

8th GESIS Summer School in Survey Methodology

Assignment

# 3MC survey design and translation methodology

Analyzing ESS, EVS & WVS

Asenjo Domínguez, Alberto Luis

Reviewed by  
Prof. Dr. Michael Braun

Course 09:

“Questionnaires for Cross Cultural Surveys”

Cologne, August 19-23, 2019

## **Introduction**

As social scientists, our goal is to improve the knowledge and comprehension of the social reality. We develop theories to do so, using empirical data to test them. Furthermore, it'll be useless to develop theories for a reduced number of subjects, but also, it'll be impossible to get data from all the subjects of the world. That's why we'd developed methods to infer the qualities of the population by getting data of just a small amount of subjects. Additionally, it's important to take our theory further away, comparing our results with other contexts. Keep our conclusions in a national context. We'll need to compare them with other countries, looking for similarities and differences, and improving by this comparison our comprehension of the social world.

That's why this assignment focuses on methodology related to 3MC surveys. In this paper, we're going to review literature related to 3MC survey instrument design and translation methodology. Then, we're going to create a comparison chart, and we're going to use it with three 3MC surveys: ESS, EVS and WVS. Lastly, we're going to analyze the results of this comparison.

## **Literature review**

Firstly, it's important to clarify what we consider a 3MC survey. We define a 3MC survey as an acronym for Multinational, Multiregional and Multicultural Context survey, although, as Harkness, Braun et al. (2010) pointed out, we may add many more "M" s to this definition related to the previous ones, as Multilingual, Multicontextual, and so on. We use 3MC term instead of "cross-cultural surveys" like in Behr & Shishido (2016) because it highlights the relevance of the multi-layering of survey designing, but both terms are right.

3MC surveys present a challenge in every one of its layers. If you are working with a small team, you will need to make many sacrifices related to quality. That's why it's very common to develop 3MC surveys with a big team, but coordinating a big team it's a really hard task to manage. Also, other difficulties managing 3MC surveys teams are the many differences between institutional frameworks for each country, the linguistic differences, and the distance. These difficulties are reduced by many improvements available nowadays (online task control software, use of English as a common language for everyone, videoconferences, and so on). When these problems have been overcome, the main task can be faced: Developing a comparable questionnaire.

In the past, there was scarce literature about how to devise a comparable 3MC survey, but now many more papers and books are explaining the methodology to do so. This is because one of the main improving techniques for 3MC surveys is the experience, and with their experience linguists have developed a substantial amount of knowledge related to harmonization between cultures, and therefore it's important to take them into account in our team. Seminal contributions have been made by Janet Harkness in establishing guidelines for

3MC survey's designers. We would want to point out her efforts to improve the standards, providing her expertise to social scientists all over the world. Her improvements will enhance the quality of the data collected by 3MC surveys. Many of our references have her name between their authors.

But, what is comparability? One of the clearest definitions we have found is the one made by Harkness, Edwards et al. (2010), that said that “(comparative researches) are to be understood as projects that deliberately design their instruments and possibly other components of the survey to compare different populations and that collect data from two or more different populations” (p.5). In social sciences, comparability is a tool for screening contextual effects on social reality and testing the capacity of a theory to explain multiple contexts.

Comparability is, for many authors (Harkness, Braun et al., 2010; Jong, Dorer, Lee, Yan & Villar, 2018; Behr & Shishido, 2016) the main goal for 3MC surveys. Behr & Shishido (2016) pointed out that without comparability, making a cross-cultural survey will make no sense. As Harkness, Braun et al. (2010) pointed out, although providing ex-post harmonization for data is an option, it isn't the most reliable one. Also, developing ex-ante harmonization methods requires less effort but returns comparable data (if you do it right). That's why comparability is located in the center of the designing process, and this approach is named "comparability-driven design".

Mohler & Johnson (2010) explained the main principles for this approach. Firstly, we've to understand that, even though our goal is full equivalence in our concepts and indicators, in practice reaching it is an impossible task. Equivalence is the ideal, comparability is the reality that we can attain. Then, we've to take into account the context. It is impossible to use data without proper comprehension of its background. Also, we've to get the relevance of translation, it may create a bias on survey meaning.

As Harkness, Edwards et al. (2010) have highlighted, making a comparative-driven design is more than a study that focuses on different cultures, explaining that “they may be designed with a single general population in mind but be fielded with other populations”. A good comparable questionnaire should take into account cross-cultural differences (Behr & Shishido, 2016). There are many methods to develop 3MC surveys, but having in mind possible problems with translation or instrument implementation will help in other stages of the project.

Instruments are one of the main parts of a 3MC survey. A good instrument design is (Mohler & Johnson, 2010) the base for a good comparison. And, with different cultures involved, indicators should be functionally equivalent. ¿What does it mean that two indicators are “functionally equivalent”? Two functionally equivalent indicators “should behave in a similar manner in statistical analysis” in different cultures, in the case of a 3MC survey (Mohler & Johnson, 2010). It requires testing to ensure comparability, so it's very common to reuse

reliable indicators during successive waves (also, they will provide longitudinal comparability).

That's why it's hard to decide between three main options related to instrument development. These three options, explained by Harkness, Edwards et al. (2010), are reuse, adapt or develop instruments. Reusing an instrument has many advantages. If your survey has used it in the past, choosing this option will provide longitudinal data, and if it has worked well in the past, it will probably do it again. Sometimes, you would even choose this option when the instrument is not functionally equivalent between cultures to ensure longitudinal comparability, and you will use ex-post harmonization. Another option is to introduce modifications in indicators developed for other contexts or targets. It's easier than developing a new one, and they are partly tested, but, on the other hand, it's hard to make them suit well in your questionnaire. The last option is to develop new indicators. Although it will require more resources than the other options, a good design team can develop an almost perfect indicator for cross-cultural purposes.

Behr & Shishido (2016) and Harkness, Villar, & Edwards (2010) also suggested the relevance of deciding between adaptation and translation. Behr & Shishido (2016) noticed different uses of the term "adaptation" in cross-cultural survey design vocabulary. The first definition (the generic one) refers to "the overall process of transferring an instrument from one language and culture to another language and culture" (Behr & Shishido, 2016). The specific sense definition is made by Harkness, Villar, & Edwards (2010), showing us that the adaptation process consists of "deliberate changes to source material to meet new needs of various kinds" (p.133). There are different domains of adaptation (culture, measurement, and language), but, in fact, 3MC surveys tend to avoid adaptation, designing the source questionnaire to be translated without having to adapt it. As Harkness, Villar, & Edwards (2010) said, "survey translations are largely expected to stay close to the source text" (p.119).

That takes us to a debate. Our target is comparability between cultures. But it's hard to design a questionnaire suitable for every culture. Sometimes, keeping the same wording will improve comparability, but it will also be less culture-specific and it may guide to misunderstandings. Another option is having different questions asking for the same construct, but it's difficult to keep harmonization with different questions. That debate leads us to three main strategies: ASQ, ASQ by decentering and ADQ (Jong et al., 2018; Harkness, Edwards et al., 2010), which will be explained as follows.

ASQ, acronym for "Ask the Same Question", is the most popular strategy. It will develop questions keeping the same wording. Its philosophy is that keeping things the same will produce comparability. A major drawback of this approach is the reduction of question's concreteness and is often not suitable for a large number of languages and cultures.

ASQ by decentering improves the basic ASQ approach by the inclusion of teams from other cultures in the designing process. It's "something like a Ping-Pong exchange" (Harkness,

Edwards et al., 2010, p.47), looking forward to keeping the “essence” while improving the understandability of the question for a language. Using this with every language involved in the survey may be expensive and tough, but “as versions are produced and compared, problematic cultural and linguistic elements are removed” (Jong et al., 2018, pp.116-117). However, it has the same lack of concreteness as basic ASQ.

ADQ is radically different from ASQ. It focuses on keeping the same latent constructs and underlying concepts, but not the wording, using a "functionally equivalent" strategy. As Harkness, Edwards et al. (2010) highlighted, “it obviates the need to translate” (p.48). It will require a high level of indicator development, and a strong theoretical model to sustain comparability. Also, Jong et al. (2018) point out that “item-by-item analyses across populations may be a concern” (p.116) due to differences in the questions.

Choosing between these approaches is a matter of budget, available team members, cultures involved, etc. But there is another option, that will mix both methods, using ADQ for country-specific questions and ASQ for general ones (Harkness, Edwards et al., 2010).

As we have explained, translation is a very important part of 3MC surveys. Without an excellent translation, data comparability will be ruined completely. A largely spread mistake made by small, medium and large scale 3MC survey teams is hiring "bilingual" members or low-skilled translators and trusting them to translate the source questionnaire. Behr & Shishido (2016) emphasized that "translators should have an excellent command of both source and target language and culture and typically they translate into their mother tongue" (p.271). We shall not elaborate on the multiple types of translations involved in a 3MC survey, but it's a complex and advanced topic that requires high-level capabilities (Harkness, Villar & Edwards, 2010).

In addition to translators, a 3MC survey design team will consist of an interdisciplinary team. The exact makeup of the team will be affected by the degree of cultural input involved in the design. Also, Harkness, Edwards et al. (2010) and Jong et al. (2018) recommended using a lingua franca (common language) for the source questionnaire and team coordination. A usual trouble with lingua franca is that, in most cases, not everyone will be native speaker, so mistakes can be made. Also, as Harkness, Edwards et al. (2010) noticed, “team members whose first language happens to be the lingua franca may unintentionally dominate discussion”. Because of that, using lingua franca isn't effortless, and it's required a high level of coordination and understanding between team members. Also, although English is a common choice between international teams, it “is known to have a larger lexicon than any other language, which means that distinctions in wording in English cannot always be replicated in target languages” (Jong et al., 2018, p. 120), so, without previous awareness, it will be hard to perform a translation.

There are different options related to team composition and how and when intercultural elements are taken into account, and also, phases and timing are different in each approach.

About team composition, Harkness (2003) identified three main profiles: translators, reviewers, and adjudicators.

The leading skill for a translator, aside from translation, is experienced working with surveys and questionnaires. Surveys don't follow the same rules and techniques in terms of translation as books or papers, so specific training is required. It's recommended to hire translators to translate into their language to get better results.

Reviewers should be skilled in translation, in questionnaire designing, and also be familiar with the research topic. They will apply their translation knowledge to the research, reviewing the translator's work and checking if it fits well into the questionnaire and if it's asking what it should ask.

Lastly, the adjudicators' task is to decide between translation options. They will need to have similar skills as reviewers, but with a better understanding of the research topic. Also, their knowledge of both target and source language should be excellent.

Focusing on staging, phases, and timing, literature has outlined these key types.

The One-to-one or single translation (Harkness, 2003) uses just one translator for each language involved in the survey. It's a basic and simple approach and has some advantages, but without a doubt, it has strong drawbacks that can't be tolerated in terms of comparability. It's a cheap method (you'll only have to hire one expert) and doesn't require coordination, but it relies too much on the subjective opinion of one translator. The quality of the translated version is determined by one single translator's skills, and there is no way to check for mistakes. That's why literature doesn't encourage the use of this methodology.

Sequential development approaches (Harkness, Edwards et al., 2010) focus on the development of a source questionnaire and then translate it into every required language. This approach "basically only addresses multicultural considerations at the translation stage" (p.45). Sometimes this is not an option, because the original survey development team didn't expect the addition of countries with different languages. The process of translating a fixed questionnaire to other languages, keeping comparability, is most of the time harder than developing a brand-new one, but, as we said before, sometimes it's not an option. Using ADQ may be an option to avoid ASQ wording issues. Also, the translation isn't developed with a team approach, because the source cannot be altered.

Team approaches (TRAPD), committee-based approaches (Harkness, 2003) or multi-step approaches (Behr & Shishido, 2016) use iterations and various kinds of professionals to improve questionnaires and instruments translatability and translation, reducing subjectivity. TRAPD is the acronym for Translation-Review-Adjudication-Pretesting-Documentation.

Translation is just one stage, the other stages check for mistakes and upgrades, also providing openness.

Back translation (Behr & Shishido, 2016) is an iterating method of translation that was "in widespread use since about the 1970s" (p.272), but, as the authors pointed out, it doesn't guarantee compatibility. The method itself consists of a translation that is then back-translated into the source language to check for differences. The problem is that the back translation's differences are more related to the translator's skills than the questionnaire's quality. So, relying just on this method is not recommended. If back translation is used, it has to be used as an addendum of other methods.

Split translation (Harkness, 2003) is considered an improvement concerning the previous alternatives. This method is cheaper in terms of time, effort and money, and it also provides a good quality translation. The questionnaire is split and randomly divided between translators, and then they will all meet with another translator to review the entire questionnaire. The major drawback of this approach is consistency because if the wording of the two following questions is different, questionnaire simplicity and clarity will be compromised. That's why reviewing meetings should harmonize questions. This method reduces bias to a similar degree as the following methods but requiring fewer resources to be accomplished.

Parallel translation (Harkness, Edwards et al., 2010; Harkness, 2003; Behr & Shishido, 2016) is considered one of the best methods regarding 3MC survey translation. It "targets cross-cultural input early in the conceptual and question development stages" (Harkness, Edwards et al., 2010, p.45). Instead of splitting the questionnaire as the previous method, the whole questionnaire will be translated independently, and then, they will compare their translation at a reconciliation meeting with an adjudicator. This will "help to uncover idiosyncratic wording or different interpretations" (Behr & Shishido, 2016, p.271). It will also provide options to the translation team, enabling them to choose between different wording alternatives. Also, national teams can suggest changes in the source questionnaire to improve its translatability.

Lastly, simultaneous translation (Harkness, Edwards et al., 2010) "targets the highest degree of cross-cultural involvement, with cultural input at every stage" (p.45). This is the fully accomplished TRAPD approach and will increase ex-ante harmonization and translatability at its maximum degree. It will also be more expensive compared to the previous methods, but it will be much more effective in terms of data quality. It's used as an ideal.

As literature points out, there is a wide range of possibilities, although the final decision will probably be related to the available budget, professionals, questionnaire length and number of cultures involved.

## **Materials and methods**

In this section, we will describe our methodology to clarify our approach and let the reader understand better our view. Here, we would like to explain that the incoming procedure will just try to show easily the procedures' choice of 3 major 3MC surveys. Teams involved in decisions related to survey methodology are high-level expertise, and, for sure, they've evaluated every possible option. But there are many researchers using data from these surveys that haven't got a translation methodology background. To improve their comprehension of how these surveys try to obtain a high degree of data comparability, we will try to simplify their decisions.

We're going to analyze this 3MC surveys:

ESS (European Social Survey) is a survey that has been performed since 2001. It's focused on researching "attitudes, beliefs and behavior patterns" in different countries across Europe. The last wave includes 30 countries.

EVS (European Values Study) researches what Europeans think about life, family, work, religion, politics, and society. It comprises 47 countries all around Europe. It started in 1981.

WVS (World Values Survey) is carried out all around the world, in almost 100 countries, so it's the largest one analyzed. It started, as the previous one, in 1981, and researches about "beliefs, values, and motivations of people throughout the world".

We will assess their guidelines. For a researcher that hasn't been involved in designing these surveys, the lack of documentation about methodology means opacity, reducing confidence in their methods. That's why we will point out the level of accessibility reached by their methodological documentation. Also, we will evaluate the methodology shown in the documentation. If the documentation is wrong, or if it's very vague, we will put ourselves in a worst-case scenario. If a survey designing team is carrying out good practices, but they aren't showing it in documentation, external researchers can't trust this process.

Here, we're first going to list the criteria used for our analysis. Then, we're going to briefly explain each criterion. In results, we're going to explain each case, showing how much information has given us each document, how clear it is, and if the information given is not clear, our conclusion about it.

The evaluation/clarification sheet will have the following criteria.

- Data availability
- Documentation
- Instrument design (ASQ, ASQ by decentering, ADQ)



- Existing translation (Reuse, modify, develop)
- Instrument development (Translation, adaptation)
- Translation team
- Lingua franca
- Framework for translation

In Data availability, we will describe how easily can anyone access translation guidelines. It has to be guidelines because we want to have the same documents as national teams. If national teams have more documentation available, but it's not included in the "translation guidelines" document, we'll consider this non-existent. From our point of view, if there are no chances to access documentation related to methodological guidelines, we won't trust results.

Documentation means how documented is the process of translation. It is very common for surveys to have changelogs, reports with issues, etc. In survey guidelines, we will look for documentation requirements, and what do they include. The more information documented, the better.

In Instrument design, as explained in the Literature review, we will consider the approach followed by this survey (ASQ, ASQ by decentering, ADQ). This terminology is broadly used, so it will be a very easily accessible item.

Existing translation is, in case of an existing translation, what guidelines tell us to do between the main three options: reuse the same question (keep it identical), modify the existing question or develop a new one. Keeping the question identical is the most common option for longitudinal surveys because changing the questions will compromise longitudinal analysis.

Instrument development focuses on the decision of translating or adapting questions between cultures involved in the questionnaire. If all countries should use the same questions with the same wording, then we will say that it's translated. If the questions are changed because of the culture or the context of any country, then we will say it's adapted.

The translation team points out which profiles are involved in questionnaire translation. Mainly, the literature recommends a translator, a reviewer, and an adjudicator, as described before.

Lingua franca is the language used for communication and coordination between national groups and the central team, and the language used in documents, guidelines, and in the source questionnaire.

With Framework for translation, we will identify which method has each survey used. The main methods are, as previously explained: one-to-one, sequential, back, split, parallel and simultaneous.

## **Results**

Here is the comparison chart.

	ESS	EVS	WVS
Data availability	Freely downloadable, short version available on website	Request via email, short version on website	Request via email, short version on website
Documentation	Yes	Yes	Yes
Instrument design	ASQ by decentering	ASQ by decentering (!)	ASQ
Existing translation	Reuse, modify if huge problems	Reuse, modify if huge problems	Reuse, modify if huge problems
Instrument development	Translation	Translation	Translation
Translation team	Translator, reviewer, adjudicator	Translator, reviewer, adjudicator	Translator, reviewer, adjudicator
Lingua franca	English	English	English (+ Spanish, Arabic, Russian)
Framework for translation	TRAPD (parallel)	TRAPD (parallel) (!)	Sequential

We're going to compare step by step each criterion.

Regarding Data availability, guidelines for EVS and WVS weren't available on their web page. We had to request them via email. With EVS, they replied to us with the document. However, with WVS, we had to request it specifically. We didn't have to identify ourselves in

both cases. ESS has its documentation available for download on their web page, and it's easy to find. All of them have a brief and non-technical explanation on their websites.

Related to documentation, they all request documentation. Guidelines related to documentation are more extended on ESS, and in WVS' documentation, they included questions to drive documentation requirements.

Instrument design is one of the biggest differences between these surveys. WVS uses ASQ, and decentering is impossible because the source questionnaire is immutable, so national teams can't suggest changes in the source questionnaire. ESS, on the contrary, uses ASQ by decentering, because all questionnaires are developed simultaneously, so national teams can suggest changes in the source questionnaire, and teams with the same languages can suggest translations between them, so there is feedback between design teams. EVS, in their guidelines, pointed out that they are following ASQ by decentering, but some countries haven't got the same timetable as the rest, so collaboration between national teams is almost impossible. So EVS uses a mix between normal ASQ and ASQ by decentering.

About existing translations, all of them keep the questions from past waves and changing them if a huge problem is detected. This is important to keep longitudinal comparability, and most surveys, not just 3MC surveys, try to keep the same questions, even if they have some problems.

The translation team's components are the same: Translator, adjudicator, and reviewer. ESS and EVS have a brief profile's definition. WVS just names the profiles, but it doesn't explain them.

Although it's not specified, English is the chosen lingua franca for all of them, but WVS has also a source questionnaire in Spanish, Arabic, and Russian. It's a good choice for a very large scale as the WVS, and with many languages involved.

Lastly, and focusing on the framework of translation, there are major differences. WVS uses a sequential approach, and even though methodologically isn't the best option, in a survey where there are 100 countries involved and many languages, a parallel approach would be extremely expensive and hard to coordinate, but it will also disable national teams' suggestions. EVS has also a problem because although they follow a parallel approach, some countries won't be able to suggest changes because of the timing, so it's a partial parallel approach.

Here, we've shown that these 3MC surveys are similar in some aspects, but there are major differences related to translation framework and instrument design. The ESS reaches the highest level between the analyzed surveys. It follows strictly the literature about survey methodology, ensuring comparability with a strong ex-ante harmonization. EVS has a similar methodology, but it doesn't follow it so strictly, and that's a problem from our point of view. The existence of exceptions compromises the credibility of their methodology. By contrast,

WVS has a classic methodological approach, but although it doesn't have the same level of ex-ante harmonization as the EVS or ESS, it follows the methodology described in their guidelines, with the added difficulty of a huge number of countries and languages.

## **Conclusions**

Using surveys' datasets requires less effort than designing and conducting a survey, but it doesn't mean that it's effortless. A high level of understanding of the survey's methodology and data collection is required for the researcher. Data has always a bias, and one of the researchers' tasks is to discover the bias and judge the data with the bias in mind. That's why training in 3MC survey methodology is important. It will enable a trained researcher to critically use data from 3MC surveys, and demand or suggest a better methodology to design teams. Improving data sources is not just a matter of design teams, it's also a matter of the researchers that use this data.

We understand that reaching a high level of comparability isn't cheap. It requires a strong theoretical base, and experience earned by previous studies. Also, it's hard to choose between including more countries or improving the methodology of the survey. Designing a survey includes many hard decisions and sacrifices, but our task as researchers isn't forgiving mistakes, it's judging data by its source. That's one of the main sources of information about social reality, and by understanding the design team's decisions, we'll be able to improve our data's usage.

ESS, EVS, and WVS are reliable sources of information about social reality. Each one has made some sacrifices with a certain amount of budget available. ESS was designed later, and its design team has decided to improve translation methodology, earning experience with new techniques, such as TRAPD. It will help to improve TRAPD methodology, improving data among 3MC surveys. WVS' design team, working with an older survey, has decided to keep their questions and their methodology, simplifying the coordination between participant countries with a simpler methodology and including more and more countries each wave. EVS' design team chose to mix both perspectives, improving their methodology and having a large number of countries across Europe at the same time.

It would be interesting for future studies to add other 3MC surveys, improving the understandability of their methodology for non-trained researchers. Also, we believe that apart from looking for translation methodology, future research should also look for the translation itself. We couldn't deepen the translation because we're not translators, but this may constitute the object of future studies. It's important to constantly review every aspect of data production.

## **References**

- Behr, D. & Shishido, K. (2016). The translation of measurement instruments for cross-cultural surveys. In C. Wolf, D. Joye, T.W. Smith, & Yang-chi Fu (Ed.), *The SAGE Handbook of Survey Methodology* (pp. 269-287). London: Sage.
- European Social Survey (2018). ESS Round 9 Translation Guidelines. London: ESS ERIC Headquarters. Retrieved from [https://www.europeansocialsurvey.org/docs/round9/methods/ESS9\\_translation\\_guidelines.pdf](https://www.europeansocialsurvey.org/docs/round9/methods/ESS9_translation_guidelines.pdf)
- European Values Study (2018). Methodological Guidelines EVS2017 (App A3: Translation). Cologne: GESIS Data Archive. <https://doi.org/10.4232/1.13090>
- Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.E. & Smith, T.W. (2010). Comparative Survey Methodology. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell & T. W. Smith (Ed.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 3-16). Hoboken, NJ: Wiley.
- Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R. & Villar, A. (2010). Designing Questionnaires for Multipopulation Research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell & T. W. Smith (Ed.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 33-58). Hoboken, NJ: Wiley.
- Harkness, J. A., Villar, A. & Edwards, B. (2010). Translation, Adaptation, and Design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell & T. W. Smith (Ed.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 117-140). Hoboken, NJ: Wiley.
- Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. van de Vijver & P. h. Mohler (Ed.), *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.
- Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. van de Vijver & P. Ph. Mohler (Ed.), *Cross-Cultural Survey Methods* (pp. 35-56). Hoboken, NJ: Wiley.
- Jong, J., Dorer, B., Lee, S., Yan, T., & Villar, A. (2018). Overview of Questionnaire Design and Testing. In T.P. Johnson, B.E. Pennell, I.A.L Stoop, & B. Dorer (Ed.), *Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC)* (pp. 115-135). New York: John Wiley & Sons.
- Mohler, P. P. & Johnson, T.P. (2010). Equivalence, Comparability and Methodological Progress. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph.

Mohler, B.-E. Pennell & T. W. Smith (Ed.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 17-32). Hoboken, NJ: Wiley.

World Values Survey Association (s.f.). Rules and Procedures in Wave 7 (2017-2018): Guidance for Principal Investigators. Unpublished manuscript.