it does not affect the conclusions. Schmidt et al. (2008) also found that the IRR correction has only a slightly larger effect on the validity of personality dimensions in comparison with the DRR correction. Consequently, FC and SS personality inventories (the latter as represented, for instance, in the meta-analyses by Barrick & Mount, 1991; Hurtz & Donovan, 2000; and Salgado, 1997) show similar validity coefficients for conscientiousness. Based on this finding, overall validity size is not the factor which should be used to decide if one type of inventory or the other should be preferred for making personnel decisions.

The percentage of variance explained by artifactual errors is small for all personality dimensions, ranging from 28% for agreeableness to 36% for conscientiousness, which suggests that the search for possible moderators is appropriate. Criterion type and FC scoring type are possible moderators of validity. We will show the meta-analytic results for these moderators in the following sections of the article.

With regard to academic performance, as was hypothesized in Hypothesis 1, conscientiousness showed to be a valid and generalizable predictor. The *ρ* value was .19, which is slightly smaller than the validity found in previous meta-analyses with SS personality inventories. For example, O'Connor and Paunonen (2007) found a coefficient of .24, Poropat (2009) found a validity of .23, Salgado (2000) found a true validity of .28, and Trapmann et al. (2007) found a true validity coefficient of .27. The results of the

current meta-analysis showed that conscientiousness generalized validity across studies, although the percentage of unexplained variance is large (48%), which suggests that moderators can affect validity (e.g., university type, country, criterion measures). For example, Trapmann et al. found that the reliability of university grades is different for oral and written examinations and the use of one type or other depends on the specific university or even the academic subject. They also found that the validity of the Big Five for predicting grades is slightly different in Australia, Europe, and North America, and very different in East Asia. Based on our dataset, we cannot conduct a meta-analysis for these kinds of possible moderators. We also found that no other personality dimension showed to be a generalizable predictor of academic performance. This last result is similar to the previous meta-analytic findings (O'Connor & Paunonen, 2007; Poropat, 2009; Salgado, 2000; Trapmann et al., 2007).

Meta-analysis by personality dimensions–criterion combinations

We have hypothesized that conscientiousness and emotional stability are valid predictors for all criterion types, and that openness is a valid predictor for training proficiency. The results of the meta-analysis of the personality dimensions and three criterion types are shown in Table 5.

For job performance ratings, only conscientiousness showed a relevant validity of .20. This coefficient is very similar to the validity coefficients found by Barrick and Mount (1991; Barrick et al., 2001), Hurtz and Donovan (2000), and Salgado (1997). It is also similar to the coefficients found by Salgado (2003) when conscientiousness is not assessed with

TABLE 5
Results of the meta-analyses of the Big Five personality dimensions—criteria combinations

Personality dimension	<i>K</i>	<i>N</i>	r_w	SD_r	r_c	ρ_i	$SD\rho$	%VE	90%CV
Job performance ratings									
Emotional stability	50	9775	.03	.09	.05	.06	.12	58	-.08
Extraversion	50	11445	.07	.12	.11	.12	.17	34	-.08
Openness to experience	42	9519	.07	.11	.10	.11	.15	36	-.07
Agreeableness	44	10611	.03	.10	.04	.05	.12	44	-.10
Conscientiousness	62	13122	.11	.11	.18	.20	.16	42	.00
Training									
Emotional stability	11	2546	.06	.11	.08	.09	.13	41	-.07
Extraversion	10	2116	.01	.05	.01	.01	.00	100	.01
Openness to experience	8	884	.16	.11	.20	.22	.10	81	.12
Agreeableness	8	993	.11	.13	.16	.17	.14	52	-.01
Conscientiousness	10	2510	.10	.10	.15	.16	.12	46	.01
Productivity data									
Emotional stability	7	760	.02	.09	.02	.03	.00	100	.02
Extraversion	5	491	.04	.10	.05	.05	.00	100	.05
Openness to experience	4	312	-.01	.17	-.02	-.02	.16	44	-.21
Agreeableness	4	401	-.24	.06	-.28	-.31	.00	100	-.28
Conscientiousness	9	1074	.18	.06	.24	.27	.00	100	.24

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

measures based on the Five-Factor Model. As hypothesized, extraversion, openness to experience, and agreeableness were not generalizable predictors of job performance ratings.

With regard to training proficiency, conscientiousness showed a ρ value of .16 and the 90% CV was positive. This validity estimate is lower than the coefficients found by Barrick and Mount (1991) and Salgado (1997), but larger than the value found by Hurtz and Donovan (2000). Emotional stability did not prove to be a valid predictor of training proficiency, but the coefficient found was similar to the values found by Barrick and Mount, and Hurtz and Donovan. With regard to openness to experience, this personality dimension was shown to be a relevant and generalizable predictor of training, with a validity of .22 (90%CV = .12), as was hypothesized. This value is similar to the figures of .25, .14, and .26 found by Barrick and Mount, Hurtz and Donovan, and Salgado, respectively. Agreeableness showed a validity coefficient similar to the coefficient for conscientiousness, but the 90%CV included 0; therefore it did not show evidence of validity generalization. Extraversion did not predict training proficiency.

The third criterion examined was productivity data (e.g., sales). The hypothesis that conscientiousness and emotional stability would predict this criterion was only fulfilled for conscientiousness, which showed a validity of .27, and also showed validity generalization, as the whole observed variability was explained by artifactual errors. Another personality dimension proved to be a predictor of productivity,

but this finding was not anticipated. Agreeableness showed a validity coefficient of -.31, similar to but slightly larger than the validity for conscientiousness, and was also generalizable. Only the meta-analysis by Barrick and Mount (1991) examined the validity of the Big Five for predicting productivity data. They found that conscientiousness was the only generalizable predictor of productivity data, with a coefficient of .17. Consequently, FC measures for conscientiousness and agreeableness were shown to be better predictors than their SS counterparts (as based on Barrick and Mount's findings).

Meta-analysis by personality dimensions—FC measure combinations

As mentioned in the introduction to this article, FC inventories are not a unique type of measure. In fact, at least three different score types can be obtained from FC inventories (i.e., normative, quasi-ipsative, and purely ipsative), and this clearly contrasts with SS inventories which directly produce only normative scores (although they can be transformed into ipsative ones). Therefore, it is crucial to examine whether the different types of measures show similar or different validity sizes. Table 6 shows the results for the combinations of the Big Five and the three types of FC. According to Hypothesis 3, quasi-ipsative measures would show larger validity than normative FC, and ipsative measures.

As can be seen in Table 6, conscientiousness was a predictor for the three types of FC measures and

showed validity generalization, as the 90%CV was positive and did not contain zero in the three cases. However, the most important finding is that quasi-ipsative measures, as hypothesized, showed very much larger validity than ipsative measures (and also than normative FC measures). The true validity coefficients were .40, .16, and .16 for quasi-ipsative, normative FC, and ipsative measures, respectively. In order to check if the difference in validity could be due to differences in job type or in study authorship (e.g., independent research vs. consultancy firms), we ran an examination of the equivalence of samples between quasi-ipsative and ipsative inventories. We found that 88.9% of samples in ipsative studies and 85.7% of samples in quasi-ipsative studies consisted of the same seven occupational categories, and similar results were obtained for the study authorship. Therefore, we reject the idea that the results could be due to these two potentially confounding sources. Comparing the validity of quasi-ipsative measures of conscientiousness with the normative FC and ipsative measures, the first is 2.5 times larger than the validity of the other two measures. Therefore, Hypothesis 3 was totally supported by this finding. The importance of the finding can be better assessed if we take into account that no previous meta-analyses of the relationship between conscientiousness and job performance had found values larger than .30. For example, comparing our results with previous meta-analyses of the Big Five, the quasi-ipsative measures showed larger validity as compared with the values found by Barrick and Mount (1991), Barrick et al. (2001), Hough (1992), Hurtz and

Donovan (2000), O'Connor and Paunonen (2007), Poropat (2009), Salgado (1997, 2003), Tett et al. (1991), and Trapmann et al. (2007). We found no differences between the validity of normative FC and ipsative measures for conscientiousness, which is small in both cases. The results for these two types of measures are slightly lower than the values found in the previous meta-analyses of SS personality inventories mentioned earlier. No other dimension showed to be a predictor of job and academic performance.

Meta-analysis by personality dimension–criterion–FC measure combinations

The findings presented in the previous section suggest that a more finely tuned examination of the validity of the three types of FC measures could be of interest in order to advance our knowledge of this issue. Therefore, our last meta-analysis shows the results of the triple combination of personality dimensions, criterion type, and type of FC measures. Due to the limitations of our database, we were not able to conduct the analysis for all possible combinations (e.g., we have no studies for emotional stability). Table 7 shows the meta-analyses conducted for the combination of personality dimension, FC type, and GPA. Table 8 shows the meta-analyses for the combinations in which job performance ratings and training proficiency were used as criteria. Nevertheless, the findings should be taken with caution, as the number of primary studies is small for normative FC and ipsative inventories.

TABLE 6
Results of the meta-analyses of the Big Five–forced-choice type combinations

Personality dimension	<i>K</i>	<i>N</i>	r_w	SD_r	r_c	ρ_i	$SD\rho$	%VE	90%CV
Normative forced-choice									
Extraversion	5	2122	-.02	.11	-.03	-.04	.18	20	.17
Openness to experience	4	2059	.05	.03	.08	.09	.00	100	.08
Agreeableness	4	2059	.04	.06	.07	.07	.07	54	-.02
Conscientiousness	8	2732	.09	.08	.14	.16	.10	50	.02
Quasi-ipsative									
Emotional stability	35	6992	.11	.17	.18	.20	.26	20	-.13
Extraversion	34	7120	.06	.16	.09	.10	.24	19	-.19
Openness to experience	21	3677	.16	.16	.23	.25	.21	25	-.02
Agreeableness	24	4738	.10	.16	.15	.16	.22	22	-.11
Conscientiousness	44	8794	.23	.16	.36	.40	.21	33	.11
Pure ipsative									
Emotional stability	39	8055	.03	.08	.05	.05	.05	86	-.01
Extraversion	39	8180	.07	.08	.11	.12	.06	82	.04
Openness to experience	37	7555	.06	.09	.08	.09	.09	61	-.02
Agreeableness	36	7695	-.01	.10	-.01	-.02	.12	48	.12
Conscientiousness	40	8669	.09	.09	.14	.16	.09	64	.03

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

TABLE 7
Results of the meta-analyses of the Big Five–criterion–forced-choice type combinations in academic studies

Personality dimension	<i>K</i>	<i>N</i>	<i>r_w</i>	<i>SD_r</i>	<i>r_c</i>	<i>ρ_i</i>	<i>SD_ρ</i>	% <i>VE</i>	90% <i>CV</i>
Normative forced-choice–GPA									
Extraversion	4	1163	−.06	.08	−.09	−.10	.09	53	.01
Openness to experience	3	1100	.00	.07	.01	.01	.07	55	−.08
Agreeableness	3	1100	.06	.02	.08	.09	.00	100	.08
Conscientiousness	3	1100	.12	.07	.16	.18	.04	83	.12
Quasi-ipsative–GPA									
Emotional stability	5	1813	.06	.07	.09	.10	.07	60	.00
Extraversion	5	1813	−.13	.04	−.19	−.21	.00	100	−.19
Openness to experience	4	1681	.19	.04	.28	.31	.00	100	.28
Agreeableness	4	1681	.01	.04	.02	.02	.00	100	.02
Conscientiousness	6	2140	.13	.08	.19	.21	.07	67	.11
Pure ipsative–GPA									
Emotional stability	9	2103	.01	.06	.02	.02	.00	100	.02
Extraversion	11	2493	−.01	.09	−.02	−.03	.10	52	.09
Openness to experience	9	2103	−.05	.09	−.07	−.07	.09	55	.03
Agreeableness	10	2364	.00	.09	.01	.01	.08	58	−.08
Conscientiousness	16	3074	.12	.11	.16	.18	.12	49	.02

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

TABLE 8
Results of the meta-analyses of the Big Five–forced-choice type–criteria combinations in occupational studies

Personality dimension	<i>K</i>	<i>N</i>	<i>r_w</i>	<i>SD_r</i>	<i>r_c</i>	<i>ρ_i</i>	<i>SD_ρ</i>	% <i>VE</i>	90% <i>CV</i>
Normative forced-choice–job performance ratings									
Extraversion	5	2122	−.02	.11	−.03	−.04	.18	20	.17
Openness to experience	4	2059	.05	.03	.08	.09	.00	100	.08
Agreeableness	4	2059	.04	.06	.07	.07	.07	54	−.02
Conscientiousness	5	2352	.07	.07	.12	.13	.09	54	.02
Quasi-ipsative–job performance ratings									
Emotional stability	21	3407	.06	.13	.11	.12	.22	37	−.14
Extraversion	17	3122	.14	.12	.27	.30	.16	52	.08
Openness to experience	11	1691	.11	.21	.18	.19	.35	17	−.23
Agreeableness	14	2752	.06	.12	.11	.12	.18	38	−.10
Conscientiousness	26	4119	.19	.13	.36	.39	.15	60	.18
Pure ipsative–job performance ratings									
Emotional stability	29	6376	.02	.07	.03	.04	.04	90	−.01
Extraversion	28	6308	.06	.08	.10	.11	.07	72	.01
Openness to experience	27	5876	.06	.08	.09	.10	.07	70	.01
Agreeableness	26	5967	.00	.09	.00	.00	.10	57	−.11
Conscientiousness	29	6515	.07	.09	.10	.12	.10	57	−.01
Quasi-ipsative–training									
Emotional stability	5	1625	.11	.10	.15	.17	.12	39	.01
Extraversion	5	1625	.00	.04	.00	.00	.00	100	.00
Openness to experience	4	586	.17	.03	.22	.24	.00	100	.22
Agreeableness	4	586	.14	.14	.18	.19	.16	35	−.01
Conscientiousness	5	1625	.09	.10	.13	.14	.12	36	−.01
Pure ipsative–training									
Emotional stability	4	298	−.05	.05	−.07	−.08	.00	100	−.07
Extraversion	5	491	.03	.08	.05	.06	.00	100	.05
Openness to experience	4	245	.10	.19	.14	.15	.20	50	−.10
Agreeableness	4	407	.07	.09	.10	.11	.00	100	.10
Conscientiousness	3	262	.22	.04	.33	.36	.00	100	.33

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

As can be seen in Table 7, conscientiousness predicted GPA with the three types of FC methods, and it showed validity generalization in the three cases. Quasi-ipsative inventories were slightly more valid than normative FC and ipsative measures (.21 vs. .18), although the validity is small in the three cases. The validity size of conscientiousness is similar to that found in previous meta-analyses (e.g., O'Connor & Paunonen, 2007; Poropat, 2009; Salgado, 2000). Another relevant result is that openness to experience, as assessed with quasi-ipsative questionnaires, proved to be the most important predictor of GPA, with a true validity coefficient of .31 and also showed validity generalization (90%CV equals .28). Another relevant finding is that extraversion predicts negatively GPA and generalizes its validity.

In Table 8, the results of the meta-analyses for the triple combination (personality dimension, type of FC measure, and occupational criterion type) can be observed. We used two occupational criteria: job performance ratings and training proficiency. In the case of job performance ratings, quasi-ipsative measures proved to be excellent predictors of this criterion. Conscientiousness showed a true validity of .39, which is practically identical to the validity found when the criterion type was not used as moderator (see Table 6). The main difference between the last finding and the present one is that artifactual errors now explain much more of the observed variance (60% vs. 30%) and, consequently, the 90%CV is larger now than in Table 6. Conscientiousness, as assessed by normative FC and ipsative questionnaires, showed a smaller validity coefficient (.13 and .12, respectively). A relevant, but unexpected, result was found for extraversion, when assessed with quasi-ipsative measures, as its true validity was .30, and it also showed validity generalization (90%CV = .08). Until now, a validity coefficient of this magnitude for extraversion had never been found. Our speculation is that the validity is due to the fact that a number of primary studies dealt with occupations in which interpersonal relationships are important (e.g., sales and managerial jobs). Future research should examine this finding with a more microanalytic orientation. Limitations of the database and the objectives of this research preclude following the analysis beyond this point.

With regard to training proficiency, the limitations of our database only allow us to conduct meta-analyses for quasi-ipsative and ipsative measures. Furthermore, the number of studies and the sample sizes are small in various combinations. Therefore, the findings should be considered with caution. For quasi-ipsative measures, conscientiousness, agreeableness, emotional stability, and openness to experience showed a small validity, ranging from .14 to .24. Openness to experience showed the largest coefficient in this

case. For ipsative measures, conscientiousness showed a validity of .36 and all observed variance was explained by artifactual errors. However, as mentioned earlier, the meta-analysis was conducted with three studies and the total sample size is only 262 individuals. Consequently, this validity coefficient may change as the number of studies and the sample grow.

DISCUSSION

Over 20 years of meta-analyses have demonstrated that personality measures are valid predictors of academic and occupational performance. Empirical evidence has also been found that conscientiousness and emotional stability show validity generalization across criteria, samples, and occupations. The past meta-analytic evidence also showed that several variables moderated personality validity, including criterion type, inventory type (e.g., Big Five, COPS), occupation type, and complexity level. Additionally, the use of the Five-Factor Model of personality proved to be a useful framework for classifying the various measures designed under different conceptual models (see Salgado & De Fruyt, 2005, for a general review).

Nevertheless, a number of researchers have questioned the utility of personality measures for academic and personnel selection decisions (e.g., Morgeson et al., 2007a; Murphy & Dziewieczynski, 2005). Among other things, personality measures were criticized because the validity size was generally small, and because individuals can fake their responses, giving the most appropriate response to be hired. On the other hand, previous meta-analyses did not distinguish between SS and FC personality measures, and they did not estimate the validity of FC personality questionnaires. Furthermore, some critics of personality measures for making academic and occupational decisions have also suggested that the criterion domain should be expanded beyond the traditional job performance ratings when the validity of personality variables is examined (e.g., K. R. Murphy, in Morgeson et al., 2007a). In line with this point of view, in the present meta-analytical effort we try to include, as far as was possible, other less used occupational criteria that reflect individual behaviours and outputs.

In connection with Hypothesis 1, our findings clearly supported the idea that conscientiousness is a valid predictor for all criteria and it generalizes its validity. The hypothesis that emotional stability would be a predictor of all criteria was not supported, as we found that it did not predict job performance ratings, training proficiency, or productivity data. Therefore, the findings suggest the role of conscientiousness as a key variable for a theory of work and academic performance. Our results indicate that

around 16% of job performance variance is due to the effects of conscientiousness. This clearly contrasts with around 5% of variance found in the meta-analyses by Barrick and Mount (1991), Hurtz and Donovan (2000), and Salgado (1997) among others (see also the reanalysis by Schmidt et al., 2008). Hypothesis 2 was also supported, as openness to experience was shown to be a predictor of training proficiency and it generalized its validity. Hypothesis 3 was also confirmed, as quasi-ipsative inventories proved to be more valid than ipsative ones.

In view of the objectives of this research, this meta-analysis makes some unique contributions. The most important contribution of this study is to show that when conscientiousness is assessed with quasi-ipsative FC questionnaires, this personality dimension is the best single personality predictor of academic and job performance. The validity size for predicting job performance is almost double that of the validity found in classic meta-analyses of personality and job performance relationships (e.g., Barrick & Mount, 1991; Hough, 1992; Salgado, 1997; see also Schmidt et al., 2008, for the results using IRR correction). Moreover, the validity for the quasi-ipsative measures of conscientiousness is slightly larger than the validity of other common predictors used for personnel and academic decisions, including assessment centres (Glauger, Rosenthal, Thornton, & Bentson, 1987), structured interviews (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994), situational judgement tests (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), and similar or slightly larger than the validity found for the integrity test and other COPS (Ones & Viswesvaran, 2001b; Ones et al., 1993).

A second contribution was to show that other personality dimensions, such as extraversion, agreeableness, and openness to experience, were also predictors of specific criteria, with validity sizes larger than the values found for SS inventories (e.g., Barrick & Mount, 1991; Barrick et al., 2001; Hurtz & Donovan, 2000; Salgado, 1997). These two contributions suggest that quasi-ipsative questionnaires are very robust procedures for assessing personality in academic and occupational settings. Until now, no previous meta-analyses had examined this issue. It was therefore unknown that quasi-ipsative questionnaires were such excellent methods for making personnel decisions.

The comparison between normative FC and ipsative inventories suggests that the latter are similarly or slightly more valid, although the number of studies is relatively small to reach a firm conclusion. Comparing the results of the previous meta-analyses, mainly conducted with SS inventories, with those of the normative and ipsative FC inventories, it cannot be concluded that the SS questionnaires are superior in validity.

Therefore, taking these findings as a whole, it can be said that FC measures of the Big Five, especially in the quasi-ipsative form, are useful tools for making academic and occupational decisions because their validity is similar or even larger than the validity of SS inventories, and because they present some singularities (e.g., the relationship between agreeableness and productivity) that can contribute to improving the validity of a compound of cognitive and personality measures.

Additionally, the findings reported in the section on methodology (see Table 1) demonstrated that the reliability coefficients of the FC inventories are similar to the coefficients found for SS inventories in previous personality validity meta-analyses (e.g., Barrick & Mount, 1991; Hough, 1992; Salgado 1997) and in meta-analysis of personality reliability coefficients (e.g., Viswesvaran & Ones, 1999). This finding is relevant as some researchers had claimed that FC inventories showed inflated coefficients (e.g., Johnson et al., 1988; Tenopir, 1988), whereas other researchers had suggested that the reliability coefficients could be slightly smaller than those of normative inventories (e.g., Bartram, 1996). Our results totally agree with Bartram's perspective. The finding is also important in connection with the possible effects on the size of validity coefficients. The effects of measurement error are practically the same for SS and FC inventories.

It is also interesting to compare the validity of quasi-ipsative measures of conscientiousness with the validity of general mental ability (GMA). Different meta-analyses have shown that the observed average validity of GMA is around .22 ~ .25 (e.g., Hartigan & Wigdor, 1989; Hunter & Hirsh, 1987; Hunter & Hunter, 1984; Salgado et al., 2003; Schmitt et al., 1984). The observed average validity of the quasi-ipsative measures of conscientiousness was .23, a value very similar to the one found for GMA. However, as Schmidt et al. (2008) showed, an important difference is that the range restriction values of GMA are very much larger than the respective values of personality measures. This last characteristic probably contributes to producing the larger validity of GMA tests. Furthermore, personality measures are usually less reliable.

Our findings as a whole, considered together with the findings by Nguyen and McDaniel (2000), according to which FC inventories are less fackable than SS questionnaires, have important implications for the conditions that FC measures should fulfil to be used in practical situations (see Hicks, 1970). First, SS inventories are more affected by faking than FC inventories. Second, the validity of FC measures of conscientiousness (especially if they are quasi-ipsative inventories) is similar to or larger than the validity of SS inventories in some cases. Thus, it can be

concluded that FC inventories can be a good alternative to SS questionnaires for making academic and personnel decisions.

Study limitations and strengths

This research also has some limitations that should be mentioned. First, some criteria were not examined (i.e., job absenteeism or tenure) because no studies reported validity for these criteria. Second, the number of studies is small in some cases or the total sample size is small (see Salgado, 1998b). Consequently, the estimates reported here could change as the number of studies or the sample size is enlarged. Third, we did not examine whether FC inventories based on the FFM have larger validity than FC inventories not based on the FFM. Previous research found that SS inventories based on the FFM showed larger criterion validities (Salgado, 2003; see also Warr, Bartram, & Brown, 2005a), and the same can be expected for FC inventories. Fourth, the current meta-analysis did not examine the effect of using FC methodology in both predictor and criteria, and some studies showed that the validity is larger when both measures used FC formats (e.g., Bartram, 2005, 2007). Fifth, readers should note that a potential limitation is that personality instruments differ in design and in terms of the variables and construct measured. An anonymous reviewer suggested that there are possibly instrument design issues confounded with FC scoring formats. The present meta-analysis cannot control these issues with the current database. Future studies should examine these potential confounding issues, using appropriate inventories and scoring systems. As the anonymous reviewer noted, the new scoring model for the OPQ32r provides the possibility of comparing the validities of FC ipsative and FC normative scores from the same instrument and the same set of data (see the OPQ32r manual for examples of rescoring; SHL, 2009). This could be the basis for future studies.

Many of the strengths of this research have already been mentioned, but we wish to emphasize here that we used a very comprehensive database of primary studies, that this is the first meta-analysis which differentiates the various types of FC measures, and that the combination of personality dimensions, criterion type, and FC type give important clarification on the role of personality dimensions at work.

Ipsative, quasi-ipsative, and normative scores in personality inventories

This meta-analysis has shown that FC measures predict academic and occupational performance. However, these findings do not mean that FC

measures can be unreservedly used for making academic and personnel decisions, as a number of issues were not answered in the present study and future research should be conducted to respond to them. For example, it will be necessary to investigate if there are any gains in using SS and FC inventories simultaneously. Additional research should also be done on the effects of faking on FC inventories, as suggested by Nguyen and McDaniel (2000). More importantly, the problem that FC measures may produce ipsative scores, and that these scores could posit difficulties for comparing individuals, is not solved with this meta-analysis. In other words, this meta-analysis does not allow us to resolve the problem of recovering normative scores from ipsative or quasi-ipsative scores. Of course, if normative information can be recovered from ipsative scores, the FC format would have an advantage over the SS format and, therefore, it would be the format of preference. However, methodological advances have been made only very recently in this direction, mainly in the use of IRT technology. In the last few years, two approaches for using IRT approaches for creating and scoring FC inventories have been suggested and both of them are very promising. The first approach relies on an ideal-point response process (Chernyshenko, Stark, Drasgow, & Roberts, 2007; McCloy et al., 2005; Stark et al., 2005). The second approach deals with items of dominance (Brown & Bartram, 2009; Brown & Maydeu-Olivares, 2011; Maydeu-Olivares & Brown, 2010); Maydeu-Olivares & Böckenholt, 2005. Both approaches are based on the assumptions posited by Thurstone (1928) for the measurement of attitudes.

These recent IRT techniques applied to ipsative and quasi-ipsative formats can also have implications for the assessment of global dimensions rather than facets. In this regard, Brown and Maydeu-Olivares (2011) pointed out that "given the same number of traits, the lower the average correlation between them the better the true scores are recovered". This may have an important consequence for the assessment of the Big Five and facets using ipsative measures. At the construct level, the Big Five are orthogonal or relatively independent of each other. Therefore, the latter condition pointed out by Brown and Maydeu-Olivares can be fulfilled with measures at the construct level. However, by definition and empirically, the facets of the personality dimensions largely correlated among themselves as a consequence of the latent variable (i.e., the Big Five). Consequently, ipsative measures may be less appropriate for recovering the true scores of the facets.

Future research should also be devoted to the important issue of whether SS, quasi-ipsative, and ipsative measures of the same factor are really assessing the same content. The question of whether

FC and SS personality measures assess different aspects or elements of the same personality construct or they essentially assess the same elements is an issue that is not examined in this meta-analysis. Consequently, based on our results, we cannot make firm conclusions about this question. Previous research did not respond to it either. Therefore, this is an open matter, relevant from both the theoretical and practical points of view. From the theoretical point of view, if SS and FC measures cover different aspects of a specific personality dimension, then it makes sense to put together the two types of measures in order to have a better assessment of personality dimensions. On the other hand, from a practical point of view, if FC and SS measures are correlated but they are relatively independent, then FC measures could show incremental validity over SS measures for predicting academic and occupational performance. In our opinion, this last issue is one that deserves future research. Both types of techniques may be linked to differences in the way that an individual processes information. In the case of FC scores, individuals would focus on comparing alternatives to each other in order to create a ranking or hierarchy. In the case of SS scores, and according to Hogan's (2001; Johnson & Hogan, 2006) conceptualization, individuals would manage their presentation to other people. Meglino and Ravlin (1998) suggested that each methodology models a different cognitive phenomenon, and even when it is technically possible to turn SS scores into ipsative ones and to recover normative data from ipsative data, this does not mean that the process recovers the cognitive phenomenon that was modelled by each kind of measure. Furthermore, as Brown and Maydeu-Olivares (2011) have suggested, FC formats are more cognitively demanding than SS formats. In the case of FC formats, the block size is positively related with the cognitive load of the responder because there are $n(n-1)/2$ binary comparisons. Consequently, comparing FC and SS responses to the same items might produce erroneous conclusions because the two formats would require cognitive processing that may be or may not be less appropriate for the goal of the study. In this view, SS and FC methodologies would be complementary rather than alternative methodologies.

CONCLUDING COMMENTS

In summary, this meta-analysis comprehensively reviewed the validity of the Big Five personality dimensions as assessed with FC measures for predicting academic performance (e.g., GPA) and occupational performance (e.g., job performance ratings, training proficiency, productivity). The results showed that the quasi-ipsative FC measures of

conscientiousness are more valid than any other type of measures, including SS, normative FC, and ipsative ones. The validity size of quasi-ipsative measures of conscientiousness is almost twice that of the validity of SS inventories, and equal to or larger than the validity of assessment centres, structured interviews, situational judgement tests, and personality composites (i.e., integrity, COPS). The results of the reliability estimates indicated that FC measures do not inflate or deflate reliability coefficients. Therefore, as a whole, quasi-ipsative FC measures can be seen as useful tools for making academic and personnel decisions, and they can be seen as a robust alternative to SS inventories because they are more resistant to faking.

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.
- *Alonso, P. (2011). *Effects of social desirability in normative and ipsative tests*. Unpublished master's dissertation, Department of Social Psychology, University of Santiago de Compostela, Spain.
- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology. Why do we need them? What might they be? *The American Psychologist*, 63, 839-851. doi:10.1037/0003-066X.63.9.839
- Anderson, N., & Sleap, S. (2004). An evaluation of gender differences on the Belbin Team Role Self-Perception Inventory. *Journal of Occupational and Organizational Psychology*, 77, 429-437.
- *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.
- *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.
- *Bachman, J. G. (1964). Prediction of academic achievement using the Edwards Need Achievement Scale. *Journal of Applied Psychology*, 48, 16-19.
- *Balch, D. E. (1977). *Personality trait differences between successful and non-successful police recruits at a typical police academy and veteran police officers*. Unpublished doctoral dissertation, US International University, Kenya.
- Baron, H. (1996). Strengths and limitation of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56.
- Barrick, M. R. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, 18, 359-372.
- 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.
- Belbin, M. (1981). *Management teams: Why they succeed or fail*. London, UK: Heinemann.
- Bennett, M. (1977). Testing management theories cross-culturally. *Journal of Applied Psychology*, 62, 578–581.
- *Bhatnagar, R. P. (1969). A study of some EPPS variables as factors of academic achievement. *Journal of Applied Psychology*, 53, 107–111.
- Brogden, H. E. (1954). A rationale for minimizing distortion in personality questionnaire keys. *Psychometrika*, 19, 151–148.
- *Brown, A., & Bartram, D. (2009, April). *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, LA.
- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71, 460–502. doi:10.1177/0013164410375112
- *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.
- Carretta, T., & Ree, M. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment*, 9, 325–335.
- Cattell, R. B. (1944). Psychological measurement: Ipsative, normative and interactive. *Psychological Review*, 51, 292–303.
- 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.
- Chan, W., & Bentler, P. M. (1993). The covariance structure analysis of ipsatives. *Sociological Methods and Research*, 22, 214–247.
- Chapman, D. W., Blackburn, R. W., Austin, A. E., & Hutcheson, S. M. (1983). Expanding analytic possibilities of Rokeach values data. *Educational and Psychological Measurement*, 43, 419–421. doi:10.1177/001316448304300211
- Chernyshenko, O., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumption of an ideal point response process: Towards increasing the flexibility of personality measures. *Psychological Assessment*, 19, 88–106.
- Cheung, M. W.-L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling*, 9, 55–77.
- *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, 355–366.
- Clemons, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs*, 14, 1–56.
- *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.
- *Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality tests and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, 16, 155–169.
- *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
- *Davis, K. R., & Banken, J. A. (2005). Personality type and clinical evaluations in an obstetrics/gynecology medical student clerkship. *American Journal of Obstetrics and Gynecology*, 193, 1807–1810.
- *Dodd, W. E., Wollowick, H. B., & McNamara, W. J. (1970). Task difficulty as a moderator of long-range prediction. *Journal of Applied Psychology*, 54, 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). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.
- Edwards, A. L. (1973). *Edwards personal preference schedule manual*. New York: Psychological Corporation.
- Feather, N. T. (1973). The measurement of values: Effects of different assessment procedures. *Australian Journal of Psychology*, 25, 221–231.
- *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 presented at the Division of Occupational Psychology conference, British Psychological Society, Bournemouth, UK.
- French, J., & Raven, B. H. (1959). The basis of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150–167). Ann Arbor, MI: University of Michigan, Institute for Social Research.
- *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.
- *Gallesich, J. (1970). An investigation of correlates of academic success of freshmen engineering students. *Journal of Counseling Psychology*, 17, 173–178.
- *García-Izquierdo, A. L. (2001). *Validación orientada al criterio de procedimientos de selección de personal de oficio para el pronóstico del rendimiento laboral y formativo en el sector de la construcción*. Unpublished doctoral dissertation, University of Murcia, Murcia, Spain.

- *Gebhart, C. G., & Hoyt, D. T. (1958). Personality needs of under- and over-achieving freshman. *Journal of Applied Psychology*, 42, 125–128.
- 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
- Glaugier, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511. doi:10.1037/0021-9010.72.3.493
- Gleser, L. J. (1972). On bounds for the average correlation between subtest scores in ipsatively scored tests. *Educational and Psychological Measurement*, 32, 759–765.
- *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.
- *Goodstein, L. D., & Heilbrum, A. B. (1962). Prediction of college achievement from the Edwards Personal Preference Schedule at three levels of intellectual ability. *Journal of Applied Psychology*, 46, 317–320.
- *Gordon, L. V. (1993). *Gordon Personal Profile-Inventory: Manual 1993 revision*. San Antonio, TX: 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, South Orange, NJ.
- *Hakel, M. (1966). Prediction of college achievement from the Edwards Personal Preference Schedule using intellectual ability as a moderator. *Journal of Applied Psychology*, 50, 336–340.
- Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academies Press.
- Hayes, W. O. (1967). *Quantification in psychology*. Belmont, CA: Brooks/Cole.
- *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.
- Heggstad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91, 9–24.
- Herscovis, M. S., Turner, N., Barling, J., Arnold, K. A., Dupré, K. E., Inness, M., et al. (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.
- *Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with “fake-proof” measure of the Big Five. *Journal of Research in Personality*, 42, 1323–1333. doi:10.1016/j.jrp.2008.04.006
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270–1285.
- 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, 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, R. T. (2001). Personality and industrial and organizational psychology. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 3–16). Washington, DC: American Psychological Association.
- Hogan, R. (2005a). Comments. *Human Performance*, 16, 405–407. doi:10.1207/s15327043hup1804_6
- Hogan, R. (2005b). In defense of personality measurement: New wine for old whiners. *Human Performance*, 18, 331–341. doi:10.1207/s15327043hup1804_1
- Hogan, R., & Hogan, J. (2001). Assessing leadership: A view of the dark side. *International Journal of Selection and Assessment*, 9, 40–51.
- Horn, J. L. (1971). Motivation and dynamic calculus concepts from multivariate experiment. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (2nd printing, pp. 611–641). Chicago, IL: Rand McNally.
- 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.
- Hough, L. M., & Oswald, F. L. (2005). They’re right, well... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, 18(4), 373–387. doi:10.1207/S15327043hup1804_4
- Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79(2), 184–190. doi:10.1037/0021-9010.79.2.184
- *Hughes, G. L., & Prien, E. P. (1986). An evaluation of alternate scoring methods for the mixed standard. *Personnel Psychology*, 39, 839–847.
- *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.
- 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, pp. 321–357). 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 ed.). 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.

- *Izard, C. E. (1962). Personality characteristics (EPPS), level of expectation, and performance. *Journal of Consulting Psychology*, 26, 394.
- Jackson, D. J., & Alvin, D. F. (1980). The factor analysis of ipsative measures. *Sociological Methods and Research*, 9, 218–238.
- Jackson, D. N. (2002). *Employee screening questionnaire manual*. Port Huron, MI: Sigma Assessment Systems, ESQ.
- *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). Spuriousness and spuriousness: the use of ipsative personality tests. *Journal of Occupational Psychology*, 61, 153–162.
- Johnson, J. A., & Hogan, R. T. (2006). A socioanalytic view of faking. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 209–231). Charlotte, NC: IAP.
- 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
- Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral roles of individual differences. *Journal of Applied Psychology*, 92, 107–127. doi:10.1037/0021-9010.92.1.107
- *Kahn, J. H., Nauta, M. M., Gailbreath, R. D., Tipps, J., & Chartrand, J. M. (2002). The utility of career and personality assessment in predicting academic progress. *Journal of Career Assessment*, 10, 3–23. doi:10.1177/1069072702010001001
- Kaplan, S., Bradley, J. C., Luschman, J. N., & Haynes, D. (2010). On the role of positive and negative affectivity in job performance: A meta-analytic investigation. *Journal of Applied Psychology*, 94, 162–176. doi:10.1037/0021-9010.94.1.162
- *Kazmier, L. J. (1961). Cross-validation groups, extreme groups, and the prediction of academic achievement. *Journal of Educational Psychology*, 52, 195–198.
- King Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, 65, 507–516.
- Knapp, D., Heggstad, 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.
- *Knauff, E. B. (1955). Test validity over seventeen-year period. *Journal of Applied Psychology*, 39, 382–383.
- Kolb, D. A. (1985). *Learning style inventory*. Boston, MA: McBer & Co.
- *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.
- Kuder, F. (1975). *Manual: Kuder E general interest survey*. Chicago, IL: Science Research associates.
- *Kusch, R. I., Deller, J., & Albrecht, A. G. (2008, July). *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.
- 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
- *Lunneborg, P. W. (1970). EPPS patterns and academic achievement in counseling clients. *Educational and Psychological Measurement*, 30, 393–398. doi:10.1177/0013164470030000223
- *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
- Maydeu-Olivares, A., & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioral Research*, 45, 935–974. doi:10.1080/00273171.2010.531231
- Maydeu-Olivares, A., & Böckenholt, U. (2005). Structural equation modeling of paired comparison and ranking data. *Psychological Methods*, 10, 285–304.
- McCloy, R., Heggstad, E., & Reeve, C. (2005). A silk purse from the sow's ear. Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, 8, 222–248.
- McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood*. New York, NY: Guilford Press.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599–616.
- *McDaniel, M. A., Yost, A. P., Ludwick, M. H., Hense, R. L., & Hartman, N. S. (2004, April). *Incremental validity of a situational judgment test*. Paper presented at the 19th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.
- 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.
- Meglino, B. M., & Ravlin, E. C. (1998). Individual values in organizations: concepts, controversies, and research. *Journal of Management*, 24, 351–389.
- *Morgan, R. R. (1975). Prediction of college achievement using the need achievement scale from the Edwards Personal Preference Schedule. *Educational and Psychological Measurement*, 35, 387–392. doi:10.1177/001316447503500217

- *Morgan, R. R. (1976). Utilization of levels of intellectual ability as a control variable in studies of non-intellectual factors in academic achievement. *Educational and Psychological Measurement*, 36, 465–472. doi:10.1177/001316447603600229
- 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.
- Moscoso, S., & Salgado, J. F. (2004). "Dark side" personality styles as predictors of task, contextual, and job performance. *International Journal of Selection and Assessment*, 12, 356–362.
- Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management*, 13, 153–200.
- *Muchinsky, P. M., & Hoyt, D. P. (1973). Predicting college grades of engineering graduates from selected personality and aptitude variables. *Educational and Psychological Measurement*, 33, 935–937. doi:10.1177/001316447303300425
- *Mukherjee, R. N. (1968). Achievement values and scientific productivity. *Journal of Applied Psychology*, 52, 145–147.
- Murphy, K. R., & De Shon, R. (2000). Inter-rater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53, 873–900.
- Murphy, K. R., & Dziewieczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18, 343–357.
- 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* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- *Nelson, C. A. (2008). *Job type as a moderator of the relationship between situational judgment and personality*. Unpublished doctoral dissertation, Capella University, Minneapolis, MN.
- *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.
- Ng, T. W. H., Sorensen, K. L., & Eby, L. T. (2006). Locus of control at work: A meta-analysis. *Journal of Organizational Behavior*, 27, 1057–1087.
- Nguyen, N. T., & McDaniel, M. A. (2000, April). *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.
- 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 ed.). New York, NY: McGraw-Hill.
- *Nyfield, G., Gibbons, P. J., Baron, H., & Robertson, I. (1995). *The cross-cultural validity of management assessment methods*. Surrey, UK: Saville & Holdsworth.
- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43, 971–990. doi:10.1016/j.paid.2007.3.017
- Oh, I.-S., Schmidt, F. L., Mount, M. K., Le, H., Guay, R. P., Takahashi, K., et al. (2011, April). *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, Chicago, IL.
- *Olson, D. A., Shultz, K. S., & Scott, G. (2000, April). *The association between personality preferences and behavioral ratings for physician leaders*. Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Ones, D. S. (1993). *The construct of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995–1027. 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, GA.
- Ones, D. S., & Viswesvaran, C. (2001a). 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. (2001b). Personality at work: Criterion-focused occupational personality scales (COPS) used in personnel selection. In B. Roberts & R. T. Hogan (Eds.), *Applied personality psychology* (pp. 63–92). Washington, DC: American Psychological Association.
- Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, 18, 389–404. doi:10.1207/S15327043hup1804_5
- 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.
- Poropat, A. E. (2009). A meta-analysis of the Five-Factor Model of personality and academic performance. *Psychological Bulletin*, 135, 322–338. doi:10.1037/a0014996
- Radcliffe, J. A. (1963). Some properties of ipsative score matrices and their relevance for some current interest tests. *Australian Journal of Psychology*, 15, 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., & De Fruyt, F. (2003). The validity of FFM personality dimensions and maladaptive traits to predict negative affects at work: A six month prospective study in a military sample. *European Journal of Personality*, 17, 101–121.
- *Rolland, J. P., & Mogenet, J. L. (2001). *Système de description en cinq dimensions (D5D). Manuel réservé aux psychologues*. Paris, France: Les Editions du Centre de Psychologie Appliquée.

- *Rothe, H. F. (1946). Output rates among butter wrappers. *Journal of Applied Psychology*, 30, 320-327.
- *Rothe, H. F. (1947). Output rates among machine operators. *Journal of Applied Psychology*, 31, 484-489.
- *Rothe, H. F. (1951). Output rates among chocolate dippers. *Journal of Applied Psychology*, 35, 94-97.
- *Rothe, H., & Nye, C. T. (1958). Output rates among coil winders. *Journal of Applied Psychology*, 42, 182-186.
- Rothmann, S., Meiring, D., Van der Walt, H. S., & Barrick, M. R. (2002). *Predicting job performance using personality measures in South Africa*. Paper presented at the 7th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.
- *Rose, R. M., Fogg, L. F., Helmreich, R. L., & McFadden, T. J. (1994). Psychological predictors of astronaut effectiveness. *Aviation, Space, and Environmental Medicine*, 65, 910-915.
- *Rozelle, R. (1968). The relationship between absenteeism and grades. *Educational and Psychological Measurement*, 28, 1151-1158. doi:10.1177/001316446802800415
- *Rust, J. (1999). The validity of the Giotto integrity test. *Personality and Individual Differences*, 27(4), 755-768.
- 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). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- *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., & Schmitt, N. (2012). On reconciling conflicting meta-analytic findings regarding integrity test validity. *Journal of Applied Psychology*, 97, 550-556. doi:10.1037/a0028167
- *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, Spain.
- 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. (2000, April). *The Big Five personality dimensions as predictors of alternative criteria*. Paper presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- 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. (2004). *Moderator effects of job complexity on the Big Five validity*. Poster presented at the annual conference of Society for Industrial and Organizational Psychology, Chicago, IL.
- Salgado, J. F. (2005). Personality and social desirability in organizational settings: Practical implications for work and organizational psychology. *Papeles del Psicólogo*, 26, 115-128.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of GMA validity for different occupations in the European Community. *Journal of Applied Psychology*, 88, 1068-1081. doi:10.1037/0021-9010.88.6.1068
- Salgado, J. F., & De Fruyt, F. (2005). Personality in personnel selection. In A. Evers, O. Schmit-Voskuyl, & N. Anderson (Eds.), *Handbook of personnel selection* (pp. 174-197). Oxford, UK: Blackwell.
- 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 Skills*, 83, 1195-1201.
- Salgado, J. F., & Moscoso, S. (2000). Autoeficacia y criterios organizacionales de desempeño [Self-efficacy and organizational criteria of performance]. *Apuntes de Psicología*, 18, 179-191.
- Saville, P., Holdsworth, R., Nyfield, G., Cramp, L., & Mabey, W. (1984). *The Occupational Personality Questionnaires (OPQ)*. London, UK: Saville & Holdsworth.
- *Saville, P., Sik, G., Nyfield, G., Hackston, J., & MacIver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in the measurement of job competencies across time and in separate organisations. *Applied Psychology*, 45, 243-262.
- *Schippmann, J. S., & Prien, E. P. (1989). An assessment of the contributions of general mental ability and personality characteristics to management success. *Journal of Business and Psychology*, 3, 423-437.
- 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., & Urry, V. W. (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, IA: 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 Psychology*, 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.
- *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.
- *SHL. (2006). *OPQ32 manual and user's guide*. Surrey, UK: Author.
- SHL. (2009). *OPQ32r user manual*. Surrey, UK: Author.
- *Slocum, J. W., Jr., & 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. Ann Arbor, MI: Employment Testing Consortium Project.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, 29, 184–203.
- Stark, S., Chernyshenko, O., Drasgow, F., & Williams, B. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25–39. doi:10.1037/0021-9010.91.1.25
- *Striker, L. J., Schiffman, H., & Ross, J. (1965). Prediction of college performance with the Myers-Briggs Type Indicator. *Educational and Psychological Measurement*, 25, 1081–1095.
- Tenopir, M. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology*, 73, 749–751.
- 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.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967–993.
- Tett, R. P., Christiansen, N. D., Robie, C., & Simonet, D. V. (2011, May). *International survey of personality test use: An American baseline*. Paper presented at the 15th conference of the European Association of Work and Organizational Psychology, Maastricht, The Netherlands.
- Thomas, K. W., & Killman, R. H. (1974). *Thomas-Killman Conflict Mode Instrument*. Mountain View, CA: Xicom.
- Thompson, B., Levitov, J. E., & Miederhoff, P. A. (1982). Validity of the Rokeach Value Survey. *Educational and Psychological Measurement*, 42, 899–905.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–554.
- Trapmann, S., Hell, B., Hirn, J.-O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift für Psychologie*, 215, 132–151.
- Travers, R. M. W. (1951). A critical review of the validity and rationale of forced-choice technique. *Psychological Bulletin*, 48, 62–70.
- Van der Walt, H. S., Meiring, D., Rothmann, S., & Barrick, M. R. (2002, June). *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, South Africa.
- *Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, 19, 175–199.
- Villanova, P., Bernardin, H. J., Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnel. *Personnel Psychology*, 60, 224–235.
- Viswesvaran, C., & Ones, D. S. (1999). Measurement error in “Big Five factors” personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60, 224–235.
- Viswesvaran, C., Ones, D. S., & 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
- *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
- *Warr, P., Bartram, D., & Martin, T. (2005b). Personality and sales performance: Situational variation and interactions between traits. *International Journal of Selection and Assessment*, 13, 87–91.
- *Weiss, P., Wertheimer, M., & Groesbeck, B. (1959). Achievement motivation, academic aptitude, and college grades. *Educational and Psychological Measurement*, 19, 663–666.
- *Whetzel, D. L., McDaniel, M. A., Yost, A. P., & Kim, N. (2010). Linearity of personality-performance relationships: A large-scale examination. *International Journal of Selection and Assessment*, 18, 310–320.
- *White, L. A. (2002, October). *A quasi-ipsative temperament measure for assessing future leaders*. Paper presented at the 14th Annual Conference of the International Military Testing Association, Ottawa, Canada.
- *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 Psychology*, 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.
- *Young, M., & Dulewicz, V. (2007). Relationship between emotional and congruent self-awareness and performance in the British Royal Navy. *Journal of Managerial Psychology*, 22, 465–478.
- *Zagar, R., Arbit, J., & Wengel, W. (1982). Personality factors as predictors of grade point average and graduation from nursing school. *Educational and Psychological Measurement*, 42, 1169–1175.
- Zibarras, L. D., & Woods, S. A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology*, 83, 499–511.

Original manuscript received May 2012

Revised manuscript received June 2012

First published online October 2012

APPENDIX A

Classification of the personality scales according to the five-factor model

<i>Inventory</i>	<i>Scale</i>	<i>Direction</i>	<i>Inventory</i>	<i>Scale</i>	<i>Direction</i>
Emotional stability			Agreeableness		
AIM	Adjustment	+	AIM	Agreeableness	+
Caliper	Resilience	+	CCSQ	Emphatic	+
CCSQ	Resilience	+	CCSQ	Participative	+
CSQ	Emotional sensitivity	+	CCSQ	Modest	+
D5D	Neuroticism	–	CES	Helping	+
EPPS	Succorance	+	CES	Concern for others	+
GPP-I	Emotional stability	+	D5D	Agreeableness	+
IPIP	Emotional stability	+	EPPS	Deference	+
OPQ	Relaxed	+	EPPS	Aggression	–
OPQ	Tough-minded	+	EPPS	Abasement	–
OPQ	Worrying	–	EPPS	Nurturance	+
OPQ	Optimistic	+	GPP-I	Personal relations	+
PAPI	Emotional stability	+	ICS	Competing	–
SDI	Self-assurance	+	IPIP	Agreeableness	+
Extraversion			MBTI	Thinking–feeling	+
AIM	Leadership	+	OPQ	Caring	+
Caliper	Extraversion	+	OPQ	Democratic	+
CCSQ	Sociable	+	OPQ	Competitive	–
CCSQ	Persuasive	+	OPQ	Trusting	+
CSQ	Need to control	+	PAPI	Agreeableness	+
CSQ	Sociability	+	Conscientiousness		
CSQ	Group orientation	+	AIM	Dependability	+
D5D	Introversion	–	AIM	Work orientation	+
EPPS	Exhibition	+	Caliper	Time management	+
EPPS	Affiliation	+	CCSQ	Competitive	+
EPPS	Intracception	–	CCSQ	Results orientated	+
EPPS	Dominance	+	CCSQ	Energetic	+
GPP-I	Ascendancy	+	CCSQ	Structured	+
GPP-I	Sociability	+	CCSQ	Detail conscious	+
IPIP	Extraversion	+	CCSQ	Conscientious	+
MBTI	Extraversion–introversion	+	CSQ	Approach to organizing	+
OPQ	Outgoing	+	CSQ	Attitude to authority	+
OPQ	Affiliative	+	CSQ	Need for results	+
OPQ	Emotional control	–	CES	Achievement	+
OPQ	Persuasive	+	CES	Honesty	+
OPQ	Controlling	+	CES	Fairness	+
OPQ	Outspoken	+	D5D	Conscientiousness	+
OPQ	Socially confident	+	EPPS	Achievement	+
PAPI	Extraversion	+	EPPS	Order	+
Open to experience			EPPS	Endurance	+
CCSQ	Innovative	+	ESQ	Conscientiousness	+
CCSQ	Flexible	+	GIT	Integrity	+
CCSQ	Analytical	+	GPP-I	Responsibility	+
CSQ	Understanding people	+	GPP-I	Cautiousness	+
CSQ	Attitude to change	+	GPP-I	Vigour	+
D5D	Openness	+	IPIP	Conscientiousness	+
EPPS	Change	+	MBTI	Judging–perceiving	+
EPPS	Autonomy	+	OPQ	Detail conscious	+
GPP-I	Original thinking	+	OPQ	Conscientiousness	+
IPIP	Open to experience	+	OPQ	Forward planning	+
MBTI	Sensing – intuition	+	OPQ	Achieving	+
OPQ	Conceptual	+	OPQ	Vigorous	+
OPQ	Artistic	+	PAPI	Conscientiousness	+
OPQ	Behavioural	+	SCT	Achievement values	+
OPQ	Innovative	+	SDI	Initiative	+
OPQ	Conventional	–	SDI	Achievement	+
OPQ	Variety seeking	+	WPQ	Need for achievement	+
PAPI	Open to experience	+			

APPENDIX B

Main codes and input values for the primary studies included in the meta-analysis

Reference	N	Inventory	Format	Criterion	ES	EX	O	A	C	r_{33}	u
Adkins and Naumann (2001)	264	CES	I	SLS	—	—	—	—	.18	.77	.94
Antler, Zaretsky, and Ritter (1967)	30	GPI	Q	JPR	.58	—	—	—	—	—	—
Balch (1977)	100	EPPS	I	TRA	-.10	.12	-.10	-.01	.20	—	—
Bartram (2007) Korea	366	OPQ	I	JPR	.17	.21	.11	.13	.22	.52	.99
Bartram (2007) S. Africa	68	OPQ	I	JPR	.02	-.07	.02	-.09	.04	.52	—
Bartram (2007) USA	86	OPQ	I	JPR	-.04	.25	.09	.18	-.05	.52	—
Bennett (1977)	45	SDI	Q	JPR	.40	—	—	—	-.05	—	—
Bennett (1977)	49	SDI	Q	JPR	.46	—	—	—	.07	—	—
Bhatnagar (1969)	261	EPPS	I	GPA	—	-.11	—	.10	.13	—	—
Brown and Bartram (2009)	835	OPQ	I	SFR	.12	.13	.08	-.01	.15	—	—
Christiansen et al. (2005) a	60	IPIP	Q	JPR	—	—	—	—	.46	—	.62
Christiansen et al. (2005) b	62	IPIP	Q	JPR	—	—	—	—	.17	—	.63
Clevenger, Pereira, Wiechman, Schmitt, and Schmidt-Harvey (2001)	207	OPQ	I	JPR	—	—	—	—	.16	.76	—
Converse et al. (2008) a	132	Caliper	Q	GPA	.04	-.06	—	—	.13	—	—
Converse et al. (2008) b	132	Caliper	Q	CAB	-.18	.15	—	—	-.24	—	—
Conway (2000)	1567	MBTI	F	JPR	—	-.06	.06	.06	.06	—	1.00
Davis and Banken (2005) a	63	MBTI	F	GPA	—	.22	—	—	—	—	—
Davis and Banken (2005) b	63	MBTI	F	JPR	—	.35	—	—	—	—	—
Fine and Dover (2005)	193	ICS	I	TRA	—	.06	—	.02	—	—	1.00
Fineman (1975) a	293	WPQ	F	JPR	—	—	—	—	.21	—	—
Fineman (1975) b	246	WPQ	F	SAL	—	—	—	—	.20	—	—
Fineman (1975) c	84	WPQ	F	PRG	—	—	—	—	.25	—	—
Francis-Smythe, Tinline, and Allender (2002)	225	OPQ	I	JPR	.08	.08	.28	—	.16	—	—
Furnham (1994) a	176	CSQ	I	CWB	-.01	.03	.00	—	-.06	—	—
Furnham (1994) b	176	CSQ	I	JPR	.12	.10	.17	—	.23	—	—
Furnham and Stringfield (1993) a	222	MBTI	F	JPR	—	-.10	.09	.05	-.01	—	1.00
Furnham and Stringfield (1993) b	148	MBTI	F	JPR	—	.05	.02	.02	.12	—	1.00
García-Izquierdo (2001)	84	OPQ	I	TRA	-.04	.02	.26	—	.28	—	—
Goffin, Jan, and Skinner (2011)	114	ESQ	Q	CWB	—	—	—	—	-.21	—	—
Goodstein and Heilbrum (1962) a	206	EPPS	I	GPA	.06	.05	-.15	.01	.17	—	—
Goodstein and Heilbrum (1962) b	151	EPPS	I	GPA	.02	-.02	-.10	.01	-.02	—	—
Gordon (1993) T4.18	94	GPP	Q	PRG	-.18	.28	.05	.17	.10	—	—
Gordon (1993) T4.20	47	GPI	Q	JPR	-.37	—	—	—	-.26	—	—
Gordon (1993) T4.21	396	GPI	Q	SAL	-.07	.22	—	—	-.11	—	—
Gordon (1993) T4.22 a	101	GPI	Q	PRG	—	.25	—	—	—	—	—
Gordon (1993) T4.22 b	103	GPI	Q	PRG	—	.28	—	—	—	—	—
Gordon (1993) T4.23	146	GPP	Q	PRG	.03	.08	.23	.05	.16	—	—
Gordon (1993) T4.23 b	100	GPP	Q	PRG	.01	.27	.22	-.31	.15	—	—
Gordon (1993) T4.24 a	200	GPP	Q	JPR	.06	.22	.24	.06	.08	—	—
Gordon (1993) T4.24 b	200	GPP	Q	PRG	.00	.20	.20	.01	.04	—	—
Gordon (1993) T4.25	97	GPP	Q	JPR	.06	.22	.26	-.09	.12	—	—
Gordon (1993) T4.27 a	78	GPP	Q	JPR	-.09	.36	.24	-.13	.30	—	—
Gordon (1993) T4.27 b	158	GPP	Q	JPR	.22	-.10	.04	.17	.22	—	—
Gordon (1993) T4.27 c	146	GPI	Q	JPR	.04	.18	—	—	.19	—	—
Gordon (1993) T4.28 a	75	GPI	Q	CWB	-.36	.18	—	—	-.41	—	—
Gordon (1993) T4.28 b	99	GPI	Q	JPR	.10	.18	—	—	.48	—	—
Gordon (1993) T4.28 c	77	GPI	Q	JPR	.30	.19	—	—	.19	—	—
Gordon (1993) T4.28 d	97	GPI	Q	JPR	.33	-.02	—	—	.43	—	—
Gordon (1993) T4.29	747	GPP	Q	CWB	-.45	.23	-.30	.29	-.51	—	1.00
Gordon (1993) T4.31 a	216	GPP	Q	TRA	.21	-.08	.15	-.01	.30	—	—
Gordon (1993) T4.31 b	55	GPP	Q	TRA	.48	-.05	.25	.38	.06	—	—
Gordon (1993) T4.31 c	185	GPP	Q	TRA	.22	.04	.19	.28	.09	—	—
Gordon (1993) T4.32 a	130	GPP	Q	TRA	.07	.08	.16	.07	.25	—	—
Gordon (1993) T4.32 b	72	GPP	Q	JPR	.09	.21	.07	-.20	.23	—	—
Gordon (1993) T4.33	90	GPP	Q	JPR	.19	-.04	.06	.20	.33	—	—
Gordon (1993) T4.36	531	GPP	Q	JPR	-.09	.13	.19	.13	.15	—	—
Gordon (1993) T4.37	29	GPP	Q	JPR	.25	.51	.27	.41	.42	—	—
Gordon (1993) T4.39	213	GPP	Q	PRR	.28	-.04	.10	.24	.32	—	—
Gordon (1993) T4.42 a	216	GPP	Q	GPA	.06	-.08	.23	.07	.20	—	—

(continued overleaf)

APPENDIX B
(Continued)

Reference	N	Inventory	Format	Criterion	ES	EX	O	A	C	r _{yy}	u
Gordon (1993) T4.42 b	292	GPP	Q	GPA	.22	-.16	.17	.07	.30	-	-
Gordon (1993) T4.42 c	1078	GPP	Q	GPA	.02	-.14	.18	-.02	.08	-	-
Gordon (1993) T4.44	95	GPP	Q	GPA	.06	-.22	.36	.08	.22	-	-
Graham and Calendo (1969)	69	SDI	Q	JPR	-	-	-	-	.02	-	-
Grimsley and Jarret (1973)	100	GPP	Q	PRG	.13	.18	.12	-.27	.18	-	.48
Guller (2003)	375	EPPS	I	JPR	.06	-.04	.00	.04	-.01	-	-
Hakel (1966)	102	EPPS	I	GPA	.16	-.04	-.06	-.01	.25	-	-
Hirsh and Peterson (2008)	196	IPIP	I	GPA	-.02	-.14	-.02	-.24	.32	-	-
Hughes and Dood (1961)	90	GPI	I	SLS	.22	-.17	-	-	.14	-	-
Hughes and Dood (1961)	90	GPI	Q	SLS	.06	-.08	-	-	.08	-	-
Hughes and Prien (1986)	49	GPI	Q	JPR	-	-	-	.15	-.03	.42	-
Iliescu, Ilie, and Aspas (2011) a	833	ESQ	Q	CWB	-	-	-	-	-.43	-	-
Iliescu et al. (2011) b	224	ESQ	Q	CWB	-	-	-	-	-.37	-	-
Izard (1962) a	3	EPPS	I	GPA	-	-	-	-	.40	-	-
Izard (1962) b	33	EPPS	I	GPA	-	-	-	-	.33	-	-
Izard (1962) c	180	EPPS	I	GPA	-	-	-	-	.28	-	-
Izard (1962) d	98	EPPS	I	GPA	-	-	-	-	.17	-	-
Jackson et al. (2000)	106	ESQ	Q	CWB	-	-	-	-	-.41	-	1.00
Kahn, Nauta, Gailbreath, Tipps, and Chartrand (2002) a	674	MBTI	F	GPA	-	-.05	.06	.07	.07	-	-
Kahn et al. (2002) b	674	MBTI	F	CAB	-	-.04	-.05	.00	-.04	-	-
Kazmier (1961)	140	EPPS	I	GPA	.15	.13	-.06	-.06	.04	-	-
Kriedt and Dawson (1961)	41	GPI	Q	JPR	-	-.38	.31	.22	.37	-	-
Kusch, Deller, and Albrecht (2008)	145	OPQ	I	SFR	.20	.23	-.07	-.05	.18	-	-
Lievens, Harris, Keer, and Bisqueret (2003)	78	OPQ	I	TRA	-.03	-.03	.31	.24	.18	.80	.59
Lunneborg (1970)	188	EPPS	I	GPA	-.10	-.05	.07	.00	-.02	-	-
McDaniel et al. (2004)	384	OPQ	I	JPR	-.06	.16	.09	-.04	.03	-	.96
Morgan (1975) a	217	EPPS	I	GPA	-	-	-	-	-.02	-	-
Morgan (1975) b	135	EPPS	I	GPA	.02	.07	.07	-.01	.11	-	-
Muchinsky and Hoyt (1973)	129	EPPS	I	GPA	-	-.24	-	-	.10	-	-
Mukherjee (1968)	50	SCT	F	PRO	-	-	-	-	.32	-	.90
Nelson (2008)	114	OPQ	I	JPR	-.11	-.02	.07	-.21	.00	-	-
Neuman (1991)	247	GPP	Q	JPR	-	.23	-.37	.26	.17	-	1.00
Neuman and Kickul (1998)	284	GPP	Q	JPR	-	-	-	-	.32	-	-
Nyfield, Gibbons, Baron, and Robertson (1995) Turkey	503	OPQ	I	JPR	.04	-.02	-.02	-.05	.04	-	1.00
Nyfield et al. (1995) USA	103	OPQ	I	JPR	-.01	.15	-.15	-.06	-.03	-	.65
Olson, Shultz, and Scott (2000)	122	MBTI	F	JPR	-	.31	-.07	-.20	-.09	-	-
Perkins and Corr (2005)	68	OPQ	I	JPR	.00	-	-	-	-	-	.93
Robertson, Gibbons, Baron, MacIver, and Nyfield (1999)	114	OPQ	I	JPR	-.03	-.02	-.05	.14	-.03	-	.83
Robertson et al. (1999)	68	OPQ	I	JPR	.24	.01	.10	-.20	.20	-	1.00
Robertson et al. (1999)	90	OPQ	I	JPR	-.09	.12	-.09	.15	-.12	-	.98
Robertson et al. (1999)	131	OPQ	I	JPR	-.02	-.03	-.16	.10	.25	-	.87
Robertson et al. (1999)	34	OPQ	I	JPR	.03	-.15	-.02	-.30	.03	-	.99
Rolland and Mogenet (2001)	415	D5D	I	GPA	.00	-.04	-.18	.00	.16	-	-
Rolland and Mogenet (2001)	415	D5D	I	CAB	.07	.01	.12	.01	-.11	-	-
Rust (1999)	432	GIT	I	JPR	-.01	.09	-	.08	.02	-	-
Sackett, Gruys, and Ellington (1998)	247	EPPS	I	PRG	-	-	-	-	.16	-	-
Salgado (1991)	189	PAPI	I	JPR	.03	-.08	-.04	-.08	.09	-	-
Salgado (1991)	118	PAPI	I	JPR	-.05	.02	.02	-.12	-.02	-	-
Saville, Sik, Nyfield, Hackston, and MacIver (1996)	440	OPQ	I	JPR	-.03	.03	.07	.00	-.02	-	-
Saville et al. (1996)	270	OPQ	I	JPR	.03	.12	.12	-.02	.08	-	-
Schippmann and Prien (1989)	148	EPPS	I	JPR	-.02	.05	.09	-.05	-.09	-	-
Schippmann and Prien (1989)	148	GPP+SDI	Q	JPR	.06	.15	.27	-.12	.14	-	-
SHL (2006) val19	79	OPQ	I	JPR	.06	-.13	.11	-.03	.24	-	.93
SHL (2006) val22	120	OPQ	I	JPR	.03	.06	.14	-.05	.01	-	.91
SHL (2006) val30	114	OPQ	I	SLS	-.04	.10	-.06	-.28	.06	.76	1.06
SHL (2006) val32	36	OPQ	I	TRA	.05	-.19	.03	.15	-.20	.90	.88
Slocum and Hand (1971)	57	EPPS	I	JPR	.17	.07	.05	.02	-.17	-	-

(continued overleaf)

APPENDIX B
(Continued)

Reference	N	Inventory	Format	Criterion	ES	EX	O	A	C	r_{yy}	u
Slocum and Hand (1971)	37	EPPS	I	JPR	-.05	.17	.05	.20	.02	-	-
Sommerfeld (1997)	332	GPI	Q	JPR	.01	-	-	-	.32	-	-
Striker, Schiffman, and Ross (1965) a	225	MBTI	F	GPA	-	-.18	-.07	.01	.24	-	-
Striker et al. (1965) b	201	MBTI	F	GPA	-	-.07	-.10	.07	.13	-	-
Striker et al. (1965) c	201	MBTI	F	CAB	-	-.02	.07	.09	.09	-	-
Vasilopoulos et al. (2006)	327	IPIP	Q	GPA	-	-	-	-	.10	-	-
Warr et al. (2005a)	119	CCSQ	I	SLS	-.03	.10	-.10	-.19	.26	.89	.90
Warr et al. (2005a)	78	CCSQ	I	SLS	.04	.05	-.17	-.15	.20	.81	.90
Warr et al. (2005a)	90	CCSQ	I	SLS	-.04	.08	.25	-.32	.21	.89	.90
Whetzel, McDaniel, Yost, and Kim (2010)	1152	OPQ	I	JPR	-.02	.07	.08	-.04	.08	-	1.00
White (2002)	613	AIM	Q	JPR	.06	.22	-	.01	.20	.53	.70
White (2002)	399	AIM	Q	JPR	.07	.06	-	-.01	.04	.59	.71
Willingham and Ambler (1963)	208	GPI	Q	CWB	-	.17	-	-	-.19	.33	1.00
Willingham, Nelson, and O'Connor (1958)	1039	GPI	Q	TRA	.05	.00	-	-	.03	-	1.00
Witt and Jones (1999)	168	OPQ	I	JPR	.01	.07	.01	.01	.03	-	-
Young and Dulewicz (2007)	261	OPQ	I	JPR	.16	.14	.11	.17	.20	-	-
Zagar, Arbit, and Wengel, (1982)	570	EPPS	I	GPA	-.01	.07	.02	.07	.02	-	-

CAB = counterproductive academic behaviour; CWB = counterproductive work behaviour; GPA = grade point average; JPR = job performance rating; PRG = progress; PRR = peer rating; SAL = salary; SFR = self-rating; SLS = sales; TRA = training; r_{yy} = criterion reliability; u = range restriction value; I = ipsative; N = normative; Q = quasi-ipsative.

WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of Copyright law. No Further reproduction and distribution of this copy is permitted by transmission or any other means.