A TEST DEVELOPMENT MODEL FOR DEVELOPING A LOCALISED TEST
TEST DEVELOPMENT AND VALIDATION APPROACH
Socio-Cognitive Model (O’Sullivan 2014; in press)
Customising tests for local needs

Locally appropriate  test  Globally Neutral
## LOCALISATION

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 0</strong></td>
<td>Aptis General (or other existing variant) in a <strong>full, four-skills</strong> package</td>
<td>User selects a four-skills package of any Aptis (General or variant) available for use.</td>
</tr>
<tr>
<td><strong>Level 1</strong></td>
<td>Options for localisation are limited to selection from a <strong>fixed range of pre-existing features</strong>, such as delivery mode and/or components</td>
<td>User is able to select the skills to be tested and/or the mode of delivery that are appropriate. For example, the Reading package (Core component + Reading component) of Aptis General, taken as a pen-and-paper administration.</td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td><strong>Contextual localisation:</strong> <em>lexical, topical modification</em></td>
<td>Development of specifications for generating items using existing task formats but with topics, vocabulary, etc. relevant for specific domains (e.g. Aptis for Teachers, projects in India, Rwanda, Saudia Arabia).</td>
</tr>
<tr>
<td><strong>Level 3</strong></td>
<td><strong>Structural reassembly:</strong> changing the number of items, proficiency levels targeted, etc., <strong>using item-bank content.</strong></td>
<td>Developing a test of reading targeted at a specific level, e.g. B1, using existing task types and items of known difficulty calibrated to the Aptis reading scale.</td>
</tr>
<tr>
<td><strong>Level 4</strong></td>
<td>Partial <strong>re-definition of target construct</strong> from existing variants. Will involve <strong>developing different task types</strong> to elicit different aspects of performance,</td>
<td>Developing new task types that are more relevant for a specific population of test-takers, while remaining within the overall framework of the Aptis test system (<em>e.g. Aptis for Teens</em>).</td>
</tr>
<tr>
<td><strong>Level 5</strong></td>
<td>The construct and/or other aspects of the test system are changed to such an extent that <strong>the test will no longer be a variant within the system.</strong></td>
<td>For example, developing a matriculation test for uses within a formal secondary educational context; developing a certification test available to individuals rather than organisations, etc.</td>
</tr>
</tbody>
</table>
LOCALISATION

Stakeholder engagement

- Needs analysis
- Pilot and pretest
- Test delivery

Assessment Development
- Test production
- Rating scales

Assessment Research
- Test analysis
- Test validation
TEST DEVELOPMENT CYCLE

Rationale for test development

Project planning

Needs analysis

Test Panel Review

Design & specifications

Revision of content & specifications

Piloting / pretesting & analysis

Item writing and quality assurance

Rater / Item Writer training

Test delivery (live administration)

Monitoring
STAKEHOLDER ENGAGEMENT
STAKEHOLDER ENGAGEMENT & VALIDITY

Messick’s 1989 unified approach to validity
The socio-cognitive model: test-taker at the core

Important to understand
- Target Language Use (TLU) domain the test-takers are being tested for
- test-taker profile
- context of test delivery

Stakeholders:

“the test designers, teachers, students, score users, governments or any other individual or group that has an interest in how the scores are used and whether they are useful for a given context” (2007, p. 14).

Fulcher & Davidson (2007)
THE CONTEXT

English Medium of Instruction (EMI) high school in China
International curriculum
Prestigious:
- likely precursors to top international academic opportunities
- places to develop bilingual or near-bilingual proficiency in academic English

3,000 applicants per year
Extremely high-performing students apply
10% acceptance rate
Battery of placement tests: English proficiency one of the key discriminating factors

Existing Aptis versions:
- Aptis for Teens
- Aptis General
- Aptis Advanced

Not fit-for-purpose:
- Too easy
- Irrelevant topics
STEPS IN STAKEHOLDER ENGAGEMENT

- Feedback and reflection
- In-depth discussion with management - Needs Analysis tool
  - decision-making criteria
  - decision-making process
  - test-administration requirements and constraints
  - teaching faculty
- Survey of teachers
- Analysis of curriculum, textbooks and examination papers
- Limited classroom observation
3. Which of the following do your students need to listen to in English for their studies? (Both live and broadcast through media)

<table>
<thead>
<tr>
<th>Classroom lectures and presentations</th>
<th>Never (not needed)</th>
<th>Rarely</th>
<th>Sometimes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptions and explanations of processes or systems</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Debates and discussions</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Announcements and instructions</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Interview and opinions</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Extended narrations and performances</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Fact-rich information in a format</td>
<td>Never (not needed)</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
</tbody>
</table>

Which of the following do your students need to read in English for their studies? (Both paper-based and on-screen)

Which of the following do your students need to listen to in English for their studies? (Both live and broadcast through media)
RESULTS OF STAKEHOLDER ENGAGEMENT

Key requirements and challenges

- *Discriminate at the higher levels (B1-C2)*
- *Reflective of actual communicative demands in the classroom*
- *Reflective of textbook topics*
- *Timing during test administration*

Survey of teachers and analysis of textbooks & exams facilitated a more granular understanding of the TLU domain:

- *Communicative tasks students are required to engage in*
- *Topics and text genres they encounter*
- *Variety of Englishes they are exposed to*
**Task:** Write an article for an online magazine using a set of graphically-presented research notes

**Task:** Write an email to a university professor about a topic related to possible future study.
ASSESSMENT DEVELOPMENT
<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>Aptis General (or other existing variant) in a <strong>full, four-skills</strong> package</td>
<td>User selects a four-skills package of any Aptis (General or variant) available for use.</td>
</tr>
<tr>
<td>Level 1</td>
<td>Options for localisation are limited to selection from a <strong>fixed range of pre-existing features</strong>, such as delivery mode and/or components</td>
<td>User is able to select the skills to be tested and/or the mode of delivery that are appropriate. For example, the Reading package (Core component + Reading component) of Aptis General, taken as a pen-and-paper administration.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Contextual localisation: <strong>lexical, topical modification</strong></td>
<td>Development of specifications for generating items using existing task formats but with topics, vocabulary, etc. relevant for specific domains (e.g. Aptis for Teachers, projects in India, Rwanda, Saudia Arabia).</td>
</tr>
<tr>
<td>Level 3</td>
<td><strong>Structural reassembly</strong>: changing the number of items, proficiency levels targeted, etc., <strong>using item-bank content</strong></td>
<td>Developing a test of reading targeted at a specific level, e.g. B1, using existing task types and items of known difficulty calibrated to the Aptis reading scale.</td>
</tr>
<tr>
<td>Level 4</td>
<td>Partial <strong>re-definition of target construct</strong> from existing variants. Will involve developing different task types to elicit different aspects of performance,</td>
<td>Developing new task types that are more relevant for a specific population of test-takers, while remaining within the overall framework of the Aptis test system (<em>e.g.</em> Aptis for Teens).</td>
</tr>
<tr>
<td>Level 5</td>
<td>The construct and/or other aspects of the test system are changed to such an extent that the test will no longer be a variant within the system.</td>
<td>For example, developing a matriculation test for uses within a formal secondary educational context; developing a certification test available to individuals rather than organisations, etc.</td>
</tr>
</tbody>
</table>
PRINCIPLED COMPROMISES

- Time to develop tests – how long does it take to develop a test?
- Length of time for test sections – where can we reduce test taker time?
- Variety of Englishes
- Item writer training
- CBT / pen & paper format – logistics
Context of Use

Consequence/Impact

Revision of content & specifications
Monitoring
Design & specifications
Rater Training
Needs analysis
Test Panel Review

Rationale for test development
Project planning
Implementation

Revises for test development

Consequence/Impact

Test Takers
Parents & Guardians
Employers
Teachers
School Principals
School Administrators
School Boards
Examination Boards
Test Administrators
Education Boards
Broader Society
Policy Makers
Test Developers
Academics
Lawyers

www.britishcouncil.org
CONCLUSIONS

• Collaboration between teams
• Localisation and engagement with stakeholders
• Principled compromises
THANK YOU

Sheryl.Cooke@britishcouncil.org
Judith.Fairbairn@britishcouncil.org
Jamie.Dunlea@britishcouncil.org
Kevin.Rutherford@britishcouncil.org
Richard.Spiby@britishcouncil.org