Received 18 May 2016
Accepted 22 May 2016

A failure before analysis: the soup to nuts of preparing for multicountry analyses

Charlotte M. Karam
Olayan School of Business, American University of Beirut, Beirut, Lebanon, and
David A. Ralston
University Fellows International Research Consortium, USA

Abstract

Purpose – A large and growing number of researchers set out to cross-culturally examine empirical relationships. The purpose of this paper is to provide researchers, who are new to multicountry investigations, a discussion of the issues that one needs to address in order to be properly prepared to begin the cross-cultural analyses of relationships.

Design/methodology/approach – Thus, the authors consider two uniquely different but integrally connected challenges to getting ready to conduct the relevant analyses for just such multicountry studies. The first challenge is to collect the data. The second challenge is to prepare (clean) the collected data for analysis. Accordingly, the authors divide this paper into two parts to discuss the steps involved in both for multicountry studies.

Findings – The authors highlight the fact that in the process of collecting, there are a number of key issues that should be kept in mind including building trust with new team members, leading the team, and determining sufficient contribution of team members for authorship. Subsequently, the authors draw the reader's attention to the equally important, but often-overlooked, data cleaning process and the steps that constitute it. This is important because failing to take serious the quality of the data can lead to violations of assumptions and mis-estimations of parameters and effects.

Originality/value – This paper provides a useful guide to assist researchers who are engaged in data collection and cleaning efforts with multiple country data sets. The review of the literature indicated how truly important a guideline of this nature is, given the expanding nature of cross-cultural investigations.

Keywords Data cleaning, Data collections

Paper type Research paper

As we seek to gain more in-depth knowledge of management practices both within multinational corporations operating in various parts of the world as well as small-to-medium enterprises operating locally, the importance of gathering data across continents is more important today than it ever has been. And, with the growing globalization of the world economy, it will become even more important in the future. To understand the dynamics of the relationships among the multiple players in today's global economy, survey research plays a pivotal role.

Presently, there exist a number of multicountry databases at the organizational level (e.g. Compustat, Data Stream International, Bloomberg). However, as we know, organizational decisions are made by individuals and groups of individuals, not by organizations. And, unfortunately, where there is a true scarcity of multicountry databases is at this individual – decision-maker, negotiator, motivator, leader, employee – level. Accordingly, there is a crucial need to build large-scale, multicountry databases of primary, individual-level, survey data so that we might answer many of the

Cross Cultural & Strategic Management
Vol. 23 No. 4, 2016
pp. 590-612

2059-5794
DOI 10.1108/CCSM-05-2016-0105

questions that we need to resolve and which can be resolved only with data about a person's perceptions, attitudes, values, beliefs, and behaviors.

Today, there are very few available large-scale research groups in the cross-cultural management area. The original one was, of course, Hofstede's (1980). However, in his seminal article, "Hofstede's model of national cultural differences and their consequences: a triumph of faith – a failure of analysis," McSweeney (2002) challenges the IBM data unpacking and discrediting a series of assumptions upon which Hofstede's later development of the cultural value dimensions were based. After reviewing the McSweeney (2002), and other critiques of Hofstede's work (e.g. Oyserman *et al.*, 2002), we would propose that the mistake of Hofstede was more a failure before analysis than a failure of analysis. That is, his data collection and data cleaning efforts were suspect and missing, respectively.

In this paper, we discuss how one might successfully develop a multicountry database that would be appropriate for the comparison of selected phenomena across a broad range of countries. As previously alluded to, there are two initial/preliminary tasks facing researchers who are interested in analyzing multicountry data. These are: building the team to collect the data; and cleaning the collected data to ensure that it is valid for analysis. Both tasks for preparing the data set are crucial; however, while the former has recently received some attention (e.g. Sargent and Waters, 2004), the latter has seemed to be substantially ignored. As such, we divide our paper into these two parts. In the former, we provide a narrative of the process that we followed. In the latter, which we see as the more interesting task due to the dearth of the comprehensive discussion of the process, we discuss the data cleaning procedures that we have found to be helpful in insuring that the data are "clean" and ready for analysis. Thus, the focus of this paper is upon what researchers need to do to develop their multicountry data sets and have them properly prepared to run the analyses to test the hypotheses of a study.

Part I: building a large-scale research team

We begin our discussion of Part I with the caveat of equifinality, meaning that similar results may be achieved by following different paths. And, while all roads may not lead to Rome, there are a variety of paths that will get you there. Likewise, there are a variety of paths that one might take to build a multimember team of international researchers. Accordingly, we do not portend that our way is the only – or even the best – way. However, we can state with certainty that our way is one way that does work even when you do not have a multimillion-dollar budget funding you.

Thus, we present our discussion of Part I as a narrative of what the second author did to create a large-scale research team, the UFIRC group, which to date has published 29 articles in quality journals. Therefore, the purpose of this first part of the paper is to provide specific steps that one might follow to build and manage a team of cross-cultural researchers with the key objective of creating a multicountry database that consists of primary data.

Why solicit members from multiple countries to create a large-scale team?

First, we contend that local expertise is needed in these large-scale data collection efforts. Because of the great disparity across countries in culture, economic level and technological sophistication, local researchers know the "topography" of the data collection terrain much better than foreign outsiders. Second, the potential participant pool will trust far more a local person collecting data than an outsider. Thus, in this scenario, multiple local researchers collect their home-country data, which they contribute to a central multicountry database. To this end, the primary question for the researcher trying to build a large-scale team is: how do I connect with local researchers and persuade them to join me in my research endeavor? We have some suggestions for your consideration.

When confronted with the idea of creating a large-scale research team, the first thought that may come to mind is not likely just do it! It is more likely, why do it? To build a large-scale team is challenging and time-consuming, by which we mean multiple years consuming. So, let us suggest a few reasons for those brave enough to consider this venture/adventure. One, when a task is very challenging and time-consuming, few will choose to take it on. However, for those who successfully take on the challenge, they will be well positioned to make their mark on the literature. They will have a unique database that no one else has … a true competitive advantage. Two, on the other side of the same coin, with the widening of business perspectives to global proportions, two or a few country studies are far less appealing today to the quality journals. Although there is certainly much merit to in-depth country studies where the researchers "go local" (Tsui *et al.*, 2007) and engage in in-depth cultural and socio-economic explorations, if we do not position ourselves to develop comparisons across multiple countries, we, as cross-cultural researchers, may lose sight of the bigger picture and the global patterns and trends. In fact, we may become as obsolete as the telephone switchboard operator or, for the younger generation, as the landline phone. Developing a large-scale research team provides job security. Of course, all of this assumes that the data collected is relevant and valid, which is a topic we address in Part II of this paper.

Trust: the key to a successful team

Before talking about what you can do to build a team, let's talk about what you, as a team leader, must bring to the table. The absolute key for success is to develop and to maintain a relationship of trust with the colleagues who will join your team. Beyond the ethicality of doing so, we academics live in a small world. That is, whether one can be trusted or not becomes well known quickly. There are too many cases, none that we will cite, of the leader(s) of a research project mistreating others in the group, typically by not giving them credit for the work that they had done. If one develops this type of reputation, no amount of perseverance will be sufficient.

Defining the expectations. It is crucial to make clear upfront what the expectations of all, including oneself, are. Creating a document that outlines these expectations is preferable to all parties trying to remember what had been said months or even years prior. This might be thought of as a contract of team involvement. Since ethicality and perceptions of appropriate behavior differ across cultures, this contract of team involvement, in essence, specifies the ethicality and appropriate behavior of team members when functioning within the team, regardless of home-cultural norms of ethicality. Thus, the more specific the contract is, the better the chances are of avoiding issues later on. An example of items that one might want to include in a "contract of team involvement" is given in the list below.

Examples of possible team leader and team members responsibilities:

(1) Team leader responsibilities:
- To develop a research project from which multiple articles can be published in top-tier and quality journals.
- To design the questionnaire to possess the potential of delivering several unique studies.

- To explain to all team members who contribute data to the research team data set that they will be offered the opportunity to be an author on all papers generated, which use the country data that they collected.
- To explain to all team members who contribute data to the research team data set that they will be offered the opportunity to be first-author on a paper from this data set, if desired by the team member. A specific procedure will need to be developed to institute this process.

(2) Team member responsibilities:

- To collect data in a country where data are not presently being collected and to be the only one collecting data in that country for the research project.
- To self-fund the data collection/data entry for his/her country.
- To have the translation – back-translation of the questionnaire completed, per Brislin (1986). Translations would be needed for all countries, except for the country(ies) that speak the language of the original questionnaire.
- To send a copy of the translated survey to the team leader(s).
- To collect the data from only professional/managerial respondents working in private industry.
- To collect data using the means prescribed (e.g. mail survey) and in the manner prescribed (e.g. voluntary participants who were provided anonymity).
- To use a data collection procedure that adheres to the sampling specifics criteria (e.g. sample size should be $\geqslant 125$ respondents; age range of respondents should be 25-55 years of age for actively employed managerial members of the workforce). Additional demographic criteria that might be included for specification are: gender distributions of the sample, educational level do respondents, workforce experience of respondents, number of respondents per any one company, size of companies in survey, industries to be included in the survey.
- To enter the data into an Excel template that will be provided.
- To serve as a resource for developing the literature review.
- To serve as a resource for developing the logic of the hypotheses for a manuscript.
- To serve as an internal reviewer for manuscripts developed from the data collection.
- To serve as an expert on the country for which the team member collected data.

Team members' expectations of the leader. The expectations that the team members will have of you, the team leader, can be summarized into two categories: ethicality and competence. Both of these are related to trust. Can the team members trust you, the leader, to be both ethical and competent? In terms of the first category, team members will look to you to take the lead in being ethical and to living up to the commitments made when presenting them with the expectations of the parties. This aspect will be substantially influenced by the reputation that you have built to-date among colleagues in your field. Obviously, the less experienced that you are, the less reputation – good or bad – you will have. And thus, the more uncertain others will be in joining your team.

In terms of the second, the team members will look to you to have the research competence to lead the team to success. This typically translates into publications in quality journals. Here you need to be prepared to oversee the various steps of the research process and to be accountable for the quality of each. This takes time, dedication, and commitment to put in the extra effort every step of the way to ensure quality completion. If you believe that you have what it takes to put in the extra effort every step of the way, then it makes sense to move on to the process of building the team. If not, this would be a good time to move onto something else.

The process of building the team

First, it is essential to develop a set of interesting and relevant research questions – deciding on what will be the focus of the research. This step is consistent with any survey research project. However, the second step is where a difference starts to emerge. Second, one needs to identify the array of countries that one sees as key to the specific research questions previously developed. While this array of countries could be global, it might also be regional or some other grouping depending upon the objectives of the research question. Third, one needs to develop a questionnaire that can be used to explore these questions. Ideally, this questionnaire should be composed of a set of cross-culturally validated measures. If it is not, the potential for data quality problems, which will be identified during the data cleaning process, become much more likely. Also, the challenge of convincing reviewers that the non-cross-culturally validated measure possesses internal validity is far more difficult. Fourth and finally, one needs to begin identifying researchers who are located in the countries of choice and who are interested in joining your team. At the same time, you may encounter colleagues who are interested in participating in this research from locations other than your key countries. Our philosophy has typically been, "the more the merrier," but this perspective is certainly not a rule. It is important to realize that this final step, as previously noted, is based on the perspective that locals – those living/working in the country that is to be studied – are better able to collect "good data" than are foreigners. The reasons for considering this approach are discussed in more detail in the collecting data section of this paper.

The search for team members. Where do I find the colleagues that I want for my research team? There is probably no wrong strategy, but the following are a few tips that we have found to be helpful. First, go to conferences to make presentations. At the end of your presentation, tell the audience that you are looking for colleagues from other countries to join the project and, if interested, to see you at the end of the session. Also, attend social events at the conferences. This presents the opportunity to talk informally about your research and the potential interest of others in becoming part of your research team. Another conference strategy is to look at the program to see who is from where and what the topics of their presentations are to, in turn, see if anyone might be a fit for your team. As an example, the second author notes, "I sat outside a meeting room, for a session that had run long, for almost a half hour waiting for an individual who had no idea I was out there or that I wanted to talk about my research. However, after an approximately twenty minute conversation over ice lemon tea, I had a new team member who could fill a key slot in my current research project."

There are opportunities, other than participating in conferences, to build your team. For example, when invited to do a presentation at a foreign university, use that opportunity to invite others to join your team. Ask colleagues who have known you and can vouch for your integrity and ability to approach their friends on your behalf. Also, you can send "cold-call" correspondence (e.g. e-mail) to colleagues, who you have identified as those you would like to have join your team, to invite them to have a conversation about your new research opportunity. This is frequently a non-successful approach; but, nothing ventured, nothing gained.

How successful you will be is also influenced by how well you are known within your academic field. The more established in your field that you are, the easier it will be to convince other colleagues to join your team. Conversely, the less experienced you are, the less likely others will want to join your team, especially established colleagues who already have a set research agenda. Nonetheless, one should not let how well-known one is to deter him or her from this venture, as it is junior researchers who most need access to a data set of this nature.

Incrementalism as a strategy for building the team. Once again, let us emphasize the principle of equifinality. We will propose one strategy for success, but we do not claim it to be the only one. However, it is one that we feel works particularly well for junior colleagues who are trying to build a team. In essence, it is incrementalism. At this point, you need to begin to nurture a trust relationship with others regarding both your ethicality and your competency. Get a few colleagues, perhaps existing friends, to join you and publish with them. They can then be your advocates on both fronts – ethicality and competency – assuming that you demonstrate both to this point. You likely will find it much more difficult getting your first-, second-, and third-team members than getting your 31, 32, and 33 team members. As a personal case in point, a few decades ago when the second author was building his first research team, it took over 12 years to assemble a team of 50+ research colleagues worldwide. Recently, for a comparable project it took him under 12 weeks. Your initial effort can continue to pay win-win dividends for the leader and the team members as long as the bond of trust remains.

Who might want to be your team members? Where do you look for the initial team members? We have already suggested that the more established colleagues in your field may be uncertain and even uninterested in joining your team. So, we recommend that you look to those who are eager to gain from working in such a team-based research endeavor. Why might colleagues be eager in this way? Some might be very, very junior (e.g. doctoral students) and therefore eager to "get out of the gates" running by participating in international research. Others might have tried and not been successful in publishing as the sole authors or leaders of a research project and are thus looking to partner with someone who is better able to publish. Still others are just insufficiently creative to come up with interesting research questions and are therefore eager to get involved in novel, emerging research projects. And, some are simply not willing or able to put in the time or hard work to publish due to a myriad of reason (e.g. personal will, pressures of work-family conflict, work-work conflict due to being employed at non-research institutions). These are the colleagues who are most likely going to be eager to join such teams, as they provide an opportunity to keep these individuals active in the academic publishing arena.

Your task, as the leader who is building a team, is to determine who would be both an individual who could make a worthwhile contribution and an individual who is eager to be part of your team. However, this is the period where your perseverance and resolve will likely be most stringently tested. You will likely spend countless hours talking with colleagues about your research idea and trying to sell them on why they should join your research team. Many will say no and some will say yes; and, of those who say yes, a few will follow through, but a number will not. At least, this is likely to be the case for the novice during his/her early stage of team building. Thus, plan to go down a multitude of "dead-end streets" with many, many colleagues in the beginning. Your success will be measured in the number who join and participate on your team, not in the number who do not.

Appropriate contribution of team members

A subject that has come up and likely will continue to come up is: what is a sufficient contribution to be an author on a research manuscript? The hard sciences and the social sciences tend to have divergent views on the answer to this question. The hard sciences such as medicine have long had extensive lists of authors on paper, which include those who have contributed to the success of the research endeavor being discussed in the manuscript (Gibbons, 1999; Onwude *et al.*, 1993; Regalado, 1995; Shapiro *et al.*, 1993; Strange, 2008, Tsao and Roberts, 2009). Some in the social science, especially those in psychology, have widely accepted stringent guidelines concerning what is needed to qualify for authorship (see American Psychological Association, 2001). We take the pragmatic perspective. If the IB discipline needs comparative research that includes more than two or three groups (e.g. countries), which we argue it does, then it is unreasonable to believe that 30 or 40 colleagues could fully collaborate in the writing of a manuscript. But, without the participation of those 30-40 colleagues, one would not be able to develop a large-scale research endeavor. Thus, other criteria need to be employed in determining authorship. These will vary depending upon the goals of the research team. Earlier in this paper, the list "Examples of possible team leader and team members responsibilities" presented what some of these criteria might be.

Collecting and consolidating the data

Once you have a sufficient number of members on your team to start collecting data, it is paramount to develop a structure for the process. While, in this section of Part I, we present specific steps for collecting and merging the data (i.e. developing the questionnaire, setting parameters for the data, designing data entry template, managing the multicountry data set), these are topics which the leader should have articulated, at least in his/her own mind, very early in the overall process.

Developing the questionnaire. As we have noted, a large-scale research project requires substantial effort. Thus, the questionnaire for it should be developed with multiple outputs (i.e. articles) planned. This requires careful planning to assemble a questionnaire that is composed of a set of instrument measures. Also, as previously noted, selecting cross-culturally validated instrument measures is the ideal. Once the measures/items of the questionnaire have been developed, the challenge becomes translating the questionnaire into the various languages relevant for the countries included in the research project.

Translation is crucial. There is strong evidence to suggest that participants found items to be more meaningful (Gibbons *et al.*, 1999) and concepts to be understood in a more refined manner (Church *et al.*, 1988) when presented in their native language thereby suggesting a key benefit to have surveys administered in a participant' s native language. And, of course, accurate translation is a key. Accuracy means not just translating words but translating the meaning that the item intends to convey. The translation – back-translation process has been well documented (Leung, 2008; Van de Vijver and Hambleton, 1996), so we will just mention the highlights of the process in this paper. One individual, who is fluent in both languages, should do the translation, and a second individual, who is also fluent in both languages, should do the back-translation into the original language version of the questionnaire. Ideally, these individuals should be translating into their native language. Any differences between the original questionnaire and the back-translated – into the original language version – questionnaire must be resolved by these two individuals. A third translator may be brought in to assist in helping to resolve the differences. At this point, the questionnaire is ready for pilot-testing. Ideally, a small ($n = 10$-15) group of participants from each country in the study should be asked to respond to the questionnaire. Upon completion, these respondents should be asked to discuss their impressions of the items with the researcher(s) from said country to investigate the potential of anomalies of interpretation within the country group. These discussions should then be shared with the other members of the research team to investigate the potential of anomalies of interpretation of the pilot groups across country groups in the research project. Revisions to the translated questionnaire items, if warranted, should now take place. Then, a re-pilot-testing of the translation(s) that needed to be revised should be conducted. Once the research leader and team are satisfied with all translations, the questionnaire is ready to be used.

Setting the parameters for the data. These parameters need to be communicated to each team member who will be collecting data in his/her respective country. These would include instructions regarding issues such as: the type of respondent desired (e.g. business professional, assembly line worker) and whether the respondent should be a native of this country or merely working and/or living there at the time of the survey, such as was the case for the Hofstede IBM data. Then, for these respondents: age distribution, gender distribution, educational distribution, organizational-level distribution, and industry distribution are desired. These are examples of likely parameters, not an exhaustive list. The parameters that should be selected will depend upon the objectives of the study.

Also to be addressed here is how the data are to be collected: paper-and-pencil; e-mail; electronic (e.g. SurveyMonkey); or other methods. This brings up a truly crucial question. Will respondents be guaranteed anonymity, confidentiality, or neither? Providing anonymity means that no one, including the researchers, can identify the respondent. This is often very important to ensure in order to secure respondents for the study. Without guaranteeing anonymity, many respondents feel less inclined to participate particularly, based on our experience, in some developing or transitioning economies where freedom of speech or human right guarantees are more fragile. Further, if respondents are concerned that their responses may be used against them, they will likely respond in a manner that they see as being the response wanted by those in power within the country or possibly the researchers. The result may be bad data due to a social desirability response bias.

Confidentially means that the researcher knows the respondent and can put him/her together with his/her questionnaire. By offering confidentiality, the researcher is promising the respondent that the respondent's answers will not be shared with anyone else. Here, the issue of trust also comes into play. Does the respondent trust the researcher to live up to this promise. Thus, confidentiality does not offer the respondent the level of safety/security that anonymity does. But, it does allow the researcher to have a better response rate because, for example, after a period of time, the researcher can contact the non-responders requesting them to respond. A final and related question is: are the responders voluntary participants or were they mandated to participate. The literature has shown that respondents that volunteer (i.e. have a genuine interest in the survey) are likely to be more careful in responding (Meade and Craig, 2012; Tourangeau *et al.*, 2000). Therefore, it is our position that granting anonymity to voluntary responders provides the most accurate data. If a researcher collects inaccurate data, the findings will provide erroneous results, and, in turn, the researcher will draw inaccurate conclusions.

Designing the data entry template. An enormous amount of work can be saved by developing a data entry template that is given to all team members collecting data to use when transcribing the questionnaire data to an electronic format (e.g. Excel). With this approach, as country data comes in to you, the team leader, you can simply merge the files since all electronic files will be using the same variable names located in the same order for each country. In our experience, Excel appears to be a reasonable electronic format to consider using. Finally, it is important to explain to your team members why you have prepared the template for them, as just discussed, and that it is important to not change it in any manner. If not, it is likely a few of your well-meaning team members will change the format or the variable names to make the template "better" in their opinion.

Managing the multicountry data set. Who should have access to the multicountry database? When the team leader negotiated a partnership with team members as they came on board, the members, in our scenario (see the list "Examples of possible team leader and team members responsibilities"), were promised that they would be coauthors of all publications that included their data. It is the team leader's responsibility to see that this commitment is met. If not, the trust relationship is broken and the team is likely to experience conflict and strife and perhaps deterioration. Thus, the team leader should be the sole keeper of the merged database and should also be the individual who controls the running of the analyses for the various studies developed from this database. This is the simple model. As the team expands, the team leader may want to form an "inner circle" of colleagues that could include an expert in statistics who is also given access to the database. But, once again, trust is paramount for the inner circle to function effectively. If trust is lost, problems will almost certainly ensue. In sum, if your multicountry database is shared with all team members, a team member could use it but not follow the contract, which the team leader and the members agreed to prior to collecting the data.

To this point, we have identified an effectual path to follow for those desiring to create a large-scale research team. We also have identified some potential pitfalls one might encounter. However, it is highly unlikely that we have identified all. So, our advice is, researcher beware. In order to move to Part II, we now assume that the data from the various countries in the study have been collected and submitted to the team leader for processing.

Part II: the data cleaning process

Based on our review of the literature and hands-on experience, we developed the following procedure for ensuring that the data are ready (clean) for analysis. Despite its importance, as the fundamental foundation for rigorous empirical research, data cleaning is often underreported raising questions about whether in fact proper attention was paid to this issue (Keselman *et al.*, 1998; Osborne, 2008). Having clean data is essential, period. Unfortunately, the potential for problems is more likely when there is a multitude of different country data sets being integrated into a single database, due typically to cultural influences on the interpretation of questionnaire items.

A reading of the cross-cultural management literature indicates that any discussion of the basic steps of data cleaning is often absent. Typically, the reporting on the data jumps from sampling, to scales, to results, leaving the essential step of data cleansing markedly absent (Osborne, 2013). Our own review of multicountry studies published in the *Journal of International Business Studies* between 2010 and 2015, which analyzes primary data that was collected by the authors using the survey method indicates that of the 31 articles identified, only two (0.065 percent) actually refer to any step relating to data cleaning. Both of these articles simply focussed on the elimination of surveys due to specific restrictions (e.g. incomplete data) and not on a detailing of their data cleaning steps or testing of statistical assumptions.

The reasons for this general absence are less clear. Perhaps, it is because editors and reviewers do not require it of the authors, which would seem to us to be an oversight. Perhaps it is because the researchers (authors) do not clean their data before running hypothesis testing analyses due to them not knowing that they needed to do so, or simply because they did not chose to go through the rigors of the process. If the reason is researcher negligence, then the oversight group (editor and reviewers) should play a role in seeing that the data cleaning process has been properly completed. To date, unfortunately, this oversight role appears to be a missing part of the review process for most IB journals.

Data cleaning takes time and effort, and for multicountry research endeavors, these same time-consuming steps need to be repeated for each country separately, making the process substantially more time-consuming and tedious. Osborne (2013, p. 12) presents a convincing argument as to why data cleaning is essential for scientific progress and how there is substantial evidence to suggest that, "Not only can problematic data points lead to violation of other assumptions (e.g. normality, variance homogeneity) but can lead to mis-estimation of parameters and effects without causing severe violation of assumptions."

We have developed our subsequent discussion based on our review of the literature and our own hands-on experience. In it, we outline the set of procedures that we have utilized to ensure that our data are clean and ready for the hypothesis testing analyses.

Overviewing the process

At the outset of delving into the data set, one clear challenge relates to the difficulty in ensuring that the constructs measured have similar meaning in the different cultural contexts examined (Peterson, 2009; Van de Vijver and Leung, 1997) and that the scales used by researchers are appropriate to capture the phenomena locally (Farh *et al.*, 2006). In a related vein, Hult *et al.* (2008) assert that failure to establish that the elements of the research design are equivalent – including construct equivalence, measurement equivalence and data collection equivalence – across the countries/cultures examined may produce biased results, and, in turn, bias the theoretical interpretations of those results.

Even when equivalence is established, an equally significant challenge exists with regards to ensuring that proper data cleaning procedures are used for each subset (e.g. country) of data collected. Thus, in the sections that follow, we provide a "recipe" of sorts listing the ingredients needed to ensure clean data. Our aim is to provide a hierarchy of steps to follow to prepare the database, which was the outcome of the Part I discussion of this paper, for analysis.

To begin, you will have a set of single country data files that you need to clean and then merge into a master database file. It is this part of the process that is the most tedious. This perhaps explains why it may be overlooked. It is, however, this part of the overall process that is of uppermost importance. To proceed, we assume that you have a single country data file opened in a statistical analysis program and that the variables are correctly labeled with valid values also defined. We recommend the following five-step approach, which should be repeated individually for each group (e.g. country). These five steps are: preparation, screening, correcting data problems, checking sample demographics, and checking factor analyses and scale reliabilities.

Preparation. What are we interested in knowing? It is important to reflect on the purpose of the overall study and how the data collected will help to answer relevant questions related to this purpose. Indeed, as part of the data cleaning process, preparation is an important first step as it sets the foundation for later steps. At this initial stage, the researcher (team leader) should gain significant familiarity or refresh one's memory regarding the type(s) of information collected, the measuring instruments used and the data file within which participant responses are entered (Smith *et al.*, 1986). Smith *et al.* (1986) suggest that researchers should re-read the data collection instrument carefully and familiarize themselves with the item-response scales. Further, it is important to take the time to become familiar with the constructs included in the data set as well as their corresponding dimensions and scoring keys. One particular point of caution is to give special attention to reverse coded items (Swain *et al.*, 2008; Weijters *et al.*, 2013) particularly for cross-cultural studies where misinterpreted phrases of negation are likely to be compounded when the scales are translated for use in other languages (Wong *et al.*, 2003).

Being familiar in this way will allow you to double check that all items, dimensions, and constructs are represented in the data set and that the corresponding data in each column has been included in the file. Familiarity also allows you to double check that the items are coded and reverse coded according to instrument specifications. It is a good idea to also scroll through the overall datasheet to ensure that the number of participants that you are expecting is in fact included.

Finally, remember that the aim is to ultimately merge the single country data files into a multicountry master database. Therefore, be sure that each single country has the minimum number of responses required to increase the between-country comparability for the later analyses. Please note that effective sample size design for cross-cultural studies involves a myriad of considerations such as the specific survey variables and the statistic of interest. Harkness *et al.* (2010) present a thorough discussion of this topic.

Screening. The main goal of screening the data is to identify problems and problematic patterns within the data set (DeSimone *et al.*, 2015). In particular, this step is important to identify whether there are bad or missing data, to assess normality of your data distribution and to identify any extreme outliers. In terms of checking for bad data, what you want to look for are basic errors in data entry (e.g. item responses outside of the acceptable range). Here, you want to keep track of the values that appear to be entered carelessly in the survey or incorrectly in the data file in order to double check with the original surveys (Meade and Craig, 2012). Further, you may be importing country data from one software program to another, perhaps from

Excel to SPSS or SAS. This transformation process may also result in errors. For example, you may receive country data in Excel format and upon examination, you may notice that some cells include data coded as “N/A” or “2&4.” When transferred into SPSS, for example, the “N/A” or “2&4” is interpreted as an error by the program.

In addition to incorrectly entered data, you will often have to deal with missing data. When you view the data with your eyes (i.e. not with statistical software), you are likely to notice mistakes. For example, a missing value might be imported into a software statistics program as “0” instead of a blank. This recoding can bias the data by changing the actual data included and potentially the variance within the variable of interest. Be sure, therefore, to review the raw data sets and to recode into values familiar to the statistical software that you are using. It is important to count the number of missing values per case (i.e. respondent’s data) as there may be a large number of values missing for one individual but not for others. It is important to decide on a criterion or a missing value threshold above which you will delete the entire case. Also, it is helpful to examine the standard deviation for the items per case and then to determine if it is less or more than 1. Generally speaking, if the standard deviation is less than 1, the case may be suspect. Scan the case further for a problematic pattern of responses (e.g. all responses with a value of 3). If this is the situation, then you will likely need to also delete the entire record for that respondent. However, we recommend waiting until the country file is merged into the master file before deleting cases from the file because when juxtaposed with other countries, the missing data may reveal important information about the instrument itself, such as unclear instructions or misplaced response items. You can choose to tabulate the occurrence of non-responses against responses for different groups to detect further patterns. If it appears that the country-level data have a large number of missing values, it may be important to contact the data collector for that country file to ask for an explanation as to why the omissions occurred. This is where the trouble log might be helpful.

Finally, when assessing normality of your data distribution, there are a number of simple steps that can be initially taken to screen your data for problems, such as creating graphic representations of the data distribution. For example, histograms are a good first step for visually detecting problems of skewness (i.e. the symmetry of the curve) and of kurtosis (i.e. height of curve in relation to its width). Other actions that should be taken as follow-up steps include comparing the actual cumulative probabilities of your data to the expected normal distribution by producing P-P plots, calculating the *Z*-skewness and *Z*-kurtosis values, or running inferential tests of normality such as the Kolmogorov-Smirnov (K-S) test. This helps indicate whether or not the data distribution differs significantly from normality. For both the *Z*-skewness and *Z*-kurtosis both values are calculated by dividing the skewness value (or kurtosis value) by the standard error of measurement of skewness (or kurtosis). Then the absolute values of both the *Z*-skewness and *Z*-kurtosis values are compared to a specific *Z*-score threshold. Commonly used *Z*-score thresholds include: 1.96 (samples of 100 or less), 2.56 (samples between 100 and 300), and 3.29 (samples of 300 or more) (see Tabachnick and Fidell, 2012).

Another method for testing normality is the K-S test with a score of 0 for a normal distribution. The further the score is from 0, the more likely it is that the data are not normally distributed. It is useful to convert the K-S scores to *Z*-scores and then to observe if the resulting absolute value of the score is greater than one of the previously mentioned *Z*-score thresholds. The *Z*-score thresholds, as mentioned, are to be chosen based on their respective conditions (Tabachnick and Fidell, 2012). If the absolute values are indeed greater than any of these thresholds, the K-S test is considered to be significant ($p < 0.05$) (see Field, 2009, p. 139). The K-S test should not be significant to assume normality.

Correcting problems in the data. As you become familiar with each data set you will note that perfectly entered data is a rarity. There may often be errors in data entry and equally as often there may be missing or incomplete data entries. It is very important to take the time to correct these errors even though some of the appropriate ways to make corrections may be time-consuming. For example, to rectify errors in data entry, the first step is to actually go back to the original surveys to cross-check data entries of the participants. This may involve checking the original datasheets (if collection online) or the paper-and-pencil printouts. If you discover that the error was actually a participant response error, you may want to consider substitution methods to rectify this error, as subsequently described. However, if it appears that multiple errors exist on the original survey (i.e. that errors appear to be the result of careless participant behavior), then the best step is to send the surveys back to the data collector for an explanation and to decide whether recollection of new surveys is necessary. Alternatively, Curran (2015) suggests that when facing data sets with evidence of careless or insufficient effort responding, it is important to employ techniques to identify and remove these data. Indeed, he offers a variety of techniques (e.g. long string analysis, inter-item standard deviation, polytomous Guttman errors, etc.) to screen survey data and to identify responders or responses that are problematic thereby reducing error in the analysis and improving the likelihood of more valid results. The use of multiple techniques together, Curran (in press) argues, is the best way to proceed in detecting such errors and that multiple techniques should be widely used as part of normal data screening and reporting.

Missing data. Missing data is another problem frequently faced by researchers, but it is one that is often not explicitly addressed (Cole, 2008; Newman, 2009; Schafer and Graham, 2002). An initial consideration is to decide what percentage of missing data from a single participant is acceptable. DiLalla and Dollinger (2006) suggest that 5 percent missing data for any given measure is a common threshold for dropping a participant. They suggest a similar 5 percent cutoff rule also be applied to individual scales that constitute multiscale inventories. Furthermore, it is very important to contemplate why the data are missing in the first place. Is there a pattern? If there is no pattern to the missing data and the data appears random, then the non-response can likely be ignored. For a good discussion of specific techniques to use to explore whether your missing data is random or not, see Osborne (2013) who also discusses the pros and cons of each technique.

Beyond determining whether the missing data is random or not, there are different techniques for dealing with missing data; and, in fact, many software programs (e.g. HLM by Raudenbush and Bryk, 2002) have a default setting to deal with missing data. This default setting is called casewise deletion and results in the case being completely removed from the data set. The major drawback here comes from the reduction of the sample size, which, in turn, leads to the loss of power and increases the likelihood of misestimating the population parameters (Schafer and Graham, 2002). The other alternative is pairwise deletion that retains the usable portions of a participant's responses. For example, if a person fills the part of the survey pertaining to Variables A, C, and D but not for B, then the data for this individual for Variables A, C, and D are retained and the blanks corresponding to Variable B are removed from the data set for this person. Although not ideal, this technique allows researchers to maintain the highest number of respondents per variables. On the downside, pairwise deletion leads to different sample sizes, which could be problematic in terms of replicability and in terms of increasing the odds of errors of inference (Cole, 2008; Osborne, 2013).

Beyond these commonly used deletion techniques for dealing with missing data, a researcher may decide not to remove the data from the sample but rather to substitute or to impute (i.e. substitute a missing observation with a likely value) observations. The object of such substitution methods is to substitute in a manner that will not bias the data as a whole (DiLalla and Dollinger, 2006). Common substitution methods include: substitution of each missing data point with the overall sample mean, or substitution of each missing data point with the imputed value. In the case of this latter, the researcher employs an equation to estimate values to replace the missing data (Osborne, 2013; Schafer and Olsen, 1998). Osborne (2013) explains that this substitution technique is similar to substitution using multiple regression in that it "uses information available in the existing data to estimate a better value than the sample average" (p. 20). Whatever technique you choose, it is important to include a clear statement describing the technique for future replication purposes. It should be noted that there is a variety of creative statistical approaches to address missing data (see Cole, 2008) and there are other more advanced techniques continuously emerging (e.g. expectation-maximization algorithm, iterative refinement technique) from various fields of study such as data analytics (see Abraham and Russell, 2004; Allison, 2002; Karmaker and Kwek, 2007). These are techniques that researchers may wish to further explore for applications in their cross-cultural management studies.

Outliers. In any data set there may exist outliers. These are data points that are problematic because they do not "hang" with (approximate) the rest of the distribution. These should be examined closely, and the decision to remove or retain should be made before the data analysis proceeds. Retention of outliers in the data set increases error variance and reduces the power of statistical tests by changing skewness or kurtosis of a variable (Osborne, 2013). Examination for both univariate and multivariate outliers is important; with the first often tested through standardized *Z*-score conversions and the second through the calculations of Mahalanobis distances of the values (Penny, 1996). Within any data set cases that are considered to be both univariate and multivariate outliers are particularly problematic and should be removed from the data set (see Tabachnick and Fidell, 2012).

Furthermore, to assess whether you have outliers in your sample, one should examine how many standard deviations above or below the mean the data points lay. According to Chebyshev's theorem, one commonly used standard is that 75 percent of good data are expected to fall within two standard deviations of the mean and 89 percent within three standard deviations (Barnett and Lewis, 1994). With larger data sets, some suggest using three standard deviations above or below the mean as a more stringent measure. DiLalla and Dollinger (2006) further recommend a method to assess what they refer to as "globally outlying participants" which involves "Creating dummy variables that reflect a participant's overall response across all items in the data set" (p. 247). If assessed to be deviant across response items, then one could decide to remove the outlier(s) completely.

The basis for examining outliers rests with the assumption that extreme values may reflect errors, or they may suggest that the extreme value was drawn from a different population (DiLalla and Dollinger, 2006). In the first case, you should remove the outlier from the data set. In the second case, you should take steps to analyze further in order to determine whether the outlier does not, in fact, come from the population that you intended. One such analysis is the Mahalanobis D statistic where the suspected outlier score is compared to the sample mean scores across all items within a survey in order to estimate the multivariate distance between the individual score and the sample mean score (see DeSimone *et al.* (2015) for a complete description of this statistic).

In cases where outliers appear to be a legitimate part of the data set, researchers may want to consider using data transformations to keep the case in the data set and simultaneously decrease error variance (Osborne, 2013). These transformations can be used to minimize the impact of the outlier on any statistical analysis undertaken (Osborne, 2013).

Checking sample demographics. Once the data has been cleaned, as described above, the next step is to compute the appropriate basic descriptive statistics to explore the sample characteristics and to get a better understanding of the groups that comprise your sample. This can be thought of as a preliminary analysis where the researcher explores "whether the sample behaved" as expected (see DiLalla and Dollinger, 2006, p. 249).

We suggest that you calculate basic statistics including: sample size (*n*), minimum and maximum, mean/mode, percentages, and standard deviation for each key variable in the survey. It is important to examine the demographics in this was to ensure that the responses for each individual fall within the sample inclusion criteria. Any response that falls outside of the parameters for that demographic should be deleted.

Checking sample demographics is additionally useful to develop an understanding of each individual country as well as for the comparability between countries. The purpose is to develop an understanding of the sample. While not essential to do at this stage since these analyses will have to be run once the data set is cleaned, we recommend running it during this preparation phase as a way to gain familiarity with the data and the demographics of the respondents who participated in the study. As you repeatedly do this while going through each group (e.g. country) data set, you will garner a better understanding of the comparability of the respondents across the various data sets. Furthermore, for the key variables, it is useful to also calculate frequency tables of the mean, median, mode, standard deviation, variance, and range. In our experience, this overview provides a helpful first-look or introduction to a specific country sample.

This is particularly important in multicountry studies, as you want to ensure that the different country samples are reasonably comparable. For example, you do not want to have an overrepresentation of individuals from one country as compared to the other countries in the data set. Indeed, it is important to have relatively close numbers in terms of sample size so that larger sample sizes do not overly dominate the results (Kankarasm *et al.*, 2011). Additionally, you want the average age (and related range) and gender distribution to be generally comparable for each country unless there is a reason as to why they should be different. For example, in some industries and in some countries we would expect there to be an unequal gender representation, such as the engineering profession in Saudi Arabia.

Checking data equivalence, factor analyses and scale reliabilities. Finally, as a last step in our recommended process, it is important to investigate the data equivalence, factor analyses, and scale reliabilities in each country data set. Hult *et al.* (2008) provide clear guidelines on the steps required to determine if various elements of the research design have the same meaning across the countries examined. These authors provide concrete and detailed steps to report data equivalence. At the simplest level, researchers should explore the extent to which the key variables of interest meet construct equivalence requirements.

Construct equivalence is explored to determine whether the variables of interest have the same item structure across contexts examined (Kumar, 2000). Factor analysis (Hult *et al.*, 2008), and the calculations of reliability scores (Cronbach, 1951; Nunnally, 1978) are common techniques employed to test for construct equivalence whether of the exploratory or confirmatory kind. The process of testing should begin with the researcher examining the correlation matrix reporting the relationships between the variables of interest. It is important to ensure that the correlations coefficients are at or above 0.30 (Hair *et al.*, 1995, Tabachnick and Fidell, 2007) to review the correlations that exist between the variables of interest. If no correlation exists, then it makes little sense to perform the factor analysis (DiLalla and Dollinger, 2006). Next, a good practice is to calculate Bartlett's test of sphericity as well as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy as an indication of the suitability of conducting the factor analysis. For the first, researchers should look for a significant χ^2 statistic and, for the second, for a KMO – index ranges from 0.0 to 1.0 – value above 0.70, which is a conservative estimate of suitability (Kaiser, 1974). An in-depth discussion of these techniques may be found in Yong and Pearce (2013).

Assuming that these preceding analyses support the decision to proceed, it is a good plan to start with an exploratory factor analysis on the items making up each of the variables/dimensions of interest in the study. The purpose is to assess whether the expected number of variables is actually extracted and whether the items proposed to constitute each dimension are found to be part of that factor/dimension. You can then follow this step with a confirmatory factor analysis. Bruce (2004) provides detailed instructions on how to run these analyses. It is important to report the results and pattern matrices for each scale and for each country to see if any serious country differences appear. Keep in mind that the sample size per country is important. The larger the sample size, the more diminished the error will be leading to a better analysis. Comrey and Lee (1992) suggest that a sample of 300 is ideal with studies that have between five and ten variables. Although there are a number of different guidelines as per the sufficient number of respondents per country (Hair *et al.*, 1995 suggest a basis of 100 or more), Yong and Pearce (2013) suggest that a 30:1 ratio of respondents-to-variables is sufficient for stability and cross-validation. Finally, an additional technique that is worthy of consideration is conducting a multigroup confirmatory factor analysis. This technique is used to evaluate measurement invariance/equivalence of a scale across different groups (e.g. countries). Byrne (2008) and Koh and Zumbo (2008) provide very good discussions of this analysis.

Following the factor analyses, once the variables and related dimensions are checked, the next step is to explore the scale reliabilities for each study variable with the data from each country. The simplest and most basic technique is to calculate and report Cronbach's α test for each dimension of each variable (Cronbach, 1951). A challenge here is that often, the decision to include items will depend on the factor analyses conducted in the factor analysis phase reported above, particularly the exploratory factor analyses. You want to include the items based on the factor analysis results and then to calculate the reliability scores (e.g. α coefficient). It is important to ensure that the αs are above 0.70 (Nunnally and Bernstein, 1994) for each variable of interest and, as importantly, that the α's are comparable to those reported by other researchers in the literature. At this point, assuming that your data have survived these tests, you can be relatively sure that your data are ready to be used to test the hypotheses of your study. Table I provides an outline of the analyses previously discussed for each of the five steps.

Table I.
Tests for cleaning multicountry data sets at each of the five steps

Step	Tests used
1. Preparation	Re-read the data collection instrument carefully and become familiar with the item-response scales. Take the time to become familiar with the constructs included in the data set as well as their corresponding dimensions and scoring keys. Give special attention to reverse coded items particularly for cross-cultural studies where misinterpreted phrases of negation are likely to be compounded when the scales are translated for use in other languages (Swain *et al.*, 2008; Weijters *et al.*, 2013)
2. Screening	Inspection: observe the data visually through the use of frequency tables, histograms, and descriptive statistics (means, standard deviations, medians, modes, maximum values, minimum values, variances) Bad data: report values that appear to be transcribed incorrectly to double check with original surveys (Meade and Craig, 2012) Missing data: scan the data manually to note "significant" or "non-significant" missing values. Do not delete cases at the country level as they may reveal patterns with the master data set. Report those that appear to be "bad data" and those systematically missing items. It is important to count the number of missing values per case (i.e. respondent's data) as there may be a large number of values missing for one individual but not for others. It is important to decide on a criterion or a missing value threshold above which you will delete the entire case. Also, it is helpful to examine the standard deviation for the items per case and then to determine if it is less or more than 1. Generally speaking, if the standard deviation is less than 1, the case may be suspect. Scan the case further for a problematic pattern of responses (e.g. all responses with a value of 3). If this is the situation, then you will likely need to also delete the entire record for that respondent. However, we recommend waiting until the country file is merged into the master file before deleting cases from the file because when juxtaposed with other countries, the missing data may reveal important information about the instrument itself, such as unclear instructions or misplaced response items Normality: observe the skewness and kurtosis values (as well as the means, maximum and minimum values, etc.) through the descriptive statistics table. The skewness and kurtosis values should be 0 in a normal distribution. The further the values are from 0, the more likely it is that the data are not normally distributed. Convert the skewness and kurtosis scores to Z-scores where: Z-skewness = (skewness − 0) ÷ standard error of skewness Z-kurtosis = (kurtosis – 0) ÷ standard error of kurtosis If the resulting absolute values of the scores are greater than 1.96, 2.56, or 3.29 (depending on whether the sample size is 100 participants or less, between 100 and 300 participants, or 300 participants or greater, respectively), then it is significant ($p < 0.05$) (Tabachnick and Fidell, 2012). However, significance tests of skewness and kurtosis should not be used in large samples because they are likely to be significant even when skewness and kurtosis are not too different from normal. In terms of the Kolmogorov-Smirnov (K-S) test, the score on the K-S test should be 0 in a normal distribution. The further the score is from 0, the more likely it is that the data are not normally distributed. It is useful to convert the K-S scores to Z-scores and then to observe if the resulting absolute value of the score is greater than one of the abovementioned Z-score thresholds. The Z-score thresholds, as mentioned, are to be chosen based on their respective conditions

(*continued*)

Step	Tests used
	(Tabachnick and Fidell, 2012). If the absolute values are indeed greater than any of the aforementioned thresholds, the K-S test is considered to be significant ($p < 0.05$) (Field, 2009). The K-S test should not be significant to assume normality. However, the K-S test is not sufficient alone as with large samples, it is very easy to get significant results from small deviations from normality. Therefore, it is also important to plot the data through the use of P-P plots, Q-Q plots, and histograms
3. Correcting problems in the data	Missing values: DiLalla and Dollinger (2006) suggest that 5% missing data for any given measure is a common threshold for dropping a participant. They suggest a similar 5% cutoff rule also be applied to individual scales that constitute multiscale inventories Outliers: graph the data with a histogram, boxplot, or scatterplot. The boxplot allows one to specify exactly which case is the outlier as appeared in the histogram or the scatterplot (Tabachnick and Fidell, 2012) Univariate outliers: convert the values to standardized *Z*-scores and compare the absolute values of these *Z*-scores to one of the following *Z*-score threshold values (Tabachnick and Fidell, 2012): (1) 1.96 if the sample consists of 100 or less participants (2) 2.56 if the sample size is between 100 and 300 participants (3) 3.29 if the sample consists of 300 or more participants If the absolute values of the *Z*-scores are greater than any of the abovementioned thresholds, the case with the original, unconverted value is considered to be a univariate outlier Multivariate outliers: calculate the Mahalanobis distances of the values and compare them to the critical values of χ^2 based on the degree of freedom (Tabachnick and Fidell, 2012). If the Mahalanobis distance is greater than the critical value of χ^2 that is being used a comparison threshold, the case with the original, unconverted value is considered to be a multivariate outlier
4. Checking sample demographics	Once the data have been cleaned, the next step is to compute the appropriate basic descriptive statistics to explore the sample characteristics and to get a better understanding of the groups that comprise your sample (DiLalla and Dollinger, 2006). We suggest that you calculate basic statistics including: sample size (*n*), minimum and maximum, mean/mode, percentages, and standard deviation for each key variable in the survey. It is important to examine the demographics in this was to ensure that the responses for each individual fall within the sample inclusion criteria. Any response that falls outside of the parameters for that demographic should be deleted
5. Checking data equivalence, factor analyses and scale reliabilities	Assumptions for conducting a factor analysis: the process of testing should begin with the researcher examining the correlation matrix reporting the relationships between the variables of interest. It is important to ensure that the correlations coefficients are at or above 0.30 (Hair *et al.*, 1995; Tabachnick and Fidell, 2007) to review the correlations that exist between the variables of interest. If no correlation exists, then it makes little sense to perform the factor analysis (DiLalla and Dollinger, 2006). Next, a good practice is to calculate Bartlett's test of Sphericity as well as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy as an indication of the suitability of conducting the factor analysis. For the first, researchers should look for a significant χ^2 statistic, and for the second, for a KMO – index ranges from 0.0 to 1.0 – a value above 0.70 is a conservative estimate of suitability (Kaiser, 1974; Yong and Pearce, 2013)

(*continued*)

Table I.

Table I.

Step	Tests used
	Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA): assuming that these preceding analyses support the decision to proceed, it is a good plan to start with an exploratory factor analysis on the items making up each of the variables/dimensions of interest in the study. The purpose is to assess whether the expected number of variables is actually extracted and whether the items proposed to constitute each dimension are found to be part of that factor/dimension. You can then follow this step with a confirmatory factor analysis. It is important to report the results and pattern matrices for each scale and for each country to see if any serious country differences appear. Keep in mind that the sample size per country is important. The larger the sample size, the more diminished the error will be leading to a better analysis. Comrey and Lee (1992) suggest that a sample of 300 is ideal with studies that have between 5 and 10 variables. Finally, an additional technique that is worthy of consideration is conducting a multigroup confirmatory factor analysis. This technique is used to evaluate measurement invariance/equivalence of a scale across different groups (e.g. countries) (Byrne, 2008; Koh and Zumbo, 2008)
	Reliability testing: following the factor analyses, once the variables and related dimensions are checked, the next step is to explore the scale reliabilities for each study variable with the data from each country. The simplest and most basic technique is to calculate and report Cronbach's α test for each dimension of each variable (Cronbach, 1951). A challenge here is that often, the decision to include items will depend on the factor analyses conducted in the factor analysis phase reported above, particularly the exploratory factor analyses. You want to include the items based on the factor analysis results and then to calculate the reliability scores (e.g. α coefficient). It is important to ensure that the α's are above 0.70 (Nunnally and Bernstein, 1994) for each variable of interest and, as importantly, that the α's are comparable to those reported by other researchers in the literature

Conclusion

It is important to choose wisely. Our discussion of the data cleaning phase should make clear how important forethought is in selecting the measures for your questionnaire. On one end of the spectrum, utilizing measures that have previously been cross-culturally validated is certainly one of the more prudent approaches when selecting measures for your study. On the other end of the spectrum, creating a number of new items that seem to represent a certain construct/dimension, without any testing, is one of the more imprudent approaches. Going back to Part I of this paper, a tremendous amount of effort will have been invested by a substantial number of individuals to get to this stage. No one, especially the organizer of this large-scale research project, wants to get to this point only to realize that the data are worthless. This depressing outcome is much more likely with the latter approach than with the former. Choose wisely.

References

Abraham, W.T. and Russell, D.W. (2004), "Missing data: a review of current methods and applications in epidemiological research", *Current Opinions in Psychiatry*, Vol. 17 No. 4, pp. 315-321.

Allison, P.D. (2002), *Missing Data*, Sage, Thousand Oaks, CA.

American Psychological Association (2001), *Publication Manual of the American Psychological Association*, American Psychological Association, Washington, DC.

Barnett, V. and Lewis, T. (1994), *Outliers in Statistical Data*, Wiley, New York, NY.

Brislin, R.W. (1986), "A culture general assimilator: preparation for various types of sojourns", *International Journal of Intercultural Relations*, Vol. 10 No. 2, pp. 215-234.

Bruce, T. (2004), *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*, American Psychological Association, Washington, DC.

Byrne, B.M. (2008), "Testing for multigroup equivalence of a measuring instrument: a walk through the process", *Psicothema*, Vol. 20 No. 4, pp. 872-882.

Church, A.T., Katigbak, M.S. and Castaneda, I. (1988), "The effects of language of data collection on derived conceptions of healthy personality with Filipino bilinguals", *Journal of Cross-Cultural Psychology*, Vol. 19 No. 2, pp. 178-192.

Cole, J.C. (2008), "How to deal with missing data", in Osborne, J.W. (Ed.), *Best Practices in Quantitative Methods*, Sage, Thousand Oaks, CA, pp. 214-238.

Comrey, A.L. and Lee, H.B. (1992), *A First Course in Factor Analysis*, 2nd ed., Lawrence Erlbaum, Hillsdale, NJ.

Cronbach, L. (1951), "Coefficient alpha and the internal structure of tests", *Psychometrika*, Vol. 16 No. 3, pp. 297-334.

Curran, P.G. (2015), "Methods for the detection of carelessly invalid responses in survey data", *Journal of Experimental Social Psychology*, Vol. 66, pp. 4-19.

DeSimone, A.J., Harms, P.D. and DeSimone, J.A. (2015), "Best practice recommendations for data screening", *Journal of Organizational Behavior*, Vol. 36 No. 2, pp. 171-181.

DiLalla, D.L. and Dollinger, S.J. (2006), "Cleaning up data and running preliminary analyses", in Austin, J.T. and Leong, F.T.L. (Eds), *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants*, Sage, Thousand Oaks, CA, pp. 241-253.

Farh, J.L., Cannella, A.A. Jr and Lee, C. (2006), "Approaches to scale development in Chinese management research", *Management and Organization Review*, Vol. 2 No. 3, pp. 1-18.

Field, A.P. (2009), *Discovering Statistics Using SPSS*, Sage, London.

Gibbons, J.L., Zellner, J.A. and Rudek, D.J. (1999), "Effects of language and meaningfulness on the use of extreme response style by Spanish-English Bilinguals", *Cross-Cultural Research*, Vol. 33 No. 4, pp. 369-381.

Gibbons, M. (1999), "The demise of the lone author", *Nature*, Vol. 450 No. 7173, pp. 1165-1172.

Hair, J., Anderson, R.E., Tatham, R.L. and Black, W.C. (1995), *Multivariate Data Analysis*, Prentice Hall, Upper Saddle River, NJ.

Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P. Ph., Pennell, B.E. and Smith, T.W. (2010), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, John Wiley & Sons, Inc., Hoboken, NJ.

Hofstede, G. (1980), "Culture and organizations", *International Studies of Management & Organization*, Vol. 10 No. 4, pp. 15-41.

Hult, G.T.M., Ketchen, D.J., Griffith, D.A., Finnegan, C.A., Gonzalez-Padron, T., Harmancioglu, N., Huang, Y., Talay, M.B. and Cavusgil, S.T. (2008), "Data equivalence in cross-cultural international business research: assessment and guidelines", *Journal of International Business Studies*, Vol. 39 No. 6, pp. 1027-1044.

Kaiser, H.F. (1974), "An index of factorial simplicity", *Psychometrica*, Vol. 39, pp. 31-36.

Kankarasm, M., Moors, G. and Vermunt, J.K. (2011), "Testing for measurement invariance with latent class analysis", in Davidov, E., Schmidt, P. and Billiet, J. (Eds), *Cross-Cultural Analysis: Methods and Applications*, Taylor and Francis Group, Psychology Press, Oxford, pp. 359-384.

Karmaker, A. and Kwek, S. (2007), "An iterative refinement approach for data cleaning", *Intelligent Data Analysis*, Vol. 11, pp. 547-560.

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donahue, B. and Levin, J.R. (1998), "Statistical practices of educational researchers: an analysis of their ANOVA, MANOVA, and ANCOVA analyses", *Review of Educational Research*, Vol. 68 No. 3, pp. 350-386.

Koh, K.H. and Zumbo, B.D. (2008), "471 Multi-group confirmatory factor analysis for testing measurement invariance in mixed item format data", *Journal of Modern Applied Statistical Methods*, Vol. 7 No. 2, pp. 471-477.

Kumar, V. (2000), *International Marketing Research*, Prentice-Hall, Upper Saddle River, NJ.

Leung, K. (2008), "Methods and measurements in cross-cultural management", in Smith, P.B., Peterson, M.F. and Thomas, D.C. (Eds), *The Handbook of Cross-Cultural Management Research*, Sage Publications, Inc., Thousand Oaks, CA.

McSweeney, B. (2002), "Hofstede's model of national cultural differences and their consequences: a triumph of faith – a failure of analysis", *Human Relations*, Vol. 55 No. 1, pp. 89-118.

Meade, A.W. and Craig, S.B. (2012), "Identifying careless responses in survey data", *Psychological Methods*, Vol. 17 No. 3, pp. 437-455.

Newman, D.A. (2009), "Missing data techniques and low response rates: the role of systematic nonresponse parameters", in Lance, C.E. and Vandenberg, R.J. (Eds), *Statistical and Methodological Myths and Urban Legends*, Routledge, London, pp. 7-36.

Nunnally, J.C. (1978), *Psychometric Theory*, McGraw Hill, New York, NY.

Nunnally, J.C. and Bernstein, I. (1994), *Psychometric Theory*, McGraw-Hill, New York, NY.

Onwude, J.L., Staines, A. and Lilford, R.J. (1993), "Multiple author trend worst in medicine", *British Medicine Journal*, Vol. 306 No. 6888, pp. 1345-1353.

Osborne, J.W. (2008), "Sweating the small stuff in educational psychology: how effect size and power reporting failed to change from 1969 to 1999, and what that means for the future of changing practices", *Educational Psychology*, Vol. 28 No. 2, pp. 1-10.

Osborne, J.W. (2013), *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data*, Sage Publications, London.

Oyserman, D., Coon, H.M. and Kemmelmeier, M. (2002), "Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses", *Psychological Bulletin*, Vol. 128 No. 1, pp. 3-72.

Penny, K.L. (1996), "Appropriate critical values when testing for a single multivartiat outlier by using the Mahalanobis distance", *Applied Statistics*, Vol. 45 No. 1, pp. 73-81.

Peterson, M.F. (2009), "Cross-cultural comparative studies and problems of coordinating international research teams", in Buchanan, D.A. and Bryman, A. (Eds), *Handbook of Organizational Research Methods*, Sage, London.

Raudenbush, S.W. and Bryk, A.S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage, London.

Regalado, A. (1995), "Multiauthor papers on the rise", *Science*, Vol. 268 No. 5207, pp. 25-32.

Sargent, L.D. and Waters, L.E. (2004), "Careers and academic research collaborations: an inductive process framework for understanding successful collaborations", *Journal of Vocational Behavior*, Vol. 64 No. 2, pp. 308-319.

Shapiro, D.W., Lasker, R.D., Bindman, A.B. and Lee, P.R. (1993), "Containing costs while improving quality of care: the role of profiling and practice guidelines", *Annual Review of Public Health*, Vol. 14 No. 1, pp. 219-241.

Schafer, J. and Graham, J. (2002), "Missing data: our view of the state of the art", *Psychological Methods*, Vol. 7 No. 2, pp. 147-177.

Schafer, J.L. and Olsen, M. (1998), "Multiple imputation for multivariate missing data problems: a data analyst's perspective", *Multivariate Behavioral Research*, Vol. 33 No. 4, pp. 545-571.

Smith, P.C., Budzeika, K.A., Edwards, N.A., Johnson, S.M. and Bearse, L.N. (1986), "Guidelines for clean data: detection of common mistakes", *Journal of Applied Psychology*, Vol. 71 No. 3, pp. 457-460.

Strange, K. (2008), "Authorship: why not just toss a coin?", *American Journal of Physiology Cell Physiology*, Vol. 295 No. 3, pp. 567-571.

Swain, S.D., Weathers, D. and Niedrich, R.W. (2008), "Assessing three sources of misresponse to reversed likert items", *Journal of Marketing Research*, Vol. 45 No. 1, pp. 116-131.

Tabachnick, B.G. and Fidell, L.S. (2007), *Using Multivariate Statistics*, Pearson Education, Boston, MA.

Tabachnick, B.G. and Fidell, L.S. (2012), *Using Multivariate Statistics*, 6th ed., Pearson.

Tourangeau, R., Rips, L.J. and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press, Cambridge.

Tsao, C.I. and Roberts, L.W. (2009), "Authorship in scholarly manuscripts: practical considerations for resident and early career physicians", *Academic Psychiatry*, Vol. 33 No. 1, pp. 76-79.

Tsui, A.S., Nifadkar, S.S. and Ou, A.Y. (2007), "Cross-national, cross-cultural organizational behavior research: advances, gaps, and recommendations", *Journal of Management*, Vol. 33 No. 3, pp. 426-478.

Van de Vijver, F.J.R. and Hambleton, R.K. (1996), "Translating tests: some practical guidelines", *European Psychologist*, Vol. 1 No. 2, pp. 89-99.

Van de Vijver, F.J.R. and Leung, K. (1997), *Methods and Data Analysis for Cross-Cultural Research*, Sage Publications, London.

Weijters, B., Baumgartner, H. and Schillewaert, N. (2013), "Reversed item bias: an integrative model", *Psychological Methods*, Vol. 18 No. 3, pp. 320-334.

Wong, N., Rindfleisch, A. and Burroughs, J.E. (2003), "Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale", *Journal of Consumer Research*, Vol. 30 No. 2, pp. 72-91.

Yong, A.G. and Pearce, S. (2013), "A beginner's guide to factor analysis: focusing on exploratory factor analysis", *Tutorials in Quantitative Methods for Psychology*, Vol. 9 No. 2, pp. 79-94.

Further reading

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W. and Gupta, V. (2004), *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*, Sage Publications, London.

Ralston, D.A., Holt, D.A., Terpstra, R.H. and Yu, K.C. (1997), "The impact of national culture and economic ideology on managerial work values: a study of the United States, Russia, Japan, and China", *Journal of International Business Studies*, Vol. 28 No. 1, pp. 177-208.

Schwartz, S.H. (2012), "An overview of the Schwartz theory of basic values", available at: http://scholarworks.gvsu.edu/orpc/vol2/iss1/11/ (accessed 12 February 2016).

Shapiro, D.W., Wenger, N.S. and Shapiro, M.F. (1994), "The contributions of authors to multiauthored biomedical research papers", *Journal of the American Medical Association*, Vol. 271 No. 3, pp. 438-442.

Corresponding author
Charlotte M. Karam can be contacted at: ck16@aub.edu.lb