Issues and Challenges in Instrument Adaptations:
Bias and Equivalence

By
Rosna Awang-Hashim, Ph.D.
Division of Education
School of Languages and Scientific Thinking
UNIVERSITI UTARA MALAYSIA

For most of the twentieth century, interest in translating/adapting tests with a second language and cultural group has been prevalent among educational and psychological researchers. The globalization of economy, tourism, migration streams, and related political changes have witnessed the steady increase of publications that concern cross-cultural differences and comparisons (see van de Vijver & Leung, 1997; van de Vijver & Tanzer, 1997). However, this healthy development is not without methodological inadequacies. Despite significant advances in the methodologies that are used to adapt instruments, there remain a number of difficult and challenging issues in this area, which might invalidate generalization of research findings. This paper will examine issues and challenges in instrument adaptations with emphasis on the issues of construct bias and equivalence. Specific reference to the International Test Commission (ITC) test adaptations guidelines (Hambleton, 1999) will be highlighted. Remedies to these problems will be discussed to enhance validity of adapted instruments.

Issues and Challenges in Instrument Adaptations: Bias and Equivalence

By

Rosna Awang-Hashim, Ph.D.

Division of Education

School of Languages & Scientific Thinking

UNIVERSITI UTARA MALAYSIA

(November 1999)

ABSTRACT

For most of the twentieth century, interest in translating/adapting tests with a second language and cultural group has been prevalent among educational and psychological researchers. The globalization of economy, tourism, migration streams, and related political changes have witnessed the steady increase of publications that concern cross-cultural differences and comparisons (see van de Vijver & Leung, 1997; van de Vijver & Tanzer, 1997). However, this healthy development is not without methodological inadequacies. Despite significant advances in the methodologies that are used to adapt instruments, there remain a number of difficult and challenging issues in this area, which might invalidate generalization of research findings. This paper will examine issues and challenges in instrument adaptations with emphasis on the issues of construct bias and equivalence. Specific reference to the International Test Commission (ITC) test adaptations guidelines (Hambleton, 1999) will be highlighted. Remedies to these problems will be discussed to enhance validity of adapted instruments.

"Advance Organizer"

☹ ☹ ☹

- Why translate/adapt instruments?
- Common errors in translation
- AERA/APA/NCME guidelines
- Issues of bias
- Issues of equivalence
- Conclusion

Why do we translate/adapt psychological tests?

- may be cheaper & faster
- the expertise to construct an instrument measuring the devised trait, skill, or ability may not exist locally
- to enhance the fairness of comparisons of individuals from different language & culture
- to allow comparative studies across groups, ethnic, & cultural groups

---

Common errors in translating/adapting instruments:

- Selection of translators- simply *bilingual* has been shown to be an unsuccessful practice
- Failure to ensure that translators selected are familiar with the *content area*
- Failure to ensure that translators selected are experienced in *test development*

Standards for Educational & Pyschological Testing:

- When a test user makes a substantial change in test format, mode of administration, instructions, *language*, or content, the user should revalidate the use of the test for the changed conditions or have a rationale supporting the claim that additional validation is not necessary or possible (AERA/APA/NCME: Standard 6.2)

(Standard 13.4)

- When a test is translated from one language or dialect to another, its <u>reliability</u> and <u>validity</u> for the uses intended in the linguistic groups to be tested should be established.

(Standard 13.6)

- When it is intended that the two versions of dual-language tests be comparable, evidence of <u>test comparability</u> should be reported.

Issues & Challenges:

- Translation/adaptation process
- Assessment process
  - construct equivalence
  - test administration
  - test format
  - test speededness

A. What is a Good Translation?

- Classical definition: ...rendering text into another language
- New & broader definition: ...producing text in a second language that is as *equivalence* as possible to the original

Who Should Translate?

Criteria to be Considered:
- (competent) bilingual
- knowledge about both languages & cultures, especially the target culture
- knowledge about content area
- knowledge about test construction

Designs for Adapting Tests:

- Forward Translation

- Backward Translation

Forward Translation:

- a group of translators adapt the test from the source language to the target language
- equivalence of the 2 version is judged by another group of translator
- revision made to the target language version

Weaknesses (Forward):

- Translators may be more proficient in one language than the other.
- Ratings of test equivalence involve judgments by persons who are bilingual, who may have used insightful guesses & may be more clever than the monolingual candidates taking the tests.
- Test developers are not in position to judge the tests equivalence themselves.

Back-translation:

- A group of translators adapts a test from the source language to the target language.
- A 2nd group of translators takes the adapted test (in the target language) & adapts it back to the source language.
- Then, the original version of the test & the back-translated version are compared & judgments are made about their equivalence.

Weaknesses (Back-translation):

- Insufficient; is only one of the many types of evidence to be compiled
- comparison of tests is carried out in the source language
- it retains inappropriate aspects of the source language (e.g.,same grammatical structure & spelling)

Challenges:

- How can we ensure that the adapted test/instrument is not seriously biased and is equivalent to the original source?
- Bias & equivalence issues

B. What is bias?

- General: dissimilarity of psychological meaning across (cultural) groups
- Practical: when cross-cultural differences do not involve target construct measured by the test
- Theoretical: when observed cross-cultural differences cannot be fully interpreted in terms of the universe of interest

(van de Vijver, 1999)

Taxonomy of Bias
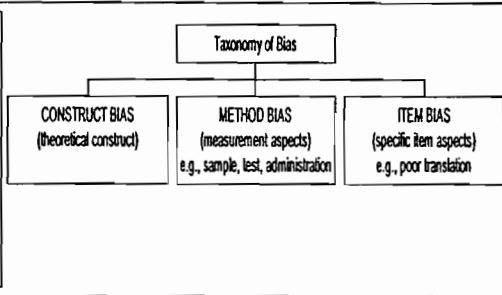
Construct Bias

■ Non-overlap of behaviors defining construct

■ e.g., "filial piety" (≈ obedience)

International Test Commission Guideline 2:

■ "The amount of overlap in the constructs in the populations of interest should be assessed."

(Guideline 2)

Types & Sources of **Method** Bias

■ Sample bias   ➜   Confounding sample differences

■ Instrument bias   ➜   Test characteristics

■ Administration bias ➜ procedural aspects

Item Bias

(DIF-Differential Item Functioning)

■ An item of a scale (e.g.,anxiety) is seemed biased if persons with same trait anxiety, but coming from different cultures, are not equally likely to endorse the item.

C. What does 'equivalent' mean?

■ Two types of equivalence:

✓ linguistic
✓ psychological

Linguistic Equivalence

■ Similarity of language features of a text

examples:
→lexical similarity
→ grammatical accuracy

Psychological Equivalence

■ Similarity of psychological meaning & scores
■ similarity in a broader sense:
→textual (e.g., words connotation, implied context of text, comprehensibility)
→Metrical (score comparability)

Translatability

A psychological test/item is
■ **well translatable** if linguistic & psychological features yield the same translation
■ **poorly translatable** if ling & psych features do not entirely converge
■ **non-translatable** if there is a complete or nearly complete overlap

What options do we have?

■ Adopt the original test
(more or less literal translation)

■ Adapt the original test
(change some/most/all aspects)

How do we decide which option?

■ Depends on the purpose of study

→comparative
→non-comparative

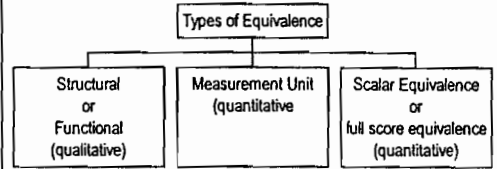(bias & challenges of equivalence expected)

Relationship between equivalence & bias

■ ↑level of comparability ↓bias

■ Highest level of equivalence obtained for bias-free measurement

ITC Guideline 21:

- "Comparisons across populations can only be made at the level of <u>invariance</u> that has been established for the scale on which scores are reported."

Taxonomy of Equivalence

Types of Equivalence		
Structural or Functional (qualitative)	Measurement Unit (quantitative	Scalar Equivalence or full score equivalence (quantitative)

Structural or Functional Equivalence (Qualitative)

- Are we measuring the same construct?
- Various statistical tools available:

→Exploratory Factor Analysis (EFA)
→Confirmatory Factor Analysis (CFA)
→Convergent & Discriminant Validity

Measurement Unit Equivalence

- Do we have equal measurement unit?
- Individual differences have a different meaning **within** and **across** cultures
- Statistical tool:

→EQS (Structural Equation Modeling)- CFA

Scalar Equivalence

- Are scores comparable (both within and across cultures)?

(especially in cross-cultural research)

Determining Equivalence

- **Deductive:** evidence for qualitative & measurement unit equivalence

- **Inductive:** evidence falsifying a particular interpretation

What options do we have?

- Focus on confirmatory evidence
- less attention for falsifying evidence
  (construct bias is infrequently studied)

Conclusion

- Combine expertise (linguist, psychologist, & psychometrician)
- Use bias & equivalence as key concepts ... (in all stages, starting from choice of concept /conceptual framework to documentation in manual)

THE END

☺ ☺ ☺
☺ ☺
☺
Thank you for listening!