



INTERNATIONAL TEST COMMISSION

International Test Commission Guidelines for Translating and Adapting Tests

Version 2010

© 2010, International Test Commission.

Please cite as:

International Test Commission (2010). International Test Commission Guidelines for Translating and Adapting Tests. [<http://www.intestcom.org>]

International Test Commission Guidelines for Translating and Adapting Tests

In 1992 the International Test Commission (ITC) began a project to prepare guidelines for translating and adapting tests and psychological instruments, and establishing score equivalence across language and/or cultural groups. Several organizations assisted the ITC in preparing the guidelines: European Association of Psychological Assessment, European Test Publishers Group, International Association for Cross-Cultural Psychology, International Association of Applied Psychology, International Association for the Evaluation of Educational Achievement, International Language Testing Association and International Union of Psychological Science. A committee of 12 representatives from these organizations worked for several years to prepare 22 guidelines, and later these guidelines were field-tested (see, for example, Hambleton, 2001; Hambleton, Merenda, & Spielberger, 2005; Hambleton, Yu, & Slater, 1999; Tanzer & Sim, 1999). Later, these guidelines were approved by the ITC for distribution to national psychological societies, test publishers, and researchers. The guidelines, organized into four categories, appear below:

Context

- C.1 Effects of cultural differences which are not relevant or important to the main purposes of the study should be minimized to the extent possible.
- C.2 The amount of overlap in the construct measured by the test or instrument in the populations of interest should be assessed.

Test Development and Adaptation

- D.1 Test developers/publishers should insure that the adaptation process takes full account of linguistic and cultural differences among the populations for whom adapted versions of the test or instrument are intended.
- D.2 Test developers/publishers should provide evidence that the language use in the directions, rubrics, and items themselves as well as in the handbook are appropriate for all cultural and language populations for whom the test or instrument is intended.
- D.3 Test developers/publishers should provide evidence that the choice of testing techniques, item formats, test conventions, and procedures are familiar to all intended populations.
- D.4 Test developers/publishers should provide evidence that item content and stimulus materials are familiar to all intended populations.
- D.5 Test developers/publishers should implement systematic judgmental evidence, both linguistic and psychological, to improve the accuracy of the adaptation process and compile evidence on the equivalence of all language versions.
- D.6 Test developers/publishers should ensure that the data collection design permits the use of appropriate statistical techniques to establish item equivalence between the different language versions of the test or instrument.
- D.7 Test developers/publishers should apply appropriate statistical techniques to (1) establish the equivalence of the different versions of the test or instrument, and (2)

- identify problematic components or aspects of the test or instrument which may be inadequate to one or more of the intended populations.
- D.8 Test developers/publishers should provide information on the evaluation of validity in all target populations for whom the adapted versions are intended.
 - D.9 Test developers/publishers should provide statistical evidence of the equivalence of questions for all intended populations.
 - D.10 Non-equivalent questions between versions intended for different populations should not be used in preparing a common scale or in comparing these populations. However, they may be useful in enhancing content validity of scores reported for each population separately.

Administration

- A.1 Test developers and administrators should try to anticipate the types of problems that can be expected, and take appropriate actions to remedy these problems through the preparation of appropriate materials and instructions.
- A.2 Test administrators should be sensitive to a number of factors related to the stimulus materials, administration procedures, and response modes that can moderate the validity of the inferences drawn from the scores.
- A.3 Those aspects of the environment that influence the administration of a test or instrument should be made as similar as possible across populations of interest.
- A.4 Test administration instructions should be in the source and target languages to minimize the influence of unwanted sources of variation across populations.
- A.5 The test manual should specify all aspects of the administration that require scrutiny in a new cultural context.
- A.6 The administrator should be unobtrusive and the administrator-examinee interaction should be minimized. Explicit rules that are described in the manual for administration should be followed.

Documentation/Score Interpretations

- I.1 When a test or instrument is adapted for use in another population, documentation of the changes should be provided, along with evidence of the equivalence.
- I.2 Score differences among samples of populations administered the test or instrument should not be taken at face value. The researcher has the responsibility to substantiate the differences with other empirical evidence.
- I.3 Comparisons across populations can only be made at the level of invariance that has been established for the scale on which scores are reported.
- I.4 The test developer should provide specific information on the ways in which the socio-cultural and ecological contexts of the populations might affect performance, and should suggest procedures to account for these effects in the interpretation of results.

The guidelines and suggestions for implementing them can be found in Hambleton, Merenda, and Spielberger (2005), Muniz and Hambleton (1997), van de Vijver and Hambleton (1996), and van de Vijver and Tanzer (1997). The best reference

for citing the guidelines is Hambleton, Merenda, and Spielberger (2005, chapter 1). These guidelines have become a frame-of-reference for many psychologists working in the test translation and adaptation area, and more general adoption of the guidelines can be expected in the coming years as the guidelines are more widely disseminated and the standards for translating and adapting tests are increased.

From a practical point of view, two major contexts can be distinguished for applying the ITC guidelines: (1) the translation/adaptation of existing tests and instruments, and (2) the development of new tests and instruments for international use. The first context refers to the situation where tests and instruments that have originally been developed in a particular language for use in some national context are to be made appropriate for use in one or more other languages and/or national contexts. Often in such cases the aim of the translation/adaptation process is to produce a test or instrument with comparable psychometric qualities as the original.. Even for non-verbal tests, adaptations are necessary not only of the accompanying verbal materials for administration and score interpretation but also of graphic materials in the test proper to avoid cultural bias. Growing recognition of multiculturalism has raised awareness of the need to provide for multiple language versions of tests and instruments intended for use within a single national context.

The second context refers to the development of tests and instruments that from their conception are intended for international comparisons. The advantage here is that versions for use in different languages and or different national contexts can be developed in parallel, i.e., there is no need to maintain a pre-existing set of psychometric qualities. The problem here often lies in the sheer size of the operation: the large number of versions that need to be developed and the many people involved in the development process.

References

- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. European Journal of Psychological Assessment, 17(3), 164-172.
- Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). Adapting educational and psychological tests for cross-cultural assessment. Hillsdale, NJ: Lawrence S. Erlbaum Publishers.
- Hambleton, R. K., Yu, J., & Slater, S. C. (1999). Field-test of the ITC Guidelines for Adapting Psychological Tests. European Journal of Psychological Assessment, 15, 270-276.
- Muniz, J., & Hambleton, R. K. (1997). Directions for the translation and adaptation of tests. Papeles del Psicologo, August, 63-70.

- Tanzer, N. K., & Sim, C. O. E. (1999). Adapting instruments for use in multiple languages and cultures: A review of the ITC guidelines for test adaptations. European Journal of Psychological Assessment, 15, 258-269.
- van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. European Psychologist, 1, 89-99.
- van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. European Review of Applied Psychology, 47(4), 263-279.

Current Version: January 17, 2010