

Adapting/Translating Psychological and Educational Assessments for Use in Multiple Languages and Cultures



Ronald K. Hambleton

Center for Educational Assessment

University of Massachusetts, Amherst

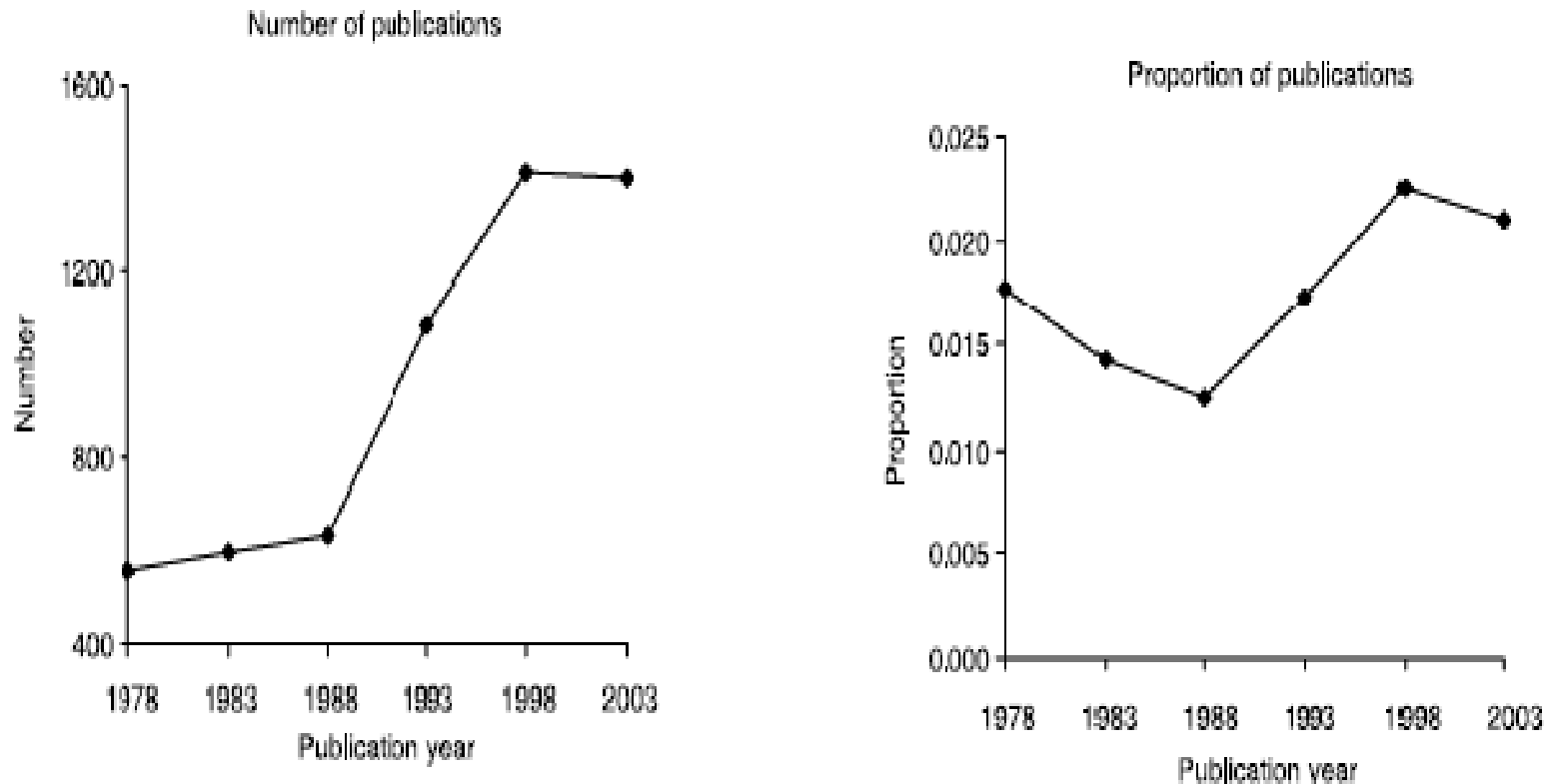
10th Conference of the Cooper Institute Series, Oct. 19, 2007.

Examples of Test Adaptations

- Intelligence, Aptitude, and Personality Tests
- Credentialing Exams (e.g., Microsoft, Novell, HP)
- Quality of Life Surveys
- TIMSS, OECD/PISA, and More
- SAT, GED, NAEP, State Educational Assessments (NCLB requires)

Cross-Cultural Studies between 1978 and 2003 (van de Vijver, 2006)

Figure 18.1 Number and proportion of publications about cross-cultural and ethnic issues from 1978 to 2003.



Three Reasons for Adapting Assessments

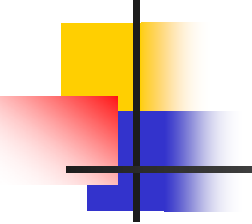


1. To increase fairness/validity of scores.
 - enables respondents/examinees to take the instrument in one of several languages. (e.g., state assessments)
2. To expand the usefulness of the assessment for doing comparative studies across languages and cultures. (e.g., NAEP, PISA, QOL measures)

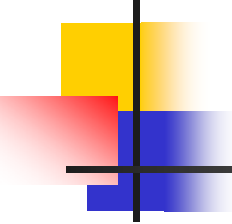


Three Reasons for Adapting Assessments

3. To permit research instruments to be used in additional languages and cultures.

- 
- Note: If goal is to offer tests in multiple languages and cultures, **test translation and adaptation is often cheaper and faster than constructing new tests.**
 - With new tests —considerable work because test equivalence cannot be assumed; new norms are needed, validity studies must be carried out, any passing scores set, etc.
 - Often, translation/adaptation preferred.

Goals of the Presentation

- 
1. To describe several myths about the test adaptation process.
 2. To highlight two steps in the test adaptation process.
 3. To (briefly) introduce the (revised) International Test Commission Guidelines for Test Adaptation.



Test Translation vs. Test Adaptation?

“Test Adaptation” is more descriptive of the process that usually takes place—directions, item formats, timing, content of passages, etc.

“Test Translation” is usually only part of the process that must take place.

Example 1 (IEA Study in Reading)



Are these words similar in meaning?

Pessimistic -- Sanguine



Pessimistic -- Sanguine

Adapted to

Pessimistic -- Optimistic

Example 2 (1995 TIMMS Pilot)

Alex reads his book for 1 hour and then used a book mark to keep his place. How much longer will it take him to finish the book?

- A. $\frac{1}{2}$ hour
- B. 2 hours
- C. 5 hours
- D. 10 hours





Example 3

Out of sight, out of mind

(Back translated from French)

invisible, insane



Example 4

Passive tense does **not** exist in all languages.

-- (e.g., There is...; There are...; it was stated that...)

Major rewrites are often needed.



Example 5

Je ne suis pas une valise.

Literal Translation: I am not a suitcase.

Appropriate Adaptation: Don't try and trick me.

Rating Scales Can Be Problematic:

In an English to Korean translation:

a Likert rating scale (strongly agree, agree, undecided,...) was difficult to translate—no literal translation. Came out as:

“Very much like me,” “like me,” “do not know or either way,” ...

Is this the same meaning??



Five Common Myths About Adapting Tests

1. Know two languages and you can be a translator.

-Not true! Poortinga commented that 80% of cross-cultural research was flawed before 1995 because of poor test adaptation.

-Many researchers say knowing the two languages may be the **least important** of several important skills.





Selection and Training of Translators

- Knowledgeable in the languages
- Knowledgeable in the cultures
- Knowledgeable in the subject matter (e.g., trapezoids)
- Knowledgeable in the principles of test/questionnaire development, item writing, scoring rubrics, etc.



Five Common Myths

2. **A good literal translation guarantees validity.**

-It does not! Format unsuitability, restrictive time limits, unclear directions, inappropriate content, and more, can be problematic, even with a good literal translation. [Needed is linguistic, psychometric, and psychological equivalence.]

Five Common Myths

3. Judgmental reviews are sufficient to identify problems in an test translation; translators know their job.

-Reality: Translators can overlook many important features.

-Analogy to item writing—field-testing turns up many problems with test items.

Five Common Myths

4. The common strategy of (1) a back-translations design, and (2) the use of a bilingual design to compile empirical data, is sufficient to validate a test for use in a second language.

[Will explain back-translation, and bilingual designs a bit later in the presentation]

They are not!

- Back translation does **not** involve looking at target language version of the test.
- Bilingual candidates are **not** representative of unilingual candidates in either language.

Five Common Myths

5. Constructs are universal, so all tests can be translated into other languages and cultures.

-They are not! Exceptions include intelligence tests, quality of life measures, health surveys. Also, with achievement tests and credentialing exams—the content domain may not be relevant in a second culture/country.



Goal Two of the Presentation

What positive steps can be taken to construct a valid test in a second language and culture?

--**Eight** steps are offered by Hambleton & Patsula (1999).

[Will focus on only **two** steps now.]

Prior Steps: If it known that tests will be adapted:

Lots of suggestions in the literature for the source language version:

1. Keep the sentences simple.
2. Use the active voice not passive voice. (Translations are better.)
3. Minimize the use of pronouns. (Pronouns complicate the translations process.)



Prior Steps: If it known that tests will be adapted:

4. Avoid culturally specific content. (e.g., Don't use passages about baseball; American television shows; or American foods, etc.)
5. Avoid vague words like "frequently" and "sometimes." (They create ambiguity for the translator.)



Steps for Adapting Tests

4. Translating and adapting the test.
 - Multiple translators are needed.
 - Several designs and variations available.



Backward Translation Design

An exam is translated (adapted) into the target language and then back-translated into the original or “source” language (by different translators). Original and back-translated versions of the exam are compared. [Use at least two translators. Costs are increased but the merits of multiple translators are substantial.]

Backward Translation Design



Advantages

1. Some problems can be identified.
2. Researchers do not need to be bilingual.

Disadvantages

1. No review of target language version!
2. Evidence that this design is NOT sufficient (e.g., Reckase & Kuncze, 2002)



Quote from Harkness and Schoua-Glusberg (1998):

- “In practical and theoretical terms (results, effort, costs, reliability, and viability), it is one of the least recommendable procedures. Ironically, the problem is that, if a study description states that a back-translation was carried out, this will be read as an indicator of quality.”



Forward Translation Design

A test is translated (adapted) into the target language and then the original (source) language and target language versions of the test are compared by bilingual experts. [use at least two translators]



Forward Translation Design

Advantages

1. Judgments made directly about language versions of a test.

Disadvantages

1. Bilingual judges make assessments of a test for monolingual examinees.



Steps for Adapting Tests

7. Carrying out a comprehensive validity study of the adapted test.

- construct equivalence
- method equivalence
- item equivalence



Data Collection Designs

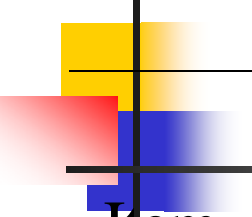
1. Bilingual Designs

2. Monolingual Designs

A. Source Language Examinees
Take Original and Backward
Adapted Versions

B. Monolingual Examinees Take
Versions in Their Own
Languages

Checking for Construct Equivalence



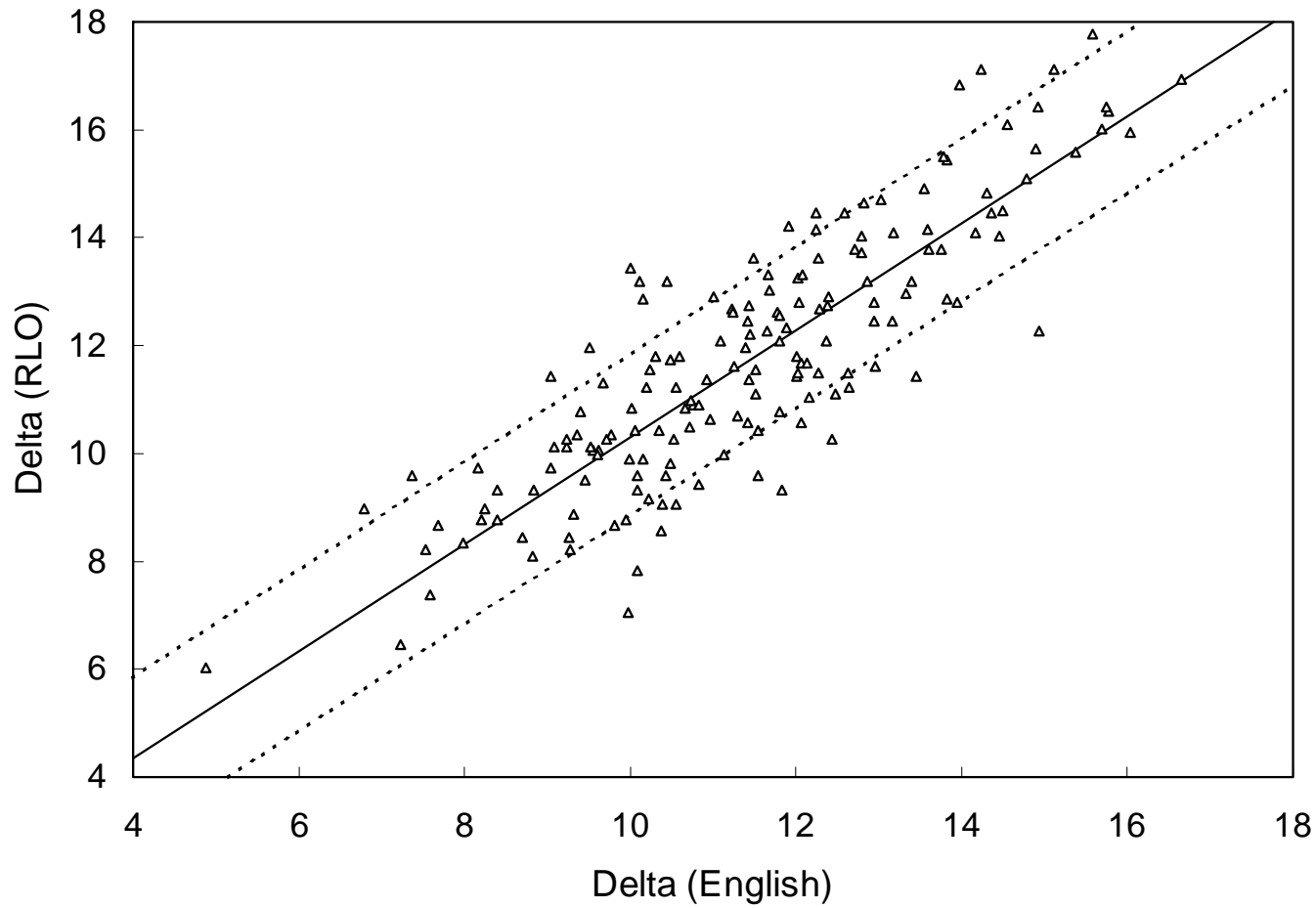
Item	Factor 1		Factor 2	
	Language A	Language B	Language A	Language B
1	.56	.49	-.07	.14
2	.49	.40	.59	.41
3	.43	.57	.11	-.04
4	.48	.52	.63	.68
5	.54	.46	.38	.45
...
19	-.03	.22	.49	.44
20	.25	.13	.53	.56
21	.29	.08	.27	-.13
22	.09	.26	.02	.22



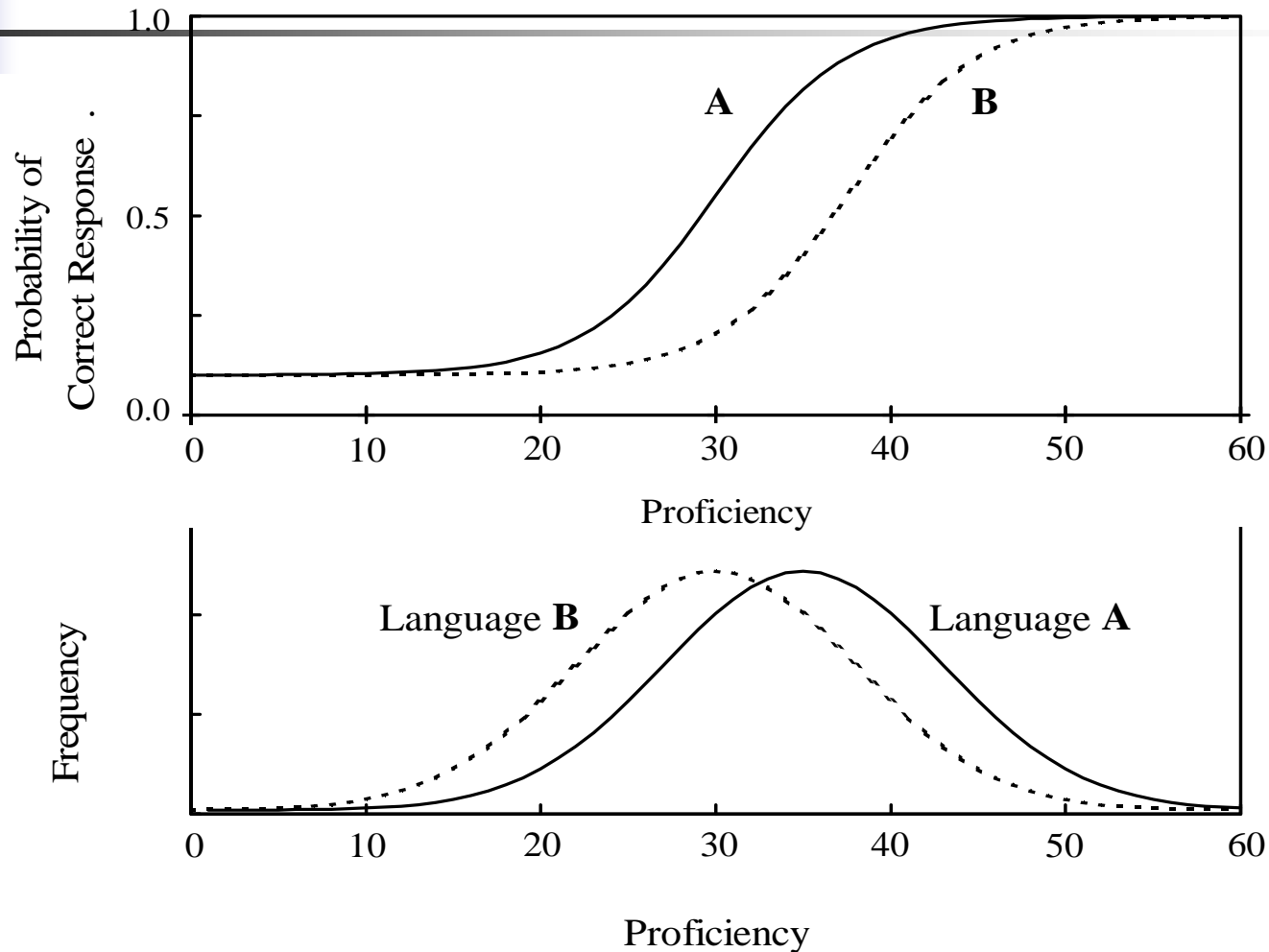
Checking for Method Equivalence

1. Any biases due to directions, test format, scoring?
2. Preparation of test administrators?
3. Equivalence of test-taking skills?

Checking for Item Equivalence: Delta Plots



Checking for Item Equivalence: IRT Methods (change in difficulty)





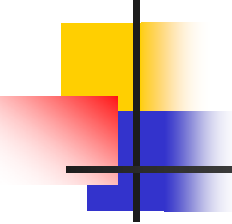
Goal 3—ITC Test Adaptation Guidelines


- 22 guidelines that range from review of the construct in multiple cultures, to test translation and adaptation procedures, to statistical analyses for checking construct, method, and item level equivalence.



Three Concluding Questions

1. Has progress been made in test adaptation methodology?
 - Definitely!
 - Moved from single unqualified translator and bilingual design to considerably more sophisticated methodologies.

- 
- Recognition today that no single procedure can provide sufficient proof to justify the use of an adapted test.
 - Much like validation initiatives—multiple sources of evidence are needed.

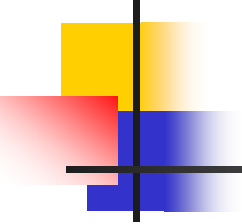


2. What lies ahead for the educational and psychological assessment field?

- Needed is a commitment of resources and time, and effort to implement existing methods.
- Belief lingers that this work is easy and can be done well in a short time and with little expense.

3. What are the most important areas for improvement in practice?

- choose **multiple** qualified translators; give them sufficient time
- aggressively implement empirical methods
- build on past experiences and knowledge base.

- 
-
- Hambleton, Merenda, & Spielberger (Eds.). (2005). **Adapting educational and psychological tests for cross-cultural assessment.** Hillsdale, NJ: Erlbaum.