

RESEARCH ARTICLE

Cross-Cultural Perspective in Romanian I/O Psychology

FONS J. R. VAN DE VIJVER
*Tilburg University, The Netherlands,
North-West University, South Africa,
and University of Queensland, Australia*

Abstract

The article describes important issues in the fledgling cross-cultural approach to I/O psychology in Romania. The first part of the paper deals with currently popular cross-cultural models of work values, and observes the relative absence of Romania in this global psychological map. The second part of the paper is more methodological and first describes a global framework of method issues in cross-cultural comparisons (bias and equivalence). I then describe specific issues in designing cross-cultural studies, such as the choice between existing questionnaires and developing novel tests, the importance of alternative interpretations of cross-cultural differences, and the need to have a team with all relevant types of expertise on board. Romanian psychology may benefit from the development of a cross-cultural approach, dealing with cultural heterogeneity in Romania, Romanians in the diaspora, and the place of Romania on the “psychological map of the world”.

Keywords

Romania, cross-cultural approach, research agenda, design, analysis

Starting point of this article is the increasing interest in cross-cultural I/O studies conducted within Romania (e.g., studies of differences in work values among different ethnic groups in Romania) or international, comparative studies involving Romania. Some cross-cultural studies have been conducted within Romania (e.g., Becker et al., 2012; Owe et al., 2013; Smith et al., 2016), but regrettably few large international I/O studies have included Romanian samples. Few international studies seem to have been initiated by researchers from Romania. Several studies have addressed well-known cross-cultural value dimensions, such as Hofstede’s (explained below in more detail), in Romania, often using existing international data as reference (and not collecting new international data). This state of affairs, in which there is much more

national than international literature on the topic, is regrettable even if it is easily understandable given the history and lack of research resources of psychology in Romania. Yet, it is clear that migration to and from Romania and globalization make a mark on Romanian society. This article is meant to explore the prolegomenon to a research agenda for the fledgling cross-cultural approach to I/O psychology in Romania. Not being a Romanian citizen and not being intimately familiar with Romanian psychology, it would be presumptuous for me to formulate a research agenda. However, from my international experience, I have a fair idea of themes of a research agenda that are relevant in other places. In order to develop a full-fledged research agenda, it will be important to add local elements and adjust the

agenda to the specific research expertise and interests present.

In the next part of the paper I describe Hofstede's (1980, 2001) model as the most influential model of cross-cultural differences in work values. Even if this model is not unique, it shares characteristics with many models in cross-cultural I/O psychology, as explained below. I briefly describe its contents and reasons why the model has become so influential in cross-cultural I/O psychology. This description includes a brief evaluation of the model, highlighting both its strengths and shortcomings. In the second part of the paper I move from substance to methods by describing the most common design and analysis issues of cross-cultural studies. This section describes the main methodological issues to be dealt with when conducting cross-cultural I/O studies. Finally, conclusions are drawn.

The First Influential Cross-Cultural Model of Work Values

Even if it was not the first in its kind, the cross-cultural study among IBM employees by Geert Hofstede (1980, 2001; see also www.geerthofstede.com) was the largest one at the time of its publication and has become the most frequently cited study in this area (Taras, Kirkman, & Steel, 2010). The study has some important characteristics that have been emulated multiple times: the study (a) is quantitative, (b) is comparative, (c) is based on a massive sample size (over 100,000 respondents were involved, which was unique in those days), (d) focuses on the structure of values by employing exploratory factor analysis, which was very popular in those days (process models became popular only much later), and (e) focuses on the country-level structure of values. This latter characteristic requires further explanation as it has become a major source of misunderstanding. Hofstede was not interested in individual data of employees in different countries, but he aggregated individual data at country level; each cell in the data matrix he analyzed referred to the average of a country on an item. As each country in his matrix constituted one observation, the data matrix that he analyzed

had very few observations (less than 40). By aggregating data Hofstede ensured that the factors that he found in his exploratory factor analysis referred to value constellations of countries, not of individuals.

In the original 1980 work, there were four dimensions. Power Distance "is the extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally. This represents inequality (more versus less), but defined from below, not from above. It suggests that a society's level of inequality is endorsed by the followers as much as by the leaders" (all definitions given here are quoted from <http://www.geerthofstede.com/dimensions-of-national-cultures>). "Individualism on the one side versus its opposite, collectivism, is the degree to which individuals are integrated into groups. On the individualist side we find societies in which the ties between individuals are loose: everyone is expected to look after her/himself and her/his immediate family. On the collectivist side, we find societies in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families (with uncles, aunts and grandparents) which continue protecting them in exchange for unquestioning loyalty." Uncertainty avoidance "deals with a society's tolerance for uncertainty and ambiguity. It indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations. Unstructured situations are novel, unknown, surprising, different from usual." Finally, "masculinity versus its opposite, femininity, refers to the distribution of emotional roles between the genders which is another fundamental issue for any society to which a range of solutions are found. The IBM studies revealed that (a) women's values differ less among societies than men's values; (b) men's values from one country to another contain a dimension from very assertive and competitive and maximally different from women's values on the one side, to modest and caring and similar to women's values on the other."

Many researchers had addressed that link between culture and psychological functioning before. Well-known examples are

studies in culture-and-personality (e.g., Bock, 1988). The culture-and-personality school has attracted many scholars, but in retrospect the intellectual yield of this tradition has been limited. The school has not produced a widely accepted model of the link between culture and personality, nor a model of personality differences across cultures. From this perspective, Hofstede's work has been pioneering. The main strength of the model is its empirical basis. Furthermore, there is a rich tradition in psychology, sociology, and ethnography in which the dimensions of individualism–collectivism and power distance have been described (e.g., Tönnies, 1887/2012; Triandis, 1995). Not surprisingly, these are the dimensions that have received the most attention in the literature.

The model and its applications have various problems. The first are methodological in nature. Firstly, the internal consistencies of the scales at country level are remarkably low. In a replication of the original study, Spector, Cooper, and Sparks (2001) found very low internal consistencies of the Hofstede scales at country level; in Romania the values ranged from .18 to .57. Secondly, individualism–collectivism and power distance are conceptualized as independent dimensions. However, a close reading of Hofstede's original 1980 publication reveals that these two were not independent dimensions but opposite poles of a single dimension in the original data analysis. It is not surprising that many studies have reported a strong negative correlation between individualism and power distance. The second problem comes from applications of the model. Despite repeated warnings in the literature, some of which coming from Hofstede himself, applications of the model have turned out to be prone to the so-called ecological fallacy. This refers to the attribution of country characteristics to individuals within a country. The probably best known example is the treatment of all East Asians as collectivistic. The main problem is that collectivism is a country-level characteristic that applies to countries in East Asia and not an individual-level characteristic of all East Asians. In case of an ecological fallacy, individuals are equated with their cultures, thereby neglecting within-culture differences.

The third problem involves the lack of global coverage of the countries in the original study. The world has 200+ countries but in the original study only about 40 countries were represented. Now, this would not be a problem if countries would have been sampled in a random manner. Obviously, random sampling was impossible as the study was limited to countries with a local IBM representation, thereby excluding large parts of the world, such as former communist countries and developing countries. So, the sample has a very strong cultural bias, with an overrepresentation of affluent countries.

It may seem unexpected that Hofstede's model has become so popular despite these shortcomings. The main reasons for the popularity are probably the novelty of the topic, the corresponding lack of competing models, the potential relevance of the four factors to many different fields (e.g., consumer behavior, management, HR, intercultural communication, and leadership behavior), and the increasing influence of globalization when the study was published 35 years ago. Other models have been proposed in the literature (e.g., House, Hanges, Javidan, Dorfman, & Gupta, 2004; Trompenaars & Hampden-Turner, 1998) but these have never become as popular as Hofstede's.

Values in Romania

Romania is poorly represented in these large cross-cultural projects. Hofstede did not have data for Romania available, therefore, he used projected values which yielded a pattern of "high power distance, collectivism, high uncertainty avoidance and a tendency towards femininity" (Neculăesei & Tătărușanu, 2008, p. 201). A fairly large number of studies have addressed organizational and national culture in Rumania, often directly or indirectly based on Hofstede's dimensions. From my (foreign) perspective, the limited overlap in the conclusions drawn in the various studies is remarkable. What could be the underlying reasons? Firstly, several studies have used the original Hofstede instrument, but data on internal consistency and replication of the four factor structure are lacking, which leaves open the possibility that replicability could be challenged by poor psychometric

characteristics of the instrument. Secondly, convenience sampling has been used in most studies. So, representativeness is hard to establish and incomparability of findings across studies should come as no surprise. Thirdly, Romania has changed a lot in the last 25 years. The country now harbors both (older) traditional, bureaucratic organizations as well as (younger) post-communist organizations that often have a very different organizational climate (Hurduzeu, 2015). As a consequence, the search for a single national organizational culture may be futile with such tremendous within-country variability. This conclusion does not make studies of organizational studies futile. Quite on the contrary, the heterogeneity of organizations (and cultures) within Romania makes the country an excellent location for such studies, examining the reasons and consequences of this diversity, which is relevant not just for Romania.

Design and Analysis of Cross-Cultural I/O Studies

The choice of measurement instrument is an important question in initial stages of a cross-cultural I/O project (Van de Vijver & Fischer, 2009). There are essentially three options: use an existing instrument, adapt an existing instrument, or develop a new instrument (Van de Vijver, 2015, 2016). What the best option will be, depends on a combination of linguistic, cultural, and psychometric criteria. Linguistic criteria involve the meaning of words and sentences (denotative and connotative meaning). Cultural criteria involve the compliance with local norms and habits and applicability of item contents. Psychometric criteria involve validity and reliability.

The first option, called *adoption*, amounts to a close (“literal”) translation, which means that an item or instrument is directly translated into another language, thereby staying very close to the original version. Adoptions are very popular, as they are simple to implement, cheap, have a high face validity, and (given favorable results of equivalence tests, as explained below) allow to compare scores obtained with the instrument across all

translations. Close translations work well for simple, straightforward questions, such as items about age and gender. However, when items refer to concepts that are more likely to have a cultural coloring that is not identical across participating groups, a close translation may be problematic. For example, items about Santa Claus and religious practices have usually a rather limited domain of cultural applicability.

The second option is called *adaptation*. When an item is adapted, parts of the contents are altered so as to improve the cultural appropriateness of the item. For example, an item about left–right wing preferences in politics may need to adjust the specific examples of left and right political parties. For example, in some countries “liberals” are more on the left side of the political spectrum (e.g., the US), whereas in other countries, like Australia, liberals are conservatives and are more on the right side. When a complete instrument (rather than an item) is adapted, it usually means that some items are closely translated whereas other items are just to improve the adequacy in the new cultural context. It may be noted that the underlying concept, left–right orientation, is deemed adequate and applicable in the countries of the study. It is only the operationalization that requires a context-specific approach.

Assembly is the third option. Assembly amounts to the development of an entirely new instrument. There is interesting work on depression in Zimbabwe; Patel, Abas, Broadhead, Todd, and Reeler (2001) found that the terms used to describe depressive feelings among Shona speakers referred to metaphors of the heart and the head. Similar metaphors do not exist in English. As a consequence, translating an instrument from English into Shona to measure depression would amount to writing new items that tap into the discourse that is used by Shona speakers. These assembled, new instruments cannot be used to compare scores of depression across countries; yet, such instruments are optimally suited to measure specific psychological constructs in specific cultures with a high level of accuracy and face validity.

Is there a single best way of translating instruments? Adoption is the most frequently employed procedure, presumably because it offers most scope for cross-cultural comparisons. Comparing scores across cultures is easiest when starting from directly translated items (adoption). It should be noted that in large-scale cross-cultural studies, such as PISA (the Programme of International Student Assessment; <http://www.oecd.org/pisa/>), items are developed in such a way that only direct translations are needed. All items that could have elements that are difficult to translate or that could have components that are specific for some or most of the participating countries are eliminated at beforehand. Many cross-cultural studies, however, start from a different position and set out to investigate the cross-cultural generalizability of an existing instrument that has been developed in a specific cultural context. For example, much work on the five-factor model of personality uses a single, originally American-English instrument that is translated in multiple languages. Item adaptations in such projects are not very likely, as researchers typically try to maintain as many original items as possible. Still, it is important to realize that modern statistical procedures do not require all items to be identical across all cultures. The Wechsler Intelligence Scales for Children have been translated in multiple languages and these language versions often deviate from each other in major ways. Georgas, Weiss, Van de Vijver, and Saklofske (2003) compared the scores obtained in various countries by identifying sets of closely translated items. Item Response Theory was used to estimate ability scores of each country using subsets of closely translated items; the non-common items were treated as country specific and were used to help define the latent underlying variable continuum.

It can be concluded that there is no single best way of translating or adapting instruments. Still, it is important that translations are checked for accuracy. A host of procedures are available to test this accuracy, such as translation – back translation procedures (Brislin, 1970), committee approach, interviewing bilinguals, preparing multiple forward translations, cognitive

interviews using their translated instrument, etc. (Harkness, 2003). Decades ago preparing translations was considered to be the expertise of linguists/translators. Thinking in cross-cultural psychology about preparing translations has progressed beyond that view. Large-scale studies tend to use teams with expertise in different domains, notably item writing/psychometrics, language, and cultures.

Bias

Bias refers to the presence of nuisance factors that challenge the comparability of scores across cultural groups. If scores are biased, their psychological meaning is culture dependent and cultural differences in assessment outcome are to be accounted for, partly or completely, by auxiliary psychological constructs or measurement artifacts. Three sources of bias are discussed here: construct, method, and item.

Construct bias occurs when the construct measured is not identical across groups. Construct bias precludes the cross-cultural measurement of a construct with the same measure. An example can be found in Ho's (1996) work on filial piety (defined as the set of psychological characteristics associated with being a good child to one's parents). In China children are expected to show love and respect to their parents but in addition, children are assumed to take the role of caretaker of elderly, needy parents. In a Western-based instrument the latter aspect (taking care of needy parents) would be less salient or even absent.

Construct bias is also apparent in commitment research. There has been a great interest in differences and similarities in behavioral commitment in East and West, particularly Japan and the US. High levels of behavioral commitment among Japanese workers (indicated by low turnover) were not strongly correlated with attitudinal commitment, as was found in the US. Therefore, the behavior of (or thoughts about) leaving one's organization was a good indicator of attitudinal commitment in the US, but not in Japan (Smith, Fischer, & Sale, 2001).

The second type of bias, called *method bias*, comes from all kinds of factors that are related to methodological and design aspects of a study, such as sample incomparability, tester and interviewer effects, and the mode of administration (comparing different modes of administration to assess a single construct).

Differences in samples can produce biased results. For example, in a study assessing solidarity in immigrant and mainstream Dutch participants, Arends-Tóth and Van de Vijver (2008) found that the expected higher family solidarity among immigrants than among mainstream Dutch was found, yet a considerable part of this difference could be statistically accounted for by socioeconomic status differences (in an analysis of covariance using socioeconomic status as the covariate). So, the expected cross-cultural differences were amplified by confounding sample differences.

Administration bias can be caused by differences in the procedures or mode used to administer an instrument. For example, when interviews are held in respondents' homes, physical conditions (e.g., ambient noise, presence of others) are difficult to control. There is evidence that personal characteristics of the interviewer have an impact on the attitudes observed among participants, a phenomenon that is known as deference. Deference to the interviewer has been reported; subjects were more likely to display positive attitudes to a particular cultural group when they are interviewed by someone from that group (e.g., Aquilino, 1994). Participants are usually more prepared to disclose personal information in self-completion contexts, such as self-administered questionnaires or online surveys, than in an interview. The probably best-known example of method bias are response styles, such as social desirability and acquiescent, midpoint, and extreme responding. It has been shown that these four response styles are systematically related to each other, together constituting a General Response Style (He & Van de Vijver, 2013). Cross-cultural studies have found consistent differences in such response styles. More specifically, ethnic groups or countries that are more affluent or more educated show lower levels of response styles. So, Western

countries, typically having high scores on affluence and education, can be expected to display much lower levels of social desirability and the other response styles than non-Western countries. Northern European countries tend to show lower levels of response style than Southern-European countries.

The third type of bias distinguished here refers to anomalies at item level and is called *item bias* or *differential item functioning*. According to a widely used definition, an item is biased if respondents with the same standing on the underlying construct (e.g., they are equally introverted), but who come from different cultures, do not have the same mean score on the item. The score on the construct is usually derived from the total test score. Of all bias types, item bias has been the most extensively studied; various psychometric techniques are available to identify item bias (e.g., Camilli & Shepard, 1994; Van de Vijver & Leung, 1997). Although item bias can arise in various ways, common sources are poor item translation, ambiguities in the original item, low familiarity/appropriateness of the item content in certain cultures, and the influence of cultural specifics such as nuisance factors or connotations associated with the item wording. In a recent study we compared well-being among (mainstream) Spanish and Dutch participants, asking their level of satisfaction with various domains in life (Benítez Baena, Van de Vijver, & Padilla García, in review). In cognitive interviews we found that in particular questions about family led to different interpretations. The word "family" was interpreted by Spanish participants as referring to their extended family (so, including grandparents, aunts, uncles, etc.), whereas Dutch participants referred to their nuclear family. Also, when asked what they did during their leisure time, Spanish participants mainly referred to activities with friends, whereas Dutch participants often referred to contacts with the extended family. Most Spanish participants often have contacts with their extended family, but did not view these activities as part of their leisure time.

Equivalence: Comparability of Constructs and Scores

Five hierarchically ordered types of (in)equivalence are proposed here (cf. Van de Vijver, 2015; Van de Vijver & Leung, 1997; Johnson, 1998, provides an excellent overview of the many different definitions of equivalence that have been proposed in the literature).

The first and most basic type of equivalence is called *construct inequivalence*; it amounts to comparing *apples and oranges* (e.g., the comparison of Chinese and Western *filial piety*, discussed above). If constructs are inequivalent, comparisons lack a common attribute, thereby precluding any comparison. There has been discussion whether a construct can exist in any specific culture if the language of that culture does not have a word for the concept. For example, there are cultures in which there is no linguistic distinction between guilt and shame. From a linguistic relativist perspective, this would mean that people in such cultures cannot distinguish these two emotions. Interestingly, this is not the case. Breugelmans and Poortinga (2006) asked Raramuri Indians in Mexico and rural Javanese, where the distinction between guilt and shame is not linguistically marked, to describe their bodily sensations in a number of hypothetical situations, some of which would be typically associated with guilt and some with shame. The patterning of their bodily reactions suggested that the participants felt different emotions in shame and guilt situations, which were very similar to the bodily reactions reported by individuals from countries with a linguistic distinction between guilt and shame. These results indicated that the concepts of guilt and shame can be used to describe emotions of Raramuri Indians in Mexico and rural Javanese, despite the absence of words in their language to describe these emotions.

Structural or functional equivalence is found if an instrument administered in different cultural groups measures the same construct in all these groups. In statistical terms, structural equivalence is supported if the same factors are found in different cultural groups. Much research in this tradition uses exploratory factor analysis and compares the

factors across the cultures involved. In many cross-cultural studies of personality a Western instrument to assess the five factors is translated and administered in a new culture (McCrae & Allik, 2002). The primary interest of researchers in this tradition is in confirming that in the new culture the same five factors are found as found elsewhere. There is indeed impressive evidence that the five factors are universal, although it has been argued that these five factors may not be exhaustive and that notably in non-Western cultures the five should be complemented by a factor denoting the social and relational aspects of personality (Cheung et al., 2001; Church, 2010).

The third type of equivalence is called *measurement unit or metric equivalence*. Instruments show this type of equivalence if their measurement scales have the same units of measurement, but a different origin. This type of equivalence assumes interval- or ratio-level scores (with the same measurement units in each culture). Measurement unit equivalence applies when the same instrument has been administered in different cultures and one or more biased items. This type of equivalence is often investigated using confirmatory factor analysis. The focus of the analysis is on the regression line, linking the latent factor to the observed scores. This regression line has two parameters: the slope, which is comparable to the factor loading in exploratory factor analysis, and the intercept, which indicates whether there is a systematic upward or downward shift in the regression line in any cultural group. The latter is an indicator of item bias. Metric equivalence has been found if all factor loadings are identical (“invariant” in the parlance of structural equation modeling) across all cultures, but not all intercepts are identical.

Only in the case of *scalar (or full score) equivalence* can direct comparisons be made; this is the only type of equivalence that allows for the conclusion that average scores obtained in two cultures are different or equal. Scalar equivalence assumes the same identical interval or ratio scales across cultural groups. In statistical terms, scalar invariance is supported if both factor loadings and intercepts are identical across cultural groups.

There is a close connection between the level of equivalence that has been observed

and the statistical inferences that can be made. If (only) structural equivalence has been supported, comparisons can be made in terms of the underlying construct. Having observed structural equivalence of the five-factor structure, it becomes possible to use the same set of personality traits to describe personality in all cultures involved. However, such an analysis does not yet warrant a comparison of mean scores in an analysis of variance or t test. Only if scalar invariance is supported, means can be compared across cultures. It is probably not superfluous to remark that many cross-cultural studies in which scores are compared across cultures do not provide evidence of scalar invariance of the scores. Without such evidence, comparisons of mean scores can only be premature. These comparisons are not necessarily inadequate, but there is just no methodological ground for such comparisons.

Preparing a Cross-Cultural Study: Some Relevant Questions

In this section I describe a number of important issues to consider when preparing a cross-cultural study. The section is not a cookbook. The reader is referred to other texts for an introduction to theories and methods. Introductory texts on cross-cultural psychology can be found in Berry, Poortinga, Breugelmans, Chasiotis, and Sam (2011) and Shiraev and Levy (2014); design and analysis issues are described in Van de Vijver and Leung (1997) and Matsumoto and van de Vijver (2011).

This section is meant to stimulate thinking about relevant topics right from the very first moment of inception of a cross-cultural study. The reasoning behind this section is that the quality of cross-cultural studies is enhanced by adequate quality management in all stages, starting from reading the relevant literature underlying the project to eventually reporting the results. There are numerous studies in the literature in which an integrated quality perspective is missing. For example, very advanced theoretical frameworks are tested using poor designs, conceptually poor studies use highly advanced statistical modeling, instruments that are being used in a cross-cultural study do not stand critical

psychometric scrutiny, etc. Cross-cultural studies are not only about conceptualization, not only about design, not only about sampling, not only about instrument administration, not only about analysis, and not only about reporting. The quality of a study is a combination of the quality of all study components.

1. Will the study advance the field?

It is good practice to think at beforehand about the question whether a study will make a difference. What is the novelty of the study? The fact that the study has not been conducted before is usually not sufficient. Suppose that someone is interested in a comparison of work motivation in Romania and the United Kingdom. The first question to consider then is why it is interesting to compare these specific countries on this specific concept. What is special about Romania and about the United Kingdom to make this comparison interesting? Is the comparison relevant for the further development of theory? Could the study be used to test (i.e., confirm or disconfirm) a specific model of cross-cultural differences in work motivation? Would it be possible to obtain very difficult to access samples that would make the study unique? Would it be useful to add another Slavic or some other culture? Clearly, it is important to go beyond the obvious reason that the proposed study has not been done before.

2. Which instruments should be used in the study?

The important consideration here is to what extent existing questionnaires or other instruments would be useful for the proposed study. If the study fits in an existing tradition, such as the studies of that five-factor model of personality described before, the choice for an existing instrument that needs to be translated is obvious. However, the choice is often more involved and requires a good evaluation of the appropriateness of existing questionnaires. For example, if the interest would be in identifying personality features that are unique for Romania, existing questionnaires may not be sufficient; additional qualitative, ethnographic studies may be needed to identify such unique aspects. If the interest

would be in the question whether relational orientation is an important personality characteristic in Romania, scales could be adopted from Chinese or South African instruments assessing this aspect of personality.

When an existing instrument is “imported”, an assessment should be made of its adequacy. This can start with a content analysis of all items, addressing the question of whether these items are adequate, applicable to Romania, and together would make a fair assessment of the concept in the Romanian context. In addition, an assessment of translatability is needed: Can all instructions and items be adequately translated in Romanian?

There is a trade-off between using existing and novel instruments. Existing instruments have been tried and tested elsewhere, thereby providing psychometric details that are not available for new instruments. In addition, the use of existing instruments makes comparisons with studies conducted elsewhere easier. However, this psychometric advantage can be outweighed by the conceptual issue of a less adequate cultural fit in the new context.

3. *Does the research team have the necessary skills to successfully conduct the study?*

The management aspects of cross-cultural studies should not be underrated. Studies involving multiple countries are often logistically complex, requiring a lot of coordination between researchers in the participating countries. Therefore, it is recommendable to think about the types of expertise that are required for a successful completion of the project and to check whether there is a match between needed and available expertise. I make the implicit assumption here that the cross-cultural study is not the work of a single person, but requires two or more people. The assumption is based on my experience in the cross-cultural field where it has become very unusual to find single-authored publications describing cross-cultural studies. Many projects will require psychological, cultural, linguistic, and methodological expertise. Few people will have all the expertise for comparative studies.

A coordination issue that has become more important over the years involves ethics approval. Even if most psychological research does not have to comply with the strict ethics rules of medical research, obtaining approval in all countries will require some coordination. In our research group, we usually draft a template that can be used in other countries, apply for ethics and submit an application to our ethics committee, explaining that the data collection is part of a larger data collection and that we seek approval for the Dutch data collection. Once we have obtained this permission, we inform researchers in other countries that can inform their ethics committees that approval was obtained for the Dutch leg of the study. What can make these procedures complicated is that forms to submit to ethics approval committees, supplementary materials to be submitted to committees, requirements to ensure privacy of data storage, expiry dates of permissions, times required by the committee to process applications, and other practical procedures typically differ across countries. Requirements of a specific country may have implications for all countries (e.g., some items may not be allowed in one country, which could lead to choosing another instrument or data collection procedure in all countries).

4. *How will the study team work together? What are the rights and duties of the research team members?*

Small-scale cross-cultural studies involving few countries usually do not use written contracts (or email threads) to specify the rights and duties of all participants. However, when groups become larger and participating researchers do not know each other very well, it is important to make clear appointments about responsibilities and entitlements early on. Thorny issues are usually not who will collect data, who will take care of project finances, and other obvious features that are usually arranged before a study starts. However, there are various additional features that require good appointments, such as ownership of the data (who is the owner of the data for how long? What will happen with data ownership once key papers have been published?), use of data for secondary analyses by non-project

members, authorship of publications (including author order), ownership of the jointly developed instruments and the question of who has or have the power to decide in case of a conflict about project-related issues.

5. *How will the comparability of data from different countries or cultural groups within a country be ensured?*

Assuming that the study is quantitative in nature, it is imperative to show that instruments administered in different countries or cultures are comparable across the samples. The kind of comparability that is required depends on the aim of the study. Some cross-cultural studies set out to compare constructs (construct-oriented studies; Van de Vijver & Leung, 1997), whereas other cross-cultural studies set out to compare scores on constructs (level-oriented studies). In the former, the aim is to check whether constructs, such as extroversion or work motivation, are applicable in all cultural groups studied. The equivalence analyses then usually boil down to conducting exploratory factor analyses in all countries and comparing the factors across the countries. In the latter, finding identical constructs across cultures is necessary, yet insufficient. It should also be shown that the measures are free from bias. In the current literature there is a focus on a single type of bias, namely item bias. So, if the aim of the study is to compare scores across different cultures, analyses are needed to show that no items are biased. Numerous procedures to conduct item bias analyses (differential item functioning analyses) are available in the literature (see, e.g., Van de Vijver & Leung, 1997). Biased items are excluded from further analyses.

The above description is straightforward, yet the application of the procedures is often not so straightforward. Notably when many cultures are involved or long instruments have been administered, bias analyses can be cumbersome, notably item bias analyses. It is very common in large-scale studies to find that many items show psychometric bias, usually item bias. Such bias would preclude any cross-cultural comparison including these items. There is not yet a clear, shared solution for this problem of prevailing item bias. Some workarounds have been proposed. For example, although most original item bias analyses focus on significance, most

statistical analyses allow for the estimation of effect size, which makes a shift to effect size analysis possible. What could be done is that the criterion for evaluating item bias is no longer significance but medium effect size. More specifically, only items with a medium bias effect size are removed from further analysis. What has also been done is that score differences between cultures are computed for scale scores based on all items and are compared to score differences between cultures computed on the basis of presumably unbiased items (i.e., those items that did not show a medium effect by size or were not significant in the item bias analysis). If the effect size of the cross-cultural differences is not much different in both analyses, the pragmatic conclusion can be drawn that the prevailing item bias did not have a major impact on the size of the cross-cultural differences, thereby making it plausible that scale scores can be used that are based on all items.

6. *Are there confounding differences? If so, how can these be dealt with?*

Campbell (1986) has argued that comparisons of two countries are often susceptible to alternative interpretations, also known as rival hypotheses (Bond & Van de Vijver, 2011). I illustrate the issue on the basis of a hypothetical example. Suppose that we are interested in work motivation in China and the Netherlands. A questionnaire to assess this motivation has been administered in both countries and significant differences in mean scores were observed, with the Chinese showing higher scores. The obvious conclusion would then be that Chinese employees show higher levels of work motivation than Dutch employees. However, what other explanations could be envisaged if we are not prepared to agree with these conclusions and would play the devil's advocate? It is not difficult to come up with a number of alternative interpretations; for example, response styles may have led to a difference in scores, there may have been differences in salary, educational level or other background characteristics between the employees in the two countries, the organizations from which employees were recruited may have been different in the two

countries, response rates may have been different in the two countries, to mention a few candidates. All these rival hypotheses have in common that they could partly or entirely explain observed differences between Chinese and Dutch work motivation. Campbell argued that it is often easy to think of alternative interpretations prior to a study. He argued that a good research design makes it possible to refute alternative interpretations. An example would be to include a measure of response styles and to examine to what extent cross-cultural differences in work motivation can still be found after statistical correction for response styles. Similarly, measures of background variables such as salary in education, can be used in the same manner.

In short, we can put more confidence in the conclusions of our cross-cultural studies if we are able to refute more alternative interpretations. In theory, this way of reasoning may seem an infinite regression as it is always possible to generate new alternative interpretations. However, in practice it works differently. Alternative interpretations are often relatively straightforward and limited in number from a practical perspective. In many cases the most obvious candidates are background variables, response styles, and sampling strategies. After controlling for the most obvious alternative explanations, we can be pretty confident that observed cross-cultural differences in scale scores can be interpreted in terms of the intended constructs rather than in terms of alternative interpretations.

Conclusion

Psychology has come a long way in Romania. Similar to other Eastern European countries, psychology is well-established as a discipline within the country; yet, the international visibility could be enhanced. In my view, there is ample opportunity to increase this visibility. The country provides tremendous resources for cross-cultural researchers; the Romanian population is heterogeneous, thereby providing a “natural laboratory” to study acculturation and intergroup relations. An interesting feature of Romania is that these ethnic groups have co-existed for long periods of time, where much acculturation research in Western Europe and the US studies groups that live in the country of

settlement for only up to three generations. The Romanian diaspora also provides an important resource. For example, in which countries do Romanian fare better and in which countries do they fare worse? What would be the reasons behind this difference? Finally, the heterogeneity of Romanian organizations is a relevant study topic. Furthermore, it is interesting to better understand the specific Romanian psychological constellation by comparing Romanian samples to samples from other countries. Addressing these issues will not only advance Romanian psychology, but it will help to advance cross-cultural psychology as a discipline. Therefore, the fledgling cross-cultural agenda will help to further the links between Romanian and international psychology.

References

- Aquilino, W. S. (1994). Interviewer mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly*, 58, 210–240.
- Arends-Tóth, J. V., & Van de Vijver, F. J. R. (2008). Family relationships among immigrants and majority members in the Netherlands: The role of acculturation. *Applied Psychology: An International Review*, 57, 466–487. doi:10.1111/j.1464-0597.2008.00331.x
- Becker, M., Vignoles, V. L., Owe, E., Brown, R., Smith, P. B., Easterbrook, M., ... & Camino, L. (2012). Culture and the distinctiveness motive: constructing identity in individualistic and collectivistic contexts. *Journal of Personality and Social Psychology*, 102, 833–855. doi:10.1037/a0026853
- Benítez Baena I., Van de Vijver, F. J. R. Padilla García, J.L. (in review). A Global global Strategy strategy for Investigating investigating Item item bias in cross-cultural studies: A mixed methods approach integrating differential item functioning and cognitive interviews.
- Berry, J. W., Poortinga, Y. H., Breugelmans, S. M., Chasiotis, A., & Sam, D. (2011). *Cross-cultural psychology. Research and applications*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9780511974274
- Bock, P. K. (1988). *Rethinking psychological anthropology: Continuity and change in the study of human action*. New York, NY: Freeman.
- Bond, M. H., & Van de Vijver, F. J. R. (2011). Making scientific sense of cultural differences in psychological outcomes: Unpackaging the magnum mysterium. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 75–100). New York, NY: Cambridge University Press.
- Breugelmans, S. M., & Poortinga, Y. H. (2006). Emotion without a word: Shame and guilt among Raramuri Indians and rural Javanese. *Journal of Personality and Social Psychology*, 91, 1111–1122. doi:10.1037/0022-3514.91.6.1111

- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216. doi:10.1177/135910457000100301
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Campbell, D. T. (1986). Science's social system of validity-enhancing collective believe change and the problems of the social sciences. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science* (pp. 108–135). Chicago: University of Chicago Press.
- Cheung, F. M., Leung, K., Zhang, J. X., Sun, H. F., Gan, Y. G., Song, W. Z., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the Five-Factor Model complete? *Journal of Cross-Cultural Psychology, 32*, 407–433. doi:10.1177/0022022101032004003
- Church, A. T. (2010). Current perspectives in the study of personality across cultures. *Perspectives on Psychological Science, 5*, 441–449. doi:10.1177/1745691610375559
- Georgas, J., Weiss, L., Van de Vijver, F. J. R., & Saklofske, D. H. (Eds.). (2003). *Cultures and children's intelligence: A cross-cultural analysis of the WISC-III*. New York, NY: Academic Press.
- 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). New York, NY: Wiley.
- He, J., & Van de Vijver, F. J. R. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences, 55*, 794–800. doi:10.1016/j.paid.2013.06.017
- Ho, D. Y. F. (1996). Filial piety and its psychological consequences. In M.H. Bond (Ed.), *Handbook of Chinese psychology* (pp. 155–165). Hong Kong: Oxford University Press.
- Hofstede, G. (1980). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Beverly Hills, CA: Sage.
- Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P.W., & Gupta V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.
- Hurduzeu, R-E. (2015). Culture and Leadership: The case of Romania. *The Romanian Economic Journal, 28*, 119–134.
- Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial, 3*, 1–40.
- Matsumoto, D., & Van de Vijver, F. J. R. (Eds.) (2011), *Cross-cultural research methods in psychology*. New York, NY: Cambridge University Press.
- McCrae, R. R., & Allik, J. (Eds.) (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer Academic/Plenum. doi:10.1007/978-1-4615-0763-5
- Neculăesei, A., & Tătărușanu, M. (2008). Romania – cultural and regional differences. *Analele Stiintifice ale Universității „Alexandru Ioan Cuza” din Iași, Tomul LV*, 198–204.
- Owe, E., Vignoles, V. L., Becker, M., Brown, R., Smith, P. B., Lee, S. W., ... & Baguma, P. (2013). Contextualism as an important facet of individualism-collectivism personhood beliefs across 37 national groups. *Journal of Cross-Cultural Psychology, 44*, 24–45. doi:10.1177/0022022111430255
- Patel, V., Abas, M., Broadhead, J., Todd, C., & Reeler, A. (2001). Depression in developing countries: Lessons from Zimbabwe. *British Medical Journal, 322*, 482–484. doi:10.1136/bmj.322.7284.482
- Shirayev, E., & Levy, D. A. (2014). *Cross-cultural psychology*. Pearson Education Limited.
- Smith, P. B., Fischer, R., & Sale, N. (2001). Cross-cultural industrial/organizational psychology. In C. L. Cooper and I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 16, pp. 147–194). New York, NY: Wiley.
- Smith, P. B., Ahmad, A. H., Owe, E., Celikkol, G. C., Ping, H., Gavreliuc, A., ... & Vignoles, V. L. (2016, in press). Nation-level moderators of the extent to which self-efficacy and relationship harmony predict students' depression and life satisfaction evidence from 10 cultures. *Journal of Cross-Cultural Psychology*. doi:10.1177/0022022116648210
- Spector, P. E., Cooper, C. L., & Sparks, K. (2001). An international study of the psychometric properties of the Hofstede Values Survey Module 1994: A comparison of individual and country/province level results. *Applied Psychology: An International Review, 50*, 269–281. doi:10.1111/1464-0597.00058
- Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multi-level, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*, 405–439. doi:10.1037/a0018938
- Tönnies, S. (2012). *Gemeinschaft—Gesellschaft*. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften. (Originally published in 1887). doi:10.1007/978-3-531-94174-5
- Triandis, H. C. (1995). *Individualism & collectivism*. Boulder, CO: Westview Press.
- Trompenaars, F., & Hampden-Turner, C. (1998). *Riding the waves of culture*. New York, NY: McGraw-Hill.
- Van de Vijver, F. J. R. (2015). Methodological aspects of cross-cultural research. In M. Gelfand, Y. Hong, & C. Y. Chiu (Eds.), *Handbook of advances in culture & psychology* (Vol. 5, pp. 101–160). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780190218966.003.0003
- Van de Vijver, F. J. R. (2016, in press). *Test Adaptations. ITC Handbook*.
- Van de Vijver, F. J. R., & Fischer, R. (2009). Improving methodological robustness in cross-cultural organizational research. In R. S. Bhagat & R. M. Steers (Eds.), *Handbook of culture, organizations, and work* (pp. 491–517). Cambridge, New York: Cambridge University Press. doi:10.1017/CBO9780511581151.019
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.