

REVIEW ARTICLE

## *Choosing an adequate design and analysis in cross-cultural personality research*

Jia He<sup>1,2</sup>, Fons J. R. van de Vijver<sup>2,3,4</sup>

1: German Institute for International Educational Research, Frankfurt am Main, Germany

2: Tilburg University, Netherlands

3: North-West University, South Africa

4: University of Queensland, Australia

The flourishing of cross-cultural personality research requires a keen eye for rigorous methodology in such research. With decades of experience in cross-cultural research methods, we have come to appreciate that methodological aspects of such studies are critical for obtaining valid findings. Ill-designed or -conducted studies may produce results that are difficult to interpret. A careful design and analysis can help to deal with various methodological problems in cross-cultural personality studies. Drawing on the extensive knowledge that has been accumulated in cross-cultural and personality research in the past decades, we describe a framework of bias and equivalence

that enables the choice of adequate research methods and the avoidance of pitfalls that endanger valid conclusions in cross-cultural personality research. Specifically, we focus on sampling issues, test adaptations, and the combination of emic and etic approaches in this short review article. We encourage researchers to use the tools and experience that are available to considerably enlarge our insights in cross-cultural differences and similarities in personality research.

KEY WORDS

bias; research design; equivalence; etic-emic

---

CORRESPONDING AUTHOR – Jia He, Ph.D., German Institute for International Educational Research, Schloßstrasse 29, 60486 Frankfurt am Main, Germany, e-mail: [jia.he@dipf.de](mailto:jia.he@dipf.de)

AUTHORS' CONTRIBUTION – A: Study design · B: Data collection · C: Statistical analysis · D: Data interpretation · E: Manuscript preparation · F: Literature search · G: Funds collection

TO CITE THIS ARTICLE – He, J., & van de Vijver, F. J. R. (2017). Choosing an adequate design and analysis in cross-cultural personality research. *Current Issues in Personality Psychology*, 5(1), 3–10.

RECEIVED 02.09.2016 · REVIEWED 07.12.2016 · ACCEPTED 21.01.2017 · PUBLISHED 20.02.2017

## BACKGROUND

Cross-cultural personality research is burgeoning. The field is however complicated by different traditions and challenged by various methodological pitfalls. On the one hand, proposed personality models have been validated in the sense that similar structures of personality can be found in various cultural contexts (e.g., McCrae & Allik, 2002; Schmitt, Allik, McCrae, & Benet-Martínez, 2007); on the other hand, nuanced and more context-dependent aspects of personality have emerged in both culture-comparative (e.g., Church et al., 2011) and indigenous research on personality (e.g., Behrens, 2004). Personality instruments developed in one culture may not travel well to another culture, which may render results that are hard to interpret (e.g., Fetvadjev & van de Vijver, 2015). With decades of experience in cross-cultural research methods, we have come to appreciate that methodological aspects of such studies are critical for obtaining valid findings (van de Vijver & Leung, 1997). A careful design and analysis goes a long way to deal with various methodological problems. In this paper, we propose that advancing the field of culture and personality requires a better integration of perspectives and procedures in cross-cultural research methods.

Below we describe the use of adequate research methods and the avoidance of pitfalls in cross-cultural personality research, particularly focusing on choosing an adequate research design. We first introduce bias and equivalence as a general framework to deal with methodological challenges. This framework serves as the backbone of cross-cultural research methods that are relevant in all stages of a cross-cultural study. The second part of the paper is more topical. We focus in this part on three methodological aspects, which are salient in current cross-cultural studies of personality, namely sampling issues, instrument adaptation, and the combination of etic and emic approaches. These issues are by no means exhaustive, yet they merit special attention in advancing personality and culture research. Finally, we draw conclusions.

## BIAS AND EQUIVALENCE

### TAXONOMY OF BIAS

Bias occurs when score differences on the indicators of a particular construct do not correspond to differences in the underlying trait or ability (van de Vijver & Leung, 1997). This incomplete correspondence means in practice that whereas a response in one culture represents a target construct (e.g., conscientiousness), responses in another culture reflect other constructs (e.g., social desirability) or additional constructs (e.g., a combination of conscientiousness and social desirability). Based on the source of invalidity,

three types of bias are distinguished, namely construct bias, method bias, and item bias.

*Construct bias* indicates that the construct measured is not identical across cultures. It can occur when there is only a partial overlap in definition of the construct across cultures, or when not all relevant behaviors associated with the construct are present and properly sampled in each culture (van de Vijver & Poortinga, 1997). For instance, self-esteem conceptualized in interpersonal contexts is considered to result from fulfillment of desires for love (affiliative quality) and/or status (social dominance) (Zeigler-Hill, 2010), whereas self-esteem in more independent contexts is more related to individual achievements. Consequently, it is important to take multiple aspects of self-esteem into consideration in comparing different cultures where the sources of self-esteem differ.

*Method bias* comprises all nuisance factors that derive from the sampling, structural features of the instrument, or administration processes. *Sample bias* results from incomparability of samples due to cross-cultural variations in sample characteristics that have a bearing on target measures, such as confounding cross-cultural differences in education levels when testing intelligence, variations in urban or rural residency, or in affiliation to religious groups. *Instrument bias* involves problems deriving from instrument characteristics, such as stimulus familiarity (in cognitive and educational tests) and response styles (in personality and attitude inventories). Individuals from different cultures tend to have different levels of familiarity with stimulus materials (e.g., pictures taken in one culture may be easier to recognize for people from other cultures than for people from the same culture), response modes (e.g., differences in familiarity with computers in computer-assisted assessment), and response procedures (e.g., working with multiple choice formats). Such cross-cultural differences in background characteristics tend to influence the scores on target measures. In personality research, response styles are perhaps the most frequently studied source of method bias. Response styles are defined as the systematic tendency to use certain categories of the answering scale on some basis other than the target construct (Paulhus, 1991). For a long time they have been considered a validity threat that should be corrected for. A case in point comes from work by Eysenck and Eysenck (1975), who proposed to interpret personality scores of respondents on the target scales of their personality inventory (extroversion, psychoticism, and neuroticism) only if their scores on a social desirability measure were below a pre-determined threshold. Recent evidence suggests that these styles are more likely to represent valid individual and cultural differences closely related to personality (e.g., He, Bartram, Inceoglu, & van de Vijver, 2014; He & van de Vijver, 2015; Uziel, 2010). Findings on effects

of response styles on personality have been rather mixed. Some studies reported that the structure, mean levels, and variance of personality measures were confounded by response styles (e.g., Danner, Aichholzer, & Rammstedt, 2015; Rammstedt, Goldberg, & Borg, 2010), whereas other studies reported negligible effects of response styles on personality measures both within and across cultures (e.g., Grimm & Church, 1999; Ones, Viswesvaran, & Reiss, 1996). Still, it is unclear whether correction for response styles results in higher validity and better comparability of personality measures. Caution is needed in the use of corrections for these response styles; methods to adjust for response styles may remove genuine cross-cultural differences if individual or cross-cultural differences in scores are not just based on response styles but on a combination of response styles and genuine personality differences (Fischer, 2004).

A final type of method bias is *administration bias*. This type of bias can come from administration conditions (e.g., data collection modes, class size), ambiguous instructions, interaction between administrator and respondents (e.g., halo effects), and communication problems (e.g., language difference, taboo topic). In their comparisons between a computerized and paper-and-pencil administration of the Eysenck Personality Questionnaire, Merten and Ruch (1996) found that both modes produced comparable results in terms of scale means and standard deviations, yet the computerized assessment seemed to result in higher reliability for the Lie Scale. In general, method bias tends to have a global influence on cross-cultural score differences (e.g., mean scores of measures vulnerable to social desirability tend to be shifted upwards or downwards). If not appropriately taken into account in the analysis of data, method bias can be misinterpreted as real cross-cultural differences.

*Item bias* means that an item has a different psychological meaning across cultures. More precisely, an item of a scale (e.g., measuring agreeableness) is said to be biased if persons with the same level of trait, but coming from different cultures, are not equally likely to endorse the item. Item bias can arise from poor translation, inapplicability of item contents in different cultures or from items that trigger additional traits or have words with ambiguous connotations. For instance, certain words (e.g., the English word “distress”) or expressions in one language (e.g., the expression “comparing apples and oranges”) exist in some languages but often involves different fruits) may not have an equivalent in a second language, which challenges the translations of an instrument.

## TAXONOMY OF EQUIVALENCE

The taxonomy of equivalence, presented below, addresses the implications of bias for the comparability

of constructs and scores. More specifically, equivalence refers to the measurement level at which scores obtained in different cultural groups can be compared. Van de Vijver and Leung (1997) proposed a hierarchical classification of equivalence, distinguishing construct equivalence, metric equivalence, and scalar equivalence.

There is *construct equivalence* in a cross-cultural comparison if the same theoretical construct is measured in each culture. Without construct equivalence, there is no basis for any cross-cultural comparison; comparing inequivalent constructs amounts to comparing apples and oranges. Construct equivalence is a prerequisite for cross-cultural comparison. Researchers need to explore the structure of the construct and adequacy of sampled items. When a construct does not have the same meaning across the cultures in a study, researchers have to acknowledge the incompleteness of conceptualization and compare the equivalent subfacets.

*Metric equivalence* means that measures of interval or ratio level have the same measurement unit but different origins. In the case of metric equivalence, scores can be compared within cultural groups (e.g., male and female differences can be tested in each culture), and mean patterns and correlations across cultural groups, but mean scores cannot be compared directly across cultures. A simple example is the distance being measured by kilometers and miles. Distances measured by kilometers can be compared directly, and so can distances measured by miles, yet without converting the two measurements to the same origin, a valid cross-group comparison is impossible.

*Scalar equivalence*, the highest level of equivalence, implies that scales have the same measurement unit and origins. Scalar equivalence is the most difficult to establish in multicultural comparisons. Only if there is scalar equivalence are scores obtained bias free and thus can be compared directly. Analyses of variance, *t* tests, and more sophisticated analyses with mean structures such as multilevel analysis and structural equation modeling are appropriate for (and only for) this level of equivalence.

It should be noted that the three sources of bias and the three levels of equivalence are closely related, but not in full correspondence. In order to achieve construct equivalence, construct bias should be addressed; method and item bias may not influence construct equivalence, but they jeopardize measurement unit and full score equivalence. Table 1 provides a list of strategies to cope with different types of bias in order to achieve equivalence. In the next section, we highlight three methodological aspects that are related to choosing an adequate design for cross-cultural personality research. We highlight methodological considerations that are particularly relevant in understanding cultural universals and specifics of personality.

Table 1  
*Strategies in dealing with bias*

Type of bias	Strategies
Construct bias	Decentering (i.e., simultaneously developing the same instrument in several cultures) Convergence approach (i.e., independent within-culture development of instruments and subsequent cross-cultural administration of all instruments)
Construct bias and/or method bias	Use of informants with expertise in local culture and language Use of samples of bilingual subjects Use of local surveys (e.g., content analyses of free-response questions) Non-standard instrument administration (e.g., thinking aloud) Cross-cultural comparison of nomological networks (e.g., convergent/discriminate validity studies, monotrait-multimethod studies, connotation of key phrases)
Method bias	Extensive training of administrators (e.g., increasing cultural sensitivity) Detailed manual/protocol for administration, scoring, and interpretation Detailed instructions (e.g., with sufficient number of examples and/or exercise) Use of context variables (e.g., educational background) Use of collateral information (e.g., test-taking behavior or test attitudes) Assessment of response styles Use of test-retest, training and/or intervention studies
Item bias	Judgmental methods of item bias detection (e.g., linguistic and psychological analysis) Documentation of "spare items" in the test manual which are equally good measures of the construct as actually used test items

*Note.* Taken from van de Vijver and Tanzer (2004).

## TOPICAL ISSUES IN THE METHODOLOGY OF CULTURE- AND-PERSONALITY STUDIES

### SAMPLING IN PERSONALITY RESEARCH

Three sampling schemes are commonly employed in cross-cultural research: *convenience*, *systematic*, and *random* sampling. These apply to the sampling of both cultural groups and individuals. A large number of personality studies so far have used convenience sampling, in which cultural groups under study are not primarily governed by conceptual considerations but by availability, such as knowing a colleague from the other culture. Experience shows that such studies tend to suffer from the same sampling bias: The affluent part of the world (e.g., Europe, North America) is overrepresented, while less affluent countries, notably in Latin America, Africa, and South-East Asia, are underrepresented. To minimize sampling bias, a systematic sampling scheme is proposed whereby the sampling of cultures should be guided by research goals (e.g., select heterogeneous cultures if the goal is to establish cross-cultural similarity and homogeneous cultures if looking for cultural differences) (e.g., Boehnke, Lietz, Schreier, & Wilhelm, 2011).

The ideal sampling is to randomly sample culturally representative respondents in a large number

of randomly selected cultures; yet, due to resources and accessibility restraints, it is rarely accomplished in cross-cultural personality studies. Few personality projects span dozens of cultural groups (McCrae, 2002; Schmitt et al., 2007). It is not surprising that our knowledge of culture-level personality differences is not very systematic and replicable.

In sampling individuals, many studies have used university students or community samples, implicitly assuming that they constitute matched samples. However, this assumption may be invalid. For example, university education quality and enrolment rates in developed and developing countries differ significantly, which can introduce selection biases in the sampling process. When participants are recruited using convenience sampling, the generalization of findings to their population can be problematic. If the strategy to find matched samples does not work, it may well be possible to control for factors that induce sample bias by assessing such factors so that their influence can be statistically controlled (e.g., by using weights or analyses of covariance to account for confounding differences).

### TEST ADAPTATION

The choice of instruments in cross-cultural personality research depends not only on the availability

of existing instruments, but also on the research aim and methodological considerations. Three options in instrument choice are available in a cross-cultural study: adoption, adaptation, and assembly (van de Vijver & Leung, 1997). When the items in the source and target language versions have an adequate coverage of the construct measured and the response formats are appropriate in various cultures, *adoption* can be used by applying a close translation of this measure in another culture (Harkness, 2003). *Adaptation* involves a combination of a close translation of certain stimuli and modifications of other stimuli when adoption of all stimuli is inappropriate for linguistic, cultural, or psychometric reasons. Nowadays, adaptation is most frequently used when a multidisciplinary, multicultural perspective is taken (Harkness, van de Vijver, & Mohler, 2003). *Assembly* refers to the compilation of a new measure when the first two options are inadequate. An assembly can maximize the cultural appropriateness of an instrument, but it makes quantitative comparisons of scores across cultures difficult. Adoption is preferred if the goal is to compare scores across cultures directly,

whereas adaptation and assembly are better to maximize the ecological validity of the instrument.

As test adaptations have become a standard method to make sure that instruments are suitable for use in a cross-cultural context (Harkness et al., 2003), we further illustrate the different types of adaptation. The proposed classification of adaptations starts with the four types of equivalence: conceptual, cultural, linguistic, and measurement (Table 2). There are two subtypes of adaptations within each type. Related classifications can be found in Harkness et al. (2003). This taxonomy was initially developed on the basis of cross-cultural studies in large-scale surveys and intelligence testing; yet, most subtypes also apply to cross-cultural research on personality.

A *concept-driven adaptation* is a change of an instrument feature, usually the contents of a question, to accommodate differences in the indicators of culture-specific concepts, such as knowledge of the name of a very well-known person in the country (as an indicator of crystallized intelligence) or applicability of a certain concept to refer to an underlying construct, such as praying as a sign of religiosity, as

Table 2  
*Types of adaptations*

Domain	Kind of adaptation	Description and example
Concept	Concept-driven adaptation	Adaptation to accommodate differences in concepts in different cultures (e.g., knowledge of name of a widely known public figure in a culture)
	Theory-driven adaptation	Adaptation that is based on theory (e.g., tests of short-term memory span should use short stimuli in order to be sensitive, which may require the use of different stimuli across cultures)
Culture	Terminological/fact-driven adaptation	Adaptation to accommodate specific culture or cultural characteristics (e.g., conversion of currency)
	Norm-driven adaptation	Adaptation to accommodate cultural differences in norms, values, and practices (e.g., avoidance of loss of face)
Language	Linguistics-driven adaptation	Adaptation to accommodate structural differences between languages (e.g., the English word “friend” can indicate both a male and a female person, whereas many languages have gender-specific nouns for male and female friends)
	Pragmatics-driven adaptation	Adaptation to accommodate conventions in language usage (e.g., level of directness of requests by interviewers)
Measurement	Familiarity/recognizability-driven adaptation	Adaptations that result from differential familiarity of cultures with assessment procedures for specific stimuli (e.g., use of differential pictures of objects, such as pictures of houses)
	Format-driven adaptation	Adaptation to formats of items or responses (e.g., adaptations in response scales to reduce impact of extremity scoring)

*Note.* Taken from van de Vijver & He (in press).

praying is not equally relevant as an indicator across religions.

*Theory-driven adaptations* are instrument changes due to theoretical reasons. An instrument that has questions with a strong theoretical basis may require extensive adaptations in order to have items that still comply with the theory. In the domain of personality testing, theory-driven adaptations are uncommon as the field of personality has not yet advanced to a stage in which such close links between constructs and their assessment can be specified.

*Terminological/fact-driven adaptations* refer to culture-specific aspects that are less known or unknown elsewhere, representing “hard” aspects of culture. This type of adaptation occurs often in cognitive testing, such as the conversion of currencies (e.g., dollars to yen) or between metric measures (gallons to liters); in personality research, measures that refer to culture-specific aspects, such as the name of cities in the country, or names of national institutions or public figures would require a similar adaptation.

*Norm-driven adaptations* accommodate cultural differences in norms, values, and practices, representing “soft” aspects of culture. An item about someone’s activity at family parties (such as being the center of the party) may have some features of extroversion in many cultures, but as roles in such a party are culturally regulated, the item suitability will differ across contexts. Items dealing with such scripts need modification when they are used in countries with different customs.

*Linguistics-driven adaptations* refer to adaptations to accommodate structural differences between languages. For example, languages differ in their differentiation of words to denote kinship, such as the presence or absence of words to refer to cousins and nephews or to paternal and maternal grandparents. Another example is that in English “friend” can indicate both a male and a female person, whereas various languages use gender-specific words for male and female friends, such as German (“Freund” and “Freundin”). Also, fuzzy quantifiers such as rather, quite a bit, and moderately may be difficult to translate.

*Pragmatics-driven adaptations* capture changes in an instrument to accommodate culture-specific conventions in language usage, such as discourse conventions. The extensive literature on politeness indicates that close translations of requests do not convey the same level of directness and politeness in different cultures (Brown & Levinson, 1987). As another example, some languages use informal and formal ways to address other persons (such as the informal “tu” and the formal “vous” in French, which in English would both be translated as “you”). The problem with translating such terms is exacerbated by the differential use of the formal form across languages. For example, many languages would use the formal form in inventories to address participants, whereas

in other languages (such as Dutch) the choice would depend on the target audience (e.g., the informal form in a student survey and the formal form in a survey for the general population).

*Familiarity/recognizability-driven adaptations*, common in cognitive tests, result from differential familiarity of cultures with assessment procedures for specific stimuli. In personality assessment it involves the use of words that differ in commonness. For example, “feeling blue” in a depression questionnaire can be hard to translate into other languages as it may be difficult to find a metaphor that is as short, clear, and common to refer to a depressed mood.

Finally, *format-driven adaptations* refer to changes in formats of items or responses to avoid unwanted cross-cultural differences. For example, differences in extremity scoring may be reduced by using more options in Likert-type response scales.

#### COMBINATION OF ETIC AND EMIC APPROACHES

It is difficult to accommodate the diversity in cross-cultural personality findings under a universal theoretical roof (etic), as theories in personality are sometimes tied to their cultural contexts (emic) that cannot be fully characterized by universal frameworks (e.g., Church, 2009; Church et al., 2011; Fetvadjev, Meiring, van de Vijver, Nel, & Hill, 2015). Both the etic and emic approaches have certain methodological advantages and disadvantages; yet, it seems difficult to escape from the impression that differences between the two approaches have been much overrated and that both approaches are more complementary than often assumed.

Reconciliation of the etic and emic approaches in cross-cultural personality research can be expected to advance the field. On the one hand, we need to be more critical about the universal frameworks and promote indigenous studies that move away from western-based and -centric traditions. On the other hand, the idea that personality is only tied to cultural contexts and that no universal features can be found is counterproductive. Coming to terms with this seeming incompatibility requires a combined emic-etic approach. Making the best use of both approaches can lead to incremental knowledge accumulation. A combined approach involves the integrated and balanced treatment of culturally universal and specific aspects with etic and emic measurements or the use of mixed methods (i.e., using both qualitative and quantitative methods). Such an integrative approach faces plenty of challenges, such as the lack of procedures to combine qualitative and quantitative procedures, and the multitude of decisions to be made relying on both pieces of evidence. An example comes from the data structure obtained in the qual-

itative and quantitative stages of the South African personality project. Nel et al. (2012) derived a personality structure from qualitative (interview-based) data, comprising nine clusters: Conscientiousness, Emotional Stability, Extraversion, Facilitating, Integrity, Intellect, Openness, Relationship Harmony, and Soft-heartedness. Subsequent quantitative work (self-reports on items derived from the qualitative structure) revealed a simpler, six-factor solution (Fetvadjev et al., 2015): Conscientiousness, Emotional Stability, Extraversion, Facilitating, Positive Interpersonal Relatedness, and Negative Interpersonal Relatedness. The combined approach is considered promising, and the integration of emic and etic studies awaits further methodological developments.

## CONCLUSIONS

It is shown above that cross-cultural personality research can draw on a rich tradition of both qualitative and quantitative studies with multiple perspectives and culturally appropriate methods. We can move forward by striking a balance between universal and culture-specific aspects of personality and combining this balance with a solid methodology. It is important that we move away from preconceptions about universality and cultural specificity and that we become open-minded in the choice of models and procedures. We have focused here on procedures, highlighting how a context-appropriate cross-cultural personality study uses a combination of design and analysis issues: design can help to pre-empt various interpretation problems afterwards, whereas a fitting analysis is crucial to exploit adequate design so as to make valid conclusions possible. So, the choice of an adequate design is crucial in the potential value of a study. It is characteristic of modern personality research that it is more pragmatic and less dogmatic about choices of models and analyses. More than ever before, we appreciate that good cross-cultural personality research requires input from multiple sources, both in terms of theories and in terms of procedures. The field has moved beyond simple cross-cultural applications of the Five-Factor Model and exploratory factor analyses. It can be expected that these developments will continue and that we will rely more on other, more ecologically valid methods of personality assessments, such as free text from social media and observations of natural behavior together with self-report data. A theory-driven, context-appropriate, and well-thought-out design and analysis will also be crucial in such studies. We have introduced in the paper the framework of bias and equivalence, which should guide our design and analysis of cross-cultural personality research. We have highlighted sampling, adaptation and the combination of etic and emic approaches as topical areas

in culture-and-personality research where important developments are taking or should take place. Mindfully applying these design features is expected to advance our understanding of this field.

## REFERENCES

- Behrens, K. Y. (2004). A multifaceted view of the concept of *Amae*: Reconsidering the indigenous Japanese concept of relatedness. *Human Development*, 47, 1–27. doi: 10.1159/000075366
- Boehnke, K., Lietz, P., Schreier, M., & Wilhelm, A. (2011). Sampling: The selection of cases for culturally comparative psychological research. In D. Matsumoto & F. J. R. van de Vijver (eds.), *Cross-cultural research methods in psychology* (pp. 101–129). New York, NY: Cambridge University Press.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, United Kingdom: Cambridge University Press.
- Church, T. A. (2009). Prospects for an integrated trait and cultural psychology. *European Journal of Personality*, 23, 153–182. doi: 10.1002/per.700
- Church, T. A., 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. doi: 10.1037/a0025290
- Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, 57, 119–130. doi: 10.1016/j.jrp.2015.05.004
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London, United Kingdom: Hodder and Stoughton.
- Fetvadjev, V. H., Meiring, D., van de Vijver, F. J. R., Nel, J. A., & Hill, C. (2015). The South African Personality Inventory (SAPI): A culture-informed instrument for the country's main ethnocultural groups. *Psychological Assessment*, 27, 827–837. doi: 10.1037/pas0000078
- Fetvadjev, V. H., & van de Vijver, F. J. R. (2015). Measures of personality across cultures. In D. H. Saklofske & G. Matthews (eds.), *Measures of personality and social psychological constructs* (pp. 752–776). San Diego, CA: Academic Press.
- Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, 35, 263–282. doi: 10.1177/0022022104264122
- Grimm, S. D., & Church, T. A. (1999). A cross-cultural study of response biases in personality measures.

- Journal of Research in Personality*, 33, 415–441. doi: 10.1006/jrpe.1999.2256
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (eds.), *Cross-cultural survey methods* (pp. 19–34). Hoboken, New Jersey: John Wiley & Sons.
- Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. P. (eds.). (2003). *Cross-cultural survey methods*. Hoboken, New Jersey: John Wiley & Sons.
- He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology*, 45, 1028–1045. doi: 10.1177/0022022-114534773
- He, J., & van de Vijver, F. J. R. (2015). Self-presentation styles in self-reports: Linking the general factors of response styles, personality traits, and values in a longitudinal study. *Personality and Individual Differences*, 31, 129–134. doi: 10.1016/j.paid.2014.09.009
- 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, NY: Kluwer Academic Publisher.
- McCrae, R. R., & Allik, J. (eds.). (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer Academic Publisher.
- Merten, T., & Ruch, W. (1996). A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Personality and Individual Differences*, 20, 281–291. doi: 10.1016/0191-8869(95)00185-9
- Nel, J. A., Valchev, V. H., Rothmann, S., van de Vijver, F. J. R., Meiring, D., & de Bruin, G. P. (2012). Exploring the personality structure in the 11 languages of South Africa. *Journal of Personality*, 80, 915–948. doi: 10.1111/j.1467-6494.2011.00751.x
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679. doi: 10.1037/0021-9010.81.6.660
- Paulhus, D. L. (1991). Measurement and control of response biases. In J. Robinson, P. Shaver, & L. Wrightsman (eds.), *Measures of personality and social psychological attitudes* (vol. 1, pp. 17–59). San Diego, CA: Academic Press.
- Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44, 53–61. doi: 10.1016/j.jrp.2009.10.005
- Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits. *Journal of Cross-Cultural Psychology*, 38, 173–212. doi: 10.1177/0022022106297299
- Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243–262. doi: 10.1177/1745691610369465
- Van de Vijver, F. J. R., & He, J. (in press). Bias and equivalence in cross-cultural personality research. In T. A. Church (ed.), *Personality across cultures*.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
- Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29–37. doi: 10.1027/1015-5759.13.1.29
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 54, 119–135. doi: 10.1016/j.erap.2003.12.004
- Zeigler-Hill, V. (2010). The interpersonal nature of self-esteem: Do different measures of self-esteem possess similar interpersonal content? *Journal of Research in Personality*, 44, 22–30. doi: 10.1016/j.jrp.2009.09.005