Cross-Cultural Issues in Personality Assessment

Filip De Fruyt and Bart Wille

Importance of Personality Assessment in a Globalized Economy

Personality assessment is an established part of many selection procedures in Western countries (Furnham, 2008; Sackett & Lievens, 2008), despite its questioned predictive validity throughout the years. Opponents (e.g., Morgeson et al., 2007a, 2007b) have mainly inquired the small magnitude of the predictive correlations and further criticized the fakability of self-descriptions in at-stake contexts such as job selection procedures. Proponents (Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007) meta-analytically reviewed validity coefficients and concluded that validities (1) are not trivial; (2) generalize across different contexts and cultures, with job characteristics acting as a moderator; (3) have demonstrated utility for selection decisions; and (4) are not necessarily worse than validities obtained with alternative methods of selection assessment (Rolland & De Fruyt, 2009). Although most authors agree that many individuals will put their best feet forward when describing their personality (in a selection context), there are varying opinions on how to handle and consider impression management. In addition to selection, personality assessment is used more and more in the context of career development and coaching, so its prominence and impact in the industrial and organizational (IO) field is steadily increasing. Given the range of criteria that are predicted by traits, it is to be expected that the frequency of personality assessments in IO professional practice will amplify in a globalized economy, where direct and indirect contacts with colleagues and customers representing diverse cultural backgrounds will be the norm rather than the exception. This multicultural context generates a series of questions and challenges that are beyond the description of personality differences for members of a single culture. With respect to *personality description*, questions at stake include the following: (1) what kind of trait model (and accompanying operationalization) should one use to describe individual's personality within and across cultural contexts, which gets at whether one can use inventories that are developed in one culture to assess applicants with a different cultural background? (2) What norms should one use when comparing individuals from diverse cultural backgrounds applying for jobs in which they will have to collaborate intensively? (3) Do applicants from diverse cultural backgrounds perceive assessment contexts differently? In other words, are self-enhancing strategies in personality descriptions in development or selection contexts perceived alike across cultural groups? (4) What about the accuracy of personality stereotypes of cultural groups? Given their potential impact in selection processes, it is important to know whether such stereotypes reflect a kernel of truth or do not match observed differences among cultural groups. With respect to the *predictive validity* of personality, a key question is whether culture acts as a moderator of personality–criterion relationships.

The current chapter will first explore the two key constructs, that is, culture and personality, examining major models describing basic dimensions of culture and introducing a model assumed to tap the common core of personality differences observable within and across cultures, respectively. The subsequent section reviews personality findings that showed to be largely universal across cultures. The next part discusses methodological and psychometric requirements for comparing personality scores across cultures, followed by an analysis of the importance of personality dimensions and mean-level personality differences among cultures. Tett and Burnett's (2003) trait-based interactionist model of job performance is subsequently discussed, taking into account the potential impact of culture-level variables. The implications of these findings for the IO professional practice are discussed in a practitioner's window. This chapter closes with a section identifying major knowledge gaps and perennial issues in the field of cross-cultural personality assessment in IO psychology.

Culture and Its Core Dimensions

The definition of culture, how to distinguish among cultural groups, and the kind of core dimensions that are necessary to describe cultures have been the subject of intensive debate and research in the past decades. Matsumoto (2000) provided an overarching description integrating different key attributes and defined culture as:

> A dynamic system of rules, explicit and implicit, established by groups in order to ensure their survival, involving attitudes, values, beliefs, norms, and behaviors, shared by a group but harbored differently by each specific unit within the group, communicated across generations, relatively stable but with the potential to change across time.

p. 24

This definition clearly acknowledges that individuals within a particular culture differ in terms of assimilating and manifesting various cultural attributes and further underscores that cultures have the potential to change over time. Both attributes affect how traits will have to be delineated from observable behavior. There have been several attempts to investigate core dimensions of cultural differences and cultural value frameworks in particular (for an excellent review, see Nardon & Steers, 2009). Two of these models were specifically developed within an IO framework and had considerable impact on this area, that is, Geert Hofstede's (1980, 2001) four-dimensional model of cultural differences and the work of the Global Leadership and Organizational Behavior Effectiveness (GLOBE) group (R. J. House, Hanges, Javidan, Dorfman, & Gupta, 2004).

Hofstede's Model

In the 1960s and 1970s, Hofstede (1980, 2001) had access to international survey data completed by a large sample of service and marketing personnel employed in 40 countries of a firm initially referred to as "Hermes" (later on revealed to be "IBM"). The survey was intended to assess and compare morale across divisions of IBM located in multiple countries. Hofstede factor analyzed aggregated scores across employees within these 40 societies and found that four major dimensions best represented the variance. The first dimension, "Power distance," reflects how societies find solutions to deal with the basic problem of human inequality. Cultures characterized by high power distance are organized very hierarchically often with a set of formal rules on how to navigate within this hierarchy. In cultures with high power distance, people accept authority and comply with orders and directions given by those higher in the hierarchy. A second dimension, "Uncertainty avoidance," describes how cultures cope with stress in the face of an unknown future. Societies characterized by high uncertainty avoidance will invest in different programs and institutions to deal with harm and

334

disaster; they value stability and do not tolerate deviant ideas and behavior. One of the most well-known dimensions of Hofstede's model is "Individualism–collectivism," describing how individuals are integrated into primary groups. In collectivistic cultures, person's identity is strongly bound to family relationships and the "in-group" to which one belongs, whereas in more individualistic societies, a person's identity is more related to individualistic strivings and achievements. In collectivistic societies, the group will take care of the person, whereas in individualistic societies, people have to take care of themselves and their direct family. Finally, the fourth factor "Masculinity–femininity" describes how societies are organized around the division of emotional roles between men and women. In more masculine societies, assertiveness, making career, and earning money are considered important, whereas more feminine societies value cooperation and getting along. Hofstede's research and dimensions were first criticized to have a strong Western bias, because African and Asian countries were underrepresented among his initial set of 40 countries. The Chinese Culture Connection (1987) challenged this ethnocentric viewpoint with a more emic research program, proposing an additional factor: "Confucian work dynamism." Hofstede (2001) later added this dimension to his model under the label "Long-term versus short-term orientation," reflecting differences among cultures in the choice of focus for people's efforts: the future or the present.

Hofstede (2001, p. 500; Exhibit A.5.1) ranked different cultures in terms of their scores on the cultural value dimensions. For example, the United States was ranked as highly individualistic (rank 1 of 53), lower on power distance (rank 38), higher on masculinity (rank 15), and lower on uncertainty avoidance (rank 43) relative to 53 other countries, whereas Japan was ranked as less individualistic (rank 22–23 of 53), somewhat more power distant (rank 33), top in masculinity (rank 1), and higher uncertainty avoidant (rank 7). These country rankings were used in numerous studies and correlated with other national-level data such as indicators of economic activity and wealth, health, and happiness, but also aggregate personality and national character ratings. However, as Matsumoto's definition of culture underscored, cultures are dynamic entities, and an update of this ranking of countries as well as a reexamination of the comprehensiveness and the content of Hofstede's model may be required after three decades of fast economical, societal, and political changes.
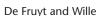
GLOBE

A second major research effort in the search for the dimensions of cultural values has been undertaken by a consortium of 160 researchers from many parts of the world under the direction of Robert J. House. The GLOBE research program (Chokhar, Brodbeck, & House, 2007; R. J. House et al., 2004; R. House, Javidan, Hanges, & Dorfman, 2002) was designed to examine implicit leadership theories and attributes of effective leadership in various cultural contexts. Data of about 17,000 managers employed in 951 organizations in 62 societies across the world were examined. In GLOBE, culture is defined as "the shared motives, values, beliefs, identities, and interpretations or meanings of significant events that result from common experiences of members of collectives that are transmitted across generations" (R. J. House & Javidan, 2004).

GLOBE defines nine major dimensions of cultural differences plus an additional six to describe leadership behavior. These nine dimensions are institutional collectivism or "the degree to which organizational and societal institutional practices encourage and reward the collective distribution of resources and collective action," in-group collectivism or "the degree to which individuals express pride, loyalty, and cohesiveness in their organizations or families," power distance or "the degree to which members of a society expect and agree that power should be stratified and concentrated at higher levels of an organization or government," performance orientation or "the degree to which an organization or society encourages and rewards members for performance improvement and excellence," gender egalitarism or "the degree to which a society minimizes gender role differences while promoting gender equality," future orientation or "the degree to which individuals

in organizations or societies engage in future-oriented behaviors such as planning, investing in the future, and delaying individual or collective gratification," humane orientation or "the degree to which members of a society encourage and reward individuals for being fair, altruistic, friendly, generous, caring, and kind to others," assertiveness or "the degree to which members of a society are assertive, confrontational, or aggressive in social relationships," and finally uncertainty avoidance or "the extent to which members of a society seek certainty in their environment by relying on established social norms, rituals, and bureaucratic practices" (R. J. House, Quigley, & de Luque, 2010, p. 118, Table 1).

These dimensions were subsequently used to cluster 61 nations participating in GLOBE according to cultural values and beliefs into 10 a priori proposed clusters: South-Asia, Anglo, Arab, Germanic Europe, Latin Europe, Eastern Europe, Confucian Asia, Latin America, Sub-Sahara Africa, and Nordic Europe. This clustering received considerable empirical support (Gupta, Hanges, & Dorfman, 2002). For example, the Arab cluster includes Egypt, Morocco, Turkey, Kuwait, and Qatar, and these societies are found to be highly group oriented, hierarchical, masculine, and low on future orientation (Kabasakal & Bodur, 2002). One of the major purposes of GLOBE was to examine leadership practices and how good leadership was perceived within these clusters. For example, in the Arab cluster, mid-level managers defined outstanding leadership as characterized by team-oriented and charismatic features, but also that an outstanding leadership style is not reflected by extreme positions on leadership traits (Kabasakal & Bodur, 2002). This example well illustrates how cultures may differentially value leadership behavior and hence value the personality traits that are associated with this competency (Bono & Judge, 2004). In addition, evidence is provided for the distinction between the cultural values and the cultural practices part in the assessment of the GLOBE dimensions, with only the values, but not the practices, being associated with features of outstanding leadership behavior. Javidan, House, Dorfman, Hanges, and de Luque (2006, p. 903) concluded that "In other words, leaders' reported effectiveness is associated with the society's cultural values and aspirations, but the society's effectiveness is associated with its cultural practices."

Both Hofstede and the GLOBE consortium have been very influential in alerting IO psychologists to the notion of cultural differences and providing the field with dimensional models to denote cultural attributes. Both approaches strongly contrast with the many "easy" operationalizations of culture that simply rely on race or nationality as markers of an individual's culture and disregard the cultural heterogeneity beyond directly accessible markers. Different comparative reviews and mutual criticisms (Hofstede, 2006, 2010; Javidan et al., 2006) have further sharpened our thinking about cross-cultural differences, its applications, and its challenges. It is clear now that cultural value dimensions have main effects on a series of outcome variables, such as emotions, attitudes and perceptions, behaviors, and job performance (Taras, Kirkman, & Steel, 2010), but culture values can also moderate relationships between other predictors (e.g., personality) and these outcomes.

One Model Fits All? The Universal Structure of Personality

Personality and culture show reciprocal relationships, with the expression of personality traits affected by culture and individuals' unique personality affecting and shaping that culture (Chao & Moon, 2005). The Five-Factor Theory distinguishes between basic tendencies and characteristic adaptations (McCrae & Costa, 1996). Basic tendencies (i.e., the traits from the Five-Factor Model, FFM) are considered as causal entities that are largely independent from cultural factors but are assumed to influence various characteristic adaptations, such as interests, motives, work competencies, and values. Individuals' value systems are also shaped by cultures' shared meaning systems. An important assumption of the Five-Factor Theory is that the structure of personality should be relatively invariant across different cultures. The question at stake becomes whether this FFM can be used for cross-cultural personality descriptions?

336

Cross-Cultural Replicability of the Big Five

The past decades, personality psychologists reached a relative consensus on the importance of five major personality dimensions, the so-called Big Five, to represent the major variance among personality descriptions. Lexical studies conducted in different languages mainly converged on the nature and number of factors suggesting that extraversion, agreeableness, neuroticism, conscientiousness, and intellect were necessary and sufficient to account for the communalities enclosed in self- and peer descriptions on large sets of personality-descriptive adjectives (Goldberg, 1982). From a different angle, Costa and McCrae (1992) introduced the FFM of personality, complementing their initial Neuroticism–Extraversion–Openness (NEO) model, already capturing the domains of neuroticism, extraversion, and openness to experience, with the domains of agreeableness and conscientiousness. Openness to experience deviates from the lexical Big Five intellect factor because it reflects a broader content including receptivity to a range of experiences and a fluid and permeable structure of consciousness that is not well represented in the natural language by trait adjectives (McCrae, 1994). They further demonstrated that this FFM was able to accommodate all main factors recurrently observable across major personality inventories. Although the terms "Big Five" and "FFM" are often used interchangeably to refer to the consensus on their importance as major constructs of personality, they have clearly different historical roots.
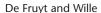
Cross-cultural psychologists have pointed our attention to the distinction between emic and etic approaches with respect to the use of psychological constructs in cross-cultural research (Matsumoto, 2000). Indeed, most FFM research has been *etic* in origin, examining the replicability of instruments that were to a large extent originally designed in Western cultures, in a multitude of countries across the globe. Such investigations have been done widely with the NEO Personality Inventory–Revised (NEO-PI-R; Costa & McCrae, 1992; Rolland, 2002) and its successor the NEO-PI-3 (McCrae, Costa, & Martin, 2005). There is massive evidence that the FFM structure is replicable in self- and peer ratings in cultures across all continents, at least when administered to people with sufficient reading command of the native language. Moreover, the FFM structure also showed to be valid across different age groups from adolescence (NEO-PI-3; De Fruyt, De Bolle, McCrae, Terracciano, & Costa, 2009) to adulthood (NEO-PI-R; McCrae & Terracciano, 2005b), making it the model par excellence to study gender and developmental trends from a cross-cultural perspective (see further in this chapter). De Fruyt, Aluja, Garcia, Rolland, and Jung (2006) further illustrated that the factor structure of the NEO-PI-R kept preserved across the IQ distribution in selection contexts, whereas Marshall, De Fruyt, Rolland, and Bagby (2005) demonstrated that the NEO structure was replicable across different administration contexts, including not-at-stake situations, career counseling (mild at-stake), and selection situations (high-stakes context). Together, these studies suggest that the FFM is applicable for cross-cultural assessment of personality in IO applications.

Dimensions Beyond the Big Five

There have been also *emic*, also called indigenous, approaches toward personality description, where researchers started within a particular culture to comprehensively sample personality descriptors bottom–up and examine their underlying structure, rather than importing (top–down) a personality inventory designed in a different culture. For example, Meiring, Van de Vijver, Rothmann, and De Bruin (2008) examined the structure of personality descriptors in 11 languages spoken in South-Africa, Church and his team (Katigbak, Church, Guanzon-Lapena, Carlota, & del Pilar, 2002) suggested additional indigenous dimensions to account for the commonality in Filipino college student personality ratings, and Benet-Martinez and John (1998) examined the personality structure in the Spanish language in Hispanic minorities. Overall, these authors demonstrated that there is evidence for a common cross-cultural personality-descriptive vocabulary as well as emic traits that

337

may have particular relevance and importance for a specific culture. Cheung and Leung (1998), however, strongly argued in favor of Chinese indigenous personality measures.

Personality psychologists have suggested additional factors beyond the Big Five as well within Western cultures. For example, Paunonen and Jackson (2000), reconsidering an initial selection of personality adjectives made by Saucier and Goldberg (1998), suggested 10 possible dimensions that are difficult to position within the Big Five: (1) religious, devout, reverent; (2) sly, deceptive, manipulative; (3) honest, ethical, moral; (4) sexy, sensual, erotic; (5) thrifty, frugal, miserly; (6) conservative, traditional, down to earth; (7) masculine–feminine; (8) egoistical, conceited, snobbish; (9) humorous, witty, amusing; and (10) risk-taking and thrill-seeking. Likewise, further elaborating within the lexical research paradigm, Ashton, Lee, and Son (2000) suggested "honesty–humility" as a sixth major factor, reflecting attributes such as fairness and sincerity. Reviewing these supplements, it is unclear whether some are to be considered as facets or blends of the Big Five or are indeed replicable major dimensions above and beyond the basic five. For example, a reanalysis of the data initially used by Ashton et al. (2004) as support for the honesty–humility dimension, by the same group of authors (except, Ashton and Lee), showed that no more than three factors, that is, extraversion, agreeableness, and conscientiousness, were replicable across 14 datasets from 12 different cultures, with "honesty–humility" turning up as a facet of agreeableness (De Raad et al., 2010). This reanalysis further demonstrated that also a well-known personality factor like "neuroticism/emotional stability," which is represented in almost every single theory or model on personality differences, was not replicable. Overall, this work by De Raad et al. (2010) well illustrates the limits of the lexical paradigm analyzing the passive personality-descriptive vocabulary to denote the major dimensions of personality.

Up until now, it is unclear whether these additional or emic-derived traits predict criteria of importance for IO psychology, beyond the dimensions and facets already enclosed in broad personality taxonomies for which there exists cross-cultural support. Contrary to the personality field, comprehensiveness is not necessarily the most important requirement for a personality-descriptive taxonomy to be used in IO psychology. For applied purposes, such as selection assessment, predictive validity is ultimately most important, and rather than being comprehensive, a personality measure should reflect those traits that are most useful to understand the criteria of interest such as job performance or leadership emergence for example. This implies that a personality measure fit for IO applications should assess not only several facets of conscientiousness, such as "self-discipline," "achievement," and "planning," but also traits that form blends between conscientiousness with other broad personality domains such as "control" (forming a blend with neuroticism) and "proactivity" (blending with extraversion) (Rolland & De Fruyt, 2009) because there is evidence that conscientiousness and related traits are predictors of work performance.

General Versus Contextualized Personality Inventories in IO Psychology

Many personality inventories in the past were developed from a clinical angle, for example, the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) or the Minnesota Multiphasic Personality Inventory (MMPI; Butcher & Williams, 2000), followed by a generation of inventories focusing at the description of trait variation observable in the general population such as the NEO-PI-R (Costa & McCrae, 1992) or the scales from Goldberg's International Personality Item Pool (IPIP; Goldberg et al., 2006). The legislation on job selection assessment in many countries, including the United States (Americans with Disabilities Act) and many European countries (e.g., France; Loi n° 92-1446 du 31 Décembre 1992), explicitly requires that assessments should have demonstrable relevance for the work context. The implication for personality assessment is that personality inventories administered in the context of job selection or career coaching should be directly relevant to judge on an individual's suitability for a particular job or contribute to an understanding

of functioning at work. General personality inventories, however, often include many items that are not immediately work related, making such instruments potentially contestable when used in IO professional practice. From a different angle, in an attempt to increase validities of personality assessments for IO applications, Lievens, De Corte, and Schollaert (2008) convincingly demonstrated that the inclusion of a frame-of-reference substantially increases the validity of personality descriptions to predict performance criteria. They showed that adding a frame-of-reference to the general instructions for personality description (e.g., "Describe how you generally behave at work or at school") or adding word tags to the items (e.g., "I am curious *at work*") leads to higher validities for predicting criteria considered important in the framed contexts.

These two evolutions led to an increased use of contextualized personality inventories specifically designed to assess personality at work either through the administration of work-related personality-descriptive items or through the addition of a "work-frame" to the instructions or a combination of these. Introducing "work context" in the items, on top of the personality behavioral descriptive part, makes such inventories inevitably more culture-bound. For example, an item like "A negative evaluation at work bothers me for days" (as an indicator of frustration tolerance) (Personality for Professionals Inventory [PfPI]; Rolland & De Fruyt, 2009) introduces an organizational and cultural practice into a personality-descriptive item. Merging context and behavioral description introduces extra challenges to demonstrate equivalence of measures across cultures (see further in this chapter).

Structure of Maladaptive/Dark Side Traits

The past years, IO psychology witnessed a growing attention for the assessment of aberrant traits and personality dysfunction (De Fruyt & Salgado, 2003; Salgado & De Fruyt, 2005; Wu & Lebreton, 2011). This transfer followed a growing awareness in human resources to pay more attention to dark side behaviors at work, partly accelerated by the multiple examples of mismanagement and the economical crisis after the millennium. Before this shift, human resources as a discipline were heavily under the influence of positive psychology, with more attention for the bright than the dark side of functioning.

There have been few attempts to assess maladaptive aspects of personality in the work context, with Robert and Joyce Hogan among the first to call the attention of IO psychologists to the dark side of personality (R. Hogan, Hogan, & Roberts, 1996). In clinical psychology and psychiatry, aberrant personality traits are described on Axis II of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association, 2000), which articulates 10 specific personality disorders, including the paranoid, schizoid, schizotypal, antisocial, borderline, histrionic, narcissistic, avoidant, dependent, and obsessive–compulsive personality disorder. Recent developments in clinical psychology, however, support the view that personality disorders do not represent qualitatively distinct categories, but should be conceived as continua of personality tendencies (Van Leeuwen, Mervielde, De Clercq, & De Fruyt, 2007) that affect broad areas in people's lives, including behavior at work (De Fruyt, De Clercq, et al., 2009). The validity of general personality-descriptive models, such as the FFM, to understand personality pathology has been extensively investigated (Costa & Widiger, 2002). This research line has demonstrated that general and maladaptive personality traits substantially overlap and that personality disorders can be described along the FFM dimensions, suggesting that differences between normality and abnormality are quantitative rather than qualitative.

Although well documented in Western countries, this assumption has not been examined widely outside North-America or Western Europe, except for a study by Rossier, Rigozzi, and Personality Across Culture Research Group (2008) replicating these associations in nine French-speaking African countries. The paradigm shift, in which personality disorders are better understood dimensionally, together with the observation that general personality traits also capture core features of personality

pathology, suggests that these constructs and assessment methodology might be applied successfully in IO psychology. The cross-cultural replicability of the FFM, together with the work by Rossier et al. (2008), is a first step in examining whether the evaluation of personality dysfunction may extend cross-culturally. Given the results of the GLOBE research group on the perception of leadership, it is to be expected that narcissistic leadership will be more perceived dysfunctional in, for example, the Arab cluster relative to Germanic European countries, where outstanding leadership is defined by team-oriented and charismatic features, in the absence of extreme positions on leadership traits (Kabasakal & Bodur, 2002).

Replicable Findings Across Cultures

Due to the relative consensus on the cross-cultural replicability of more structural aspects of personality, considerable progress has been made in the past decade to examine cross-cultural patterns of gender and age differences. Data have been accumulated not only through meta-analytic summaries of convenience samples, but also via targeted sampling across various cultures using a single comprehensive personality inventory. The major advantage of this last approach is that one circumvents the necessity to classify different scales assumed to assess a similar construct when compiling the meta-analytic database. The Personality Profiles of Cultures (PPOC; McCrae & Terracciano, 2005a, 2005b) and the Adolescent Personality Profiles of Cultures Project (APPOC; De Fruyt, De Bolle, et al., 2009), a consortium of international research partners collecting data with the NEO-PI-R (Costa & McCrae, 1992) or its more reader and adolescent friendly version, the NEO-PI-3 (Costa, McCrae, & Martin, 2008; McCrae, Martin, & Costa, 2005), have considerably contributed to this field. Given their comprehensive and hierarchical character, as well as their replicability across a range of cultures, the NEO measures are well suited to examine gender and age differences across the globe.

Universal Gender Differences

In a follow-up on previous narrative (Maccoby & Jacklin, 1974) and meta-analytic reviews (Feingold, 1994) of gender differences on a more narrow set of traits, Costa, Terracciano, and McCrae (2001) investigated gender differences in NEO-PI-R self-ratings obtained in 24 samples of adults and 14 samples of young adults across the FFM domains and their 30 facets. They further examined gender differences as a function of socioeconomic status (SES) indicators of cultures, including Hofstede's (2001) dimensions, in addition to gross domestic product, female literacy, life expectancy, and fertility rate, indicated by the number of children. Although convenience samples largely taken from Western cultures and often with undergraduates serving as young adult samples, the data lent itself to an examination of gender differences due to the replicable factor structure of the NEO-PI-R across countries. Observed gender differences were further compared with gender stereotypes assessed with the Bem Sex Role Inventory (Bem, 1974) to investigate whether stereotypes have some "kernel of truth."

Costa et al.'s (2001) findings can be easily summarized as follows: (1) At the FFM domain level, females score higher on neuroticism and agreeableness, and the orientation of these differences also generalizes to their facets. For extraversion and openness, gender differences seem to cancel out against each other at the domain level, but there are consistent gender differences at the facet level. Men score higher on E5: excitement-seeking and E3: assertiveness, whereas women have on average higher scores on E1: warmth, E2: gregariousness, and E6: positive emotions. Men further obtain higher scores on O5: openness to ideas, whereas women score higher on O2: aesthetics, O3: feelings, and O4: actions. Negligible gender differences are observed for conscientiousness. Important from the perspective of the current chapter is that these patterns generalize within (across young and older adults) and across cultures, suggesting stable cross-cultural patterns. (2) If gender differences

are observed, they are usually limited to half a standard deviation, with most differences reflecting a quarter standard deviation. (3) There is a strong agreement between gender stereotypes (Bem, 1974) and observed gender differences, underscoring the "kernel of truth" hypothesis regarding gender stereotypes. (4) Both nature and size of the differences are largely consistent with previous literature on a more limited set of traits and meta-analytic evidence described by Feingold (1994). The findings further suggest that gender differences also generalize from young to late adulthood. (5) If gender differences show up to some extent in personality ratings of one trait, the size of these differences generalizes across the other traits, suggesting that gender differentiation orientation generalizes within a culture. This finding inspired Costa et al. (2001) to rank societies in terms of gender role differentiation, showing that Zimbabwe had the lowest gender role differentiation, with small to negligible gender differences among traits, whereas Belgium showed the largest differences across the FFM. (6) Costa et al. (2001) correlated this ranking of sex role differentiation with the criteria characterizing cultures and found that if gender differences are observed, they are more sizeable in countries with a larger gross domestic product, literacy and life expectancy of women, and lower fertility. These findings are intriguing and surprising because also the Scandinavian countries such as Norway, Sweden, and Denmark are at the top end of the observed gender differentiation ranking. These countries were among the first Western societies in action to reduce gender inequality and glass ceiling effects, and especially, in these countries, gender differences are more pronounced. These findings obtained from convenience samples and self-ratings were largely confirmed in research by the PPOC and APPOC research teams examining gender differences in a much broader set of cultures (50 different countries across all continents) in which individuals were requested to describe somebody they knew well (McCrae & Terracciano, 2005b, p. 553, Table 4).

Universal Age Differences

A parallel route was followed accumulating the findings on cross-cultural age trends for personality ratings, first starting with analyses of mainly convenience samples obtained from a limited set of societies, followed by a more systematic description of age effects across a broad range of cultures by the (A)PPOC research teams. McCrae et al. (1999) started to examine whether the age trends observed in the normative NEO-PI-R sample generalized across five additional cultures (Germany, Italy, Portugal, Croatia, and South-Korea) in an attempt to figure out whether these age trends reflect common maturation processes (in the case of similar patterns across cultures) or whether age patterns were more culture-bound (in the case of different age trends). In line with the patterns observed in the United States, neuroticism, extraversion, and openness showed average declines with age in adulthood, whereas agreeableness and conscientiousness showed mean-level increases. The magnitude of these changes was small to moderate. These age trends were further confirmed at the FFM domain level for extraversion, openness, and conscientiousness in a broad set of 50 cultures by the PPOC research team, underscoring the notion that these age patterns reflect either common maturation patterns showing up relatively independent of cultural differences (McCrae & Costa, 1996) or are bound to common cultural processes that assert relatively common influences on traits across cultures.

One Method Hits All? Measurement Challenges When Comparing Cultures

A series of measurement issues and bias have to be taken into account before constructs and measures can be meaningfully compared across cultures. Cross-cultural researchers have distinguished among construct, method, and item bias (Van de Vijver & Leung, 1997a, 1997b), and the absence of bias is referred to as equivalence or invariance. Church (2010) provides an excellent introduction to the terminology and measurement challenges within cross-cultural measurement.

Church (2010) describes that construct or conceptual bias occurs: "when the definitions of the construct only partially overlap across cultures" (p. 154). For example, the content of a construct like intelligence is in some cultures constrained to cognitive functioning, whereas it includes more social competences in other cultures. The personality trait of assertiveness has a more negative connotation in the Netherlands, Belgium, and Germany, though it is considered mainly as a desirable and extraverted attribute in the United States (De Fruyt, Mervielde, Hoekstra, & Rolland, 2000). Church (2010) distinguishes among three forms of method bias: sample, instrument, and administration bias. Cross-cultural comparisons may be distorted through sample differences on possible confounding factors, design characteristics of the instrument (e.g., the use of Likert scales or the sorting of items across a Q-sort format may be familiar in one culture, but less frequently adopted in another culture), and finally the way the data are administered may be experienced differently by cultural groups and induce response differences (Church, 2010). For example, selection assessments may be perceived as more threatening in individualistic countries with a high power distance. A third kind of bias is item bias or differential item functioning (DIF): "DIF occurs when individuals with the same level or amount of a trait, but from different cultural groups, exhibit a different probability of answering the item in the keyed direction" (Church, 2010, p. 154). In a recent study, Church et al. (2011) examined DIF in factor loadings and intercepts from a multigroup confirmatory factor analysis (CFA) of NEO PI-R data obtained in the United States, the Philippines, and Mexico, showing that 40%–50% of the items exhibited some form of DIF. Moreover, DIF at the item level also affected the facet level, suggesting that the comparison of mean-level facet and domain scores across cultural groups should be done with caution.

In addition, Church (2010) defines different forms of equivalence, including conceptual, linguistic, and measurement equivalence. Different degrees of overlap between how constructs are defined across cultures are indicated by conceptual equivalence, whereas linguistic equivalence points to the accuracy of translations. For example, the NEO-PI-R item "I wouldn't enjoy vacationing in Las Vegas," as a reverse indicator of E5: Excitement-seeking, may have to be amended to better fit a local culture, when one would consider to use the NEO-PI-R in let us say, Iran. Finally, different levels of measurement equivalence or measurement invariance will have to be demonstrated (Vandenberg & Lance, 2000). In line with the CFA framework, weak factorial or configural invariance is demonstrated when the same number of latent constructs and the same pattern of salient and nonsalient factor loadings is presented across (cultural) groups. Metric invariance (strong factorial invariance) can be concluded when factor loadings (slopes) can be constrained to be equal across cultures without significant loss of model fit (Church, 2010). Finally, scalar invariance can be demonstrated when the item intercepts are also equal across cultural groups.

Steenkamp and Baumgartner (1998) have argued that mean scores of (cultural) groups are only meaningfully comparable when configural, metric, and scalar invariances have been established, showing that the factorial structure (configural), the scale intervals (metric), and the zero point of the scale are the same across different groups. Scalar equivalence is about the meaning of scores for different groups (Van de Vijver & Leung, 1997a), in other words: does a particular raw score indicates the same level of a trait in different groups and has the same interpretation in all cultures? If scalar equivalence is demonstrated, then we can derive meaningful conclusions from such comparisons. The demonstration of some form of scalar equivalence is hence a prerequisite for making comparisons among any groups (McCrae & Terracciano, 2008).

The determination of scalar equivalence is hotly debated among cross-cultural personality researchers (McCrae & Terracciano, 2005a). A main group of cross-cultural psychologists uses multigroup CFAs (MCFAs) to examine scalar equivalence, although the requirements for CFA are very stringent. Alternatively, Item-Response Theory (IRT)–based methods to examine DIF can be used to establish scalar equivalence (Reise & Henson, 2003), but large sample sizes are required and analyses become more complex when items with Likert scales have to be analyzed. Adopting these

methods for comparing sets of personality-descriptive item across cultures learns that many items show DIF, but also that DIF forwards to the facet level, and does not cancel out across multiple items compiling a facet (Church et al., 2011). A second way to demonstrate scalar equivalence is through the use of bilingual retest studies, in which bilingual respondents administer a personality inventory twice. Under the condition of equivalence, means for the two language versions of the inventory should be equal (McCrae & Terracciano, 2005b). The MCFA approach is further criticized because it generally assumes that the indicators of a trait are interchangeable, and this is usually not the case.

McCrae and Terracciano (2005a, 2008) proposed a different route to demonstrate scalar equivalence and argue that scalar equivalence is not an absolute property, but a matter of degree for which a pattern of evidence should be demonstrated, preferably via several of the previously suggested methods, because all have their specific drawbacks. Instead, they suggest a top–down approach where the group means are considered as scale scores, and their construct validity is investigated. A potential difficulty here is that one needs data from a large number of cultures and appropriate criteria at the culture level. McCrae and Terracciano (2008) argue that if one is able to pinpoint a nomological network of convergent and discriminant validity for a culture-level construct, the mean scores must have some degree of scalar equivalence (see further in this chapter).

Self-Reports Versus Multi-Informant Ratings

Although observer ratings have been used frequently in personality research (Hofstee, 1994), this source of assessment input has been underresearched and underutilized in IO psychology (Connelly & Ones, 2010; see Chapter 20, this volume). There are two reasons to assume that reports by knowledgeable others (peers, supervisors, or subordinates) will be used progressively more, that is, evidence for increased validity above and beyond self-descriptions and the expanding use of 180° or 360° feedback in the course of career development and coaching trajectories.

Barrick, Mount, and Strauss (1993) were among the first to report in the international literature that observer ratings of the FFM had validities to predict work performance that were almost twice the validities of self-ratings in sales people. Similar findings were reported by Oh and Berry (2009) using 360° ratings of managerial performance. Operational validities for supervisor ratings predicting task and contextual performance were significant for four of the FFM dimensions, except for agreeableness, and generally increased when combined with peer and subordinate ratings. When further complemented with self-ratings, the operational validities ranged from .23 (agreeableness) to .45 (openness to experience) for task performance and from .37 (openness to experience) to .50 (extraversion) for contextual performance. The adjusted multiple Rs for the FFM dimensions rated by all raters were .53 and .58 for managerial task and contextual performance, respectively. These findings suggest that the inclusion of observer ratings increases validity coefficients and that this increase is also a function of the different rater perspectives. In comparison with studies relying on self-ratings, the other FFM dimensions also show up as significant dimensions explaining facets of work performance. Oh, Wang, and Mount (2010) meta-analytically summarized validity coefficients available in 16 studies reporting on 20 independent samples, with observer personality ratings and work performance criteria rated by different sources. For all FFM dimensions, including openness to experience, validities for observer ratings were higher than for self-ratings and increased with the number of raters available. The meta-analytic work by Connelly and Ones (2010, Table 11) shows convergent results and also pleas to involve multiple raters to improve reliability and validity.

Despite evidence that observer ratings have incremental validity beyond self-descriptions and that validity of the assessments increases with the number of observers, it is not clear whether subordinate or 180° ratings are easy to obtain in all cultures. This is a highly underresearched area in IO psychology. One can assume, for example, that in cultures characterized by large power distance, inviting

employees to rate the attributes of their supervisor may be perceived as odd, whereas in more collectivistic cultures, peers may have difficulty to perceive a target as an independent self, describing the target's personality more in terms of fulfilling (work) roles and relationships with significant others in the in-group (Heine, 2001). Such cultural attributes may have a profound impact on the personality descriptions and induce different forms of administration, construct, and rating bias when working with observer ratings.

Impression Management Tendencies and Culture

In general, personality psychologists agree that candidates will put their best feet forward in selection assessments affecting the means of personality scales. To reckon this phenomenon, De Fruyt, De Clercq, et al. (2009) argued to take the assessment context into account and compare an individual's score relative to others' scores obtained under similar assessment conditions. For IO applications, the implication is that test developers will have to provide different norm sets obtained in low-, mild-, or high-stakes assessment contexts (for more coverage on faking personality tests, see Chapter 11, this volume). Personnel coaching and development programs are usually considered in Western societies as mild at-stake situations, whereas selection assessment is usually conceived as a high-stakes situation, although it remains an empirical question whether these conditions are perceived likewise across the globe. Anyway, for within-culture comparisons, locally built norms are necessary, and there should be convergence between the context of application and the context in which the normative data have been collected.

There is further evidence that cultures differ in terms of motivation for self-enhancement. A recent meta-analysis across 91 cross-cultural comparisons by Heine and Hamamura (2007) showed an average effect size of .84 in self-enhancement between Western and East-Asian samples. These differences can be partly explained due to the different cost–benefit ratio for self-enhancement for North-Americans versus East-Asians. Self-enhancement contributes to self-esteem and generates positive feelings among North-Americans, but negatively impacts East-Asians threatening their within-group integration and relationships (Heine & Buchtel, 2009). There is further evidence that East-Asians hold more dialectical views about themselves, including positive and negative views, whereas North-Americans underscore the positive views. Whether these self-enhancing tendencies also differentially operate in selection contexts is unclear.

Mean-Level Personality Differences Across Cultures

Aggregate Personality Ratings and Geographical Patterns

There is a long tradition of speculation about a geographical distribution of personality traits; in other words: "where one lives reveals what one is like" (Allik & McCrae, 2004, p. 13), although there are hardly empirical studies comparing personality ratings across multiple cultures. The main reason is that several requirements (see previously in this chapter) must be fulfilled before mean trait ratings can be meaningfully compared. Although personality traits have been mainly studied at the level of individuals, the past years witnessed a growing attention for aggregate ratings of personality, that is, a mean computed for a trait across a sample of individuals living in a particular culture that is subsequently used as a variable characterizing that cultural group. The level of analysis hence shifts from the individual to the culture level. If such differences across cultures would be replicable, systematic, and valid, then such aggregate ratings may be of considerable value for IO psychological applications. Assume, for example, that there would be systematic differences across cultures in terms of aggregate ratings of conscientiousness; one could examine whether such differences are associated with culture-level variables such as gross domestic product, wealth, productivity, or absenteeism data.

Likewise, the demonstration of meaningful average personality differences among U.S. states would necessitate the compilation of specific norms per region.

Primary evidence for the existence of regional personality differences in the United States has been provided already some decades ago by Krug and Kulhavy (1973) using Cattell's Sixteen Personality Factor Questionnaire (Cattell, Eber, & Tatsuoka, 1970) and more recently by Plaut, Markus, and Lachman (2002) using a measure of the Big Five. Corroborating this research line, Rentfrow, Gosling, and Potter (2008) examined regional personality differences in scores on the Big Five Inventory (John & Srivastava, 1999) in an impressive sample of near to 620,000 Internet respondents. A comparative analysis across these three studies learns that aggregate trait levels are to a considerable extent consistent across different geographical locations for neuroticism and openness to experience and somewhat consistent for extraversion and agreeableness, despite differences in sampling, measures, and a time frame of 30 years (Rentfrow, 2010). No consistent patterns for conscientiousness were observed. Moreover, regional personality differences were associated with important culture-level variables, including social connectedness (social capital), political orientation, and health. For example, state-level agreeableness was correlated .35 ($p < .05$), conscientiousness $-.44$ ($p < .05$), and state-level neuroticism $-.52$ ($p < .05$) with social capital, and people living in left-leaning states were higher in openness and lower in conscientiousness relative to right-leaning civilians (Rentfrow, 2010; Rentfrow et al., 2008).

Taking a cross-cultural angle, Allik and McCrae (2004) analyzed NEO-PI-R self-reports from 27,965 college student and adult men and women from 36 different cultures. Allik and McCrae (2004) considered means comparable because scalar equivalence was roughly demonstrated through a set of bilingual studies showing similar personality profiles across translations, together with evidence for the construct validity of within-culture aggregate personality ratings (see further in this chapter). They found that standard deviations for the 30 NEO-PI-R facets were systematically larger among European cultures than among Asian and Black African cultures. Multidimensional scaling showed personality traits to be geographically distributed, with neighboring countries exhibiting more similar personality profiles. A multidimensional scaling plot of 36 cultures, rotated toward a horizontal dimension positively associated with extraversion and openness and negatively with agreeableness, and a vertical axis associated with neuroticism and low conscientiousness, showed a clear separation between European and American cultures on the right from Asian and African cultures on the left. The United States and Canada were located near the bottom on the right of the plot, together with a Baltic and the Scandinavian countries. Although geographical proximity grouping was not perfect, it was certainly not random. A similar analysis on data collected with observer ratings in the course of PPOC (McCrae & Terracciano, 2005a), followed by a rotation to maximize associations with extraversion (horizontal axis) and neuroticism (vertical axis), again showed a plot of cultures that were historically and ethnically related. Summarizing the patterns across these two studies, the first relying on self-reports and the second on observer ratings, shows that Europeans and Americans are higher in extraversion and somewhat higher in openness compared to Asians and Africans.

In addition to comparing means across cultures, one can also factor analyze aggregate personality ratings from multiple cultures, also called ecological factor analysis (EFA). McCrae and Terracciano (2005a) adopted EFA on aggregated personality ratings of individuals from 51 cultures and showed that four of the FFM, neuroticism, openness, agreeableness, and conscientiousness, replicated the individual-level structure, with extraversion showing close approximation, loaded by five extraversion facets and some other facets that did not load the individual-level extraversion factor. They concluded that the FFM is not only applicable at the individual level, but also that there is a culture-level FFM, with a specific culture-level extraversion factor that is somewhat different from the individual-level dimension.

Finally, Stankov (2011) used hierarchical linear modeling (HLM) to examine individual, country, and societal cluster differences on Big Five personality traits, attitudes, values, and social norms in a

sample of 2029 students from across the globe. Instead of computing an average per culture, HLM enables to decompose observed variance across different nested levels. Individuals were nested under 45 countries (level 2), and these were nested in nine societal clusters (level 3) culled from GLOBE. Both personality traits (7.41% of the variance) and values (7.48%) were only slightly affected by country and societal cluster differences, and variance was mainly explained at the level of the individual, ranging from 87.23% for openness to 95.77% for agreeableness. Social norms were assessed with the nine GLOBE dimensions. Also their variance was to a large extent explained at the level of the individual (average of 84.07%), with 5.97% and 9.95% accounted for by the country and societal cluster level, respectively. The results reported by Stankov (2011) suggest that cultural influences on Big Five personality trait scores are limited, although results should be interpreted with caution because sample size is limited, especially at levels 2 and 3 of the analysis.

Do Aggregate Ratings Predict Something Meaningful?

In the course of the PPOC-project, McCrae et al. (2005a) correlated aggregate personality observer ratings obtained with the NEO-PI-R (Costa & McCrae, 1992) with culture-level variables. Aggregate personality ratings turned out to be replicable within cultures and showed meaningful associations with Hofstede's dimensions, values (Inglehart & Norris, 2003; Schwartz, 1994), well-being, gross domestic product, and the human development index. Several of these associations were replicated in APPOC (McCrae et al., 2009). Aggregate observer means converged with aggregate self-reports for the domains of neuroticism, extraversion, and openness to experience, but not for agreeableness and conscientiousness, although significant convergent associations were found for four agreeableness and four conscientiousness facets.

The validity of aggregate traits and the nature of the previously described significant associations with culture-level criteria have been subject of intense debate (for a discussion on culture-level criteria associations with conscientiousness, see Heine, Buchtel, & Norenzayan, 2008). As a reply, Mõttus, Allik, and Realo (2010) examined associations between self-reports on conscientiousness facets and a broad range of culture-level criteria across 42 cultures, including the 36 cultures from McCrae (2002), expanded with 3 African cultures, Lithuania, Poland, and Finland. The associations with the observer ratings reported in PPOC (McCrae & Terracciano, 2005a) were also examined. They provided clear a priori hypotheses about the expected relationships, examined associations at the facet instead of the domain level, and used a range of criteria (e.g., atheism, smoking, democracy, obesity, alcohol consumption, and gross domestic product) that were really representative of the culture and its population. Without correcting for gross domestic product, 29% of the correlations were significant at $p < .01$. Controlling for national wealth reduced the number of significant correlations by almost half. Confirmation of hypotheses was different across the six conscientiousness facets and self- versus observer ratings.

National Character Ratings

National character ratings are a different type of culture-bound personality ratings, that is, descriptions of the personality of a typical individual representing a national or a cultural group. Such descriptions can be auto or hetero stereotypes with the first reflecting ratings provided by in-group members, whereas hetero stereotypes are provided by people of a different culture. These national character ratings are subsequently compared to observed descriptions of in-group members to examine whether such ratings have validity or are just stereotypes in the eye of the beholder without a kernel of truth.

Terracciano et al. (2005) examined the correspondence between national character ratings on a measure of the FFM and APPOC observer ratings on the NEO-PI-R (McCrae & Terracciano,

346

2005a) across 49 cultures. There was no correspondence between the two sets of ratings. For example, Indonesia, Nigeria, Turkey, Poland, and Japan obtained the highest national character scores for neuroticism, though the observed means (expressed in T-scores) on the NEO-PI-R for neuroticism for these countries ranged from 47.8 (Nigeria) to 51.4 (Turkey). The authors concluded that national character ratings appeared to reflect unfounded stereotypes.

The inaccuracy of geographical personality stereotypes has been further confirmed in studies by McCrae, Terracciano, Realo, and Allik (2007) with Northern and Southern Italians and by Realo et al. (2009) comparing Russian self-reported averages with perceptions by civilians of neighbor countries. Rogers and Wood (2010), however, did find that Americans' geographical personality stereotypes for openness to experience and neuroticism show considerable accuracy when compared with the results reported by Rentfrow et al. (2008), with above-chance accuracy for agreeableness and extraversion. They further show what regional indicators might contribute to accuracy such as population density and political voting patterns. Rogers and Wood (2010) conclude their work by saying that geographical personality stereotypes may have some accuracy under certain conditions.

Does "Big Five" Also Reflect Universal Importance and Validity?

Although there is considerable support for the factors that are minimally necessary to structure personality traits and help defining their nomological net, this evidence does not imply that personality traits as concepts are perceived equally important in all cultures (Heine & Buchtel, 2009), and all Big Five dimensions are equally important for understanding personality at work across cultures.

People from different cultures do not equally weigh personality information. There is evidence that people from more collectivistic societies rely more on situational factors and are less inclined to use personality information than people from individualistic cultures to explain differences in behavior (Heine & Buchtel, 2009; Morris & Peng, 1994). Although the factor structure of traits seems to be roughly replicable across cultures, this does not imply that all Big Five dimensions are equally important within each single culture. For example, within Western–industrialized and individualistic societies, getting ahead traits such as extraversion and conscientiousness may be considered more important, whereas in more collectivistic cultures, more communal and getting along traits like agreeableness may be valued more. Likewise, it can be hypothesized that more interpersonal traits such as extraversion and agreeableness will be esteemed differently as a function of the power distance level of a culture. These examples clearly show that replicability of factor structure across cultures and importance of factors within specific cultures are two different questions, and actually, there is a dearth of studies examining the importance of the Big Five dimensions across cultures. Moreover, the significance of Big Five dimensions within a particular culture may change over time. For example, Western–industrialized countries in which traits like extraversion and conscientiousness were considered important dimensions for adaptation and functioning may notice a shift toward increased importance of openness to experience related traits such as innovation, creativity, and self-reflection. Finally, the importance of personality traits relative to other individual differences such as intelligence, attitudes, skills, and values may change across time in a rapidly transforming world economy. The current meta-analyses on predictor–criterion validities summarize validity coefficients reported in individual studies published across a broad time period, often decades ago. Given the largely changing economies of the past 20 years, it might be interesting to examine cohort differences in validity coefficients.

Validity generalization is a crucial issue for IO applications and practices that are similarly applied in different cultures. The majority of the meta-analyses on the predictive validity of trait measures relied on individual studies conducted with Westerners (Barrick & Mount, 1991; Connelly & Ones, 2010; J. Hogan & Holland, 2003; Oh et al., 2010; Salgado, 1997). As far as we know, there is no meta-analytic evidence that validities of personality measures generalize to non–Western cultures. Such

347

confirmation is not only absent for the FFM, but is also deficient for indigenous traits. For example, it would be interesting to examine whether traits resulting from indigenous personality research in China (Cheung et al., 1996), such as "interpersonal relatedness," predict aspects of job performance, such as contextual performance, better than the FFM. At the level of the FFM, it would be interesting to investigate whether the same traits predict similar criteria across cultures, and whether the magnitude of these predictive validities is moderated by cultural characteristics. For example, Heine and Buchtel (2009) recently suggested that personality may be less predictive of behavior in collectivistic cultures, due to the presumed larger impact of norms, prescribed roles, and pressure from the social network on the person's behavior.

Tett and Burnett's (2003, p. 503, Figure 1) trait-based interactionist model of job performance can be used to better understand how culture may impact the validity of traits to predict work behavior and job performance. They distinguish work behavior from job performance, because the latter involves an evaluation within a specific context that may be valued differently across cultures. Work behavior may lead to intrinsic rewards for the individual, due to the possibility to express his/her personality, whereas job performance leads to extrinsic rewards such as salary, feedback, and recognition from others.

Tett and Burnett's (2003) trait-activation theory further distinguishes moderators of the trait–work behavior relationship at the task, social, and organizational levels (for more coverage of trait-activation potential, see Chapter 5, this volume). For example, orderliness as a trait may be positively related to job performance for accounting tasks (task level) in a team valuing precision and punctuality (social level) and in a detail and outcome-oriented company (organization level), but fail to predict performance in task, social, and organizational environments with a different focus. Moreover, personality expression may be further affected by job demands (tasks and duties inherent in the job), distractors (factors interfering with performance), constraints (factors restricting the manifestation of the trait), releasers (factors counteracting a constraint), facilitators (factors making triggers more salient that are already in the situation). An accounting job includes many tasks demanding orderliness, as too much small talk with colleagues during working hours may distract from the primary tasks; the increased use of information technology may constrain the impact of personality, whereas an unforeseen bug in a program may counteract such constraint, making individual differences again more salient. Finally, dealing with a file of a highly valued customer may act as a facilitator for precision and attention to detail.

Reviewing this model, it is clear that culture may impact upon the task, social, and organizational moderators affecting the personality trait–work behavior relationship. Moreover, culture will also affect the evaluation of work behavior and the extrinsic rewards associated with good work performance. For example, in societies with many strong norms and low tolerance of deviant behavior, the so-called tight cultures versus loose cultures with weak social norms and higher tolerance of deviant behavior (Gelfand et al., 2011), one can expect that bad performance leads to lower extrinsic rewards and negative feedback. This tendency may be strengthened in individualistic societies, holding persons more accountable for their individual contribution and strivings. In addition, it may be expected that tight (Gelfand et al., 2011) and high uncertainty avoidant and more feminine cultures (Hofstede, 2001) will put more constraints on the expression of individual differences, hence impacting the strength of the trait–work behavior relationship that can be observed.

Finally, validity generalization studies often and correctly pay a lot of attention to the predictor side of the equation. However, one should be also thoughtful about the nature and construct validity of the criteria that one wants to predict. Job performance indicators may be perceived very differently across the globe. For example, "waiter service" in a restaurant is defined and perceived very differently across cultures due to divergence in the way labor is organized and multiple cultural expectations. What is considered as good performance in most restaurants in the United States (speed of service, removing plates as soon as one person around the table has finished her/his meal, asking

whether the meal is good and fits your expectations multiple times in a time frame of 20 min) leads to dismissal in Western Europe where eating is considered as a social event requiring time to enjoy the food and company, and where you make a reservation for a table for the entire evening. In the United States, people are lining up until a table is free and multiple shifts of service have to be completed at a single table on an evening. This cultural difference is also reflected at the financial level: In many U.S. restaurants, you are financially punished when one has to serve more than five people at a table, whereas you may have a discount in Western Europe. This example well illustrates how (job) performance criteria may be perceived very differently across cultures.

Knowledge Gaps and Perennial Issues

The previous review has made clear that considerable progress has been made in the past 20 years with respect to cross-cultural personality assessment, though it is also obvious that these developments emerge at a slow pace and most often follow rather than precede calls and questions emerging from the applied field of personality assessment. Four major challenges can be identified requiring immediate attention.

Indigenous Versus Universal Traits

It is clear that a common set of traits, integrated within the FFM, can be used to denote personality differences across the globe, but it remains to be investigated whether indigenous traits predict variance in IO criteria above and beyond the more universal traits. Despite the universality of this trait taxonomy, we know almost nothing about the importance within particular cultures of the major factors enclosed in the FFM. A similar problem arises with respect to validity generalization. Most validation studies have been conducted with Westerners, but validities remain to be demonstrated in, for example, African and South-American cultures. For IO applications, comprehensiveness of trait taxonomies will be less important, though inventories will have to include those traits that are most useful for predicting IO criteria. Studies examining the moderating role of culture on personality–criterion relationships could be conducted along Tett and Burnett's (2003) trait-based interactionist model of job performance described previously, distinguishing the major variables affecting this relationship.

General Versus Work-Related Personality Inventories

Legislation within several countries and recent research recommends the use of work-related over general personality inventories for IO applications. The addition of a "work" frame-of-reference to instructions and/or items (Lievens et al., 2008) and complementing self-descriptions of personality with, preferably multiple, observer ratings (Connelly & Ones, 2010; Oh et al., 2010) enhances reliability and validity of the assessments in Western cultures. The adoption of work-related personality inventories, including items referring to observable work behavior, will facilitate the involvement of multiple raters such as subordinates, direct colleagues, or supervisors. It remains to be examined, however, under what circumstances such observer ratings add validity, and whether societal and organizational cultures moderate such relationships. In addition, scalar invariance will have to be demonstrated within cultures, before self- and observer ratings can be meaningfully compared and integrated.

Heterogeneity Within Cultures

Although migration and various forms of intercultural transmission have been universal phenomena throughout history, the way in which cultural differences are perceived and have to be handled in societies has dramatically changed in the recent past. Whereas immigrants were previously assumed

to adapt and assimilate as quickly and profoundly as possible to the language, habits, and culture of the receiving society, Western societies nowadays consider diversity and a plethora of cultural backgrounds as a strength that should be taken into account, respected, and sometimes preserved. As a result, societies have definitely become more heterogeneous in terms of the cultural backgrounds of their members. In addition, individuals within a particular society may belong to different (cultural) groups at the same time or cultural boundaries may have become permeable and fuzzy. For example, children from Moroccan immigrants born in Germany may share characteristics with the host German culture, but will also resemble features and values from their Moroccan roots. Moreover, people within a culture may be member of multiple groups at the same time, that is, reflecting a different cultural heritage and background, SES (raised in a low-SES family and via upward mobility moved to a higher class, or the other way around), and gender. These different group attributes will interact, and personality inventory developers and assessment practitioners will have to face this complex reality. Practitioners and researchers will have to disentangle, for example, whether poor psychometric problems observed in a heterogeneous group within a single society are attributable to problems with understanding particular items (due to insufficient language command of the visiting's culture native language) or reflect measurement inequivalence.

Differences Between Cultures and the Feasibility of Multicultural Norms

Several studies were reviewed in this chapter suggesting that personality traits show a geographical distribution within the United States (Rentfrow, 2010) and across the globe (McCrae & Terracciano, 2005a; Stankov, 2011). There are diverging opinions, however, with respect to the comparability of such means, requiring the demonstration of some form or a degree of scalar equivalence. The hetero geneity of cultural backgrounds represented within societies and the fact that individuals often belong to multiple groups (e.g., age, gender, and an ethnic minority group) at the same time introduces very complex "equivalence" questions to be dealt with. Given the increasing cultural diversity of the workforce, the global economy, and contacts with customers from a broad range of cultures, it is to be expected that the importance of personality traits will increase and it is not an understatement to conclude that we are just at the beginning of a flourishing field of research and consulting. The challenge for academia and research will be to take the lead in this debate and provide the applied field with recommendations and workable suggestions.

Practitioner's Window

Given the increasing multiculturalism within individual societies and the steadily growing number of contacts across nations in the global economy, human resources practitioners will be faced more and more with questions on culture's consequences for the description and comparison of personality. The previous overview has made a number of points clear that may help the practitioner facing these questions.

(a) The trait structure represented by the FFM is valid to describe general personality traits across different cultures. More indigenous dimensions may supplement this description. Major age trends for the FFM traits are largely culturally universal, and gender differences seem to be more pronounced and generalizable in Western cultures.

(b) There are a series of FFM or Big Five inventories available in different languages (academic and commercially), although these are not substitutes for each other and cannot be used interchangeably. For comparative purposes, practitioners should use the same Big Five/FFM inventory across

---

cultures, examining whether these translations/adaptations meet (part of) the requirements for making such comparisons.

(c) There are no compelling data on the importance of FFM traits across cultures. For example, extraversion may be considered more important in individualistic societies, whereas agreeableness may be valued more in feminine-oriented cultures.

(d) In addition to culture, one should also take into account the assessment context. There is massive evidence that the assessment context (low versus medium or high stakes) impacts the personality scores within Western cultures, necessitating specific norms obtained in similar assessment contexts to make meaningful comparisons.

(e) It is inconclusive whether self-enhancement/impression management strategies are used differently across cultures and across contexts within these cultures.

(f) Cultures do not differ dramatically in terms of mean-level personality scores. Differences between varying at-stake contexts have probably a larger impact on the distribution of personality scores than differences between cultures.

(g) Individuals' personality descriptions should be preferable compared against normative distributions obtained from individuals from the same cultural background administered the inventory in the same (low-, medium-, or high-stakes) assessment context. In the absence of such norms, preference should be given to norms taking into account the assessment context, given the smaller magnitude of differences between cultures.

(h) For multicultural selection, such as in the case of ex-patriots, it is also recommended to compare individuals' scores to the normative distributions obtained in the host culture (as a supplement to point g). Likewise, for the selection of applicants from diverse cultural backgrounds who have to work together, it is recommended to assemble a cross-cultural normative set, representing the different cultural groups.

(i) There are not enough studies in non-Western cultures to conclude that the validity of personality traits to predict various forms of work behavior and performance is universal in nature and strength. Tett and Burnett's (2003) trait-based interactionist model of job performance provides a valuable framework to understand how culture may moderate this relationship.

(j) In addition to paying attention to the predictor side of the equation, that is, personality traits, practitioners should also carefully analyze the nature of the criterion. Like specified in Tett and Burnett's (2003) model, not all work behaviors are valued equally across cultures.

(k) There is an increased use of contextualized and maladaptive personality measures, in addition to general traits. Also the use of observer ratings in addition to self-ratings is highly encouraged. Whether these new assessment practices are generalizable across cultures remain an open question.

(l) Finally, aggregate personality ratings make sense and are replicable, although do not correspond to national character stereotypes. Practitioners should hence be very cautious relying on stereotypes of cultural groups.

References

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits—Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35*, 13–28.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., & De Raad, B. (2004). A six-factor structure of personality–descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*, 356–366.

Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: Correlations with Machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality, 14*, 359–368.

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., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives—Test of the mediating effects of goal-setting. *Journal of Applied Psychology, 78*, 715–722.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 115–162.

Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multi method analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729–750.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology, 89*, 901–910.

Butcher, J. N., & Williams, C. L. (2000). *Essentials of MMPI-2 and MMPI-A interpretation* (2nd ed.). Minneapolis: University of Minnesota Press.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Chao, G. T., & Moon, H. (2005). The cultural mosaic: A metatheory for understanding the complexity of culture. *Journal of Applied Psychology, 90*, 1128–1140.

Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology, 29*, 233–248.

Cheung, F. M., Leung, K., Fan, R. M., Song, W., Zhang, J.-X., & Zhang, J.-P. (1996). Development of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology, 27*, 181–199.

Chinese Culture Connection. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology, 18*, 143–174.

Chokhar, J. S., Brodbeck, F. C., & House, R. J. (2007). *Culture and leadership across the world: The GLOBE book of in-depth studies of 25 societies*. Mahwah, NJ: Lawrence Erlbaum.

Church, A. T. (2010). Measurement issues in cross-cultural research. In G. Walford, E. Tucker, & M. Viswanathan (Eds.), *The Sage handbook of measurement* (pp. 151–177). Los Angeles: Sage.

Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the revised NEO Personality Inventory. *Journal of Personality and Social Psychology, 101*, 1068–1089.

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

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

Costa, P. T., McCrae, R. R., & Martin, T. A. (2008). Incipient adult personality: The NEO-PI-3 in middle-school-aged children. *British Journal of Developmental Psychology, 26*, 71–89.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.

Costa, P. T., & Widiger, T. A. (2002). *Personality disorders and the five-factor model of personality* (2nd ed.). Washington, DC: American Psychological Association.

De Fruyt, F., Aluja, A., Garcia, L. F., Rolland, J. P., & Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants. *International Journal of Selection and Assessment, 14*, 101–112.

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., & Costa, P. T. (2009). Assessing the universal structure of personality in early adolescence: The NEO-PI-R and NEO-PI-3 in 24 cultures. *Assessment, 16*, 301–311.

De Fruyt, F., De Clercq, B. J., Miller, J., Rolland, J. P., Jung, S. C., Taris, R., & Van Hiel, A. (2009). Assessing personality at risk in personnel selection and development. *European Journal of Personality, 23*, 51–69.

De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J. P. (2000). Assessing adolescents' personality with the NEO PI-R. *Assessment, 7*, 329–345.

De Fruyt, F., & Salgado, J. F. (2003). Editorial: Personality and IWO applications: Introducing personality at work. *European Journal of Personality, 17*, S1–S3.

De Raad, B., Barelds, D. P. H., Levert, E., Ostendorf, F., Mlacic, B., Di Blas, L., . . . Katigbak, M. S. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology, 98*, 160–173.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *The Eysenck Personality Questionnaire*. Sevenoaks, UK: Hodder & Stoughton.
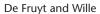
Feingold, A. (1994). Gender differences in personality—A meta-analysis. *Psychological Bulletin, 116*, 429–456.

Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment, 16*, 300–305.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., & Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*, 1100–1104.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. D. Butcher (Eds.), *Advances in personality assessment* (*Vol. 1*, pp. 203–234). Hillsdale, NJ: Erlbaum.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Gupta, V., Hanges, P. J., & Dorfman, P. (2002). Cultural clusters: Methodology and findings. *Journal of World Business, 37*, 11–15.

Heine, S. J. (2001). Self as a cultural product: An examination of East Asian and North American selves. *Journal of Personality, 69*, 881–906.

Heine, S. J., & Buchtel, E. E. (2009). Personality: The universal and the culturally specific. *Annual Review of Psychology, 60*, 369–394.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19*, 309–313.

Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. *Personality and Social Psychology Review, 11*, 4–27.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G. (2006). What did GLOBE really measure? Researchers' minds versus respondents' minds. *Journal of International Business Studies, 37*, 882–896.

Hofstede, G. (2010). The GLOBE debate: Back to relevance. *Journal of International Business Studies, 41*, 1339–1346.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8*, 149–162.

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

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions—Questions and answers. *American Psychologist, 51*, 469–477.

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: An introduction to Project GLOBE. *Journal of World Business, 37*, 3–10.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organization: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

House, R. J., & Javidan, M. (2004). Overview of GLOBE. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 9–26). Thousand Oaks, CA: Sage.

House, R. J., Quigley, N. R., & de Luque, M. S. (2010). Insights from Project GLOBE: Extending global advertising research through a contemporary framework. *International Journal of Advertising, 29*, 111–139.

Inglehart, R., & Norris, P. (2003). *Rising tide: Gender equality and cultural change around the world*. New York: Cambridge University Press.

Javidan, M., House, R. J., Dorfman, P. W., Hanges, P. J., & de Luque, M. S. (2006). Conceptualizing and measuring cultures and their consequences: A comparative review of GLOBE's and Hofstede's approaches. *Journal of International Business Studies, 37*, 897–914.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York: Guilford Press.

Kabasakal, H., & Bodur, M. (2002). Arabic cluster: A bridge between East and West. *Journal of World Business, 37*, 40–54.

Katigbak et al., 2002Katigbak, M. S., Church, A. T., Guanzon-Lapena, M. A., Carlota, A. J., & del Pilar, G. H. (2002). Are indigenous personality dimensions culture specific? Philippine inventories and the five-factor model. *Journal of Personality and Social Psychology, 82*, 89–101.

Krug, S. E., & Kulhavy, R. W. (1973). Personality differences across regions of the United States. *Journal of Social Psychology, 91*, 73–79.

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

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Marshall, M. B., De Fruyt, F., Rolland, J. P., & Bagby, R. M. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. *Psychological Assessment, 17*, 379–384.

Matsumoto, D. (2000). *Culture and psychology: People around the world*. Belmont, CA: Wadsworth/Thomson Learning.

McCrae, R. R. (1994). Openness to experience—Expanding the boundaries of Factor V. *European Journal of Personality, 8*, 251–272.

McCrae, R. R. (2002). NEO-PI-R: Data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 105–125). New York: Kluwer Academic/ Plenum.

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality inventories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York: Guilford Press.

McCrae, R. R., Costa, P. T., de Lima, M. P., Simoes, A., Ostendorf, F., Angleitner, A., & Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology, 35*, 466–477.

McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment, 84*, 261–270.

McCrae, R. R., Martin, T. A., & Costa, P. T. (2005). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment, 12*, 363–373.

McCrae, R. R., & Terracciano, A. (2005a). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407–425.

McCrae, R. R., & Terracciano, A. (2005b). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547–561.

McCrae, R. R., & Terracciano, A. (2008). The five-factor model and its correlates in individuals and cultures. In F. J. R. Van de Vijver, D. A. Van Hemert, & Y. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 249–283). Mahwah, NJ: Erlbaum.

McCrae, R. R., Terracciano, A., De Fruyt, F., De Bolle, M., Gelfand, M. J., & Costa, P. T. (2009). The validity and structure of culture-level personality scores: Data from ratings of young adolescents. *Journal of Personality, 78*, 815–838.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, H. (2007). Climatic warmth and national wealth: Some culture-level determinants of national character stereotypes. *European Journal of Personality, 21*, 953–976.

Meiring, D., Van de Vijver, F., Rothmann, I., & De Bruin, D. (2008). Uncovering the personality structure of the 11 language groups in South Africa: SAPI project. *International Journal of Psychology, 43*(3–4), 364.

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

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

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*, 949–971.

Mõttus, R., Allik, J., & Realo, A. (2010). An attempt to validate national mean scores of conscientiousness: No necessarily paradoxical findings. *Journal of Research in Personality, 44*, 630–640.

Nardon, L., & Steers, R. M. (2009). The culture theory jungle: Divergence and convergence in models of national culture. In R. S. Bhagat & R. M. Steers (Eds.), *Cambridge handbook of culture, organizations, and work* (pp. 3–22). Cambridge, UK: Cambridge University Press.

Oh, I. S., & Berry, C. M. (2009). The five-factor model of personality and managerial performance: Validity gains through the use of 360 degree performance ratings. *Journal of Applied Psychology, 94*, 1498–1513.

Oh, I. S., Wang, G., & Mount, M. K. (2010). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*, 762–773.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995–1027.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality, 68*, 821–835.

Plaut, V. C., Markus, H. R., & Lachman, M. E. (2002). Place matters: Consensual features and regional variation in American well-being and self. *Journal of Personality and Social Psychology, 83*, 160–184.

Realo, A., Allik, J., Lonnqvist, J. E., Verkasalo, M., Kwiatkowska, A., Koots, L., . . . Renge, V. (2009). Mechanisms of the national character stereotype: How people in six neighbouring countries of Russia describe themselves and the typical Russian. *European Journal of Personality, 23*, 229–249.

Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment, 81*, 93–103.

Rentfrow, P. J. (2010). Statewide differences in personality toward a psychological geography of the United States. *American Psychologist, 65*, 548–558.

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science, 3*, 339–369.

Rogers, K. H., & Wood, D. (2010). Accuracy of United States regional personality stereotypes. *Journal of Research in Personality, 44*, 704–713.

Rolland, J. P. (2002). The cross-cultural generalizability of the five-factor model. In R. R. McCrae & J. Allik (Eds.), *The five factor model of personality across cultures* (pp. 7–28). New York: Kluwer Academic/Plenum.

Rolland, J. P., & De Fruyt, F. (2009). *PfPI: Inventaire de Personnalité au Travail*. Paris: ECPA.

Rossier, J., Rigozzi, C., & Personality Across Culture Research Group. (2008). Personality disorders and the five-factor model among French speakers in Africa and Europe. *La Revue Canadienne de Psychiatrie, 53*, 534–544.

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

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

Salgado, J. F., & De Fruyt, F. (2005). Personality in personnel selection. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection* (pp. 174–198). Oxford, UK: Blackwell.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the big five? *Journal of Personality, 66*, 495–524.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (*Vol. 18*, pp. 85–119). Thousand Oaks, CA: Sage.

Stankov, L. (2011). Individual, country and societal cluster differences on measures of personality, attitudes, values, and social norms. *Learning and Individual Differences, 21*, 55–66.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90.

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*, 405–439.

Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, L., Ahn, C., Ahn, H. N., . . . McCrae, R. R. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science, 310*, 96–100.

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

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

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Van de Vijver, F. J. R., & Leung, K. (1997a). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Van de Vijver, F. J. R., & Leung, K. (1997b). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Theory and method* (*Vol. 1*, pp. 257–300). Boston: Allyn & Bacon.

Van Leeuwen, K. G., Mervielde, I., De Clercq, B. J., & De Fruyt, F. (2007). Extending the spectrum idea: Child personality, parenting and psychopathology. *European Journal of Personality, 21*, 63–89.

Wu, J., & Lebreton, J. M. (2011). Reconsidering the dispositional basis of counterproductive work behavior: The role of aberrant personality. *Personnel Psychology, 64*, 593–626.