The Art and Science of Language Testing

Kathy Tinios-Rose, M.S.
Jeanné Makiney, Ph.D.
Overview

- Getting to know you . . .
- Language Proficiency and Interpreting Skill
- Employment vs. Certification
- Test Types
- Language Software
- Cultural Norms
- Scoring
- Candidate Feedback
- Test Security
How can we help you?

- What experiences have you had with language testing?
- What are your language testing needs?
- What information do you want to take with you after leaving this tutorial?
Language Proficiency

Candidate’s knowledge, skill and ability with language

- **Native / Heritage Speakers**
  - Grew up speaking it in the home

- **Learned Language Speakers**
  - Formal education in the States
  - Formal education in the target country
Interpreting Skill

- Candidate’s skill or ability to accurately render information from one language to another while maintaining equivalent language level (Register)

- Candidate’s skill or ability to interpret at the appropriate speed for the required profession (if speed is a factor)
Crossover

- English language skill at college graduate level
- Target language skill at 8th grade
- Able to understand English at high level but renders it at a lower level due to lack of language proficiency
- Target language skill at college graduate level
- English language skill at 8th grade
- Cannot understand proceedings in English; therefore unable to render information accurately

Maintaining Language Register can be critical in high-stakes professions such as court interpreting.
Employment vs. Certification

Purpose of the examination – *What* are you attempting to measure and *Why*?

- Identifying level of language proficiency required is critical
- Is the skill of interpreting necessary or do you need to measure language skills in just one language (English or Target)?

**JOB ANALYSIS AND LANGUAGE ANALYSIS ARE THE CRITICAL COMPONENTS TO EITHER TYPE OF TESTING**
Cognitive and Physical Requirements of the Job

What do they do?
- Read written documents
- Hear
- Speak or Translate
- Write / Type

What terminology is specific to that occupation?
- Medical / Legal / Technical
Type of Test Needed

Match the exam(s) to the skills

- **Written tests** effectively measure
  - reading comprehension, grammar, vocabulary, choosing translations (antonyms, synonyms, sentences) and job specific knowledge (e.g., ethics standards)

- **Oral tests** effectively measure
  - Spoken language proficiency
  - Interpreting skills

- Most likely will use a combination of both written and oral
Written vs. Oral Testing

Getting the most value from test administration funds

- **Written tests**
  - Easy to administer to large groups of candidates
  - Less expensive
  - Easy to score

- **Oral tests**
  - One proctor per candidate
  - Multiple raters
  - Objective criteria difficult to define

- Consider using *multiple-hurdle approach*
Test Development Challenges

- Multiple Languages
- Software
- Cultural Norms
- SME Characteristics
Maintaining consistent standards across multiple languages

- Passage in one language may have context inappropriate for another language
- Back translations may appear artificial
- Not all languages have gender pronouns
- Idiomatic language
- Maintaining register/level
  - Some languages match structure of English (e.g., Spanish) closer than others (e.g., Cantonese)
  - Different amounts of lag time before translations can commence
Language Software

- Need software compatible with language experts’ software
  - Allows language experts to do some work from home
  - Language experts able to use compatible software during development meetings
  - What programs are used by your language experts?

- Microsoft Multilanguage Pack flexible for multiple languages but doesn’t work with all languages
  - May need companion software (e.g., Armenian keyboard software, Twin Bridges helpful for many Asian languages)
  - Alternate text entry via electronic pen and pad (Cantonese)
Cultural Norms

- Issues that affect communications between language experts
- Based on culture and known implicitly
- Often based on age, gender and national origin of the experts
- Impact of norms is evident in some languages but not others
- Based on the composition of any given expert team
“Americanized”

- Experts are culturally savvy to American ways
- Gender issues are non-existent
  - In Japan, the wife walks 10 steps behind her husband. In America, the wife walks all over her husband.
- Age issues are non-existent
  - Young people are seen, but not heard.
  - Young people are viewed as fresh perspective.
Problem Children

- Team consisted of three females
  - Ages include 54, 35, and 23
- Clashes noted between 54 year old and 35 year old
- 23 year old is quiet most of the time; often agrees with both so as not to make any enemies.
- Introduced one mid-fifties male with Ph.D. in that language
- 54 year old female defers to male
- 35 year old female is Americanized and rejects cultural expectations
- Conflict occurs – Male states: “Do you understand who I am?”
Illness and Remedies

- Recognizing a problem with your experts
  - Single SME dominating conversation, overriding others in group
  - Body language, tone of voice, facial expressions
  - Lack of progress toward team goals

- Orientation and instruction
  - Helping vs. hindering group dynamics
    - Caution: May move cultural issues underground
  - Passive-aggressive behaviors

- Change the makeup of team members
Commonalities Across Languages

“Personality traits”
- (e.g., passionate about language, outgoing, gregarious)

Often debate endlessly over “minor” word issues

Always looking for another way to say any given thing
Scoring Options

Holistic scoring
- Score using more global criteria (i.e., Language Proficiency, Interpreting Skills)
  - Need well-defined rating scales
  - Can take into account overall impressions such as structure and connectedness between words

Quantitative scoring
- Score on key words/units of meaning only
  - Easier to score objectively
  - Lose ability to take into account connecting structure between words
  - Easier to compromise (keywords are shared by candidates)
Scoring Options

Component scoring

- Score by component or across components
  - Cognitively less demanding to score by component
  - Scoring by component is non-compensatory, candidates must pass scales for each component

<table>
<thead>
<tr>
<th>Language Proficiency</th>
<th>English</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Language</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Interpreting Skills</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Scoring Options

Passing standards

- **Certification**
  - Gatekeeper for entry to profession
  - Need to ensure candidates will be minimally competent and able to function day 1 after certification

- **Selection**
  - Rank order candidates?
  - Pass/Fail?
Candidate Feedback

- Purpose intended for testing instrument
  - Employment
  - Licensure and Certification
  - Training / Skills Improvement
- Written test
  - Subtest scores
- Oral test
  - Rater comments that don’t compromise test security
Protecting your investment

- Maintaining Test Security
- E-mail test files to/from experts with password protection
- Maintain control of materials during expert meetings
  - Take inventory before and after sessions
  - Collect notes written on notepads
Test Security During Administration

- Candidate hides videotape recorder in jacket sleeve
- Candidate repeatedly failed the sight interpretation component
Wrapping Up

- Questions
- Comments
- Additional language testing stories
THANK YOU!