

Commentary on the Guidelines for Translation and Adaptation of Technology-Based Assessments

(ITC-ATP Guidelines for Technology-Based Assessment: Guidelines 11.1 to 11.12)

Steve Dept

Author Note

Correspondence concerning this article should be addressed to cApStAn Linguistic Quality Control, Chau. de la Hulpe 268, 1170 Watermael-Boitsfort, Belgium.

Acknowledgements: these twelve guidelines immensely benefited from the knowledgeable and insightful feedback of some of the sharpest minds in the industry: Dr Kadriye Ercikan, Dr Avi Allalouf and Dr Maria-Elena Oliveri. All three have published about comparability issues, all three have a different mother tongue (and it is not English) and provided invaluable input. This manuscript was reviewed by Stephen G. Sireci and John Weiner, editors of ITC-ATP Guidelines for Technology-Based Assessment.

Measurement of knowledge, skills, abilities and other characteristics (KSAOs) is a field that has evolved considerably in education, workplace and professional credentialing applications: all or most stages of the testing lifecycle have become increasingly technology-enabled and thereby increasing the reach and global application of assessment programs. In parallel, there is a consensus among stakeholders, test owners, test developers, psychometricians, test users and legislation bodies that technological advances should not be leveraged at the expense of reliability, validity, fairness, security, and privacy. This is a tall order, and as field practitioners continue to shape, test, improve, and apply good practice, it is necessary to document new

developments. As part of this documentation, good practice for assessment delivery in multilingual and multicultural contexts remains a complex and sensitive component.

The International Test Commission (ITC) published *Guidelines for Computer-Based and Internet Delivered Testing* (2005) three years after the Association of Test Publishers (ATP) had published their *Guidelines for Computer-Based Testing* (2002). New technologies have emerged since and assessment practice underwent significant changes, which led the ITC and the ATP to join forces and produce an up-to-date set of *Guidelines for Technology-Based Assessment*. These were published in November 2022, at the same time OpenAI released their ChatGPT interface (which will most likely lead to new practices and further updates). The ATP (2002) and ITC (2005) editions did not contain a section on translation and adaptation, and although the ITC has published *Guidelines for Translating and Adapting Tests* (ITC, 2018), these guidelines did not focus on technology-based assessment in particular. Considering this gap, the new *ITC-ATP Guidelines for Technology-Based Assessment* include a chapter on Global Considerations, which naturally hosted the guidelines for translation and adaptation of technology-based assessments that I was invited to author.

Having engineered the linguistic quality assurance component of the transition from pencil-and-paper test to computer-delivered tests in OECD's Programme for International Student Assessment (PISA), I witnessed that attempting to replicate existing practices in a completely new medium (and tweaking that medium to accommodate the said practices) is a flawed approach. To go digital meant a shift in paradigm for translation and adaptation, and that hasn't always been incorporated in test platform development. At the end of the golden age of pencil-and-paper tests we had reached consensus about the need to embed translation in the test development process. Paradoxically, when technology-based assessment went mainstream, translation often regressed to being an afterthought. In this sense, it is essential to continue to raise awareness of the complexity and the importance of translation and adaptation in multilingual testing, as it can be an important source of construct-irrelevant variance in test performance.

A sound translation and adaptation design can mitigate the risk of bias, and such a design requires some level of sophistication. The twelve guidelines for translation and adaptation of technology-based assessments function as an annotated collection of references to recognized good practice so as to help multiple stakeholders in the assessment process navigate this sophistication.

Multidisciplinarity and multi-stage development

Translation of any assessment collection instrument is geared towards providing comparable measurement across languages and cultures and, in many assessments, across modes and over time. To achieve this for technology-based assessments, it is necessary that the teams in charge of item development, platform development and translation/LQA, respectively, communicate and agree on principles, procedures, workflows, cut-off points and coordination

before the items are developed, before workflows are defined, and before the translation process begins. This approach follows the trend to increase upstream linguistic quality assurance efforts to reduce downstream disparities and related corrective action (Hambleton & Zenisky, 2011; Harkness & Behr, 2008).

Aspects such as interoperability, interdependency, and traceability have gained importance in technology-based assessments, and all require a multidisciplinary approach (**Guidelines 11.1 and 11.11**). Regardless of the translation approach, the following prerequisites can be regarded as essential for translation of technology-based assessments:

1. As far as possible, the measurement characteristics of the source version should be documented, i.e., the way a stimulus is expected to prompt a response strategy should be described; universality or at least generalizability of the constructs should be established (**Guideline 11.2**)
2. The source version needs to be made available in a format that translators can use. This should include an extraction mechanism, i.e., a way to break the source down into translatable and untranslatable components—such as structural markup—and, if possible, hide or lock untranslatable inline components (**Guideline 11.10**).
3. When preparing the source version for translation, attention should be given to facilitating the translation task. One noteworthy way to achieve this is to split the translatable parts of the source version into smaller units, usually referred to as segments.
4. There needs to be a simple process to preview both the source version and the translated version of the assessment or questionnaire at any stage of the process, preferably without file manipulation. Translators and reviewers need to preview the assessment items or questions in context because they need to understand the task that respondents are asked to perform or the question they are asked to answer.
5. Although this is not a prerequisite in a strict sense, there should be a structured repository for translation notes, linguists' comments, and other meta-data, which should either travel together with the translation package—if the translation is exported out of the platform and imported back—or inside the platform, preferably at the segment level (**Guideline 11.03**).

A testing platform that accommodates multiple language versions but does not feature translation technology at all is not translation environment. Non-professional translators may find it practical to enter (or copy-paste) translations in the free text fields of a platform, but trained translators find it unprofessional to work without adequate tools.

The number of steps in a quality assurance design do not necessarily correlate with quality of the final output. It is crucial to draw up precise specifications for each step and have a standardized documentation protocol.

The Guidelines emphasize that no translation and adaptation design would be complete without some form of piloting or field testing. It is not sufficient to pilot the source version. One

needs to collect data on the translated and adapted versions of the instrument to measure equivalence (**Guidelines 11.7 and 11.8**).

The guidelines are informative, not prescriptive, and the editors recognize that it may not be possible or necessary for testing organizations to adhere to all the suggested guidelines: the purpose is not to specify mandatory practices but rather to inform users about issues and considerations in applying technology-based assessment.

Special thanks to Prof. Stephen G. Sireci and Dr John Weiner for their thoughtful and considerate input for this Commentary.

The *Guidelines* can be downloaded free of charge from the ITC resource page <https://www.intestcom.org/page/16> or from the ATP resource page <https://www.testpublishers.org/white-papers>. Navigate to Chapter 11 : Global Testing Considerations. The twelve *Guidelines for Translation and Adaptation of Technology-Based Assessments* are numbered 11.1 to 11.12.

References

- Association of Test Publishers (2002). ATP Guidelines for Computer-Based Testing. Author.
- Hambleton, R. K. & Zenisky, A. L. (2011). Translating and Adapting Tests for Cross-Cultural Assessments. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-Cultural Research Methods in Psychology*. Cambridge: Cambridge University Press, pp. 46–74.
- Harkness, J. A. & Behr, D. (2008). How instrument quality affects translation: Insights from a film of a team translation review. Naples, Italy: International Sociological Association Research committee 33.
- International Test Commission. (2005). Guidelines for computer-based and internet delivered testing. International Test Commission.
- International Test Commission. (2018). ITC Guidelines for Translating and Adapting Tests (Second Edition, 2018). *International Journal of Testing*, 18:2, pp. 101–134. <https://doi.org/10.1080/15305058.2017.1398166>