Welcome!

Translating test items. There’s more to it than meets the eye!
To ask questions, use the “Q&A” feature
Presenters today

Steve Dept
- CEO and Founder of cApStAn
- 25 years of experience in test adaptation and linguistic quality control

John Kleeman
- Executive Director and Founder of Questionmark
- 30 years of experience in the assessment industry
Straightforward translation of test items doesn’t always work well.

This webinar will help you understand good practice so that your tests measure the same thing in other languages and cultures.
About Questionmark

Background

• Founded in 1988
• Assessment technologies and portal solutions that enable organisations to measure knowledge, skills, and attitudes securely for certification, regulatory compliance and improved learning outcomes
• ISO/IEC 27001 Certified (Learn more at www.questionmark.com/trust)

Customers

• Enterprise customers worldwide
• Business, Government; Higher Education; Certification Providers; Awarding Bodies
Questionmark Assessment Management System

- Reporting to analyze results and inform stakeholders
- Collaborative, secure authoring of surveys, quizzes, tests, and exams
- Blended delivery to support a broad range of devices and security needs
- Multilingual: 36 Languages
- Global 24/7 Support
About cApStAn

Background

• Founded in 2000
• Translation quality control and equivalence checks in international large scale assessments to maximise cross-language comparability in assessments measuring knowledge and skills
• Close cooperation with Academic world

Customers

• International organisations: OECD, UN, EU
• National organisations: universities, research institutes, education ministries
• Private sector: testing industry, talent management companies, pollsters
A holistic approach to assessment translation and adaptation

We shape, streamline, implement, improve and promote best practices in test localisation
Agenda

- Translation or adaptation?
- Planning, the most important step
- Designing your process
- Translation process in Questionmark
- Translation technologies
- Managing translation vendor
- How to pilot and check equivalence
- Summary
Translation or Adaptation

Can’t one just translate and administer the test?

What do you mean by Adaptation?
Translation or Adaptation

- In translated tests
  - Meaning shifts (language-driven) may lead to measurement errors
  - Perception shifts (culture-driven) may lead to bias

  - Early versions of IQ test designed to identify most productive white-collar employees => timed assessments with speed as a component of the score
  - Speed and time management play a (much) smaller part in African cultures => low performers on speeded IQ tests

Without adaptation, no fairness (and no comparability)
Literature & empirical findings tell us how and why different forms of adaptation (to local context and usage) affect measurement.

Working definition of adaptation in assessments:

ADAPTATIONS ARE DESIRABLE IN ALL OCCURRENCES WHERE DIRECT TRANSLATION COULD PUT RESPONDENTS FROM THE TARGET GROUP AT AN ADVANTAGE OR A DISADVANTAGE.
Why should you avoid packing a lighter refill in your hand luggage?

Typical response in some cultures:
- Inflammable, explosive, dangerous

In other cultures:
- Banned, prohibited, forbidden

Experts in test localization are aware that the concept of legality (or regulations) can **blur a construct** and elicit a different cognitive strategy in certain cultures.
Commonly used in English but (very) hard to render in (a number of) other languages

- Idiomatic English, neologisms
  - Earmarked for...
  - Stagflation
  - Trade-off
  - Greenwashing
  - To look forward to
  - To value
  - To have fun
Planning, the most important step
Two approaches to planning

You are designing a new test and want to make it applicable to 2 or more languages/cultures

You are working with an existing test that is already deployed in one language and need to translate/adapt it in other languages
Planning for a new test

Start with purpose
- Why is test being developed?
- What is construct or domain being measured?
- What is done with pass/fail or test scores?

Identify target cultures/languages
- What is population?
- What languages is it reasonable for them to take the test in?
- Are there different cultures (e.g. Indian English vs UK English vs US English)?

Seek advice
- Whether the construct/domain has any issues in any of languages/cultures
- Whether item formats need adjusting for languages/cultures (e.g. right to left)

Design with this in mind
- Define objectives/topics
- Progress exemplar items
- Define translation process

Translatability assessment after initial design
- Check items and principles work before going onto full development
Translating an existing test

- Identify which languages to translate into
  - Base this on the test purpose – what you are trying to measure
  - Identify target population and fairness/equity and balance vs cost
  - Choose languages and cultures

- Ensure translation makes sense
  - Do the constructs and domains apply in the language/culture
  - How much adaptation is needed, how much translation?

- Translatability assessment
  - Especially useful if translating into several languages
  - Find issues before doing the actual translation
Designing your process

Making informed decisions about the translation, adaptation and linguistic quality assurance design
Typical translation/adaptation design for international large scale assessments

Too complex for most assessment ecosystems
A. Source Optimization

A1 controlled writing
A2 file preparation: parsing, segmentation, locking untranslatable content
A3 Translatability Assessment
A4 create project-specific rules
  B1 create glossaries
  B2 create style guides
  B3 create language-specific rules
  B4 translation & adaptation notes
  B5 trend mgt (content transfers)
  B6 monitoring tool, documentation
  B7 train translators (& verifiers)

B. Preparation of Translation & Adaptation Process

C. Translation and Adaptation Process

D. Linguistic Quality Control Process

D1 automated checks
D2 translation verification
D3 errata and updates mgt
D4 review of verification feedback
D5 post-verif final checks
D6 reports, updated TMs, post mortem

C1 double or single translation
C2 reconciliation
C3 (team) adjudication
C4 consultation (domain experts)
C5 proofreading
EXAMPLE A4. Create Project-specific Rules

- Rules that can be linked to attributes of the source version
  - Define max. number of characters in captions
  - Forced adaptations (e.g. imperial vs metric)
  - Identify the need for dynamic text (e.g. gender)
  - Determine how text styles must be treated
  - Proportional length of response choices

Compliance with these rules can be checked automatically
What is the purpose of the question at the end of “Disaster in the making”?

A. To encourage readers to agree with the writer.
B. To make readers doubt the value of being part of a community.
C. To suggest that no-one understands this topic completely.
D. To encourage readers to form their own opinion on the topic.

Distractor B is much longer → More attractive

Proportional length of key and distractors can be checked automatically
EXAMPLE B1. Create (and maintain) glossaries

- Creating glossaries is best practice in (pre-)translation, but remains underrated and underused.
- bilingual or multilingual glossaries rather than explanatory glosses in the source language.
- Creating glossaries involves some work, brings immense quality gains, particularly in domain-specific assessments.
Glossaries may contain

- any terms and words that need to be translated consistently (technical concepts, cultural entities, ...)
- fictitious names
- terms or expressions that involve some research
- terms or expressions that can be translated in more than one way
- recurring prompts
- acronyms
- names of months and days
- names of questionnaire sections or modules
- names of buttons, tables and other UI elements

Adherence to the glossary can be checked automatically
EXAMPLE B4. Translation and Adaptation Notes

Solution:

- Automate what can be automated
- Prepare clear, concise translation & adaptation notes
  - Item by item
  - Ideally: appear in the TEnT (Translation Environment Tool)
  - Draw attention to translation traps and psychometric traps
  - Specify which adaptations are required, desirable, or ruled out
- Set up a system whereby
  - Translators need to confirm they have read the T&A note
  - Verifiers need to confirm that it has been addressed
Translation process in Questionmark
Technical translation process

- **Create in one language**

- **Export into XLIFF**

- **Translator translates into other languages**

- **Import translated XLIFF**

- **Review in system**

- **XLIFF**
  - XML Localization Interchange File Format
  - Structured text with tags
  - International standard from Oasis (who create open data standards)
  - Reasonably well supported by translation tools
Export XLIFF for Translation

Download XLIFF

Source language: English (United States)
Target language: -- Select Language --

- All text
- Only text not previously downloaded
- Include sub-topics

[OK] [Cancel]
Manage Translations of Assessments

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Assessment ID</th>
<th>Created by</th>
<th>Created on</th>
<th>Modified on</th>
<th>Last edited</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRCE Exam</td>
<td>Test</td>
<td>0229312000229312</td>
<td><a href="mailto:brian@questionmark.com">brian@questionmark.com</a></td>
<td>8/17/2018</td>
<td>8/17/2018</td>
<td></td>
</tr>
<tr>
<td>Electrical Motors and Switch Gear Test</td>
<td>Test</td>
<td>0068320000068820</td>
<td><a href="mailto:brian@questionmark.com">brian@questionmark.com</a></td>
<td>6/29/2017</td>
<td>8/9/2018</td>
<td></td>
</tr>
</tbody>
</table>
Translation technologies
CAT TOOL ≠ MACHINE TRANSLATION

- A specialized program that helps a translator
  - increase productivity
  - obtain a higher degree of consistency
- by leveraging language assets such as translation memories and term bases.
  - Translation quality depends on human translator

- Artificial Intelligence uses
  - rules
  - large sets of data
  - statistics
  - machine learning algorithms
- To produce computer-generated translations
  - Translation quality depends on the quality of sets of data + algorithms

- (some providers are pivoting towards machine translation + post editing models)
Language assets

Prehistoric approach

- Source files (in Word, Excel)
- Key Terms definitions (in Word)
- Bilingual glossary (in Excel or a Word table)
- Guidelines (in Word, Excel)
- Reference files:
  - Materials from the previous cycle (in Word, Excel)

Computer assisted approach

Project download link / Project package

- Source: .xliff
- Termbase: .tbx
- Guidelines: .tmx
- Translation memory from previous cycle: .tmx

Standard: .tmx
Managing translation vendor
Vendor management advice

- Determine your approach and requirements yourself, don’t let translation vendor talk you into theirs
- Require translation memories as part of deliverables
- Each step in the workflow should be documented
- Translators should only translate into their mother tongue
- Ask to see vendor document on how they recruit and evaluate their linguists
- For large or high stakes projects use independent linguistic QA
- Set up and impose secure file transfer (including within translation vendor)
- Ensure NDAs with all translators to protect question security (often overlooked)
Before and after the translation vendor

Translation industry best practices:
- Optimal preparation of projects for translation
- Optimal working environment and language assets
- Linguistic quality assurance

Platform architects
- Translation files (XLIFF)

Project Management
- Guidelines & glossaries

Translators
- Target versions of assessment

Reviewers
- Final target versions
Translation Verification

- Segment by segment comparison of source and target
- Compliance with translation and adaptation notes
- Focus on semantic equivalence and on level of difficulty
  (while we’re at it, check linguistic quality)
- Use standardized verifier intervention categories to report issues
- Describe all issues in a centralised monitoring tool
- Propose corrective action directly in the target files
## Translation Verification

<table>
<thead>
<tr>
<th>Source</th>
<th>Translation</th>
<th>Verification Guidelines</th>
<th>Verifier Version (if different from original)</th>
<th>Verifier Intervention Category</th>
<th>Verifier Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please agree or disagree with the following statement: During the last week, there were people who appreciated me.</td>
<td>過去1週間、自分を評価してくれる人がいた。</td>
<td>&quot;appreciated&quot; may be translated as &quot;praised&quot; or &quot;complimented&quot; in your language if more appropriate. It should not be understood as &quot;evaluated&quot; or &quot;increased the value of&quot;.</td>
<td>次の記述についてどう思いますか：過去1週間、自分の良さをくれる人がいた。</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Verification vs. Proofreading

**Verification**
- Focus on maintaining same difficulty level and ensuring correct item functioning
- Less leeway as regards form (especially in key parts)
- Preferential changes to be avoided
- Source should not be improved
- Literal matches/synonymous matches

**Proofreading**
- Focus on linguistic fluency and correctness
- More leeway as regards form
- More room for preferential changes
- Improving source not a problem
- Repetition / consistency
How to pilot and check equivalence
Piloting

- Build in a pilot to be carried out on a small(er) scale, e.g. a dry run in (some of) the target language(s)
- Or cognitive pre-test (need trained interviewers in target language)
- Finalise translation of test items before pilot
- Analysis of pilot results = a wealth of information
- Can be used to perform a **focused** linguistic and formal verification before going live
Analysis methods

- Qualitative
  - Feedback from pilot test-takers

- Quantitative
  - Run item analysis
  - Compare means of difficulty in different languages
  - (If enough data) run item analysis by language and identify discrepancies
Propose corrective action

- Items that have behaved anomalously on any of the criteria are flagged (dodgy items)
- Dodgy item reports examined by testing organisation, by linguistic reviewer and possibly by local subject matter expert
- Explain why such results may have occurred.
- Test developers propose corrective action to be implemented by linguists
- Linguists propose corrective action to be endorsed by test developers
- And then...
More information

- International Test Commission Guidelines on Adapting Tests
  - [https://www.intestcom.org/page/16](https://www.intestcom.org/page/16)

- PISA guidelines (OECD program for international student assessment)

- Translation and Cultural Appropriateness of Survey Material in Large-scale Assessments
  [http://www.capstan.be/share/#publication-4](http://www.capstan.be/share/#publication-4)

- Useful blogs
  - Questionmark: [https://blog.questionmark.com](https://blog.questionmark.com)
Translation/adaptation can (and should be) embedded in test design

Involve specialists at the earliest stages of a translation project

Time spent (i) optimising the source version; and (ii) preparing the translation/adaptation process drives quality of translated tests

Export out of your assessment system to allow translators to work with state of the art technology (importing back in should be tested first)

Organise (at least some) independent quality checks of the translations
Your questions
Know your translation needs, know your options

Before embarking on the journey of producing multiple language versions of your assessment:

- Request a free translation needs & requirements analysis by a senior cApStAn consultant at info@capstan.be

You already have a language service provider and you already have translated tests:

- Ask us to organise a sample-based translation quality evaluation and equivalence check at projects@capstan.be
How to Evaluate Questionmark

- **Request a one-on-one demo**
  The Questionmark team will contact you to arrange a demonstration tailored to your needs and questions
  [www.questionmark.com/go/request-demo-enus](http://www.questionmark.com/go/request-demo-enus)

- **Request a 30-day trial of Questionmark OnDemand**
  Complimentary technical support during trial
  [www.questionmark.com/go/od30us](http://www.questionmark.com/go/od30us)
Upcoming Webinars

November 28th
9 Risks to Test Security (and what to do about them)

December 5th
Tips for Writing Good Test Questions

December 18th
10 Quick Tips to Improve your Tests and Exams

Weekly
Introduction to Questionmark (Live product demos)

www.questionmark.com/webinars
Thank you for attending!

Keep up to date with Questionmark at blog.questionmark.com and with cApStAn at www.capstan.be/news/