Developing professional standards for EFL testing in China: Contexts, considerations, and challenges

Presenter: Jinsong Fan, Ph.D.
Fudan University
Email: jinsongfan@fudan.edu.cn
Overview

1. Standards in language testing and assessment
2. Language testing in China: What’s special?
3. Standards development and implementation in other testing contexts
4. Fundamental considerations for developing professional standards in China
5. Challenges facing standards development and validation
6. Conclusions and future studies
1. Standards in language testing and assessment

A dictionary definition of ‘standard’:
‘Standard’ refers to a level of quality, skill, ability or achievement by which someone or something is judged, that is considered necessary or acceptable in a particular situation. (Longman Advanced American Dictionary, 2000, p. 146)
Two interpretations of ‘standards’ in language testing

**Definition 1:** The skills and/or knowledge required to achieve mastery and proficiency levels leading to mastery, along with the measures that operationalize these skills and/or knowledge and the grades indicative of mastery at each level (Davies, 2008, p. 437) – e.g. cut-score, CEFR, ACTFL

**Definition 2:** An agreed set of guidelines which should be consulted and, as far as possible, heeded in the construction or evaluation of a test (Alderson, Clapham, & Wall, 1995, p. 236) – e.g. *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), *ETS Standards for Quality and Fairness* (ETS, 2002)
1. Developing a language test and ensuring that it is valid or useful is extremely difficult (see e.g. Bachman, 1990; Alderson, Clapham, & Wall, 1995; Bachman & Palmer, 1996; Henning, 2001; Fulcher & Davidson, 2007, etc.).
2. The important role that language tests play in modern society: Results on language tests are often used to make a variety of high-stakes decisions such as admissions, employment, promotion, immigration, and citizenship (see, e.g. Spolsky, 1995; Shohamy, 2001a, 2001b; McNamara, 2005, etc.).
3. The call for better accountability, transparency and fairness in language testing practices (see e.g. Kunnan, 2000, 2004; Shohamy, 2001a, 2001b; Bachman, 2005; Bachman & Palmer, 2010; Xi, 2010, etc.).
The need for those involved in the testing act to assume responsibility for the tests and their uses (Shohamy, 2001b)

(Accountability) requires shared authority, collaboration, involvement of different stakeholders – test takers included – as well as meeting the various criteria of validity... The cost of this approach is high... it takes more time, it involves more people, it requires greater resources... But...the cost is worth paying in order to demonstrate the ethicality of the profession (Shohamy, 2001b, p. 161-2)
Is test-related information equally accessible to all test candidates?

Can test users access relevant information to make informed decisions about test candidates?

Is information about the test quality available to all stakeholders?

Can independent researchers access relevant test data to investigate the quality of language tests?
Fairness

Traditional conceptualization of test fairness: Fairness as absence of bias

The relationship between test validity and test fairness: ‘A test has to be valid to be fair’ or ‘A test has to be fair to be valid’?

Theoretical frameworks of test fairness (Kunnan, 2000, 2004) and approaches to investigate test fairness (AERA, APA, & NCME, 1999; Xi, 2010)
2. Language testing in China: What’s special?

1) A huge testing country with a large number and variety of English language tests developed, administered and used at different levels and for different purposes.
2) Large scale, high stakes, and strong washback effects on English teaching and learning

**Large scale:** Huge test population

**High stakes:** The results of English language tests are often used to make important decisions such as admissions into higher education, employment, the awarding of academic degrees, application for the permanent residential permit in major cities, etc.

**Strong washback effects:** What is tested becomes what is taught: “the assessment tail wagging the educational dog” (Briggs, 1992, p. 11)
Almost all tests are developed and administered by the relevant examinations authorities.

The quality and authority of language tests are seldom questioned or challenged.

Language testing is considered as more administrative than academic behavior (Yang & Gui, 2007).

Test developers are believed to be solely responsible for the fairness of all testing operations.

Stakeholders such as test takers and teachers cannot participate in the testing process as equal partners (see also Shohamy, 2001a).
3. Standards developed and implemented in other testing contexts


A collection of 110 standards from all over the world with 58 standards identified as guidelines of good testing practice.

Many standards in the collection have been revised or updated since the completion of the project (e.g. AERA, APA, & NCME, 1985).
Language testing standards in use: A few examples

International Language Testing Association (ILTA)
- *ILTA Code of Ethics* (ILTA, 2000; see also Davies, 1997; Boyd & Davies, 2002)
- *ILTA Guidelines for Practice* (ILTA, 2007)

- *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999)
- *ETS Standards for Quality and Fairness* (ETS, 2002)
Association of Language Testers in Europe: *The ALTE Code of Practice* (ALTE, 1994; see also Avermaet, Kuijper, & Saville, 2004)

European Association of Language Testing and Assessment: *EALTA Guidelines for Good Practice in Language Testing and Assessment* (EALTA, 2006)

Japan Language Testing Association: *The JLTA Code of Practice* (JLTA, 2007; see also Thrasher, 2004)
Consideration 1: Why a new set of standards?

Why reinvent the wheel, particularly if there is an excellent wheel already? (McNamara & Roever, 2006, p. 146; see also Alderson, Clapham, & Wall, 1995).
Towards a new set of standards

1) The special features of language testing in China: Applicability and validity of other sets of standards

2) The product and process perspectives
   The *product* perspective: A set of guidelines for language testing in the Chinese context;
   The *process* perspective: Awareness-raising, more discussions about quality and professionalism, transparency and fairness, etc.
Consideration 2: Who are the targeted audiences/users of the standards?

In any testing context, be it centralized or decentralized, test validity and fairness will eventually be compromised without the collaboration of all stakeholders in the testing process, including test design, administration, and use (Fan & Jin, 2013).
Figure 1: The targeted audiences of the standards
Consideration 3: What purpose(s) do the standards serve?

**Primary Purposes** (Educational & Aspirational)
- To enhance the awareness of quality among test developers
- To promote among the stakeholders the basics of language testing

**Targeted Audiences**
- Test developers
- Test takers
- EFL teachers
- Test users
- Educational officials and administrators and other stakeholders

**Expected Outcomes**
- Provide a benchmark of good testing practices
- Enhance professional awareness
- Promote dynamics between test developers and the other stakeholders
- Pursue better washback effects

**Figure 2: Purposes, audiences, and expected outcomes**
Consideration 4: How to generate the standards?

Three possible approaches to standards development

- The top-down approach: An organization develops the standards and imposes the standards on all.
- The bottom-up approach: The standards reflect the consensus of the individual test developers.
- The interactive approach: A combination of the top-down and bottom-up approaches
Theoretical Frameworks
- Good practices in language testing
- Validity and validation
- Test administration and use
- Test fairness, etc.

Language Testing Standards

Contextual Features
- Identify the features of the local context
- Investigate the current situation of language testing practices

Considerations for validity

Considerations for validity
5. Challenges facing standards development and validation

Challenge 1: The collection, review, and critique of the standards in existence

What standards to include, and what standards to exclude: The criteria for standards selection

How to review the standards in the collection: Theoretical frameworks and research methodologies

How to build upon the existent research on language testing standards in the generation of our own standards?
How to determine the macro-structure of the standards?

If some qualities (e.g. practicality) are more relevant to the local contexts, shall these qualities be prioritized over others in the standards?

If some areas are identified as particularly problematic in the current language testing practices, shall these areas be emphasized in the guidelines?

If requirements of the local features run into conflict with what is generally held as proper in language testing (e.g. protecting the privacy of test candidates’ scores), how shall these requirements be reflected in the guidelines?
The ILTA Code of Ethics (ILTA, 2000): The failure to uphold the Code may have serious penalties, such as the withdrawal of the ILTA membership on the advice of the ILTA Ethics Committee.

ETS Standards for Quality and Fairness (ETS, 2002): The audit requirements – help to ensure that ETS products and services will be evaluated with respect to a uniform, rigorous set of standards through a well-documented process.
Are enforcement mechanisms feasible?

Are standards without enforcement mechanisms not meaningful/valuable?

Can the standards without enforcement mechanisms be adopted and adapted by individual test developers so that effective local enforcement mechanisms can be built?

How to develop the enforcement mechanisms and to ensure their implementation in language testing practices?
Challenge 4: Standards validation

- Are the standards applicable to the testing context?
- How to validate the standards?
- How to continuously improve the standards based on the validation research?
- Shall standards validation include the investigation of the impact and consequences of the standards?
- Who shall be held responsible for investigating the validity of the standards?
- Who shall get involved in the process of standards validation?
The language testing standards developed and implemented in different parts of the world have clearly indicated the pursuit for better quality and professionalism in language testing and assessment (see also Davies, 1997, 2004; Boyd & Davies, 2002).

The salient features of language testing in the China call for a set of professional standards which can cater to the needs and circumstances of language testing in the Chinese context (see also Yang & Gui, 2007; Fan & Jin, 2011, 2013).
A new set of standards will help to raise the awareness of professionalism among test developers and enhance the involvement of other stakeholders in the process of language testing.

The standards shall be targeted at both test developers and the other stakeholders, including test takers, EFL teachers, test users, publishers, curriculum designers, and officials with educational or examinations authorities, etc.

An interactive model shall be adopted in the generation of the standards with a view to both reflecting the theoretical frameworks in language testing and local features.
Future studies

Building a corpus of language testing standards:
Standards selection, review and critique; distilling useful experience and avoiding pitfalls (see Fan & Jin, 2010, 2012).

Investigating the current language testing practices in the Chinese context: Identifying the gap between the current practices and the practices as reflected in the professional standards (see Fan & Jin, 2013).

Investigating stakeholders’ views and perceptions of good testing practices: Involving stakeholders in standards development.

Investigating the validity of the standards: Applicability, usefulness, impact and consequences, etc.
7. References


8. Appendix: Standards in this presentation

The preparation of this presentation was supported by the National Social Sciences Fund (Guojia Sheke Jijin) under the project of “The development and validation of standards in language testing” (Project No. 13CYY032).
New Directions English: Language Assessment in China
“英语新方向”系列研讨会：中国英语语言评测
New Directions English: Language Assessment in China

Jinsong (Jason) FAN, Ph.D.
jinsongfan@fudan.edu.cn