

## The Translation of Measurement Instruments for Cross-Cultural Surveys

Behr, Dorothée; Shishido, Kuniaki

Postprint / Postprint

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Behr, D., & Shishido, K. (2016). The Translation of Measurement Instruments for Cross-Cultural Surveys. In C. Wolf, D. Joye, T. W. Smith, & Y.-c. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 268-286). Los Angeles: Sage Publications. <https://doi.org/10.4135/9781473957893.n19>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der  
  
Leibniz-Gemeinschaft

# The Translation of Measurement Instruments for Cross-Cultural Surveys

Dorothee Behr and Kuniaki Shishido

## INTRODUCTION

Cross-national surveys typically collect data by using an (almost) identical set of questions across different countries. The goal of cross-national surveys, that is, comparing countries or regions on various dimensions, requires that these questions are equivalent. Otherwise methodological artefacts might be taken as real similarities or differences between countries. Equivalence needs to be addressed ex-ante by adequate source questionnaire development, translation, and related testing; furthermore, it needs to be addressed ex-post by quantitative and/or qualitative assessments of questions. This chapter concentrates on the role of translation in the endeavor to produce equivalent questions in cross-national studies; it thus sheds light on what can be done ex-ante to address and ensure equivalence.

The focus of this chapter will be on cross-national, multilingual surveys that are designed for the purpose of cross-national

comparisons. Despite this focus, much of the chapter also applies to questionnaire translation beyond the survey context, such as when personality inventories are translated; to questionnaire translation within a single country, such as when questionnaires are translated for different linguistic groups or migrant populations; or to questionnaire translation in the context of adopting a questionnaire originally developed for one country for the use in another country. In addition, while questionnaires will be the main concern, many findings and good practice approaches also apply to the translation of assessment instruments, such as those assessing numeracy or literacy skills. In fact, all these specific fields contribute to a large extent to the literature on questionnaire translation.

The chapter is set up as follows. First, the importance of good questionnaire design for high-quality translation is addressed. Second, various translation and translation assessment methods are introduced. Third, the concepts of translation and adaptation are delineated.

Forth, the challenges of translating attitude and opinion items, with a special focus on response scales, will be presented. This will be done using the example of cross-national survey research in Asia. Finally, further developments, research desiderata as well as recommended readings complete the chapter.

Following a naming convention in translation studies, the original questionnaire, language or culture will be called source questionnaire, language or culture. Its counterpart will be the target questionnaire, language or culture. The terms questionnaire and instrument will both be used to refer to a question–answer-based measurement form.

### **GOOD QUESTIONNAIRE DESIGN AS A PRECONDITION FOR TRANSLATION QUALITY**

Producing comparable questionnaire translations is no longer discussed only in the context of appropriate translation and assessment methodology. Fact is now that the production of comparable questionnaires presupposes adequate source questionnaire design that incorporates different layers of cross-cultural collaboration and input (Smith, 2004, see also Chapter 4 by Johnson and Braun, this Handbook). These different layers include, taking the example of the European Social Survey (2014a), cross-cultural questionnaire development teams, involvement of all national teams at various stages throughout the process, qualitative and quantitative pre-testing as well as advance translation and piloting in several countries. This way it shall be ensured that on the conceptual level questions are equally relevant and valid for the participating countries. Moreover, increased cross-cultural cooperation shall make sure that linguistic particularities do not prevent or unnecessarily impede good translations later on.

Translation itself has become a valuable part of cross-cultural questionnaire design:

So-called advance translations are now carried out on fairly advanced though pre-final source questionnaires (Dept, 2013; Dorer, 2011). The goal is to identify issues of concern for cross-cultural implementation and translation early on, such as culturally inappropriate assumptions or linguistic problems (ambiguous terms, overly complex wording, etc.). The background to advance translation is that many problems related to source questionnaire concepts and formulations only become apparent – even to experienced cross-cultural researchers – if a concrete attempt is made to translate a questionnaire (Harkness et al., 2003). The feedback received from advance translation – similar to what happens with all other feedback received from commenting and testing – may either lead to modifications of source questions or to annotating the questionnaire specifically for translation. Translation annotations provide necessary background information to translators, amongst which background information on concepts, explanation of terms or phrases, or specific instructions for translators (Behr and Scholz, 2011). In general, translation annotations have become a valuable tool for ensuring comparable translations that measure what they are supposed to measure. Remarkably, this type of documentation in view of translation has already been suggested in the late 1940s (Barioux, 1948).

### **TRANSLATION AND TRANSLATION ASSESSMENT**

Ensuring comparable translations is most often discussed in the context of choosing the appropriate translation and assessment methodology. Even though there are a multitude of approaches in social science survey research, cross-cultural psychology, and the health sciences, good practice in these various disciplines shares a set of common features. These are summed up in Table 19.1 and include a multi-step approach, the

**Table 19.1 Core features of good practice translation and assessment methodology**

<i>Multi-step approach in general</i>	<i>Multi-step approach more in detail</i>	<i>Involvement of various persons with different skill sets and backgrounds</i>	<i>Documentation</i>
Production of the translation, including first-hand versions and <i>judgmental</i> assessment	Parallel translations Team-based review and adjudication	<ul style="list-style-type: none"> <li>• Skilled translation practitioners</li> <li>• Skilled translation practitioners</li> <li>• Survey and subject matter expert(s)</li> <li>• Other person(s) of relevance</li> </ul>	Problems, comments Problems, decisions, adaptations
Testing the translation as a measurement instrument, including <i>empirical</i> assessment	Qualitative testing, such as cognitive interviewing Quantitative testing, such as a pilot study	<ul style="list-style-type: none"> <li>• Cognitive interviewers</li> <li>• Members of the target group</li> <li>• (Interviewers)</li> <li>• Survey researchers/statisticians</li> <li>• Members of the target group</li> </ul>	Problems, decisions, adaptations Problems, decisions, adaptations Of entire process

involvement of various persons with different skills sets and expertise, and documentation of both the overall process and individual decisions and findings (e.g., Acquadro et al., 2008; Harkness, 2003; Hambleton, 2005).

A multi-step approach includes (1) the process of producing a translation, including judgmental assessments, and (2) the process of testing the translation as a measurement instrument among members of the target group. Each of these processes can further be broken down.

### ***Production of a Translation***

Good practice includes, as a first step, the production of parallel translations, that is, two independently produced translation versions (e.g., Harkness, 2003). Parallel translations help to uncover idiosyncratic wording or different interpretations (e.g., ‘feeling anxious’ either in the sense of worry or in the sense of anxiety). Furthermore, they offer stylistic variants (e.g., different syntactical structures) and help identify clear-cut errors that inevitably occur in translation, even among experienced translators (e.g., omission of important elements). It should be self-evident, however, that the parallel translation approach should not be taken as a remedy against a weak translator but rather

draw on two persons with (potentially complementary) expert translation skills. In a subsequent team-based review session, the two translation versions are then reconciled to result in a final translation. Reconciling translations can include choosing one or the other translation for a given question, combining the two translations, modifying a given translation, or generating a completely new one. Arriving at a final translation always involves thorough decision-making and should thus not be limited to just selecting the ‘better’ out of the two translations that are offered. Aspects that will play a role are meaning, conciseness, fluency, questionnaire design conventions, consistency, amongst others (Behr, 2009). Thus, translation always is a multi-dimensional decision-process.

In what has been called one of the ‘major misunderstandings in the field’ (Hambleton and Zenisky, 2011: 66), translations have all too often been made by a friend, a colleague or a partner simply because they happen to be ‘bilingual’. Nowadays there is a broad consensus that the skills and background of the personnel employed are crucial in determining the outcome. Translators should have an excellent command of both source and target language and culture and typically they translate into their mother tongue. Other requirements include a combination of (questionnaire) translation experience, knowledge

of the study topic and of questionnaire design principles. In social science survey research the view is taken that the 'people most likely to be good questionnaire translators are people who are already good translators and who learn or are trained to become questionnaire translators' (Harkness et al., 2004: 463). Supporting this position, research from the field of translation studies suggests that translation practitioners, when compared to translation novices, act more sense-oriented, take into account larger segments of text, are more likely to attend to the needs of prospective users of a text and more so exploit their cultural and world knowledge in the process of translation (Jääskeläinen, 2010; Shreve, 2002). Nevertheless, if translation practitioners are not cognizant of do's and don'ts in questionnaire translation, they will need to be briefed and trained on what it means to retain measurement properties and design principles in translation (e.g., characteristics of response scales, balanced wording, etc.). Translators may in fact also be trained on the job in team review sessions, as described below.

Team-based review, the follow-up to parallel translation, is a recommended way to spread needed skills and expertise among several people (Harkness, 2003). A team-based review brings the translators together with survey and subject matter experts as well as other persons that are deemed helpful for the task. A team approach can thus ensure that linguistic and measurement aspects are taken on board when making decisions. While the pooling of expertise is the great advantage of the team-based approach, there are factors in this team set-up that might negatively impact on the outcome, such as defending a version for personal reasons or not wanting to criticize each other (Brislin, 1980). These challenges are likely to be mastered by providing information on the chosen translation approach and on review 'rules' (the quality of the translation is the focus, not any personal assessment) to all parties prior to team selection. Thus, the participants of a team

approach can 'mentally' prepare for an interdisciplinary exchange. Moreover, Harkness et al. (2010a) recommend testing translators in terms of their review performance and team suitability prior to hiring, which is certainly easier to implement for longer instruments than for shorter ones. In addition, the actual review process can be supported by allocating enough time for the process as well as involving skilled personnel to chair the review session. Review leaders are crucial for the overall outcome; they should be knowledgeable both of questionnaire translation and of study and measurement characteristics. Often, they have good knowledge of the subject matter and/or of survey methods and as such they guarantee that the measurement perspective is adequately taken into account during the translation process.

After review, a so-called adjudication step may be necessary for adding further expertise, clarifying remaining uncertainties, and for final decision-making (Harkness, 2003). In addition, enough time should always be set aside for copy-editing, including consistency or fluency checks, as well as proof-reading in terms of spelling, grammar, and completeness.

### ***Special Case: Back Translation***

In the above descriptions of good practice in questionnaire translation the method of back translation has deliberately been omitted. Back translation, in widespread use since about the 1970s (Brislin, 1970), is controversially discussed in the research community (Harkness, 2003; Leplège and Verdier, 1995). Essentially, it involves translating the 'actual' translation of a questionnaire back into the source language and the subsequent comparison of the two source-language versions with a view to identify discrepancies. Even though (gross) errors can be detected using back translation, the method itself is no guarantee that the 'actual' translation is indeed comparable, working as intended, intelligible

or fluent. In one study, for instance, back translation did not identify that 'feeling downhearted and depressed' was translated as 'clinically depressive', since the back translation came back as 'depressed' and as such suggested no problem. Additionally, if wrongly implemented, the use of back translation fosters a translation that stresses equivalence of form over equivalence of meaning and thus it can even be detrimental to translation quality. Furthermore, much of the detection capability of a back translation depends on the skills and instructions given to a back translator and also on the background of the person who eventually compares the back translation to the original questionnaire. Especially if back translation is relied on as a sole quality check, which is in fact what already Brislin (1970) warned against, a low quality of the questionnaire translation is likely. While major survey programs or centers have discarded back translation altogether, focusing on target text-centered appraisals of the questionnaire instead (European Social Survey, 2014b; Ferrari et al., 2013), back translation is still a recommended method in many fields in the health sciences and cross-cultural psychology. However, also here efforts are underway to critically evaluate the method, notably by comparing psychometric properties and user preferences based on questionnaires produced according to different translation methods. Even though it needs to be acknowledged that the methodological set-ups of these studies differ in more than in the inclusion of back translation or not, first results suggest that at least in terms of user preference the 'back translation version' falls behind other methods (e.g., Hagell et al., 2010).

### ***Empirical Assessment of a Translation***

The recommended multi-step approach to questionnaire translation is not only reflected

in the subsequent steps of translation, review, and adjudication but also in the supplementary empirical assessment of the translated questionnaire. The particularity of this additional layer lies in the assessment of the questionnaire as a measurement instrument, which is, after all, its ultimate goal. Empirical assessment brings in the intended target group.

Qualitative assessment, typically in the form of cognitive interviews, is a way to gain in-depth knowledge of how respondents understand individual questions and arrive at their answers. Based on the respondent explanations, conclusions can be drawn on whether the translated instrument measures what it is intended to measure (see Chapter 24 by Willis in this Handbook).

Quantitative assessment is typically based on quantitative pretests or larger pilot studies, which may test pre-final questionnaires as such but also different question versions in split-ballot manner. The exact nature of statistical analyses heavily depends on the sample size, on how many items are used to measure a construct or on research traditions in the different disciplines (see Chapter 39 by Ciecuch et al. in this Handbook).

The types of analyses that are possible also depend on the timing of translation and empirical assessment. If cognitive interviewing or pretest or pilot studies are implemented simultaneously in different languages and cultures, equivalence across countries and cultures can be assessed in addition to other 'national' testing routines.

The question of when translation and its empirical assessment takes place within the survey cycle also has an effect on whether or not the source questionnaire can be modified based on the results. Once a source questionnaire has been finalized in a cross-national study, any feedback received from empirical assessment of the translated version can solely contribute to improving the translation itself (such as replacing a word by another) but not to improving the source questionnaire in general. If, however, translation and



its empirical assessment take place as part of comparative questionnaire design, modifications of both the translation itself and the pre-final source questionnaire are possible (Fitzgerald et al., 2011).

Even though qualitative and quantitative assessment of the translation can also identify clear-cut translation mistakes (e.g., ‘waiting’ translated as ‘wanting’), these types of errors should ideally be eradicated at earlier steps. After all, this is the rationale of thoroughly implemented parallel translation and team review/adjudication. Empirical assessment should primarily tackle more subtle issues of translation, such as connotations or misunderstandings, as well as cultural problems or generic design problems.

### **Translation Documentation**

The general translation and assessment approach should be documented, not only in terms of steps implemented but also in terms of personnel involved. Furthermore, the individual steps should be documented: Translator documentation may include problems or alternative formulations (e.g., a word used in the translation may be difficult to understand). Reviewer and/or adjudicator documentation may include problems and (adaptation) decisions (e.g., ‘spending a school term abroad’ translated as ‘spending several months abroad’ to take into account different lengths of exchanges for different countries). Documentation pertaining to empirical testing may include information on identified problems with the translation and/or the source questionnaire and decision based upon these findings (e.g., ‘romantic partner’ too closely translated. It is perceived as odd. It should be rendered in a more sense-oriented way). Documentation on both the general and the specific level gives future instrument and data users a first indication as to the quality of the translation and also a source to turn to in case of unexpected statistical results. During the translation process

itself, documentation helps to inform later steps in the process, thus making the entire process more efficient.

### **Harmonization**

Harmonization is receiving increased attention in cross-national studies. On the one hand, it refers to the process of developing a common version for different varieties of a ‘shared language’. This would apply, for instance, to a common French version for France and Belgium. Differences between shared-language questionnaires should only occur where this is culturally or linguistically needed. The rationale is to remove any unnecessary variation that might impact on the comparability of data (Harkness, 2010; Wild et al., 2009).

On the other hand, harmonization may refer to the process of fostering consistency in translation decisions and thus comparability independent of the respective language. Harmonization of this type is warranted in countries that simultaneously need to produce several language versions of a questionnaire, such as Switzerland needing to translate cross-national source questionnaires into French, German, and Italian. Additional efforts are needed at the adjudication, that is, finalization stages to harmonize the different linguistic versions, possibly by members of staff mastering the different languages. Furthermore, the call for consistent translation decisions should be extended to *all* countries in a study. This wider type of harmonization can either be helped by (in-person) meetings in which representatives from each target language take part to discuss problematic issues (Acquadro et al., 2012). Alternatively, FAQ lists with country queries on the meaning or scope of terms and developer feedback that are regularly updated and circulated among translation teams may serve the purpose (Furtado and Wild, 2010). In the same vein, proactive translation assessment of selected languages at an early stage

can identify issues that should be clarified for all countries (Wild et al., 2005). All these efforts show that questionnaire translation methodology in a cross-national study is increasingly shifting from a vertical perspective that only looks at one target questionnaire in relation to the source to a horizontal perspective that looks at several target questionnaires simultaneously. Especially in-person meetings require additional time and money and may also be logistically difficult to implement. Moreover, countries need to be more or less at the same stage within the process to make it work. Ongoing harmonization efforts by updated FAQ lists are thus a powerful alternative; these lists can also be accessed by countries which join a survey at a later stage.

### ***Quality in the Hands of the Translation Commissioner***

While the quality of a questionnaire translation is usually linked to the aforementioned factors, that is, source questionnaire quality and related documentation, suitable personnel and appropriate translation and assessment methods, the role of the translation commissioner should not be ignored (Chesterman, 2004). Translation commissioners, which are often synonymous to the national translation project managers, determine or provide the production conditions which then impact on the translation (quality). Production conditions include deadlines, overall process planning, payment, translation files, translation tools, briefing and training. Briefing in terms of study goals, target group, and implementation mode is vital to producing good translations. After all, translation involves decision-making taking into account these factors. Beyond briefing, training may become particularly important if hired personnel need to be trained on the particularities of measurement instruments. By providing an adequate financial, temporal and content-wise framework,

the commissioner can thus improve translation quality.

## **TRANSLATION VS ADAPTATION**

When it comes to the use of questionnaires in cross-national or cross-cultural contexts, the terms translation and adaptation dominate the relevant literature. The following is an attempt to shed light on what is meant when people refer to adaptation rather than translation. On the one hand, adaptation may be used as a generic term for the overall process of transferring an instrument from one language and culture to another language and culture. On the other hand, the term may be used to describe deliberate changes to specific questions. In the following, these two different perspectives – the generic and the specific one – will be presented.

### ***Adaptation in a Generic Sense***

In particular in cross-cultural psychology and the health sciences, using the term adaptation for the overall process is very popular (Acquadro et al., 2012; van de Vijver and Leung, 2011). Adaptation in this sense signals that ‘pure’ translation may not be sufficient and that cultural adaptation at various levels and for various questions may be needed to make an instrument suitable for a new context. Given that in these disciplines many questionnaires that were originally developed with only one culture in mind are now transferred to a new culture, the preference for the term adaptation becomes understandable. Using the term adaptation in a generic sense may also stress the need for psychometric testing to ensure that the new instrument works as intended. Rigorous statistical testing as part of the ‘adaptation’ process of an instrument is more common in cross-cultural psychology and the health sciences than in the social sciences; this may be



due to the different types and numbers of items measuring a construct as well as to issues of copyright and commercial distribution (Harkness et al., 2004). In the social sciences, it is rather the researcher working with the final data set who is responsible for statistical testing. Eventually, one may wonder whether the term adaptation also has become so immensely popular as an overall term because translation itself is often reduced to a mere word-by-word replacement or a 'literal' translation. Such an understanding testifies to a misconception of what translation involves and also to a lack of knowledge of where translation researchers have been heading since the 1950s (Bolaños-Medina and González-Ruiz, 2012).

### ***Adaptation in a Specific Sense***

Apart from adaptation in the generic sense, adaptation in the specific sense is used to refer to changes to specific questions. Harkness et al. refer to these changes as 'deliberate changes to source material in order to meet new needs of various kinds' (2010b: 133). One can look at these changes from different angles: (1) domains of changes; (2) type of 'material' affected; (3) type of changes, and (4) topics potentially triggering changes.

Following a slightly modified classification approach by van de Vijver and Leung (2011), one can differentiate between adaptations in the domain culture, measurement, and language (see Table 19.2). A clear cut between these domains is not always given, though, so that some adaptations types may concurrently be assigned to different domains.

Adaptations in the domain *culture* can be subdivided into terminological-/factual-driven adaptations and norm-driven adaptations. The former deal with the 'hard' aspects of culture, whereas the latter cover its 'soft' aspects (van de Vijver and Leung, 2011). Terminological-/factual-driven adaptations accommodate factual, often obvious

**Table 19.2 Overview of adaptation domains and types**

<i>Type</i>	<i>Domain</i>	<i>Type</i>
1	Culture	Terminological/factual-driven
2	Culture	Norm-driven
3	Measurement	Familiarity-driven
4	Measurement	Format-driven
5	Language	Comprehension-driven
6	Language	Language-driven
7	Language	Pragmatics-driven

*Note:* classification slightly modified from van de Vijver and Leung (2011)

differences between countries. For instance, references to political systems (American president vs British prime minister) or school systems (British A-levels vs German Abitur) will have to be adapted. Norm-driven adaptations cater for less tangible, often less obvious differences between countries, notably as regards norms, practices or values. Asking respondents whether they have recently worn a campaign badge or sticker in countries where badge and sticker are not elements of political participation is certainly not suitable. Also aspects of social desirability or sensitivity in a given culture need to be considered. In Japan, for instance, questions involving the assessment of one's own or others' earnings as just or unjust are socially inappropriate and thus cannot be asked in a general survey (Harkness et al., 2003).

Adaptations in the domain *measurement* can be subdivided into familiarity-driven adaptations and format-driven adaptations. Familiarity-driven adaptations are needed to accommodate different familiarity with measurement instruments. Surveys may be carried out in populations that have had no (or hardly any) prior exposure to surveys. In these cases, the survey experience may need to be brought to these populations by adding explanations or instructions on how to use the survey instrument. In addition, verbal scales may be adapted or supplemented with pictorial aids to make measurement more accessible to these survey respondents (Harkness

et al., 2010c). Format-driven adaptations take into account differential response effects or styles. A Japanese agreement scale may thus label the extreme categories of an agree-disagree scale as 'agree' and 'disagree', whereas the source scale uses the labels 'strongly agree' and 'strongly disagree'. This modification would take into account the Japanese predisposition to avoid response categories with strong labels (Smith, 2004; see also 'Attitude and opinion items in translation' below).

Adaptations in the domain *language* pertain to comprehension-driven adaptations, such as when certain concepts or wordings may need to be supplemented by definitions to help adequate understanding. Also for populations with lower levels of education compared to source text respondents the wording and vocabulary of the target instrument as a whole may need to be simplified. Apart from these adaptations, various authors also list language-driven and pragmatic-driven adaptations among the adaptation types (Harkness, 2008; Harkness et al., 2003; van de Vijver and Leung, 2011). Language-driven adaptations include the array of changes that inevitably happen in translation, such as changes in the sequence of information, in sentence structure or in word class. Van de Vijver and Leung (2011) illustrate language-driven adaptations with the gender-neutral English word 'friend' and its gender-specific counterparts in German (Freund/Freundin) or French (ami/amie). Pragmatics-driven adaptations take into account that language use in social contexts differs between languages and cultures. Different discourse norms may call for modifications. For instance, the required degree of explicitness of a request may differ or the way how deference or politeness is expressed. Neglecting the peculiar discourse norms and instead rendering the source text too closely may mean that questions come across as strange or even inappropriate, with potential effects for measurement. We would argue that in particular language- and pragmatic-driven adaptations are the backbones

of 'translation' itself. This does not make them less difficult to implement, but this view would free the activity of 'translation' from a mere mechanical replacement activity and highlight the changes that are necessarily inherent in 'translation' (Baker, 2011). In this view, a meaningful distinction between translation and adaptation can be made and the term adaptation be reserved for activities that change the stimulus in more significant ways, notably in the domains culture and measurement. If all changes were called adaptation, even the most basic ones as required by different language systems (e. g., change of word order), the distinction between adaptation and translation would become futile; moreover, it would be difficult to inform data users on how a translation compares to the source version and what this could mean for statistical analyses. Of course, there will always be grey areas in the domain language of what should be called a translation and what should rather be called an adaptation. In terms of necessary documentation of decisions, it seems advisable in any case to document the more significant types of changes, especially where impact on comparable measurement can be expected, no matter whether these changes should technically be called a translation or an adaptation.

Potential candidates for adaptation are individual questions and their answer scales. However, also the visual presentation of a questionnaire may be affected, such as when colors or pictures need to be changed, modes of emphasis switched from capitalization to underlining, or answer text boxes regrouped or resized in view of typical answer patterns or conventions. Furthermore, also the layout and direction of response scales may be affected when different writing systems (left-right/right-left/up-down) are involved.

The various types of changes include addition or omission of questions or parts thereof as well as substitution of different kinds (content, pictures, and colors).

Topics potentially triggering adaptations include socio-economic topics (such

as income, housing), religion, sports, foods and drinks, drugs, activities, holidays, music, family ties, school system, health system, political system, history, name, address, and date formats, knowledge questions, and measuring units (Dean et al., 2007; Harkness et al., 2010b).

In sum, the transfer of a questionnaire from one language and culture to another language and culture requires a great deal of sensitivity towards cultural issues. The degree to what extent adaptations are needed will largely depend on whether a source questionnaire has deliberately been designed for cross-cultural use and thus tries to avoid cultural particularities right from the start or whether it was originally developed for one particular culture. Adaptations needs, notably in the domains culture and measurement, are typically taken into account and circumvented in deliberate cross-national questionnaire design or, to a lesser degree, they are also accommodated.

### **ATTITUDE AND OPINION ITEMS IN TRANSLATION – EXPERIENCES FROM CROSS-CULTURAL SURVEYS IN ASIA**

This section discusses translation issues regarding attitude and opinion items that require special attention during translation. The discussion will be based on the experience of the East Asian Social Survey (EASS), which is a cross-cultural survey exclusively conducted in Asian countries and regions. It is more difficult to obtain translation equivalence for survey items which ask about attitudes and opinions than it is for survey items which ask about the respondent's background characteristics (gender, age, education, etc.) and behaviors (number of hours of TV watched per day, frequency of exercise per month, etc.) (Behling and Law, 2000; Tasaki, 2008). Survey items which ask about attitude and opinions gauge individual values and are of high interest for psychological and

sociological survey researchers. Generally, these survey items measure answers given by respondents by using scales with 2–5 categories on dimensions such as 'good-bad' or 'agree-disagree.'

Highly abstract concepts tend to be included among attitude and opinions items and these often cause problems for translation. The response will change based on how the highly abstract concept is translated. In addition, the design and translation of response scales has a more direct impact on the response than the translation of the question itself. If response scales are adopted from other surveys and 'merely' translated, this may create an especially significant impact on survey data.

### ***Harmonization in View of Conceptual Equivalence***

The most basic and important requirement in comparative survey research is to measure the same concept across countries. When the concept to be measured deviates between cultures, conceptual equivalence is impaired. Failure in translation is one of the causes that reduce conceptual equivalence.

In 2008, an East Asian Social Survey (EASS), based on the theme of Culture and Values in East Asia, was carried out in Japan, South Korea, China, and Taiwan. Among the most difficult survey items to translate were the survey items regarding 'preferred qualities of friends' as shown in Table 19.3. The reasons why these survey items were difficult to translate included (1) the concepts to be measured being highly abstract, (2) multiple appropriate translations with different nuances being available, and (3) a lack of useful information that could clarify meaning (for example, context of preceding and following questions or notes to the translator).

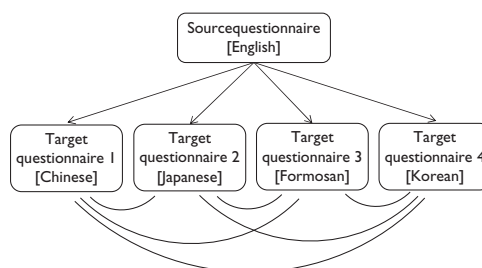
In the EASS, meetings to develop questionnaires are carried out in English, and the source questionnaire is developed in English.

**Table 19.3** Survey item for preferred qualities of friends

Q When you associate with your personal friend, how important is each of the following qualities?					
	Very important	Important	Neither Important nor unimportant	Not important	Not important at all
a. Honest	(1)	(2)	(3)	(4)	(5)
b. Responsible	(1)	(2)	(3)	(4)	(5)
c. Intelligent	(1)	(2)	(3)	(4)	(5)
d. Cultured	(1)	(2)	(3)	(4)	(5)
e. Powerful	(1)	(2)	(3)	(4)	(5)
f. Wealthy	(1)	(2)	(3)	(4)	(5)
g. Loyal	(1)	(2)	(3)	(4)	(5)
h. Warm-hearted	(1)	(2)	(3)	(4)	(5)

However, the actual survey is carried out in Chinese, Japanese, Formosan, and Korean in the respective regions. Discussions regarding the translation of cross-cultural surveys involve numerous discussions concerning how to appropriately translate the English source questionnaire. There is the assumption that if the translation from English to the languages of the multiple survey target regions can be done appropriately, comparability between the multiple survey target regions is established. However, this assumption does not apply to the translation of highly abstract survey items in particular. For example, even if the translation from English to Japanese or from English to Chinese is carried out appropriately, this does not guarantee that there is comparability between the Japanese and the Chinese version. There are multiple appropriate translations with different nuances for the Japanese translation of the term 'honest', which have meanings such as *shouzikina* (sincere), *seizituna* (faithful), *zicchokuna* (trustworthy), *socchokuna* (straightforward), *honmonono* (genuine), *zyunseino* (pure), *kouheina* (fair), *kouseina* (impartial), *nattokudekiru* (satisfactory), *ukeirerareru* (acceptable), *sobokuna* (unembellished), *mie-wo-haranai* (non-ostentatious), and *kazarinonai* (unvarnished). Likewise, there are also multiple appropriate translations with different nuances for the Chinese translation of the term 'honest.'

For concepts which are highly abstract and can have multiple appropriate translations, it is not only necessary to determine the appropriateness of the translation from English to Japanese or from English to Chinese, but it is also vital to establish conceptual equivalence between the languages of the survey target regions and carry out harmonization among these languages (Figure 19.1). Researchers and translators who understand two languages such as Japanese and Chinese are necessary for this harmonization. In East Asia, communication using *kanji* characters, which are moderately common in these regions, has been effective in increasing conceptual equivalence in addition to English. In cases when there are numerous survey target regions, harmonization in all survey items becomes difficult work and is not realistic. Harmonization is recommended, if not for the

**Figure 19.1** Harmonization between survey target regions.

entire questionnaire, at least for some items that require special attention in translation. In addition, item-specific translation annotations which specify the intended meaning of these highly abstract concepts can support harmonization.

### **Response Styles**

Generally, survey items which measure opinions and attitudes base their measurement on graded scales. The scales may range, for instance, from '1 = very important' to '5 = not important at all' or from '1 = strongly agree' to '5 = strongly disagree.' Depending on the target region and culture, there will be a disposition towards response patterns such as midpoint responding and extreme responding. These are called response styles and they create methodological artefacts which jeopardize the comparability between cultures (Tasaki, 2008; van de Vijver and Poortinga, 1997). Cross-cultural surveys should design response scales which take response styles of the target region into consideration in order to improve comparability.

East Asia is located in the Confucian cultural sphere. The standard of conduct among the teachings of Confucianism is moderation, and value is placed on being neutral and not changing. In addition, the idea of taking a moderate course is taught in Buddhism, which is followed in parts of East Asia, with value being placed on distancing oneself from too extreme ways of thinking. Added to this is collectivism: According to Hofstede (1995), Western countries such as the US, Australia, England, and Canada have strong individualistic ways of thinking while countries and regions in East Asia such as Japan, Hong Kong, South Korea, and Taiwan have strong collectivistic ways of thinking. According to Triandis (1995), collectivism aspires for everyone to think, feel, and act in the same manner while individualism prefers people to clarify their position through discussion.

The collectivistic way of thinking in Confucianism and Buddhism are thought to have an impact on response styles in social surveys. This way of thinking creates a tendency to avoid extreme responses and instead choose midpoint responses. Si and Cullen (1998) found that East Asian people (China, Japan, Hong Kong) have a greater tendency than Western people (United States, Germany, United Kingdom) to choose middle response categories when offered an explicit midpoint response category. Besides, East Asian people are less likely than Western people to select either end-point categories. Midpoint responding is remarkable especially in Japan (Hayashi and Hayashi, 1995). Japanese people have a tendency to value group-oriented culture. Group-oriented culture signifies favorably maintaining personal relationships within a group and placing importance on the order and harmony of the group. Likewise, Japanese people have a tendency to regard expressing individual opinions and emotions as shameful, and they voice their individual opinions based on the situation. There is a strong attitude towards being vague without saying your opinion instead of expressing your opinion and disturbing the situation. In social surveys which only take place in Japan, scales for opinion items are intentionally created as '1: agree, 2: somewhat agree, 3: somewhat disagree, 4: disagree', which takes the response style of Japanese people into consideration. By avoiding adverbs such as 'strongly' as well as midpoint categories such as 'neither agree nor disagree', the scale ensures a spread of responses and it further clarifies if the respondent agrees with the opinion or not.

### **Design of Response Scales**

Shishido et al. (2009) compared the questionnaires from the World Value Survey (WVS) and the International Social Survey Programme (ISSP), which are cross-cultural surveys carried out on a global scale, with the

East Asia Value Survey (EAVS), the East Asia Barometer Survey (EABS), and the Asia Barometer Survey (ABS), which are all cross-cultural surveys carried out only in Asian regions. These comparisons clarified the characteristics of response scales of attitudinal items. The characteristics common to all survey projects are (1) the frequent use of a verbal and bipolar scale, and (2) the infrequent use of a scale with more than five points. The ISSP project, which started in the 1980s as a cross-cultural survey among Western countries, very frequently uses a five-point scale that includes categories with a strong adverb and a midpoint category. In contrast to the ISSP, the WVS, which since its inception extensively covers heterogeneous cultural zones, is characterized by its frequent use of two-point and four-point scales that have no midpoint. In surveys that focus only on Asia, scales without a midpoint are used relatively frequently.

There has been a dispute as to which scale should be adopted. Smith (1997) discussed that the bipolar scale including a midpoint has a smaller risk of mistake in terms of positioning one's opinion on a response scale, and is therefore more desirable for cross-national comparison than the unipolar scale or scales without a midpoint. Klopfer (1980) and Krosnick et al. (2008) also suggested that offering a midpoint is desirable because omitting the middle alternative leads respondents to randomly select one of the moderate scale points closest to where a midpoint would appear.

On the other hand, Converse and Presser (1986) suggested that a middle alternative should not be explicitly provided because providing a midpoint leads to the loss of information about the direction in which people lean. It is better not to offer the middle point in response scales if the direction in which people are leaning on the issue is the type of information wanted (Payne, 1951; Presser and Schuman, 1980).

Cross-cultural surveys which target Western countries with individualistic

cultures, where people often clearly express their opinions, include midpoint categories in the response scales and use strong adverbs (strongly, absolutely, etc.) on both extremes of the response scales. This is not considered too much of a problem. However, in cross-cultural surveys which include East Asia with countries such as Japan which prefer vague responses, responses often seem to concentrate on the midpoint or the area around midpoint responses. Therefore, incorporating midpoint categories and strong adverbs into the scale should be considered carefully. It would also be necessary to consider whether off-scale options (such as 'Can't choose', 'It depends on the situation', 'I don't know') are incorporated into the scale or not, because response patterns to midpoint categories are similar to response patterns to off-scale options.

Figure 19.2 shows results of the EASS 2006 family module. Response distribution of 18 attitudinal items regarding family in Japan, South Korea, China, and Taiwan were compared. The response categories were on a seven-point scale which included a midpoint category. The Japanese team was against including a midpoint category in the scale, but a scale with a midpoint category was eventually adopted based on the request of other teams who placed importance on comparability with the ISSP. There is a notable proportion of midpoint responses in all regions, but midpoint responses were especially high in Japan. Taking these striking results into account, it seems advisable for researchers conducting global surveys to take into consideration also the response styles of non-Western regions when designing response scales.

### ***Translation of Response Scales***

The issues of how to design response scales and how to translate response scales are closely related. Shishido et al. (2009) compared the target questionnaires of the ISSP



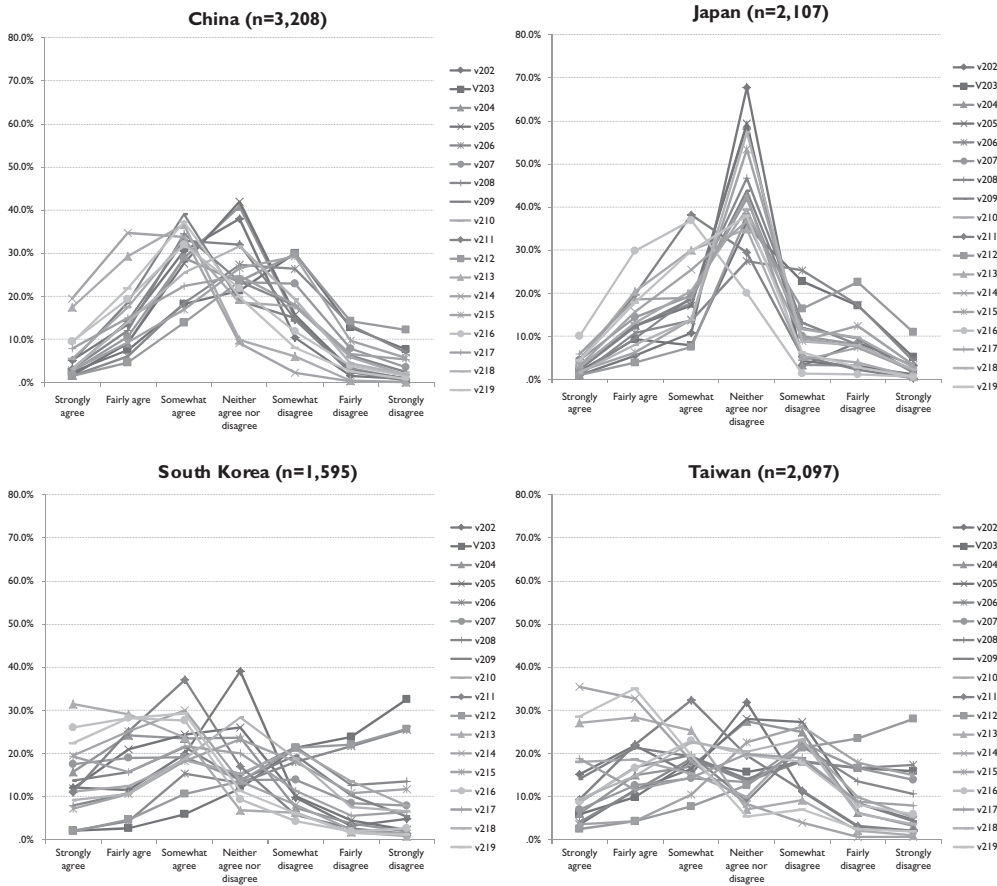
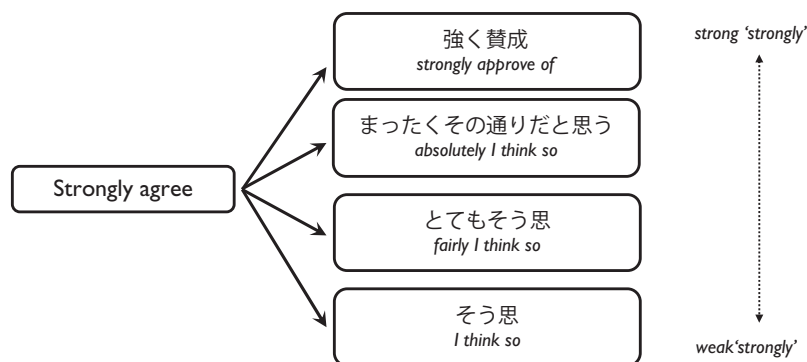


Figure 19.2 Response distribution of 18 survey items.

and the WVS and found the following: (1) there are multiple regions where the translation of the two surveys differed even though labels of response categories in English were identical; and (2) the difference in translations of response categories had an impact on response distribution. The two surveys produced different results even though the survey items themselves were identically translated. When looking at a number of cross-cultural surveys which are carried out in Japan, there are multiple translations for a category such as ‘strongly agree’ (Figure 19.3). The translations of this response category are different in the ISSP and the WVS in Japan, which produces different response distributions for the

same survey items in different surveys. The WVS directly translates ‘strongly agree’ as *tsuyoku sansei* (‘strongly approve’) while the ISSP freely translates ‘strongly agree’ as *sou omou* (‘I think so or merely ‘agree’). As a result of such translation, there were few respondents who answered with ‘strongly agree’ in the WVS; on the other hand, there were many respondents who answered with ‘strongly agree’ in the ISSP. Difficulties in translating response categories do not exist only in Japan when assuming that there are many regions which have different translations of response categories.

Given the differences in responses distribution possibly caused by translation



**Figure 19.3** Examples of Japanese translations of 'strongly agree'.

differences in response categories, identical English response scales should be translated identically among cross-national surveys (ISSP, WVS, etc.) in order to be able to compare results across these surveys. It is desirable to seek a way to harmonize translation by sharing information among local research agencies.

Countries and regions participating in cross-national survey projects should make an effort to examine the intensity of adverbial expressions and to enhance equivalence of the expressions across participant countries and regions. As for research into adverbial expressions in agreement scales, the Research into Methodology of Intercultural Surveys (MINTS) project shall be named as a reference. This project examined the equivalence of the agreement scale based on a direct rating approach that quantified and measured the impressions of respondents when confronted with different response categories (Mohler et al., 1998). A continuation of this work, including more languages and cultures, is highly desirable.

## FUTURE DEVELOPMENTS

While the research community continues to refine translation and assessment methodology as well as the integration of translational

aspects into questionnaire design, surprisingly as it is little research has been done so far about the impact of different translation or adaptation versions on data comparability. Which differences matter and which do not, and for which type of concepts or items do such findings hold, what are particularly robust items? Here, further research is urgently needed to inform translation and adaptation practices.

A related area of heightened interest is corpus linguistics and in particular the question as to how large language corpora may help substantiating translation decisions or understanding questionnaire design principles in different languages. In the latter regard, the research community should evaluate to what extent design guidelines and principles developed predominantly based on the English language and Anglo-Saxon culture really apply to other languages and cultures, or whether some re-thinking is urgently needed.

Furthermore, while much research, especially in the context of the US, deals with translation for migrant groups, systematic work on differences between questionnaire translation for migrant groups and questionnaire translation for different countries is missing. What are particular requirements and challenges that need to be addressed when translating a questionnaire for a migrant group, the particularity of which is to bring two cultures and two languages to the response process? The

growing importance of including migrants in surveys or even focusing on them makes this topic particularly sought-after.

The role of computer-aided translation tools will also become more important, especially in light of increased electronic data collection and the potential to produce the translation in a format that is directly usable by a given survey software or a data archive. The DASHISH project (2014), which aims, amongst others, at the development of a cross-cultural questionnaire design, translation, and documentation tool, can be named as a case in point.

Furthermore, closer cooperation and exchange of information between all parties concerned is likely to shape comparative projects more so than before. The harmonization of translation decisions is a key factor in this regard.

Last but not least, slowly but steadily findings, theories, and applications from translation research are finding their way into cross-cultural survey research (Behr, 2009; Bolaños-Medina and González Ruíz, 2012; Chidlow et al., 2014). This interdisciplinary exchange needs to grow to help substantive and survey researchers better understand the complexities, possibilities, and limits of translation and to improve cooperation between the fields. The same learning process, of course, is needed among translation practitioners and translation researchers when it comes to survey research and measurement issues.

## RECOMMENDED READINGS

To conclude this chapter, following works shall be recommended for further reading: two seminal works by Harkness, who tremendously influenced the field of questionnaire translation (Harkness, 2003; Harkness et al., 2010b); a chapter on cross-cultural questionnaire design by Smith (2004); an article on cross-cultural cognitive interviewing (Fitzgerald et al., 2011), and finally a general book

on translation, written by an acknowledged translation researcher (Baker, 2011).

## ACKNOWLEDGEMENTS

East Asian Social Survey (EASS) is based on Chinese General Social Survey (CGSS), Japanese General Social Surveys (JGSS), Korean General Social Survey (KGSS), and Taiwan Social Change Survey (TSCS), and distributed by the EASSDA.

## REFERENCES

- Acquadro, C., Conway, K., Giroudet, C., and Mear, I. (2012). *Linguistic Validation Manual for Health Outcome Assessments*. Lyon: Mapi Institute.
- Acquadro, C., Conway, K., Hareendran, A., and Aaronson, N. (2008). Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value in Health*, 11, 509–521.
- Baker, M. (2011). *In Other Words: A Coursebook on Translation* (2nd edn). London & New York: Routledge.
- Barioux, M. (1948). Techniques used in France. *Public Opinion Quarterly*, 12, 715–718.
- Behling, O., and Law, K.S. (2000). *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. Thousand Oaks, CA: Sage.
- Behr, D. (2009). *Translationswissenschaft und international vergleichende Umfrageforschung: Qualitätssicherung bei Fragebogenübersetzungen als Gegenstand einer Prozessanalyse*. GESIS-Schriftenreihe, 2, Bonn: GESIS.
- Behr, D., and Scholz, E. (2011). Questionnaire translation in cross-national survey research: on the types and value of annotations. *Methoden – Daten – Analysen*, 5, 157–179.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216.
- Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In

- H. C. Triandis and J. W. Berry (eds), *Handbook of Cross-Cultural Psychology: Methodology*, Vol. 2 (pp. 389–444). Boston, MA: Allyn and Bacon.
- Bolaños-Medina, A., and González Ruíz, V. (2012). Deconstructing the translation of psychological tests. *Meta*, 57, 715–739.
- Chesterman, A. (2004). Functional quality. Retrieved from <http://www.youtube.com/watch?v=IJW1Y6rAB1I> [accessed 2014-04-08]
- Chidlow, A., Plakoyiannaki, E., and Welch, C. (2014). Translation in cross-language international business research: beyond equivalence. *Journal of International Business Studies* (2014) (advance access)
- Converse, J. M., and Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. London: Sage Publication.
- DASHISH – Data Service Infrastructure for the Social Sciences and Humanities (2014). Activities: Data Quality. Retrieved from <http://dasish.eu/activities/> [accessed 2014-04-08]
- Dean, E., Caspar, R., McAvinchey, G., Reed, L., and Quiroz, R. (2007). Developing a low-cost technique for parallel cross-cultural instrument development: the Question Appraisal System (QAS-04). *International Journal of Social Research Methodology*, 10, 227–241.
- Dept, S. (2013). Translatability assessment of draft questionnaire items. Paper presented at the conference of the European Survey Research Association (ESRA), Ljubljana, SI.
- Dorer, B. (2011). Advance translation in the 5th round of the European Social Survey (ESS). *FORS Working Paper Series 2011*, 4.
- European Social Survey (2014a). ESS pre-testing and piloting. Retrieved from [http://www.europeansocialsurvey.org/methodology/pre-testing\\_and\\_piloting.html](http://www.europeansocialsurvey.org/methodology/pre-testing_and_piloting.html) [accessed 2014-04-08].
- European Social Survey (2014b). *ESS Round 7 Translation Guidelines*. London: ESS ERIC Headquarters, Centre for Comparative Social Surveys, City University London.
- Ferrari, A., Wayrynen, L., Behr, D., and Zabal, A. (2013). Translation, adaptation, and verification of test and survey materials. In OECD (eds), *Technical Report of the Survey of Adult Skills (PIAAC) 2013* (pp. 1–28). PIAAC: OECD Publishing.
- Fitzgerald, R., Widdop, S., Gray, M., and Collins, D. (2011). Identifying sources of error in cross-national questionnaires: application of an error source typology to cognitive interview data. *Journal of Official Statistics*, 27, 569–599.
- Furtado, T., and Wild, D. J. (2010). *Harmonisation as Part of the Translation and Linguistic Validation Process: What is the Optimal Method?* Retrieved from [http://www.oxfordoutcomes.com/library/conference\\_material/posters/translations/TRANSLATION%20Furtado%20Wild%20Optimal%20Methods%20handout.pdf](http://www.oxfordoutcomes.com/library/conference_material/posters/translations/TRANSLATION%20Furtado%20Wild%20Optimal%20Methods%20handout.pdf) [accessed 2014-04-08].
- Hagell, P., Hedin, P.-J., Meads, D. M., Nyberg, L., and McKenna, S. P. (2010). Effects of method of translation of patient-reported health outcome questionnaires: a randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. *Value in Health*, 13, 424–430.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, and C. D. Spielberger (eds), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., and Zenisky, A. (2011). Translating and adapting tests for cross-cultural assessments. In D. Matsumoto, and F. van de Vijver (eds), *Cross-Cultural Research Methods* (pp. 46–70). Cambridge: Cambridge University Press.
- Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (eds), *Cross-Cultural Survey Methods* (pp. 35–56). Hoboken, NJ: Wiley.
- Harkness, J. A. (2008). Comparative survey research: goal and challenges. In E. de Leeuw, J. J. Hox, and D. A. Dillman (eds), *International Handbook of Survey Methodology* (pp. 56–77). New York: Lawrence Erlbaum Associates.
- Harkness, J., Behr, D., Bilgen, I., Córdova Cazar, A. L., Huang, L., Lui, A., Stange, M., and Villa, A. (2010a). VIII. Translation: finding, selecting, and briefing translation team members. In Survey Research Center, *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of

- Michigan. Retrieved from <http://www.ccsgr.umich.edu/> [accessed 2015-05-22].
- Harkness, J., Pennell, B.-E., and Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer (eds), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 453–473). Hoboken, NJ: Wiley.
- Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., and Villar, A. (2010c). Designing questionnaires for multipopulation research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, and T. W. Smith (eds), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 33–57). Hoboken, NJ: Wiley.
- Harkness, J., van de Vijver, F. J. R., and Johnson, T. P. (2003). Questionnaire design in comparative research. In J. Harkness, F. J. R. van de Vijver, and P. Ph. Mohler (eds), *Cross-Cultural Survey Methods* (pp. 19–34). Hoboken, NJ: Wiley.
- Harkness, J. A., Villar, A., and Edwards, B. (2010b). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, and T. W. Smith (eds), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 117–140). Hoboken, NJ: Wiley.
- Hayashi, C., and Hayashi, F. (1995). Comparative study of national character. *Proceedings of the Institute of Statistical Mathematics*, 43, 27–80.
- Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*. London: McGraw Hill.
- Jääskeläinen, R. (2010). Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies. In G. M. Shreve, and E. Angelone (eds), *Translation and Cognition* (pp. 213–227). Amsterdam: John Benjamins.
- Klopper, F. J. (1980). The middlemost choice on attitude items: ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin*, 6, 97–101.
- Krosnick, J. A., Judd, C. M., and Wittenbrink, B. (2008). The measurement of attitudes. In C. Roberts and R. Jowell (eds), *Attitude Measurement, Volume 2: Designing Direct Measures* (pp. 1–85). London: Sage Publications.
- Lepège, A., and Verdier, A. (1995). The adaptation of health status measures: methodological aspects of the translation procedure. In S.A. Shumaker, and R. A. Berzon (eds), *The International Assessment of Health-Related Quality of Life: Theory, Translation, Measurement and Analysis*, (pp. 93–101). Oxford: Rapid Communications.
- Mohler, P. P., Smith, T. W., and Harkness, J. A. (1998). Respondents' ratings of expressions from response scales: a two-country, two-language investigation on equivalence and translation. *ZUMA-Nachrichten Spezial No. 3: Cross-Cultural Survey Equivalence*, 159–184.
- Payne, S. (1951). *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.
- Presser, S., and Schuman, H. (1980). The measurement of a middle position in attitude surveys. *Public Opinion Quarterly*, 44, 70–85.
- Shishido, K., Iwai, N., and Yasuda, T. (2009). Designing Response Categories of Agreement Scales for Cross-national Surveys in East Asia: The Approach of the Japanese General Social Surveys. *International Journal of Japanese Sociology*, 18, 97–111.
- Shreve, G. (2002). Knowing translation: Cognitive and experiential aspects of translation expertise from the perspective of expertise studies. In A. Riccardi (ed.), *Translation Studies: Perspectives on an Emerging Discipline* (pp. 150–171). Cambridge: Cambridge University Press.
- Si, S. X., and Cullen, J. B. (1998). Response categories and potential cultural bias: Effects of an explicit middle point in cross-cultural surveys. *The International Journal of Organization Analysis*, 6(3), 218–230.
- Smith, T. W. (1997). Improving cross-national survey research by measuring the intensity of response categories. *GSS Cross-National Report*, 17. Chicago, IL: National Opinion Research Center, University of Chicago.
- Smith, T. W. (2004). Developing and evaluating cross-national survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, and E. Singer (eds), *Methods for Testing and Evaluating*

- Survey Questionnaires* (pp. 431–452). Hoboken, NJ: Wiley.
- Tasaki, K. (2008). *Cross-Cultural Research Methods in Social Sciences*. Kyoto: Nakanishiya Press.
- Triandis, H. C. (1995). *Individualism & Collectivism*. Boulder, CO: Westview Press.
- van de Vijver, F. J. R., and Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13*, 21–29.
- van de Vijver, F.J.R., and Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto, and F.J.R. van de Vijver (eds), *Cross-Cultural Research Methods in Psychology* (pp. 17–45). New York: Cambridge University Press.
- Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., Gawlicki, M., et al. (2009). Multinational trials – recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force Report. *Value in Health, 12*, 430–440.
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., and Erikson, P. (2005). Principles of good practice for the translation and cultural translation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*, 94–104.