



Cognitive Testing of the Washington Group Translated Questions

Introduction

The question modules developed by the Washington Group (WG) – Short Set, Extended Set and Child Functioning Module – have all undergone thorough and rigorous cognitive and field-testing in multiple countries (often in multiple languages) prior to their finalization and adoption. In the absence of a gold standard for disability measurement, cognitive testing can provide the best evidence available that the questions are valid and reliable.

Cognitive testing¹ is a process of qualitative question evaluation that is used to understand and document how respondents comprehend and arrive at a response to a question in order to reduce response error and improve question response validity. This is known as the **question-response process**. It helps us understand the ways in which a question performs, including eliciting varied patterns of interpretation and dimensions of response error. This is important evidence in ensuring that questions capture the information intended by the survey designers.

The cognitive testing of a questionnaire is a vital step in assuring the collection of high-quality survey data. In multilingual studies, however, cognitive testing most often occurs exclusively in the language in which the instrument was designed (often English). The underlying assumption is that, if the question performs as intended in the **source language** then translators will be able to produce a version of the question in the **target language** that also ‘works’ well. This assumption is not always true.

The WG develops question modules in English and each module is cognitively tested in English prior to translation and cognitive testing in multiple languages, followed by field testing and eventual adoption. As it is not feasible to translate and test WG survey modules in all, or even most, languages it is then up to individual countries to take on these tasks.

The Washington Group Implementation Documents

cover the tools developed by the Washington Group on Disability Statistics (WG) to collect internationally comparable disability data on censuses and surveys. The documents address best practices in implementing the Short Set, Extended Set, Short Set – Enhanced, the WG / UNICEF Child Functioning Modules for children 2-4 and 5-17 years of age, and the WG / ILO LFS Disability Module, as well as other WG tools. Topics include translation, question specifications, analytic guidelines, programming code for analyses, the use of the tools for the purposes of disaggregation, and more.

To locate other WG Implementation Documents and more information, visit the Washington Group website:
<http://www.washingtongroup-disability.com/>.

¹ Explanations of the terminology in **red** may be found in the Glossary at the end of this document.

One of the main goals of the WG is to achieve comparability in disability statistics by having all countries use the same questions. Comparability can, however, be jeopardized by a number of problems related to how a translated question works. Among these are: problems arising from translation choices; cultural factors influencing interpretation (even in ‘perfectly’ translated questions); and lack of congruence in the **construct** of the question between the source and target versions.

The cognitive testing of translations will aid in assessing how questions work in the target language, what types of errors may arise from translations, and whether comparability has been achieved.

The goal of comparability is to have each question capture the same construct in *both* the source and target languages. Assessment of comparability is in fact embedded in each step of the preferred translation procedure, the TRAPD model (Translation, Review, Adjudication, Pretesting, and Documentation).

Translators must consider:

- Is “x” the best way to translate this term/phrase/question?
- Will respondents understand this translated question?
- Is the type of language appropriate for the intended audience?
- Are the nuances of the original question text maintained in the translation?
- Does the wording of the question feel natural?
- Are the response choices likely to be understood and used as in the original question?

Each of these considerations will help in producing a translated version that captures the same context and constructs as the original, source language version.

For a more complete discussion on the translation of the WG tools, refer to *Implementation Document #3: Translation of the Washington Group Tools*.

Translation-Related Errors in Questions

Having achieved a translated version of the questionnaire that is believed to preserve the integrity of the source language version, the next step is to cognitively test that translation. The goal of this exercise is to determine whether the translated version does in fact capture the same constructs as the original version, once it is put to practical use. As with the initial testing of newly developed questions, the main and primary goal of cognitively testing translated questions is to uncover patterns of interpretation and respondents’ answering strategies, and to evaluate the adequacy of the survey instrument in light of the original source language version.

Some potential problematic issues include:

- *Differences in social desirability bias*. Some questions can be perceived as sensitive whether due to linguistic or cultural reasons. For example, the WG self-care question was found to be problematic in some Arab-speaking countries, not due to issues of translation, but for cultural reasons. It was thought to be inappropriate for male interviewers to ask this question of female respondents. While not a translation issue, this kind of question problem would be identified when testing the translated question.

- *Differences in translation of the same language in different countries.* Translations of the same language used in different countries might differ. For example, in Spain the word “parent” was translated as “progenitor” while in Mexican Spanish the translation was “padre”. The former is less frequently used and may not be understood by respondents in Mexico or by Spanish-speaking immigrants to the USA, while the latter has a stronger male connotation and would be understood as “father” in other countries. Either choice has implications on how a question is interpreted and finally answered and different translations might be needed for different countries or different populations within the same country.
- *Differences in naturalness of language.* Translated questions, while semantically faithful, might not sound idiomatic and may be perceived as stilted, unduly formal or complicated. This may happen if the translator followed the structure or the expressions of the original version too closely, thus making it sound like a translation, and therefore making the question sound unnatural and thus changing how the question is perceived.
- *Differences in how response options are used.* A specific culture’s comfort or discomfort with selecting extreme response categories, with expressing disagreement to an unknown person, or with admitting not knowing the answer to a question can affect how questions are interpreted and answered, which may, in turn, influence data quality. The choice of words for answer categories can sometimes alleviate these cultural factors.

Ideally, cognitive testing of a translated question is meant to uncover patterns of interpretation of a question or text comparable to those in the original language, and both of these should match the interpretation that the researcher intended. Testing should help establish the translated question’s **construct validity**, that is, how well the question measures the concept(s) it is intended to measure. As will be outlined below, fewer resources are required for the testing survey translations than the testing of a newly developed set of questions.

Objectives of Cognitive Testing

1. Uncovering Translation Mistakes

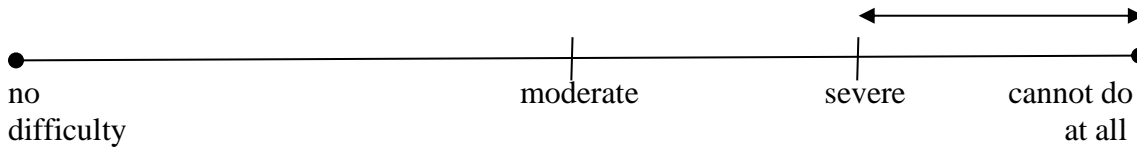
One of the main goals in the cognitive testing of a translation is to uncover mistakes in translation. These may include:

- expressions that are ‘unnatural’ in the target language, and
- additions, omissions and the use of wrong terms.

These can result in a change in the context of the question and responses that are out of scope.

Translations of questionnaires often focus on translating the question stem. While this is critical, of equal importance is the correct translation of the answer categories. This is particularly important for the Washington Group questions which attempt to capture a continuum of functioning. Improper translation of the answer categories will skew the resulting distribution. ‘No difficulty’ and ‘cannot do at all’ are the extremes of the distribution, anchoring its endpoints, and are clear concepts that are unambiguous and straightforward for translation. ‘Some difficulty’ and ‘a lot of difficulty’ are less definitive. For example, ‘some’ and ‘a lot’ should not be translated using words interpreted as ‘moderate’ and ‘severe’ respectively. The words used to translate ‘some’ and ‘a lot’ should divide the continuum into three relatively equal parts.

Visualizing a poor translation:



Using ‘severe’ as a translation for ‘a lot of difficulty’ makes this category closer to ‘cannot do at all’ so when using the recommended cut-off (a lot of difficulty or cannot do at all) many of those with difficulties of interest will be missed. This translation would also leave those with less than moderate – but some difficulty – in a position of potentially choosing ‘no difficulty’. The result is that the proportion of the population with a disability (in the 3rd or 4th categories) will be smaller than it would be had the terms been translated appropriately.

Visualizing a correct translation:



The distribution above creates four points equally distributed among the continuum that allows respondents to more easily discriminate among options. The recommended cut-off correctly identifies those with the intended level of difficulties. The translation of ‘some’ and ‘a lot’ should identify the population with the same functional status as that identified by the cognitively tested English version. Whether this has been successful can be determined by cognitively testing the translated questions.

2. Uncovering Regional Variation in Terms

Languages are sometimes spoken in different variations (dialects/colloquialisms) depending on regional differences. It is therefore important that translators have a close knowledge not only of the main target language, but also of these local variations.

Translations need to be appropriate for all regions in a country or different translations may be needed for different regions.

3. Uncovering Lack of Familiarity with Vocabulary, Idioms and Culturally Unknown or Irrelevant Concepts

Sometimes a translated question that *semantically* perfectly captures the source version is not understood by the target population because specific words or terms are used that are foreign to the target language population(s).

Translation and cultural appropriateness are closely related. If a question written in the source language uses words or terms that are foreign to the target language and culture, steps should be taken to address this situation.

4. Uncovering Cognitively Difficult Questions

When translated, some questions may be cognitively difficult to process due to either, the length of the question, syntactic complexity or grammatical structure. Consider, for example, the situation where a single word in English require a phrase in the target language – adding to the complexity of the question.

In general, it is best to keep questions (in both source and target languages) as simple and concise as possible.

Conducting the Cognitive Testing

While testing a translated version should not be very different, from a practical standpoint, from testing the source language version, a few practical aspects should be considered.

1. Cognitive Interviewers

The same qualities are needed in a cognitive interviewer in any language. However, a few important factors are worth remembering: All cognitive interviewers should have experience in qualitative research and be either native speakers of the target language or extremely fluent and familiar with the culture of the target population. Both of these considerations are important for conducting cognitive interviews as these interviews require the ability to **probe** responses in order to uncover inconsistencies responses and corresponding narratives.

2. Respondent Selection

Respondents should have the characteristics of the target population. For example, respondents with a range of difficulties and types of disabilities (including none) is important when testing WG questions. Additionally, when selecting participants for cognitive testing, special care must be put into including members from all relevant subgroups: e.g. gender, socio-economic status, educational status.

For the most part, testing will be based on the translated version of an already finalized and ‘approved’ module. Therefore, fewer respondents are needed than were required when the original source questions were developed and tested – but it would be wise to plan for no fewer than 20.

4. Protocol

The protocol for testing the translation can be based on the protocol used for the testing of the original, source version of the questionnaire. To that end, the WG will make available the original testing protocol for the original version of any of the WG modules that are being translated and tested. Note that the ‘protocol’ is largely based on the questions themselves and a semi-structured **comparative methodology**.

5. Implementation and Process

The same aspects of question interpretation and processing will be considered in all translated versions, and it will be possible to examine whether interpretation patterns for questions and responses are the same as the source across versions (source and translation).

6. Analysis

It is important that each translated version be tested against the researchers' intended meaning of each question. This can be accomplished through textual assessment – a comparison of the narratives that are the output of the cognitive test against the 'accepted' source version of the question. When respondents interpret a question in a way that is different from the designer's original intent, this signals the presence of a specification error – where the translation of the question resulted in an unintended shift of meaning. For example, translating the question "Do you have difficulty remembering" as "Do you have difficulty memorizing" changes the intent of the questions. *Remembering* and *memorizing* are two distinct and quite different activities. The translation question "Do you have difficulty *memorizing*" will elicit responses that reflect these different activities and will not be comparable. It will become important to determine whether the source of the problem is an error in translation [choice of word or phrase], or perhaps related to culturally specific issues.

The process of data analysis examines responses within interviews, across interviews (by question) and across subgroups with the goal of identifying thematic patterns in question interpretations and response errors. Six incremental steps illustrate this process:

1. Through the interview process, obtain the ways in which a respondent interpreted and formulated answers to the survey questions.
2. Synthesize the interview text into summaries, detailing how respondents formulated their answers, including events or experiences considered as well as any difficulties answering the questions.
3. Compare the summaries across respondents to identify common themes and to develop a **thematic schema** that details phenomena captured.
4. Compare those themes across subgroups to identify ways in which different groups may process questions differently depending on their differing experiences and socio-cultural backgrounds.
5. Compare themes to those that characterize the source question and the designer's intent.
6. Draw conclusions based on the thematic schema that depict how each question performs as well as providing explanation for the performance.

Summary

Cognitive testing provides survey designers and methodologists with evidence that respondents from different cultural and socio-economic backgrounds are interpreting and providing responses in a consistent manner and in line with the goals and intentions of the question design. Not only is consistency necessary among same language groups across an array of socio-cultural dimensions within a country – but consistency is also required across languages within the same country or across countries. As in the case of the latter, the situation becomes more complicated when surveys are designed to be implemented cross-nationally; the questions designed for eventual translation into multiple languages.

This note will assist survey methodologists intent on implementing a survey designed in a source language in one or more target languages in their own country and their own language(s).

Additional Resources

For a more comprehensive discussion of cognitive interviewing methodology readers are referred to the reference below; and for an overview of cognitive interviewing as it pertains to the testing of the WG questions, readers are referred to an on-line blog: <https://www.washingtongroup-disability.com/wg-blog/cognitive-interviewing-for-the-washington-group-72/>, and the following link on the Methodology page: <https://www.washingtongroup-disability.com/fileadmin/uploads/wg/Documents/ResultsoftheTestingoftheESCAP-WGQuestionSetonDisability.pdf>.

Schoua-Glusberg A & Villar A. Assessing Translated Questions via Cognitive Interviewing. In: Cognitive Interviewing Methodology. K Miller, S Willson, V Chepp, J-L Padilla (Eds.). John Wiley & Sons, Inc., New Jersey, 2014.

Glossary

Cognitive testing: A process of qualitative question evaluation that is used to understand and document how respondents comprehend and arrive at a response to a question in order to reduce response error and improve question response validity.

Comparative methodology: An inductive and iterative process in which the analyst continuously moves back and forth from raw text data, themes and emerging conceptual claims. In cognitive interviewing, the analyst compares data across survey questions, both within a single respondent interview and among all interviews, in order to arrive at a complete understanding of the question-response process and the construct measured by the question.

Construct: The actual concept (e.g. hearing – as opposed to listening) being measured by the question.

Construct validity: the extent to which a survey question captures what it is intended to measure.

Probe: A cognitive interviewing technique in which an interviewer asks follow-up questions of the respondent in order to understand the question-response process and the construct being measured by the question.

Question-response process: The cognitive and social processes in which respondents engage when answering survey questions. Cognitive interviewers seek to replicate this experience through the cognitive interview in order to understand the processes by which respondents answer survey questions.

Source language: the original language of the survey.

Target language: the language into which the survey has been translated.

Thematic schema: Patterns of question interpretation that are analytically identified to explain the question-response process and the construct being.