

The role of translators:

Adaptation of mental health tests beyond clinical research

Karol Tapia de Moya

English-to-Spanish translator, Malaga, Spain

doi: 10.56012/jfuz3231

Correspondence to:

Karol Tapia de Moya

psikapia@gmail.com

Abstract

This article explores the role of translation and translators within the process of test adaptation. With similarities to the linguistic validation process in clinical research, test adaptation is a complex multistep and iterative process in which linguistic and cultural nuances are an essential aspect. In the context of mental health tests, the impact on individuals can be significant, with far-reaching consequences across various life domains. In this specialised field, translators play a crucial role, contributing their unique expertise and insights on language and culture to help shape the test adaptation.

One of the most exciting fields for translators involves translating instruments, such as questionnaires, to measure mental health aspects in the context of clinical trials. However, this is not the only field where mental health testing is relevant.

This article describes the role of translators and translations in an assessment industry that extends beyond the pharmaceutical and medical device sector. Its applications span behavioural and social science research, education, clinical practice, forensics, organisational psychology, and various other domains.

When we venture outside the field of clinical research, it's impossible not to notice that the terms "clinical outcome assessments" and "linguistic validation" are replaced with the terms "tests" and "test adaptation". Knowledge tests, questionnaires, scales, and inventories are all

examples of tests: tools with items, standardised procedures, and established evidence of reliability/precision, validity, and fairness for measuring behaviour and mental attributes.¹⁻³ Test adaptation, in simple terms, refers to "moving a test from one language and culture to another".²

You might have spotted that the term "translation" is not used. This is because experts in the field prefer "test adaptation" to prioritise intent over content (adaptation) rather than content over intent (what they consider translation to be limited to).⁴ For them, the aims of adaptation go beyond aesthetics (how well it

reads) and readability (how easy it is to understand). While these aspects are important and desirable, the test adaptation process also evaluates appropriateness in terms of psychometric characteristics.

In general, the test adaptation process aims to produce a test in the target language and culture, measuring the same construct as the original test, maximising fairness, minimising bias, ensuring sufficient reliability/precision, and maintaining validity for the proposed uses.⁵ With this in mind, the process includes everything from determining whether an adaptation is needed or even



possible, to choosing translators and conducting validation studies.

What can go wrong?

Let's pause a moment here to paint an image of how this process can fail spectacularly. The best example is seeing what happens if someone administers a test that is not adapted for a certain country/language:

Once, when I was practising as a psychologist in Colombia, my colleagues and I were instructed to administer a language test, specifically a naming test, as part of a routine neuropsychological assessment. This particular test simply involved showing patients images they had to identify by name, and this is probably why the people in charge thought it would be acceptable to administer the test using the US English version in our Colombian context.

The result? Many patients were unable to recognise the images. "Is that a rat?" was a

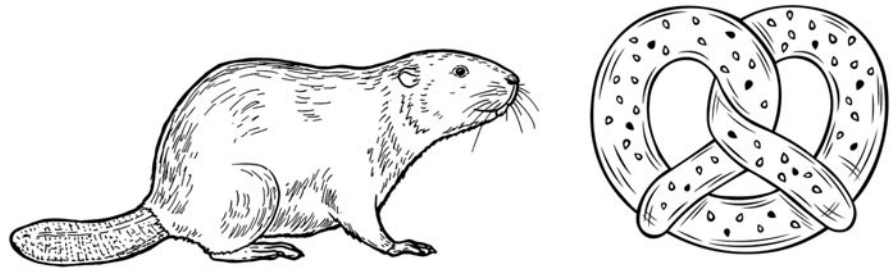


Figure 1. Images of a beaver and a pretzel similar to the ones in the naming test

common answer for the beaver and, better still, some identified the pretzel as a snake (Figure 1)! Granted, some people were able to identify the images – these were affluent patients with previous exposure to US culture. We could have mistakenly concluded that there was a widespread epidemic of

Let's pause a moment here to paint an image of how this process can fail spectacularly. The best example is seeing what happens if someone administers a test that is not adapted for a certain country/language.

aphasic patients in the lower-income parts of our city, but what was actually happening was that these items were not truly measuring naming ability; rather, they were reflecting other aspects like socioeconomic status. Any interpretations taken from the scores could not support any kind of diagnosis or recommendation at all.

How to prevent testing disasters

If we want to prevent a testing disaster like in the example above, there's a process that should be implemented and followed.

First, a need should be identified and defined. A test is adapted to a different language or culture because it needs to be available for an applied use (e.g., clinicians need it to assess people with a specific health condition in another country) or to support small- or large-scale research (e.g., international social/educational research).⁴ In the case of our example, it was needed as part of a routine neuropsychological assessment.

Second, a vital step involves deciding whether an adaptation is the best way to proceed. If a test was originally validated in English for English-speaking test takers, accompanied by English instructions, and administered by English-speaking test administrators in a specific culture, its interpretations are valid only within this particular context. Acknowledging this is essential when deciding whether an adaptation is appropriate.

Continuing with our example, one alternative could have been to just create a naming test from scratch and not bother adapting it. But there are many compelling reasons to adapt a test:



1. Adaptation makes it easier to compare different cultural and language groups;
2. Adaptation can be a cost-effective and time-saving approach compared with developing an instrument from scratch);
3. Adaptation provides an invaluable option when the target country lacks psychometric expertise in the specific subject matter to develop a test locally; and
4. Adaptation expands and leverages the reputation of an established instrument created in another language.

Third, if it's decided that adaptation is the way to go, how should we go about it? What if my colleagues had just decided to search for some alternative images, "adapted" the test themselves to suit their context, and started applying it?

Unlike in research, where there's more flexibility with reliability/precision and validity criteria,³ in an applied context, this is not an option. Simply translating the test does not automatically make it comparable to the original, and this is not something that can be left as an assumption – it needs to be proved.

Another thing to consider is the characteristics of the test we want to adapt, as all tests are not made equal. To illustrate this, let's take a look at some real examples of mental health tests.

On the one hand, the Beck Depression Inventory-II⁶ is a short and straightforward self-report inventory that measures attitudes and symptoms of depression, so would fit with what

If we want to prevent testing disasters, there's a process that should be implemented and followed.

people generally consider to be a mental health test. On the other hand, the NEPSY-II⁷ is a lengthy test made up of six domains, each assessing the neuropsychological status in children through diverse tasks, such as imitating hand positions, constructing models with blocks, and looking at faces in photographs. On a completely

different dimension, the KIDSCREEN questionnaire⁸ measures health-related quality of life in children (like psychological well-being, moods and emotions, and bullying) as part of a European large-scale research project.

These examples paint a good picture of the diversity of the mental health testing landscape: a short questionnaire that a clinician will use as a

Table 1. Linguistic/cultural components of test content

Linguistic/cultural components	Examples of interventions
Currencies	Changing the currency, changing the amount itself, changing denominations (e.g., understanding of dimes, cents, or pennies is heavily influenced by culture and a straightforward currency conversion will not always work).
Measures	Changing measurement system (imperial, metric, US customary system), converting measures, rounding figures up/down (e.g., replacing teaspoons with millilitres or grams in a recipe-related item).
Sport	Referencing sports that are familiar to the test takers (e.g., asking about sumo rules might not be appropriate for European test takers).
Geography	Changing geographical references to more culturally relevant ones (e.g., the number of continents varies depending on the culture and might range from 4 to 7).
Numbers	Considering how numbers are written, how they are verbalised, and their complexity in different languages (e.g., using commas or periods to separate numbers: 20.000 could be a twenty with three decimal zeroes or could be read as twenty thousand).
Art	Using artistic references relevant to the specific culture (e.g., Fernando Botero is a renowned Colombian painter who might not be as well-known in other countries).
History	Changing historical references to more culturally relevant ones (e.g., asking about the Mexican Revolution is an easier question for students in Mexico compared with students in Nigeria).
Daily life	Many items might require significant adaptation or complete replacement as they reference specific education systems, food, animals, activities, etc. (e.g., with regard to opening hours for businesses and shops – in Colombia they typically open at 7 am, whereas in Spain they tend to open at 9 am).

Adapted from Iliescu D. *Adapting Tests in Linguistic and Cultural Situations (Educational and Psychological Testing in a Global Context)*. Cambridge: Cambridge University Press; 2017.

brief screening tool in their practice is not the same as a test with a clear diagnostic aim or a test that is going to be used for international public health research. Aside from the inherent characteristics of the test, there are other variables that play a crucial role in shaping our approach to the adaptation process, especially from a translator's perspective. These include the target population (e.g., children, young adults with intellectual disability, adults looking for a job), the construct being measured, the geographical context of its application, and the mode of administration (e.g., computer-based or paper and pencil).

Fourth, at this point, you might be wondering who determines the correct approach or if there are objective criteria for overseeing this process. The answer is; there are several sets of standards designed for this purpose:

- The *Standards for Educational and Psychological Testing*¹ are considered the gold standard, and they provide definitions for essential concepts such as validity, fairness, reliability/precision, and ethical aspects. This framework also holds relevance in the clinical research context, as evidenced by the *Food and Drug Administration Draft Guidance for Clinical Outcome Assessments*,¹⁰ which references the standards as one of its foundations.
- The International Test Commission (ITC) has established a series of guidelines and checklists for the test adaptation process.² In collaboration with the Association of Test Publishers, the ITC has also produced translation/adaptation guidelines for technology-based assessments.⁵
- The *Guidelines for Best Practice in Cross-Cultural Surveys*¹¹ provide guidance on the questionnaire design and translation/adaptation process for multinational, multicultural, and multiregional surveys. Additionally, Dept, Ferrari, and Halleux¹² explore the various methodologies employed in the translation of large-scale assessments, such as Programme for International Student Assessment (PISA) surveys.

Finally, to check whether the translation ticks the boxes and meets the objectives of the test adaptation process, there are quality control procedures. Outside the clinical trial context, the

Tests that are not equivalent across languages and cultures can lead to incorrect conclusions and interpretations, with far-reaching consequences on the individual lives of test takers.

approaches implemented go beyond the widely used back-translation: the designs tend to have a more hybrid approach and involve multiple iterations with various steps and checks. This marks the end of the translation phase, but the adaptation process continues with various other activities such as statistical analysis, norming, and publishing, which fall outside the translator's scope.

The role of translators in test adaptation

Test adaptation is susceptible to three distinct sources of error or

invalidity:

1. cultural/language differences,
2. technical issues, designs, and methods, and
3. interpretation of results.⁹ In terms of their hands-on work, translators are instrumental in mitigating these errors, especially those that pertain to the second category. Within this category, three aspects are of special importance to translators: the test content, the translators, and the process of translation.

The first aspect – test content – refers to the themes, wording, and format of the items, tasks, and questions, and also encompasses administering and scoring.¹ Our initial example with the beaver and the pretzel is a good illustration of content (stimuli) that needs to be adapted to the target culture and language. Table 1 shows other cultural and linguistic components that might require attention.

The other two aspects – translators and the translation process – pertain to the potential sources of bias and errors that could be introduced by translators. To mitigate this risk, translators must possess expert knowledge of their source and target language and culture, as well as a deep understanding of the content of the test and the principles of testing,² as seemingly straightforward decisions can be more complex than they appear at first glance.

For example, the simple act of translating an item that says “tree” can have unforeseen consequences because using the

same word in another language does not guarantee equivalent familiarity, frequency, difficulty (e.g., tree goes from a one-syllable word in English to a two-syllable word in Spanish), or even location of the item in the target language.

Whether “small” details like this are relevant or not depends on the intent behind the item, the underlying construct, and the instructions given to translators. The process of test translation requires numerous nuanced decisions like these. Translators should recognise, address, and communicate these sometimes small but significant details, rather than assuming they are insignificant, failing to document any changes implemented, or making unilateral decisions without consulting the team.

The real impact: Who bears the consequences?

Now that we've looked at tests, test adaptations, and how translators fit into this process, there's one final yet crucial aspect to consider – the context. In research settings, such as clinical trials, a poor-quality translation of a questionnaire can affect the study's ability to assess the efficacy of an intervention or the differences between control and intervention groups, but the results are typically analysed at an aggregate level, not an individual level. Conversely, in the applied mental health field, it's important to consider that poorly adapted tests can significantly affect individuals who are often in vulnerable positions, making the impact all the more real and personal for them.

Tests that are not equivalent across languages and cultures can lead to incorrect conclusions and interpretations, with far-reaching consequences on the individual lives of test takers. Many life-altering decisions, such as those granting access to accommodations, benefits, diagnoses, treatment, economic aid, employment, scholarships, and adoption, often depend on the results of a test.

For translators, we should never lose sight of this as we work.

Conclusion

Tests are highly specialised tools, and mental health tests in particular have a wide range of significant, real-world applications.

For this reason, when adapting a test, it's imperative to have a robust procedure in place

Many life-altering decisions, such as those granting access to accommodations, benefits, diagnoses, treatment, economic aid, employment, scholarships, and adoption, often depend on the results of a test.

and to comply with established guidelines.

Translators represent just one part of test adaptation, but it's an integral part. Our expertise enables us to pinpoint critical language and cultural issues that can decisively inform certain aspects of the process.

The pursuit of validity, reliability/precision, and fairness is an ongoing endeavour, and even after implementing all measures and linguistic quality control procedures, there's no guarantee that equivalence will be achieved. For translators, our primary contribution to the adaptation process lies in minimising the introduction of errors, thoroughly documenting our decisions, and reporting potential issues to other professionals, including subject matter experts and psychometricians, to aid them in their decision-making.

Disclosures and conflicts of interest

The author declares no conflicts of interest.

References

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington D.C.: American Educational Research Association; 2014. Available from: <https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf>.
2. International Test Commission. ITC guidelines for translating and adapting tests. 2nd ed.; 2017. Available from: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf.
3. International Test Commission. ITC statement on the use of tests and other assessment instruments for research purposes; 2014. Available from: https://www.intestcom.org/files/statement_using_tests_for_research.pdf.
4. Iliescu D. Adapting tests in linguistic and cultural situations (Educational and psychological testing in a global context). Cambridge: Cambridge University Press; 2017. doi:10.1017/9781316273203,



5. International Test Commission, Association of Test Publishers. Guidelines for technology-based assessment; 2022. Available from: <https://www.intestcom.org/upload/media-library/guidelines-for-technology-based-assessment-v20221108-16684036687NAG8.pdf>.
6. Beck AT, Steer RA, Ball R, et al. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess*. 1996;67(3):588–97. doi:10.1207/s15327752jpa6703_13.
7. Korkman M, Kirk U, Kemp S. NEPSY—Second Edition (NEPSY-II). San Antonio (TX): Harcourt Assessment; 2007. Available from: <https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1963582>.
8. Ravens-Sieberer U, Gosch A, Rajmil L, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: Psychometric results from a cross-cultural

survey in 13 European countries.

Value Health. 2008;11(4):645–58.

doi:10.1111/j.1524-4733.2007.00291.x.

9. Hambleton RK. Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In: Hambleton RK, Merenda PF, Spielberger CD, editors. *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah (NJ): Lawrence Erlbaum Associates, Inc; 2004. doi:10.4324/9781410611758.
10. U.S. Department of Health and Human Services, FDA, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, and Center for Devices and Radiological Health. Patient-focused drug development: selecting, developing, or modifying fit-for purpose clinical outcome assessments [draft guidance]; 2022. Available from: <https://www.fda.gov/media/159500/download>.
11. Survey Research Center, Institute for Social Research, University of Michigan. Guidelines for best practice in cross-cultural surveys. 3rd ed.; 2011. Available from: https://ccsg.isr.umich.edu/wp-content/uploads/2020/02/CCSG_Guidelines_Archive_2010_Version.pdf.
12. Dept S, Ferrari A, Halleux B. Translation and cultural appropriateness of survey material in large-scale assessments. In: Lietz P, Cresswell JC, Rust K, Adams R, editors. *Implementation of large-scale education assessments*. Newark: Jhon Wiley & Sons; 2017. doi:10.1002/9781118762462.ch6.



Author information

Karol Tapia de Moya, MSc, is an English-to-Spanish medical translator and a licensed psychologist specialising in clinical research and psychology-related topics. Aside from working in two languages, she has experience practising as a psychologist in different cultures within her native language. You can learn more about her at PsyTranslations.com.