

Test Quality Check: Selecting the Right Test for the Right Use

Dr Victoria Clark
Assessment Development Manager, East Asia
British Council
Victoria.Clark@britishcouncil.org.hk

Indonesia, March 2018



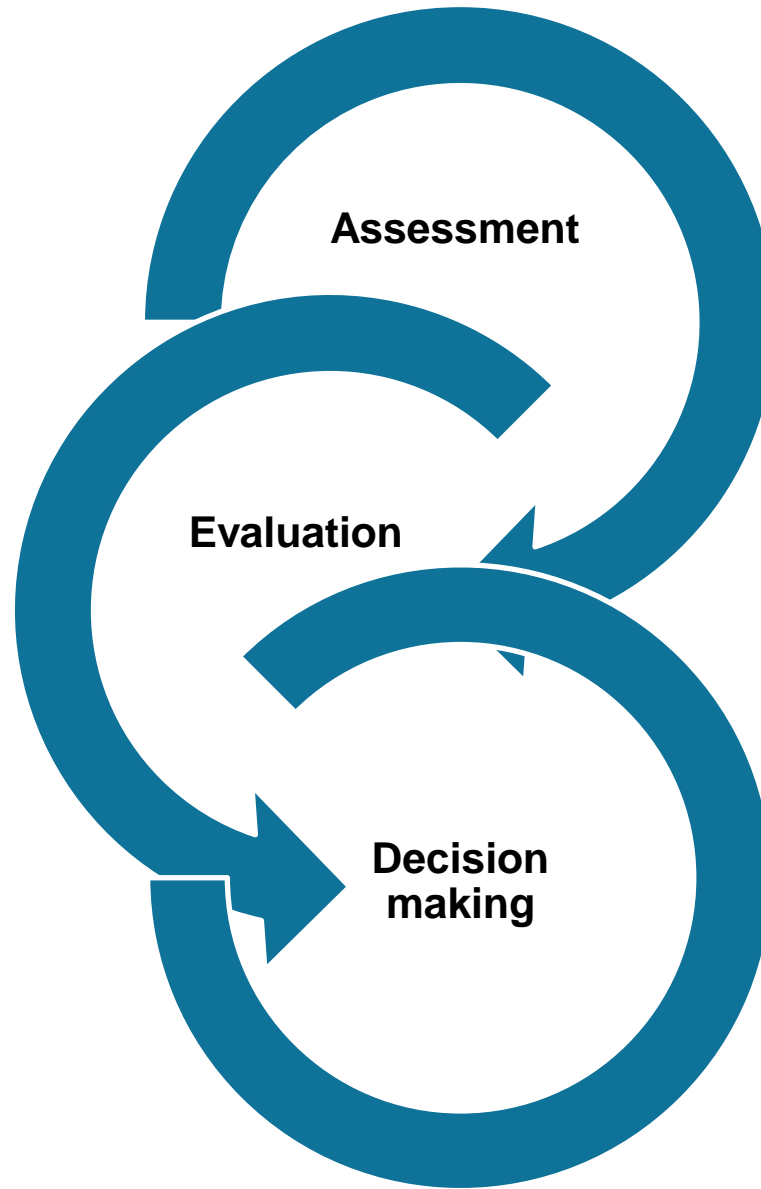
PART 1

INTRODUCTION

Assessment

In the context of language teaching and learning,
'assessment' refers to the act of **collecting information**
and **making judgments** about a language learner's
knowledge of a language and ability to use it.

(Carol Chapelle and Geoff Brindley)



Test

An educational test is a procedure **designed to elicit**
certain behaviour from which one can **make**
inferences about certain characteristics of an
individual

(J. B. Carroll, 1968, p. 46)

Test as a Tool



PART 2

CRITERIA

Criteria to consider when selecting tests

1.

Test purpose

2.

**Test taker
characteristics**

3.

**Target language use
domain**

4.

Skills/competencies

5.

Reliability

6.

Validity

7.

Score reporting

8.

**Test administration
and practicality**

9.

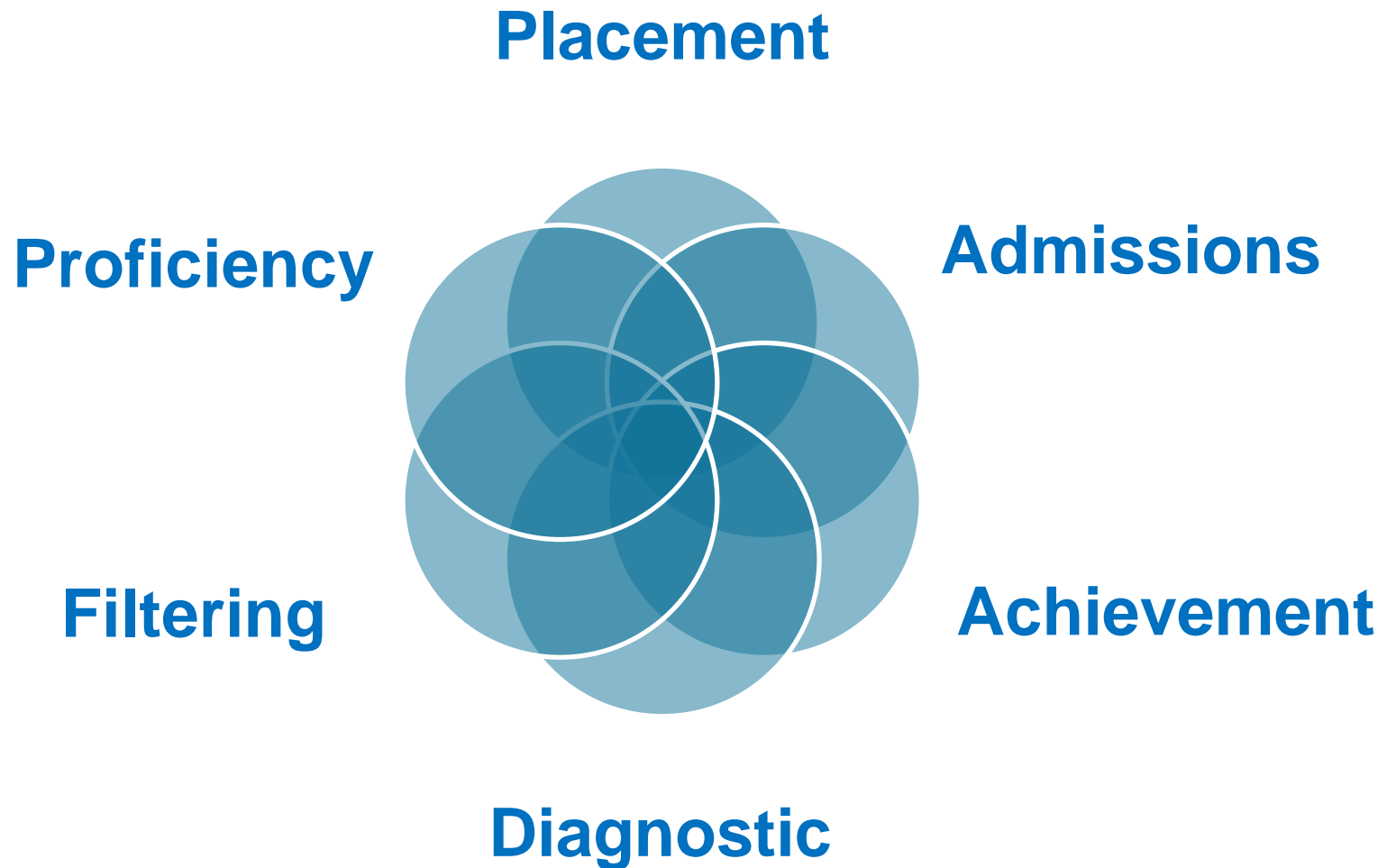
Impact

1.
Test purpose

2.
**Test taker
characteristics**

3.
**Target language use
domain**

Test purpose



Test taker characteristics

Age

Gender

Level of proficiency

First language background

Cultural background

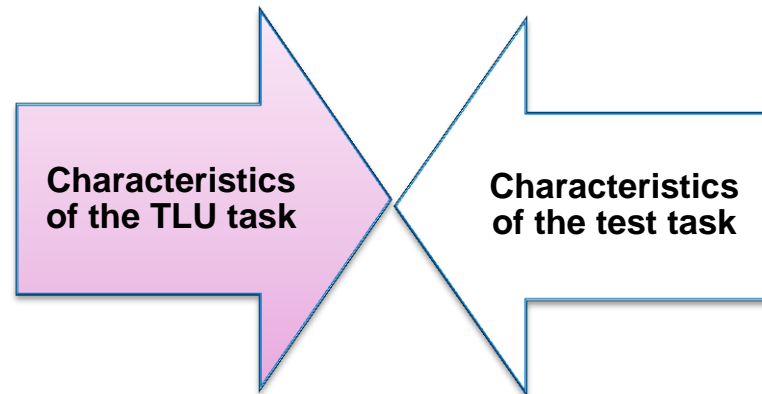
Educational level

Background knowledge



Target language use domain

TLU domain is a “situation or context in which the test taker will be using the language outside of the test itself” (Bachmann and Palmer 1996: 18).



4.

Skills/competencies

5.

Reliability

6.

Validity

Skills and competencies

What is being measured by the test?

Construct =
theoretical definition of
what it is we want to assess



Reliability

“Reliability can be defined as the extent to which the same results would be obtained on a different occasion, in a different context, or by a different assessor”(Bartmaan et al., 2007, p12).

Threats to Reliability



Administration

"Rubrics"



Scoring

Test

Validity

Validity refers to the appropriateness of the inferences and uses that result from the assessment.

- ✓ Is the assessment measuring what it is supposed to measure?
- ✓ Is the information you have collected the right kind of evidence for the decisions you need to make or the intended use?

“Validity is concerned with the clarification and justification of the intended interpretations and uses of observed scores” (Kane, 2001, p.339)

Validity = Evidence

Involves accumulating evidence to provide sound basis for the interpretations of the test scores

Publications

Research and test expertise

Assessment Advisory Board

Assessment Research Awards
and Grants

Assessment Research Group

Publications

Research projects

The Davies Lecture





Technical Report
Aptis Test Development Approach
TR/2015/001

Aptis Test Development Approach

The Aptis test system is designed to offer users an alternative to currently available high-stakes certificated examinations.



Technical Report
Linking the Aptis Reporting
Scales to the CEFR
TR/2015/003

Linking the Aptis Reporting Scales to the CEFR

The Aptis development project marked a new era for the British Council, even though it had been involved in a number of test development projects in the past.



Technical Report
Aptis General Technical Manual
Version 1.0
TR/2015/005

Aptis General Technical Manual Version 1.0

TR/2015/005 This manual describes the content and technical properties of Aptis General, the standard English language assessment product offered within the Aptis test system.



Looking into test-takers cognitive processes

Looking into test-takers cognitive processes while completing reading tasks - Brunfaut and McCray



Validating the Core Inventory for General English

The Core Inventory for General English aims to inform teachers about the levels at which learners of English master certain aspects of the language.



Effectiveness of Training Aptis Raters Online

VS/2015/001 Knoch, Fairbairn and Huisman. Aptis is an online English language assessment for adults developed by the British Council.

7.
Score reporting

8.
**Test administration and
practicality**

9.
Impact

Score reporting

When?

How are the scores reported?

- Numerical scores

- CEFR

- Subtest scores

- Overall score



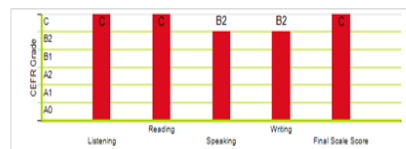
Candidate Report

Candidate Name: _____ Test Date: _____
Date of Birth: _____ Test Package: NEW Aptis Listening
Reading Speaking and Writing Package

Scale Score

Skill Name	Skill Score
Listening	40/50
Reading	48/50
Speaking	40/50
Writing	42/50
Final Scale Score	170
Grammar and Vocabulary	32/50

CEFR Skill Profile



Please turn over for CEFR Skill Descriptors

Description/explanation to aid interpretation?

CEFR Skill Descriptors

Listening

C	Has no difficulty in understanding any kind of spoken language, whether live or broadcast, delivered at fast native speed.
B2	Can understand the main ideas of proportionally and linguistically complex speech on both concrete and abstract topics delivered in a standard dialect, including technical discussions in his/her field of specialisation.
B1	Can understand straightforward factual information about common everyday or job related topics, identifying both general messages and specific details, provided speech is clearly articulated in a generally familiar accent.
A2	Can understand enough to be able to meet needs of a concrete type provided speech is clearly and slowly articulated.
A1	Can follow speech which is very slow and carefully articulated, with long pauses for him/her to assimilate meaning.
A0	Not enough to allow for any meaningful inferences about the candidate's ability.

Reading

C	Can understand and interpret critically virtually all forms of the written language.
B2	Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively.
B1	Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.
A2	Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language.
A1	Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.
A0	Not enough to allow for any meaningful inferences about the candidate's ability.

Speaking

C	Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.
B2	Can give clear, systematically developed descriptions and presentations on a wide range of subjects related to his/her field of interest, with appropriate highlighting of significant points, and relevant supporting detail.
B1	Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.
A2	Can give a simple description or presentation of people, living or working conditions, daily routines like schedules, etc. as a short series of simple phrases and sentences linked into a list.
A1	Can produce simple descriptions on mainly personal topics.
A0	Not enough to allow for any meaningful inferences about the candidate's ability.

Writing

C	Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.
B2	Can write clear, detailed texts on a variety of subjects related to his/her field of interest and shows an ability to use different registers within written texts.
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence.
A2	Can write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'.
A1	Can write simple isolated phrases and sentences.
A0	Not enough to allow for any meaningful inferences about the candidate's ability.

Test administration and practicality

Practicality: best use of resources, time and costs

- Ease of administration
- Instructions
- Invigilated?
- Computer-based/pen and paper
- Number of versions
- Preparation and test practice



Impact



Criteria to consider when selecting tests

1.

Test purpose

2.

**Test taker
characteristics**

3.

**Target language use
domain**

4.

Skills/competencies

5.

Reliability

6.

Validity

7.

Score reporting

8.

**Test administration
and practicality**

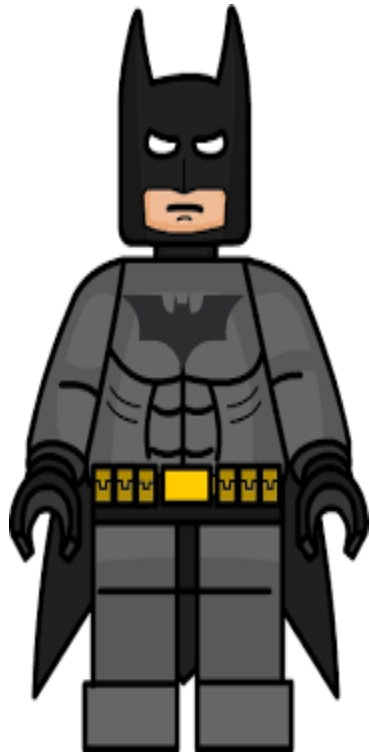
9.

Impact

PART 3

CONCLUSION

Fitness for purpose



Thank You

Victoria.Clark@britishcouncil.org.hk