Test specifications and item writer guidelines in a multilingual world: issues and solutions

John Tucker
British Council
john.tucker@britishcouncil.pl

ALTE
April 2014
Paris
Galileo Galilei

Often referred to as 'the father of modern physics', Galileo Galilei was born in Pisa, Italy, in 1564, the son of a mathematician and musician.

He attended university in Pisa but had to leave due to a lack of funds, and later taught sciences at the University of Padua. It was during his time there that Galileo did a large number of experiments, the most famous involving dropping balls of different sizes from various heights to determine the law of acceleration of falling bodies. Indeed, he is credited with several important scientific discoveries, and is still considered a great genius.

Unfortunately, however, he died in prison in 1642, whilst serving a life sentence for publishing works suggesting that the earth moved around the sun – something that went against accepted thinking at the time.
Outline of this talk

• Glossary – terms used in this talk

• Who are our item writers?

• Different types of specifications for different audiences

• Training for item writers within a Language Assessment Literacy context

• Examples of a specification pro forma and specifications for one task/level

• Example of how specifications help in writing and reviewing a test item in development

• Where multilingualism comes in: attempts to avoid cultural bias
Glossary – some terms used in this talk

**items** in a language test - questions, tasks

**input text** - stimulus content (reading text/listening text/s, single words or phrases (vocabulary), structural/functional content (grammar)

**distractor** - (incorrect) options in a multiple choice test that are meant to be plausible

**rubric** - instructions to explain what the task requires the test taker to do

**target** – the response that a test taker needs to give to get an item right

**lexical profile / K lists** – how difficult a text is lexically based on frequency lists

**Quality Reviewers** – people who check that test items meet specifications and are suitable for a particular test

**Specifications** – “a generative explanatory document for the creation of test tasks” (Davidson and Fulcher, 2007)
item writers

part-time writers who are commissioned to produce particular test items within certain parameters

teaching or exams background

currently represent more than a dozen countries

Diverse L1 backgrounds
What information should specifications include? Why do we need them?


The specifications which are finally produced should give detailed information on each paper or part of the test. They will cover the following areas:

• how long the test lasts
• what is being tested, e.g. use of grammatical rules
• how many items there are in each section
• text types used in input
• marks given for each item and total marks available
• total and individual length in words of texts used
• where and when the test can be taken
• format and length of tasks
• details of criteria for assessing free writing tasks and tests of oral production
• the focus of the item, e.g. showing detailed comprehension of a text
More about specifications

• Specifications can, and should, be written at several levels of specificity depending on the audience, with those for item writers and test developers being the most detailed.

• Specifications should provide item writers of various L1 backgrounds with a clear reference when they come to produce questions and tasks.

• Candidate Guide is a version of test specifications with less detail that helps candidates and teachers in preparing to take the test.

• A variety of other resources are available, e.g., Compleat Lexical Tutor [http://www.lextutor.ca](http://www.lextutor.ca)
The Complet Web VP!

Two languages, five list frameworks - NOW INCLUDING BNL! - for clear comparisons

Note that BNL, Coca-Core, and one part of Classic (the AWL) are not full 1000-family lists.

How to make specific list comparisons? See Demo 8 here.

Input mode A: Type or paste smaller text (<50,000 words) below and click Submit_window.

Title: Untitled

INSTRUCTIONS: Type or paste your text here and click the SUBMIT_WINDOW button. VocabProfile will tell you how many words the text contains from frequency bands as determined through analysing a number of research corpora. For a demonstration, enter one of the sample texts below.

TEXT SET-UP
General: Include an empty space after every comma or full stop.
Research: Deal with spelling errors and proper nouns.

SIZE LIMITS: Web input is normally limited to 30,000 characters - use UPLOAD method below for larger files (up to 150,000 words Oct 2012); must be .txt;
(send in straight from your own hand drive). Text is NOT stored on Web VP's server.

Input mode B: Upload larger text files (150,000 words= 30 secs, Sept 2013) >> LARGER TEXT = LESS RICH INFO (FAMS+VP: Neg up to 100,000). PROFILE ONLY THEREAFTER.

Upload space (for - txt files, encoded UTF8 if French)

1. Choose file, no file chosen
2. Choose framework at top; then submit file

Frequency frameworks by Nation, Davies, Gardner, Neufeld, Bilorglu, Lonsdale, Le Bras, - coded in PERL for Web by T. Cobb, UQAM Canada.
Training for British Council item writers

1 selected from a pool of competent education professionals based around the world

2 graduate level assessment module face-to-face (in the UK)

3 specific item writer training

4 hands-on training with mentoring and person-specific feedback

5 Commissioned item writers have their items reviewed by experienced quality reviewers

6 system will evolve further
contracted item writers (as of December 2013) - 30

in-house item writer training - 65

Roehampton course - 119
# B1 reading specifications (extract)

<table>
<thead>
<tr>
<th>Features of the Input Text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word count</strong></td>
</tr>
<tr>
<td><strong>Domain</strong></td>
</tr>
<tr>
<td><strong>Discourse mode</strong></td>
</tr>
<tr>
<td><strong>Content knowledge</strong></td>
</tr>
<tr>
<td><strong>Cultural specificity</strong></td>
</tr>
<tr>
<td><strong>Nature of information</strong></td>
</tr>
<tr>
<td><strong>Presentation</strong></td>
</tr>
<tr>
<td><strong>Lexical Level</strong></td>
</tr>
<tr>
<td><strong>Lexical Level; further criteria</strong></td>
</tr>
<tr>
<td><strong>Grammatical level</strong></td>
</tr>
<tr>
<td><strong>Avg sentence length</strong></td>
</tr>
<tr>
<td><strong>Topic</strong></td>
</tr>
<tr>
<td><strong>Text genre</strong></td>
</tr>
<tr>
<td><strong>Writer/Reader relationship</strong></td>
</tr>
</tbody>
</table>
# Extract of Quality Reviewer’s check list

## QR check list for reading

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R10</strong></td>
<td>correct structural range for level</td>
</tr>
<tr>
<td><strong>R11</strong></td>
<td>lexical profile (K lists and Flesch Kincaid)</td>
</tr>
<tr>
<td><strong>R12</strong></td>
<td>distractors are relevant</td>
</tr>
<tr>
<td><strong>R13</strong></td>
<td>only one answer is correct</td>
</tr>
<tr>
<td><strong>R14</strong></td>
<td>language is authentic</td>
</tr>
<tr>
<td><strong>R15</strong></td>
<td>Question fits purpose (eg B2: <em>demonstrate general comprehension of a single complex text.</em>)</td>
</tr>
<tr>
<td><strong>R16</strong></td>
<td>Item is appropriate for the level (CEFR)</td>
</tr>
</tbody>
</table>
Galileo Galilei

Often referred to as 'the father of modern physics', Galileo Galilei was born in Pisa, Italy, in 1564, the son of a mathematician and musician.

He attended university in Pisa but had to leave due to a lack of funds, and later taught sciences at the University of Padua. It was during his time there that Galileo did a large number of experiments, the most famous involving dropping balls of different sizes from various heights to determine the law of acceleration of falling bodies. Indeed, he is credited with several important scientific discoveries, and is still considered a great genius.

Unfortunately, however, he died in prison in 1642, whilst serving a life sentence for publishing works suggesting that the earth moved around the sun – something that went against accepted thinking at the time.

Example:

referred during lack at serving experiments
Galileo Galilei

Often [EXAMPLE] to as 'the father of modern physics', Galileo Galilei was born in Pisa, Italy, in 1564, the son of a mathematician and musician.

He attended university in Pisa but had to leave due to a [1] of funds, and later taught sciences at the University of Padua. It was [2] this epoch that Galileo conducted multiple [3], the most famous involving dropping balls of different sizes from [4] heights to determine the law of acceleration of falling bodies. Indeed, he came to be credited with several groundbreaking scientific [5], and is still considered a great genius.

Unfortunately, however, he perished in prison in 1642, whilst [6] a life sentence on a charge of heresy for publishing works positing that the earth moved [7] the sun. [119 words]

<table>
<thead>
<tr>
<th>Example</th>
<th>referred</th>
</tr>
</thead>
<tbody>
<tr>
<td>around [7]</td>
<td>taking</td>
</tr>
<tr>
<td>taking</td>
<td>lot</td>
</tr>
<tr>
<td>lot</td>
<td>shortage</td>
</tr>
</tbody>
</table>
Often [EXAMPLE] to as ‘the father of modern physics’, Galileo Galilei was born in Pisa, Italy, in 1564, the son of a mathematician and musician.

He attended university in Pisa but had to leave due to a [1] of funds, and later taught sciences at the University of Padua. It was [2] his time there that Galileo conducted a large number of [3], the most famous involving dropping balls of different sizes from [4] heights to determine the law of acceleration of falling bodies. Indeed, he is credited with several important scientific [5], and is still considered a great genius.

Unfortunately, however, he died in prison in 1642, whilst [6] a life sentence on a charge of heresy for publishing works suggesting that the earth moved [7] the sun – something that went against accepted thinking at the time.
Problems with item identified by specifications

- Item was too short

- Overall lexical content was too difficult: contained too high a percentage of words from K lists above the correct level

- Target words included key words above appropriate lexical level

- Text included structural (grammatical) elements from outside the relevant range

- More than one option was possible for one of the spaces

- “the earth moved [7] the sun” – since only one preposition (around) was available as an option, does the item test grammatical knowledge rather than reading skills?
Attempts to deal with cultural bias

Development of on-line space for item writers to ask/answer questions and collaborate – exists, for specific training needs

A small team of Senior Item Writers based around the world will review items for cultural bias - planned
Wiki site for item writers

Lexical Resources

last edited by John Tucker 1 week, 3 days ago

1. Compleat Lexical Tutor
Input text and lextutor instantly gives you the K list for individual words and the
a) change default setting to BNC-20 and
b) tick the box for ‘ALL mid-sentence capped words 1k.
http://www.lextutor.ca/vp/bnc/

2. BYU-BNC
(A searchable interface for the British National Corpus developed by staff at Brigham Young
Vocabulary-Collocation tasks at B1 and B2. It can help with collocations, synonym
http://corpus.byu.edu/bnc/
References


Specifications and test validation

Module 2
The test production process
Specifications
The production process
Commissioning
Vetting and editing
Pretesting
Item analysis
Item calibration
Establishing a difficulty scale
Anchoring
Score interpretation
Item Banking
Item Banking and test construction
Test construction without an item bank
Exercises
Suggested answers to the exercises
Appendix A – Rejected texts
Appendix B – MicroCAT examples for exercise 4
Appendix C – Further Reading
Appendix D – An example of the test development process
Specifications – theory meets practice
Any Questions?

john.tucker@britishcouncil.pl
Localization requires the test users to work with the developers.

Different levels of localization will require different amounts of resources.

The greater the differences to the general version of Aptis, the more time and resources will be required.

What is Aptis?

Locally appropriate

Resource intensive

test

Global Neutral

Cost-efficient
The main, or general version of Aptis was launched in August 2012. It is a test of general English proficiency for adults (16 years and older). The test consists of 5 components: Grammar & Vocabulary, Reading, Listening, Speaking, and Writing. Various modes for delivery are available: computer, pen and paper, and telephone.
Task specs: part of validation

Socio-cognitive framework for language test development and validation (Weir, 2005; O’Sullivan & Weir)