

Implementation & Validation of Language Proficiency & Interpreter Readiness Tests

Jean Turner, Ph.D.
Monterey Institute of International Studies &
Second Language Testing, Inc.
Expert Panel on Community Interpreter Testing and Certification
Minneapolis, Minnesota
June 12 - 15, 2007

A program of four tests....

- Language Proficiency Test (Spanish, Cantonese, Hmong)
- English Language Proficiency Test
- Initial Interpreter Readiness Test
- Final Interpreter Readiness Test

The Language Proficiency Tests

- Spanish, Cantonese, and Hmong Language Proficiency Tests
- Designed & developed by Claudia Angelelli, Ph.D. (San Diego State) for the Connecting Worlds partnership in 2003—funded by The California Endowment;
- English Language Proficiency Test (Developed by Renee Jourdenais & Jean Turner)—based on the Language Proficiency Test design
- Original purpose—to determine whether an applicant for the training as community medical interpreter offered by one of the Connecting World partners has the language proficiency needed.

Features of the Language Proficiency Tests

- 6 tasks tap listening, reading, and speaking skills (register, vocabulary, accuracy)
- Example tasks—
 - respond to comprehension questions regarding a medicine bottle label (reading) or a patient's description of his/her medical condition (listening);
 - listen to a health care provider describe a medical procedure then explain the content to a patient using language the patient will understand
- Audio/CD delivery; 35 minutes administration time
- Scored by two trained raters

The Interpreter Readiness Tests (Spanish, Cantonese, Hmong)

- Designed by Claudia Angelelli, Ph.D. for the Connecting Worlds partnership in 2003—funded by The California Endowment;
- Original Purposes
 - Initial Interpreter Readiness Test
 - Given as a screen or pretest, prior to interpreter training
 - To determine whether an individual has the requisite language and interpreting skills to benefit from a medical interpreting training program
 - Final Interpreter Readiness Test
 - Given following training
 - To determine whether an individual has the requisite language and interpreting skills to function as a community medical interpreter

Features of the IR tests

- Both IR tests consist of 4 short videotaped segments of simulated interactions between a healthcare professional and a patient.
- The patient is Spanish-speaking patient and the healthcare professional is English-speaking.
- Examinees take the role of interpreter as they view the videotape.
- Responses are recorded for scoring (by 2 trained scorers).

Scoring

- Tests scored by trained raters who are proficient in English and Spanish
- Scoring rubric includes the script for each segment organized by turns
- Critical components of turns are scored separately for accuracy (1 pt = accurate, 0 = inaccurate)
- Not all turns are scored (introductions, greetings, and irrelevant information)

Initial IR tasks

- Patient makes and verifies appointment
- Patient changes appointment
- Appointment with doctor
- Urgent Care appointment request

Final IR Tasks

- Patient makes appointment
- Patient sees a specialist
- Sight translation of consent form and patient admitted to hospital
- Patient's post-op appointment

First phase of implementation & validation work (completed in 2006)

- **Goals—**
 - To support implementation of the tests at each of the five Connecting Worlds community based organizations (CBOs)
 - To identify & train a pool of raters for each site
 - To determine rater reliability
 - To perform initial investigations of tests' validity
 - Impact of training on test performance
 - Impact of amount of experience on test performance

Reliability—SLTI master rater and site rater

Test	N	Reliability
English LPT	51	.835
Spanish LPT	47	.750
Cantonese LPT	18	.670
English-Spanish Initial IR	38	.818
English-Spanish Final IR	24	.840
English-Cantonese Initial IR	19	.655
English-Cantonese Final IR	6	.550

Impact of training (Spanish IRs only)...

- Relationship between amount of training and score on Initial or Final IR Test.
 - Amount of training = 40 hrs or less OR 50 + hrs
 - Independent-samples *t* tests
- No significant differences in scores of participants in the two groups on either the Initial or Final IR
 - Perhaps no differences were apparent because of the small differences in the amount & type of training for the two groups.

Impact of amount of experience (Spanish IRs only)...

- **Relationship between amount of experience and score on Initial or Final IR**
 - Independent-samples *t* tests
 - Amount of experience = 1 yr or less OR 1+ yrs
- **Significant difference ($p < .05$) on the Initial IR**
- **No significant difference between groups on the Final IR**
 - Small contrast in experience may explain the outcome – larger sample with more varied amount of experience needed for more conclusive findings

Second phase of SLTI validation work...

- The California Endowment shared the Spanish tests with *Hamblamos Juntos*, which supported the development of a computer-delivered version of the Spanish Language Proficiency test and The Initial and Final IR tests.
- Using the larger data set from the Hamblamos Juntos organization's work, SLTI will:
 - Conduct an item analysis and remove any unnecessary scoring units from the test.
 - Make revisions to the computer-based versions of the tests to make them more user-friendly and to make stimuli more authentic.
 - Fix all problems with computer based forms.
 - Revise and record the scripts for Form 2 of the tests and implement them in computer-based form.

Thoughts on the experience...

- 129 Empirical basis for test may provide convincing evidence of the content validity of a test...
- 130 But collecting evidence to support the usefulness of a test (its value for screening purposes, for example) is challenging...
- 131 as is investigating the meaningfulness of test scores (What is a good score? At what score can one be assumed to be 'qualified' or a 'master'?).
- 132 When thinking about recruiting field test participants, I ask myself—
 - Who likes to take tests anyway? (Are these people typical??)
 - Does \$50 alleviate test trepidation?
- 133 When thinking about the characteristics of potential raters, I ask myself—
 - What are the characteristics of good potential scorers/raters?
 - How can a pool of raters be maintained?
- And many other theoretical, logistic, and practical concerns related to designing and conducting research on the reliability and validity of tests such as these.