Developing a language assessment literacy survey for use across different contexts

Luke Harding (Lancaster University)
Benjamin Kremmel (University of Nottingham & University of Innsbruck)
Outline

• The initial idea
• Developing version 1.0
• Re-thinking the approach
• Developing version 2.0
  • Expert review
  • Pre-testing
• Where next?
The initial idea ...

- Generate empirical (descriptive) evidence to support prescriptive claims for LAL across different stakeholder groups
- Specifically, investigate stakeholder perceptions of needs and lacks
- Starting point: Taylor’s (2013) LAL profiles
Taylor’s LAL profiles (2013, p. 410)

Test writers

University admins

Teachers

Professional language testers/researchers
Version 1.0

• Simplified definitions of 8 LAL dimensions (Taylor, 2013)
• 6 stakeholder groups
  • Language teachers
  • Language test developers
  • Language testing researchers
  • Applied linguists
  • Policymakers (educational politics, uni admissions, immigration, etc.)
  • Test takers
**Knowledge of theory**

- Knowledge about models of language and proficiency as well as about the theoretical constructs of the language sciences

<table>
<thead>
<tr>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language teachers</td>
</tr>
<tr>
<td>Language test developers (professionals who create tests, write questions, develop scoring guides)</td>
</tr>
<tr>
<td>Language testing researchers (professionals who conduct research on language testing matters)</td>
</tr>
<tr>
<td>Applied linguistics researchers (professionals who conduct research on language and language learning)</td>
</tr>
<tr>
<td>Policy-makers (government officials, immigration officers, etc. who might use language test scores)</td>
</tr>
<tr>
<td>Test takers (learners who might need to take a language test)</td>
</tr>
</tbody>
</table>

**Kremmel & Harding (2015)**

- Pilot study ($N=52$; 11 countries)
- Feedback:
  1. Survey generally comprehensible
  2. Some difficulty with choosing labels
  3. Some categories (e.g., personal beliefs) less clear
Re-thinking the approach

- Single-item approach lacks validity
- Need to find out whether these are actually separable “dimensions” of LAL
- Multi-item approach would have greater diagnostic potential
- Survey would need to be:
  - Comprehensive but feasible
  - Intelligible across a range of stakeholders
Developing version 2.0
Early example items (version 2.0)

Knowledge of theory

Knowledge about ...

• the structure and use of language
• how foreign/second languages are learned
• what it means to be a proficient user of a language
• the different language skills (reading, writing, speaking)
• ...

Aim for ...

• Simple language
• Short items
• Glosses where necessary
Initial challenges (versions 2.0 – 2.3)

- Elaborating / dividing domains
- Describing group/professional identity labels
- Refining scale wording
- Balancing coverage with feasibility
# Expanding / dividing domains

<table>
<thead>
<tr>
<th>Taylor's (2013) domains</th>
<th>Revised domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of theory</td>
<td>Theoretical knowledge about language and language learning</td>
</tr>
<tr>
<td>Technical skills</td>
<td>(A) Language assessment construction</td>
</tr>
<tr>
<td></td>
<td>(B) Language assessment administration/scoring</td>
</tr>
<tr>
<td></td>
<td>(C) Language assessment evaluation</td>
</tr>
<tr>
<td>Principles and concepts</td>
<td>Principles and concepts</td>
</tr>
<tr>
<td>Language pedagogy</td>
<td>Language pedagogy</td>
</tr>
<tr>
<td>Sociocultural values</td>
<td>Impact and sociocultural values</td>
</tr>
<tr>
<td>Local practices</td>
<td>Local practices</td>
</tr>
<tr>
<td>Personal beliefs/attitudes</td>
<td>Personal beliefs/attitudes</td>
</tr>
<tr>
<td>Scores and decision making</td>
<td>Scores and decision making</td>
</tr>
</tbody>
</table>
### Describing groups/professions

<table>
<thead>
<tr>
<th>Version 1.0</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Language teacher</td>
<td>1. Language teacher</td>
</tr>
<tr>
<td>2. Language test developer</td>
<td>2. Professional examiner or rater</td>
</tr>
<tr>
<td>3. Language testing researcher</td>
<td>3. Language test developer (professional who creates tests, writes questions, develops scoring guides, etc.)</td>
</tr>
<tr>
<td>4. Applied linguist</td>
<td>4. Language testing researcher (professional who conducts research on language testing matters)</td>
</tr>
<tr>
<td>5. Policymaker (educational politics, uni admissions, immigration, etc.)</td>
<td>5. Policy-maker (government official who sets educational goals and assessment policies)</td>
</tr>
<tr>
<td>6. Test taker</td>
<td>6. Test score user (university admissions staff, immigration officer, etc. who might use language test scores for decision making)</td>
</tr>
<tr>
<td></td>
<td>7. Test taker (learner who might need to take a language test)</td>
</tr>
</tbody>
</table>
Refining scale wording

0 = illiteracy
1 = nominal literacy
2 = functional literacy
3 = procedural and conceptual literacy
4 = multidimensional literacy

(Pill & Harding, 2013)
Refining scale wording

• Version 1.0 scale:

0 means no knowledge at all and 4 means a comprehensive or expert knowledge.

• Version 2.3 scale

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>no knowledge/skill at all</td>
<td>a small level of knowledge/skill</td>
<td>a moderate level of knowledge/skill</td>
<td>a high level of knowledge/skill</td>
<td>A very high/fully comprehensive level of knowledge/skill</td>
</tr>
</tbody>
</table>
Balancing coverage with feasibility

• Version 2.3:
  • 70 items ($M = 7$ per domain, min. 5 – max. 12)
  • 5-10 minutes to complete all questions
Expert review

• 6 x professors / senior academics in the field of language testing / assessment

• Asked to complete (paper-based) survey and feedback on:
  • anything odd/out-of-place
  • any glaring omissions within each domain category
  • any less relevant items which could be taken out
  • any other general views on the questionnaire
Summary of expert comments

I think you've done a terrific job of articulating the various areas identified as making up AL, without making it too long or complex that no one is going to want to complete it.

- Glosses for technical terms
- Grammatical cohesion across whole survey
- Suggestion to include parents of test-takers as another group

I feel that it might be a bit exhaustive for two of the stakeholder groups: language teachers and test takers. I piloted it with a couple of my pre- and in-service teachers and uni students who confirmed this. Also policy-makers and test score users will also find some parts of it difficult to understand or identify with.
For your chosen group/profession, please indicate how comprehensive you think your group/profession's knowledge or skill needs to be regarding each aspect of language assessment listed below, where 0 means no knowledge/skill at all and 4 means a very high level of knowledge/skill.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no knowledge/skill at all</td>
<td>a small level of knowledge/skill</td>
<td>a moderate level of knowledge/skill</td>
<td>a high level of knowledge/skill</td>
<td>A very high level of knowledge/skill</td>
</tr>
</tbody>
</table>

What level of knowledge do people in your chosen group/profession need to have about ...

1) how to use assessments to inform learning or teaching goals
2) assessment traditions in the relevant local context
3) what a particular score indicates about an individual's language ability
4) how to communicate assessment results and decisions to students, parents, other lay audiences
5) how to recognize unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.
6) how assessments might influence classroom teaching and learning
7) how to interpret figures which show measurement error
8) their own beliefs/attitudes towards language assessment
Pre-testing

• In progress …
• 14 participants so far:
  • 7 x language teachers
  • 2 x test score users
  • 4 x test takers
  • 1 x parent of test taker
• Low request to completion ratio
• Evidence of incomplete surveys in Qualtrics system
Pre-test feedback

- How clear was the survey?

  1 = not clear at all – 5 = very clear
Pre-test feedback

• How confident were you in your answers?
  Sliding scale: 0% – 100%
  Range = 55 – 95
  Mean = 81.65 (SD=12.06)
Some reported challenges

- Interpretation
  - I had to re-read several of the statements to interpret what was being asked. (Teacher)
  - I teach MFL in secondary schools and I have never heard of the CEFR, hence that score is low. (Teacher)
Some reported challenges

• Selecting a group/profession

- Answered this as a "senior user" with a background knowledge of ELT and testing (used to be an English teacher). As such a senior user, I feel able to input an opinion on which tests we should use for the purpose of admission onto a degree programme. If the same questions were asked of admissions assistants, for example, the questions may be answered very differently. (Test score user)
Some reported challenges

• Format
  - Some statements are repeated from previous pages and I find myself trying to remember what score I gave previously. (Teacher)
  - THE DIFFERENCE BETWEEN QUESTIONS 28 AND 36. (Test taker)
  - Also, the table that is right at the top of this page - not sure of its purpose/couldn't enter the answer immediately (wanted to say 4 - language teachers need a very high level of the knowledge). (Teacher)
Emerging pre-test results (teachers)

- the concept of validity (how well an assessment measures what it is intended to measure) ($M = 4.86$)
- how assessments might influence classroom teaching and learning ($M = 4.86$)
- how to provide useful feedback on the basis of an assessment ($M = 4.71$)
- how different language skills develop (e.g., reading, listening, writing, speaking) ($M = 4.71$)
Emerging pre-test results (teachers)

- the history of language assessment ($M = 2.16$)
- how language assessments might impact upon society (either positively or negatively) ($M = 2.86$)
What next?

• Continue to improve clarity based on feedback:
  • Clearer instructions and formatting
  • Optional examples for selecting group/profession
  • Textual enhancements in stems

• If you would like to be involved in trialling the survey please let us know!