SURVEYS · TECHNOLOGY AND THE FUTURE

Multilingual and Cross-Cultural Survey Translation

Surveys in more than one language must ask the same questions in every language, which is harder than it looks. This article explains the goals of survey translation, the team approach known as TRAPD, why back translation is not enough, how to adapt questions across cultures and how to test the result.

ModQu research article · 1,602 words · 17 references (APA 7th edition)

Many surveys are conducted in more than one language. International studies compare dozens of countries; national surveys include speakers of minority and immigrant languages; companies survey customers and employees in many markets. In each case, the results can be compared only if respondents in every language answer questions that mean the same thing.

Achieving this is harder than it looks. Words rarely have exact equivalents, concepts carry different associations in different cultures and response scales may be interpreted differently. A translation can be grammatically correct yet ask a different question. This article explains what good survey translation aims to achieve, how it is organised in practice and how it can be tested.

What translation needs to achieve

The aim of survey translation is not word-for-word correspondence but **functional equivalence**: the translated question should measure the same concept, in the same way, as the original (Harkness, 2003). This involves several kinds of equivalence. The words should carry the same meaning. The question should have the same purpose and be understood in the same way. Response options should divide the range of possible answers in the same places. And the question should be equally natural and acceptable in each language.

Some differences cannot be solved by translation alone. A question about "the government" may refer to different levels of government in different countries. Agreement scales may be used differently by respondents from different cultures, regardless of translation (Johnson et al., 2005; Baumgartner & Steenkamp, 2001). Researchers therefore distinguish **translation**, which carries meaning from one language to another, from **adaptation**, which changes the content of a question deliberately to suit a different context, such as replacing a currency or an example that does not apply.

Common translation problems

Problems found in translated questionnaires fall into a few recurring types. Some are **lexical**: a word has no exact equivalent, or the nearest equivalent is more formal, more emotional or narrower in meaning. Some are **grammatical**: languages differ in whether they mark gender, formality or number, so a neutral English question may require choices in

another language, such as whether to address respondents formally or informally. Some are **cultural**: an example, activity or institution named in the question does not exist or has a different meaning in another country. And some concern **layout**: translated text may be much longer than the source, which affects how questions and response options fit on a screen.

In the terms used by van de Vijver and Leung (1997), these issues are forms of **bias**, and they distinguished problems with the concept itself, with the method of measurement and with particular items. Translation affects mainly item bias, but concept and method problems can appear at the translation stage and should be reported to the questionnaire designers rather than solved silently by translators.

The team approach: TRAPD

Best practice in cross-national surveys is a team-based approach summarised by the letters **TRAPD**: Translation, Review, Adjudication, Pretesting and Documentation (Harkness, 2003; Survey Research Center, 2016). Figure 1 shows the steps.

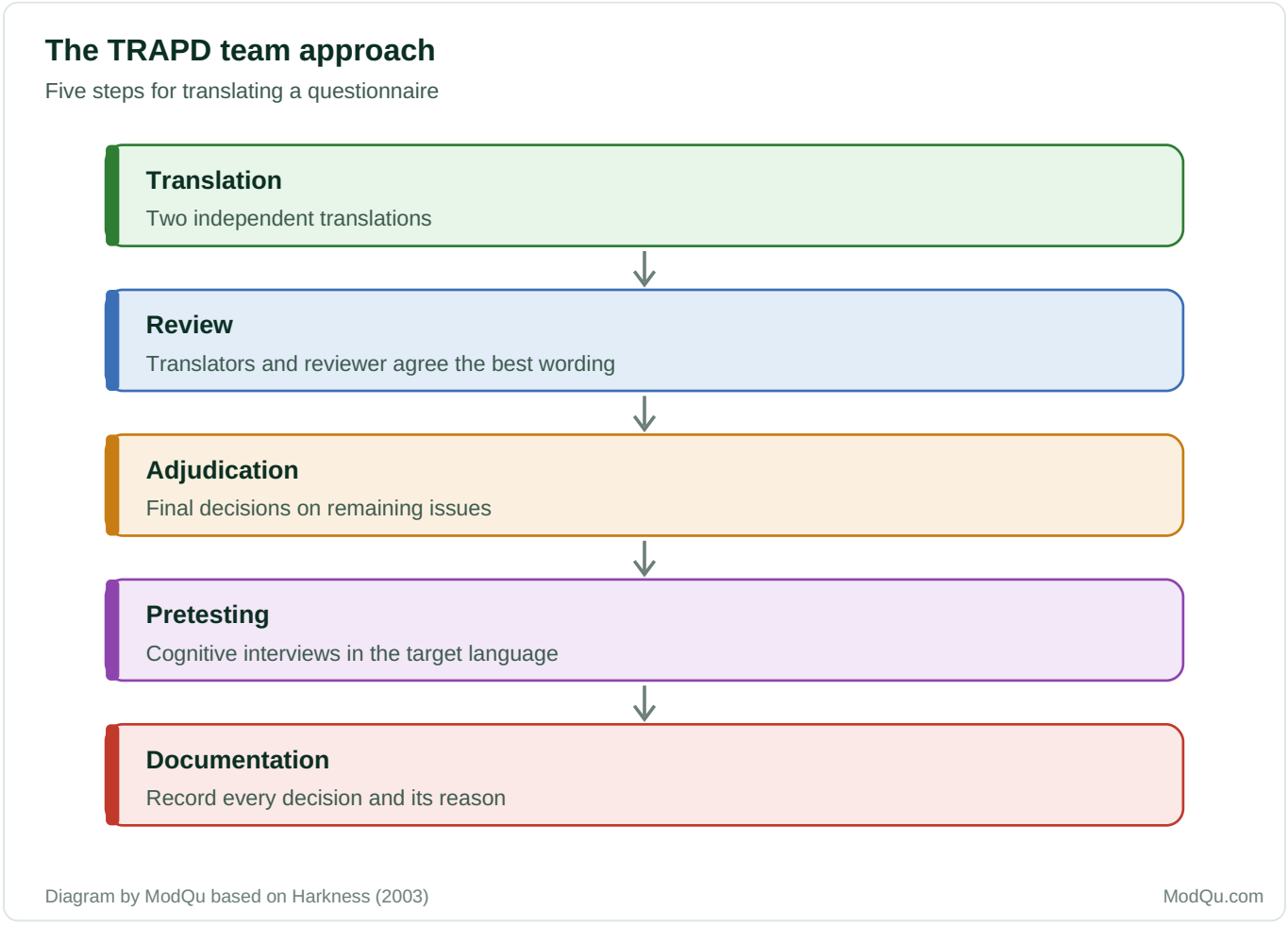


Figure 1. The TRAPD team approach to survey translation. Diagram by ModQu.

Translation. Two translators independently produce full translations. Parallel translations reveal different ways of expressing the same question and highlight difficult passages.

Review. The translators meet with a reviewer, who has expertise in survey methods and the subject matter, to compare the versions and agree on the best wording for each question.

Adjudication. An adjudicator, often the national survey lead, makes final decisions where disagreements remain and checks consistency with the source questionnaire.

Pretesting. The translated questionnaire is tested with respondents who speak the target language, for example in cognitive interviews, to find misunderstandings.

Documentation. Every decision is recorded, including problems, alternatives considered and reasons for choices, so that analysts can understand any differences later.

This approach is used in major international surveys and is recommended in guidance for cross-cultural research (Harkness et al., 2010; Survey Research Center, 2016). Similar team-based processes are recommended for translating health questionnaires (Wild et al., 2005).

Why back translation is not enough

A traditional method of checking a translation is **back translation**: an independent translator translates the target version back into the source language, and the researcher compares the result with the original. Brislin (1970) described this method, which became widely used because it allows a researcher who does not speak the target language to check the work.

Back translation has serious limits. A poor translation can produce a good back translation if the translator keeps closely to the source wording, producing an awkward or unnatural target version. Conversely, a good, natural translation may back-translate imperfectly. Back translation focuses attention on words rather than on whether target-language respondents will understand the question as intended. Behr (2017) analysed its use and concluded that it is a weak test of translation quality and that team review and pretesting are more informative. Back translation can still play a supporting role, but it should not be the main quality check.

Designing questions for translation

Translation is easier when the source questionnaire is written with it in mind. The source text should use plain language and avoid idioms, metaphors and culturally specific references. Each concept should be explained clearly so that translators know what is meant. Many international projects provide **translation notes** that explain the intended meaning of key terms and the purpose of each question.

Some projects test a draft source questionnaire for **translatability** before finalising it, asking translators from several languages to identify words and constructions that will be hard to translate. Changes at this stage are far cheaper than fixing problems in many languages later (Smith, 2004).

Response scales need particular care. Verbal labels such as "somewhat" or "quite" may not have exact equivalents, and the distances between labels may differ between languages. Using fully labelled scales with carefully chosen equivalents, and documenting the choices, reduces these problems (Harkness et al., 2010).

Testing translated questionnaires

Translated questionnaires should be pretested in each language, not only in the source language. **Cross-cultural cognitive interviewing** asks respondents in each language to explain how they understood questions and arrived at their answers. Willis and Miller (2011) describe how such interviews can show whether questions are understood in comparable ways and also reveal genuine cultural differences that affect interpretation.

After data collection, statistical tests of **measurement invariance** examine whether questions function equivalently across language groups (Davidov et al., 2014; Steenkamp & Baumgartner, 1998). If a translated item behaves differently from the others, the translation is one possible cause to investigate.

Figure 2 illustrates how a translation problem can appear in the data.

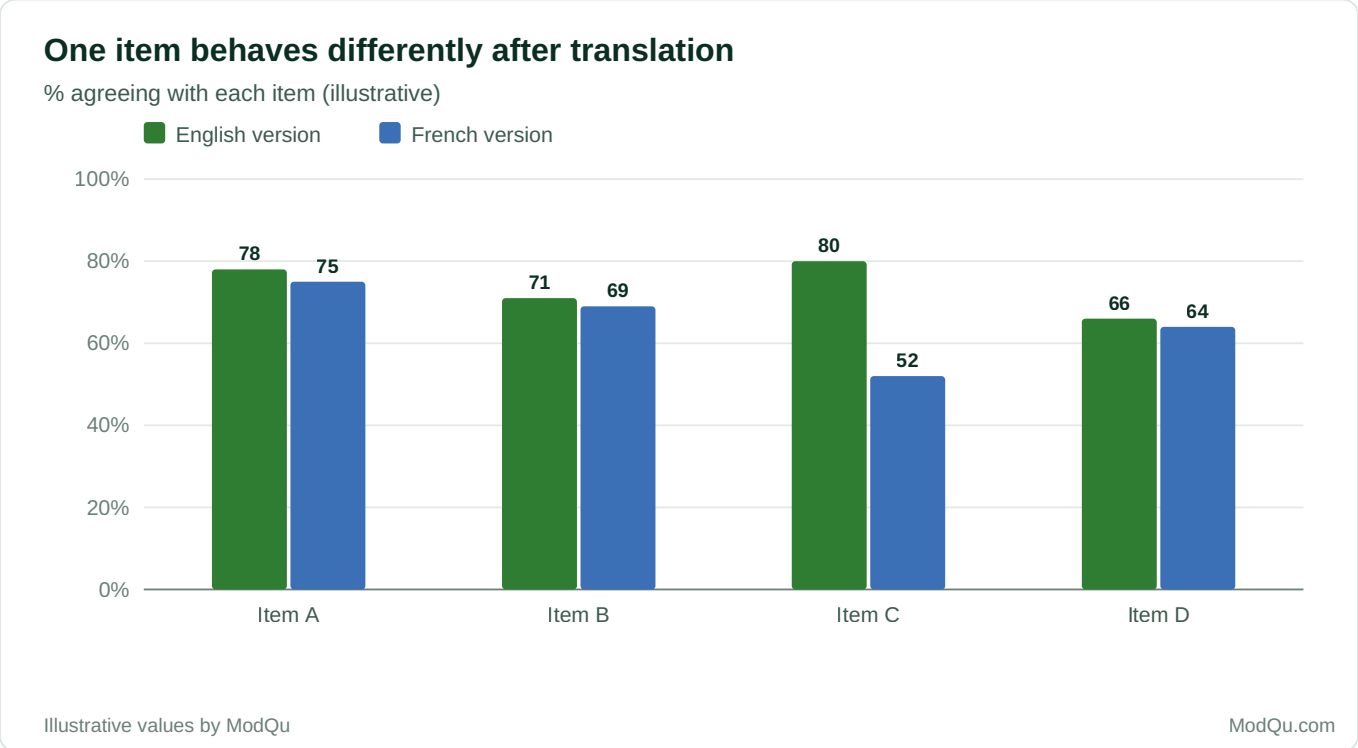


Figure 2. Agreement with four items in two language versions in a hypothetical survey; item C stands out (illustrative values). Chart by ModQu.

Time, cost and resources

A team approach costs more than a single translation. It requires at least two translators, a reviewer and an adjudicator for each language, plus time for meetings, pretesting and documentation. For small projects, organisations sometimes reduce the process, for example by using one translator and an experienced reviewer. Any reduction should be a deliberate choice, recorded in the documentation, rather than a default.

The cost of a poor translation is usually higher. Errors found after fieldwork cannot be corrected, and data from a language version with a faulty question may have to be excluded from comparisons. Health researchers adopted structured translation processes for this reason, after finding that questionnaires translated without them often failed to work equivalently (Beaton et al., 2000). Methods for predicting the measurement quality of questions can also be applied to each language version to check whether translation has changed the quality of an item (Saris & Gallhofer, 2014). Contributors to the field's main handbooks consistently argue that translation should be planned and budgeted from the start of a cross-cultural project, not added at the end (Harkness et al., 2003).

Translation within one country

Translation is not only an issue for international studies. National surveys often need versions in several languages to include immigrant and minority communities. Without them, people who do not speak the main language well are excluded or answer with poor understanding, which biases results. Guidelines from national statistical agencies recommend the same team-based approach for such translations, with attention to the regional varieties of a language spoken by the target population (Pan & de la Puente, 2005).

Machine translation

Machine translation tools now produce fluent drafts in seconds. They can be useful as one of the initial translations in a team process, but they should be reviewed by human translators who understand the survey's concepts. The use of machine translation in multilingual surveys is discussed in a separate article in this series.

An example

A retailer runs a customer satisfaction survey in English and French. One item asks whether staff were "helpful". In the French version, a single translator chose a word closer to "obliging". In the results, French-speaking customers rate staff much lower on this item than on related items, while English-speaking customers do not show this pattern. A review by two new translators and a few cognitive interviews show that the chosen word suggests servility to some respondents. The retailer revises the wording, documents the change and notes in its report that results on this item are not directly comparable with the previous year (illustrative example).

Conclusion

Good survey translation aims for equivalent meaning, not identical words. A team approach with parallel translations, review, adjudication, pretesting and full documentation produces better questionnaires than a single translator checked by back translation. Writing source questions with translation in mind, testing in every language and checking for equivalence in the data together allow surveys to generate knowledge that is comparable across languages and cultures.

References

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology, 20*(6), 573–584. https://doi.org/10.1080/13645579.2016.1252188

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*, 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Wiley.

Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. Ph., Pennell, B.-E., & Smith, T. W. (Eds.). (2010). *Survey methods in multinational, multiregional, and multicultural contexts*. Wiley. https://doi.org/10.1002/9780470609927

Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. Ph. (Eds.). (2003). *Cross-cultural survey methods*. Wiley.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264–277. https://doi.org/10.1177/0022022104272905

Pan, Y., & de la Puente, M. (2005). *Census Bureau guideline for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed* (Research Report Series, Survey Methodology No. 2005-06). U.S. Census Bureau.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research* (2nd ed.). Wiley. https://doi.org/10.1002/9781118634646

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, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 431–452). Wiley.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–90. https://doi.org/10.1086/209528

Survey Research Center. (2016). *Guidelines for best practice in cross-cultural surveys*. Survey Research Center, Institute for Social Research, University of Michigan. https://ccsg.isr.umich.edu/

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. SAGE.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

Willis, G. B., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods, 23*(4), 331–341. https://doi.org/10.1177/1525822X11416092