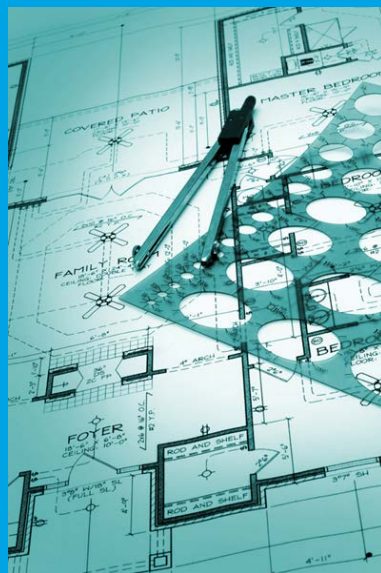


Test specifications and item writer guidelines in a multilingual world: issues and solutions

John Tucker
British Council
john.tucker@britishcouncil.pl



ALTE
April 2014
Paris

B1 reading item (from preparation materials, retired from a live test)

- 1
- 2
- 3
- 4



READING

Read the text and complete each gap with a word from the list at the bottom of the page.

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 of funds, and later taught sciences at the University of Padua. It was his time there that Galileo did a large number of , the most famous involving dropping balls of different sizes from heights to determine the law of acceleration of falling bodies. Indeed, he is credited with several important scientific , and is still considered a great genius.

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

Example:

referred

various

discoveries

taking

lot

around

during

lack

at

servicing

experiments

Preferences

Tutorial

Introduction

Previous

Flag Question

Next

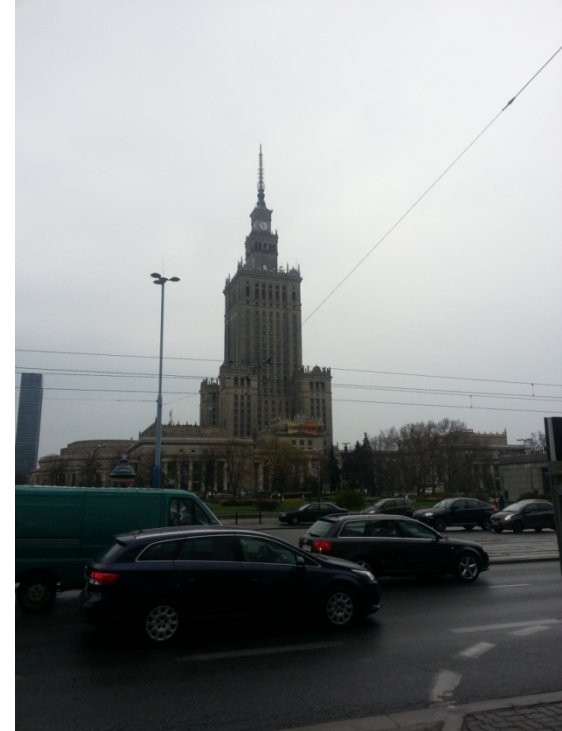
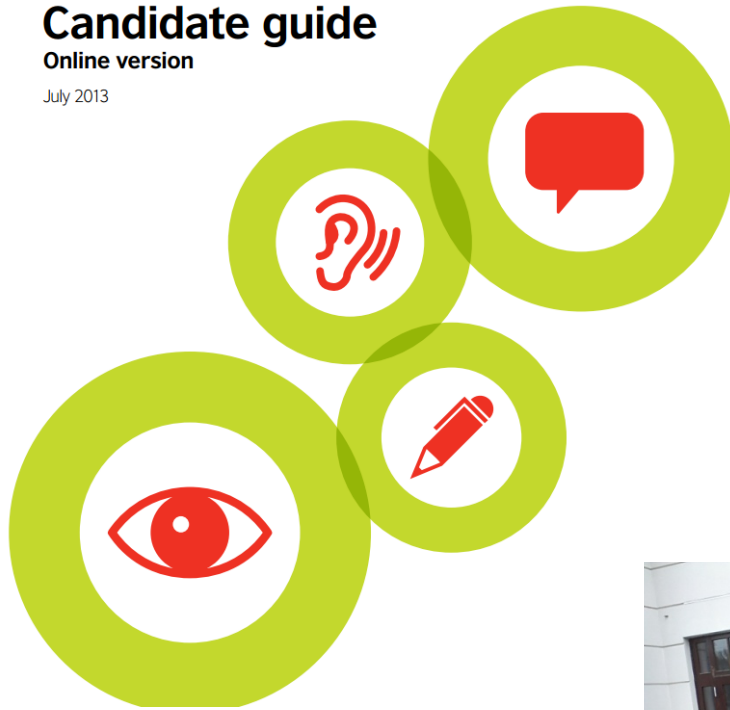
Some background

Aptis
Forward thinking
English testing

Candidate guide

Online version

July 2013



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

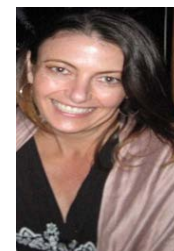


part-time writers who are commissioned to produce particular test items within certain parameters

teaching or exams background



currently represent more than a dozen countries



Diverse L1 backgrounds

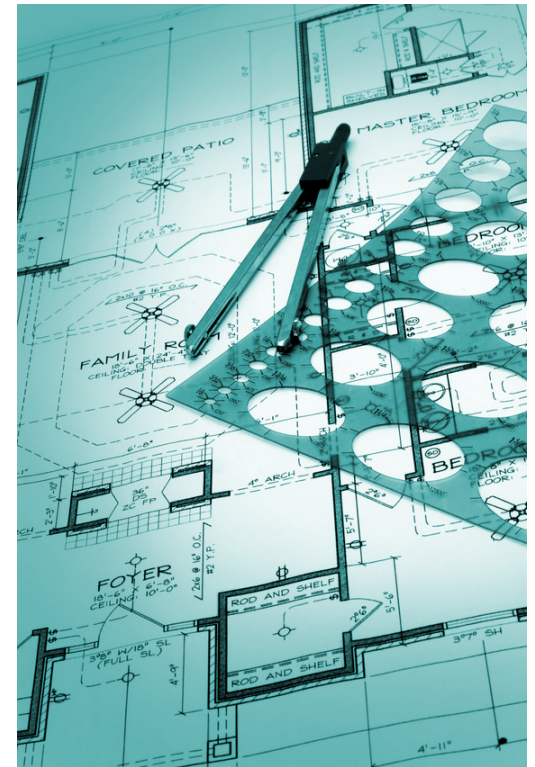


What information should specifications include? Why do we need them?

ALTE - MATERIALS FOR THE GUIDANCE OF TEST ITEM WRITERS (1995, UPDATED JULY 2005)

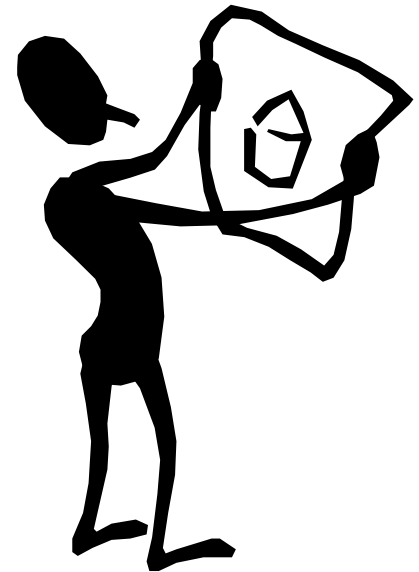
The specifications which are finally produced should give detailed information on each paper or part of the test. They will cover the following areas:

- *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
eg Compleat Lexical Tutor <http://www.lextutor.ca>



The Compleat 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](#)

- S
C
H
E
M
E
S
- CLASSIC [Lists](#)
 - BNL [\[?\] Lists](#)
 - BNC-COCA [\[?\] Lists](#)
 - Core-4
 - BNC-20 [Lists](#)
 - BNC-COCA-25 [\[?\] Lists](#)
 - FRENCH v.3, 1-5k [\[?\] Lists](#)

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

Title: **French Cognates in English Analysis** **Edit-to-a-Profile** **NEW Sentence Count** (needs std. punct.)

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 this text, or 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 hard drive). Text is NOT stored on Web VP's server

Demos : | Speech [Adults](#) [Kids](#) | [Rex M.](#) | AWL in other schemes: [AWL](#) Legal lingo: [LEGAL](#) | [New!](#) [French](#)

R E C A T	Words to reclassify => 1k (type or dbl-click) (E.g. Cognates, specific proper nouns)	ALL mid-sentence capped words	Stray proper at sentence boundary => 1k (E.g., "_Paul Martin")
	<input type="text"/> <input type="button" value="Select"/>	=> 1-k <input type="text"/> => ignore <input type="text"/> (new Jan '14)	<input type="text"/> <input type="button" value="Select"/>

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)**

DEMO SPACE for larger uploads

UPLOAD SPACE (for ~.txt files, encoded UTF8 if French)

1. No file chosen

2. 3.

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



Language Assessment Literacy at the British Council



contracted item writers (as of December 2013) - 30

in-house item writer training - 65

Roehampton course - 119

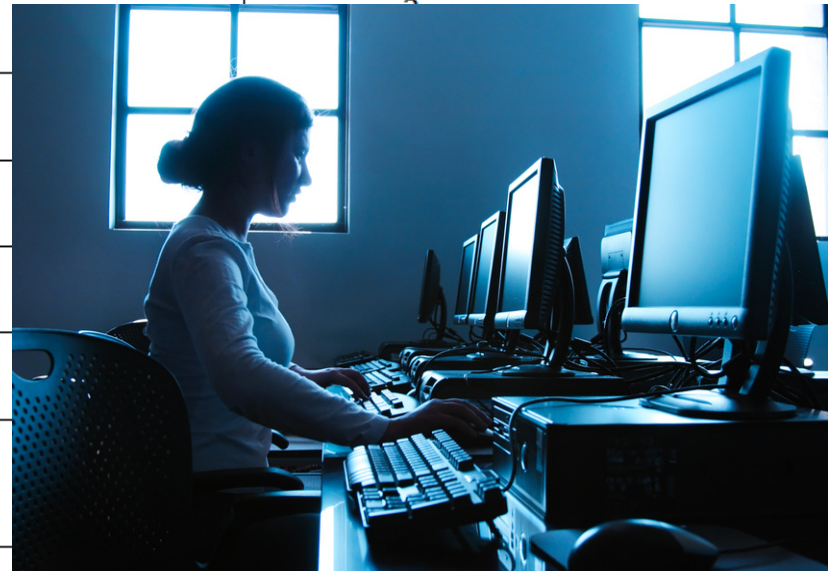
B1 reading specifications (extract)

Features of the Input Text												
Word count	125-135 words (including target words for gaps)											
Domain	Public			Occupational			Educational			Personal		
Discourse mode	Descriptive			Narrative			Expository			Argumentative		Instructive
Content knowledge	General									Specific		
Cultural specificity	Neutral									Specific		
Nature of information	Only concrete			Mostly concrete			Fairly abstract			Mainly abstract		
Presentation	Verbal				Non-verbal (i.e. graphs)				Both			
Lexical Level	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10		
Lexical Level; further criteria	The cumulative coverage should reach 95% at the K3 level. No more than 5% of words should be beyond the K3 level. (See Guidelines on Adhering to Lexical Level for more information).											
Grammatical level	A1-B1 Grammatical exponents (See Guidelines on Adhering to Grammatical Level)											
Avg sentence length	13-15 (This is an average figure. Individual sentences will span a range above and below the average)											
Topic	From topic list for B1.											
Text genre	Magazines, newspapers, instructional materials (such as extracts from textbooks describing important events or people). The texts are adapted to the level. Although not intended to be authentic, they should reflect features of relevant texts from the TLU domain. It should be possible to answer the questions: <i>where would a reader be likely to see a text like this outside the test?</i> and <i>is the genre relevant to TLU tasks important for Aptis General test takers at B1 level?</i>											
Writer/Reader relationship	The relationship is not specified. The texts will typically be written for a general audience, not a specific reader.											

Extract of Quality Reviewer's check list

QR check list for reading

R10	correct structural range for level	
R11	lexical profile (K lists and <u>Flesch Kincaid</u>)	
R12	distractors are relevant	
R13	only one answer is correct	
R14	language is authentic	
R15	Question fits purpose (eg B2: <i>demonstrate general comprehension of a single complex text.</i>)	
R16	Item is appropriate for the level (CEFR)	



B1 reading item (from preparation materials, retired from a live test)

- 1
- 2
- 3
- 4



READING

Read the text and complete each gap with a word from the list at the bottom of the page.

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 of funds, and later taught sciences at the University of Padua. It was his time there that Galileo did a large number of , the most famous involving dropping balls of different sizes from heights to determine the law of acceleration of falling bodies. Indeed, he is credited with several important scientific , and is still considered a great genius.

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

Example:

referred

various

discoveries

taking

lot

around

during

lack

at

servicing

experiments

Preferences

Tutorial

Introduction

Previous

Flag Question

Next

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]

Example	referred
	lack [1] during [2] experiments [3] sundry [4] discoveries [5] serving [6] around [7] taking lot shortage



Changes made (revised item)

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

Sandpit

↓ Delete Move Rename folder Delete folder

Name

- Reading Items for Peer Review
- Writing Items for Peer Review
- Listening Items for Peer Review
- A1 Grammar Items for Peer Review

Reading

↓ Delete Move Rename folder Delete folder

Name

- Reading A1 Cloze Item Writing Tips
- Reading A1 Cloze Sample Item
- Reading A1 Cloze Specifications
- Reading A2 Cohesion Item Writing Tips
- Reading A2 Cohesion Sample Item
- Reading A2 Cohesion Specifications
- Reading B1 Selective Deletion Cloze Item Writing Tips
- Reading B1 Selective Deletion Cloze Sample Item

☆ 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 c

- 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 University. It can help with collocations, synonymy and Vocabulary-Collocation tasks at B1 and B2. It can help with collocations, synonymy

<http://corpus.byu.edu/bnc/>

References

Alderson et al (1995). Language Test Construction and Evaluation. CUP.

ALTE. Item Writer Guidelines. http://www.alte.org/attachments/files/item_writer_guidelines.pdf

Fulcher, G. and Davidson, F. (2007). Language Testing and Assessment. Routledge

Lynch, B. and Davidson, F. Criterion-Referenced Language Test Development: Linking Cumcula, Teachers, and Tests. TESOL QUARTERLY Vol. 28, No. 4, Winter 1994

O'Sullivan, B. British Council. Aptis test development approach. Aptis technical report (ATR-1). August 2012 <http://www.britishcouncil.org/sites/britishcouncil.uk2/files/aptis-test-dev-approach-report.pdf>

O'Sullivan, B (2011) Language Testing. In James Simpson (ed) Routledge Handbook of Applied Linguistics. Oxford: Routledge.

Nation, P (2005). RANGE and FREQUENCY - Programs for Windows based PCs.
<http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Weir, CJ (2005) Language Testing and Validation: an evidence based approach. Oxford: Palgrave.

Specifications and test validation

Materials for the Guidance of Test Item Writers

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

TE WHARE WĀNANGA O TE EPOKO O TE IRA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

SCHOOL OF LINGUISTICS AND APPLIED LANGUAGE STUDIES

Home About Us Study Research Centres & Institutes Resources and Publications

You are here: Home > About Us > Staff Directory > Paul Nation

Paul Nation

- Contact Us
- News
- Events
- Staff Directory

Paul Nation

School of Linguistics and Applied Language Studies
paul.nation@vuw.ac.nz

Phone: 04 463 5628
Location: Room 415, Von Zedlitz Building,
Kelburn Pde, Kelburn Campus



TESOL Quarterly
Volume 28, Issue 4, pages
727–743, Winter 1994

BRITISH COUNCIL

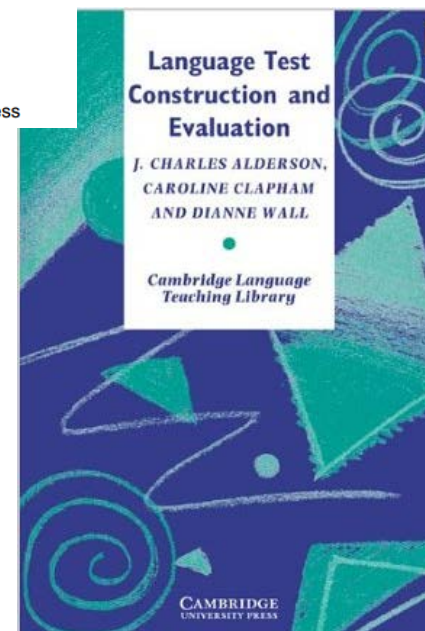
Aptis

Forward thinking
English testing

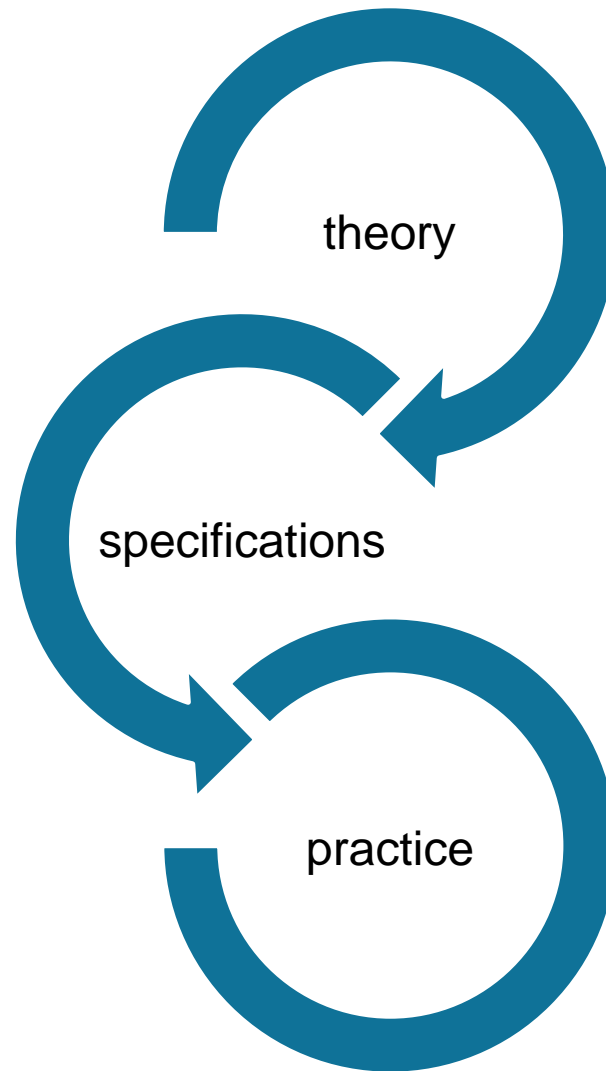
Aptis test development approach

Aptis technical report (ATR-1)

Barry O'Sullivan, British Council
August 2012



Specifications – theory meets practice

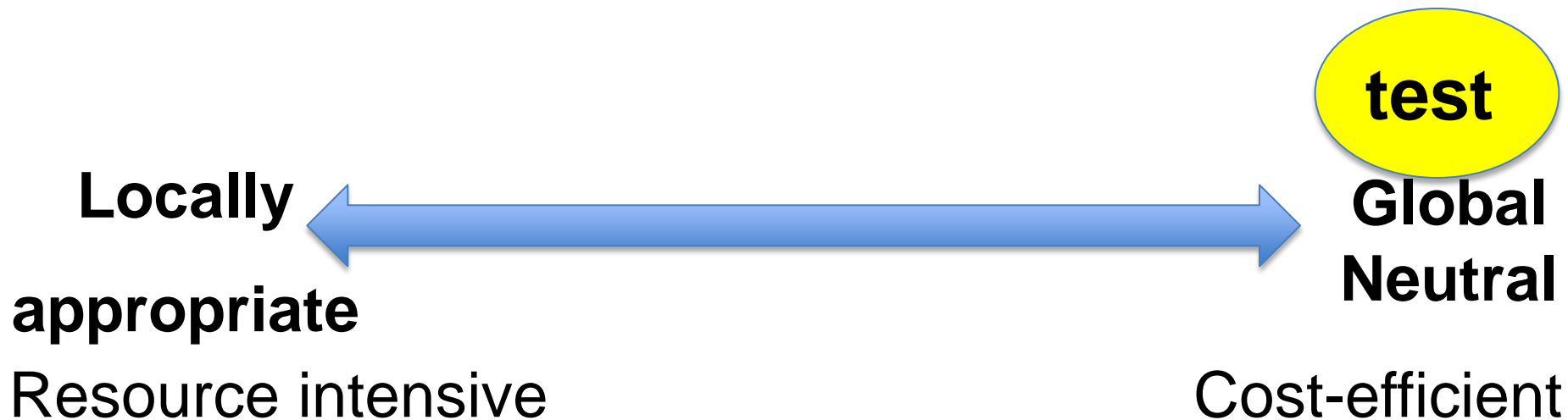


Any Questions?

john.tucker@britishcouncil.pl

What is Aptis?

- 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?

- **The main, or general version of Aptis was launched in August 2012**
- **test of general English proficiency**
- **for adults (16 years and older)**
- **5 components: Grammar & Vocabulary, Reading, Listening, Speaking Writing**
- **Various modes for delivery: computer, pen and paper, telephone,**

Task specs: part of validation

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

