

# Using test specifications to inform teaching

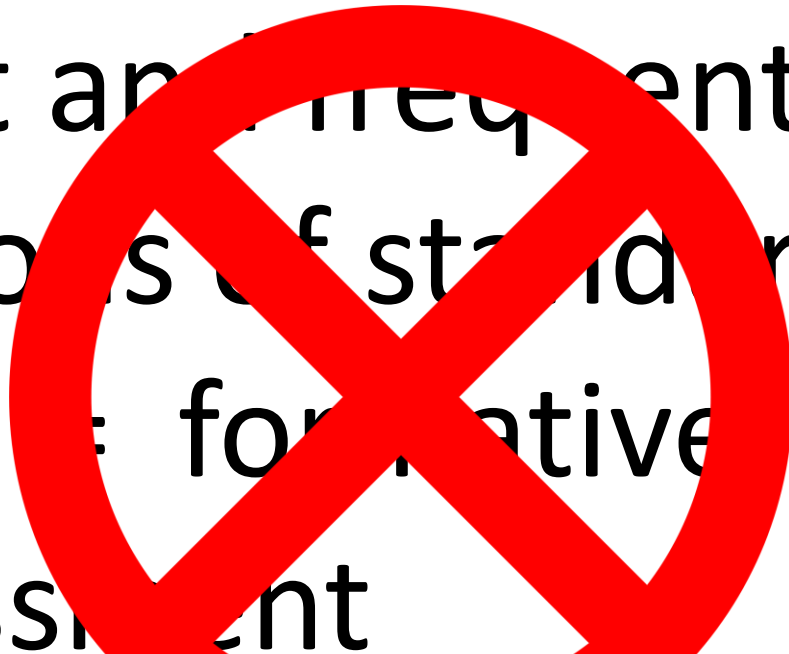
***Vietnam Language  
Assessment  
Symposium  
Ho Chi Minh City  
May 23 2017***

Jamie Dunlea  
Assessment Research  
Group  
British Council

# Starting point

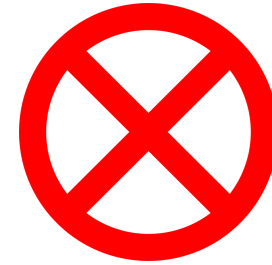
- A lot of effort is invested in large-scale, high-stakes tests
- A great deal of research is carried to support the uses and interpretations
- If a test is well-designed, it will be targeting a clear and coherent construct

# Cautions

- BUT....
  - Short and frequent versions of standardized tests are for formative assessment
- 

# Guiding principles

Teaching to the test



Teaching to the construct



# Guiding principles

- Green (2007)
- Intensive IELTS preparation classes did not result in higher test scores, compared to groups doing a wider range of EAP tasks
- Teach to the construct



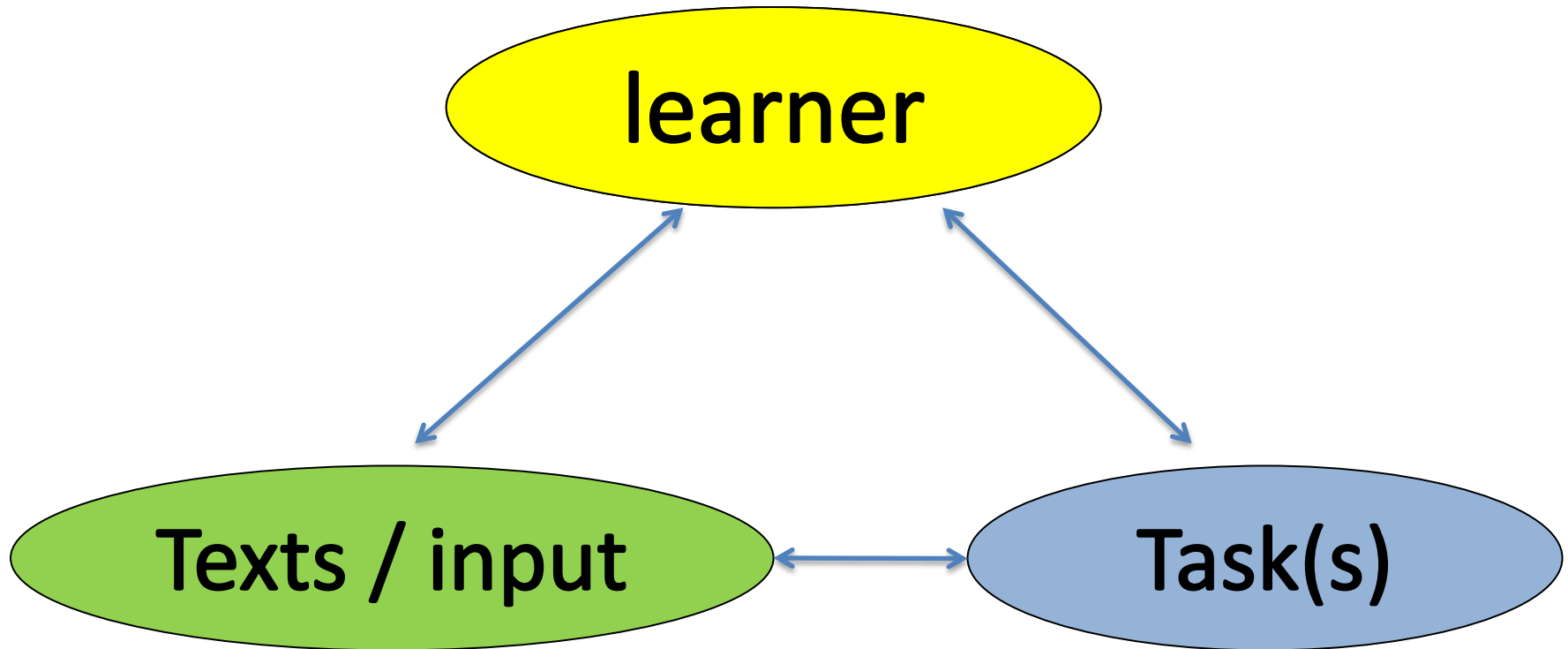
# Guiding principles

- We need to deconstruct the test tasks
- We need to understand what is being targeted by these tasks
- What is the construct?

# Guiding principles

- We can use the tools that have been developed to build well-designed test tasks
- We can take these tasks apart, break them down to understand what parts might be useful in the classroom

# Guiding principles





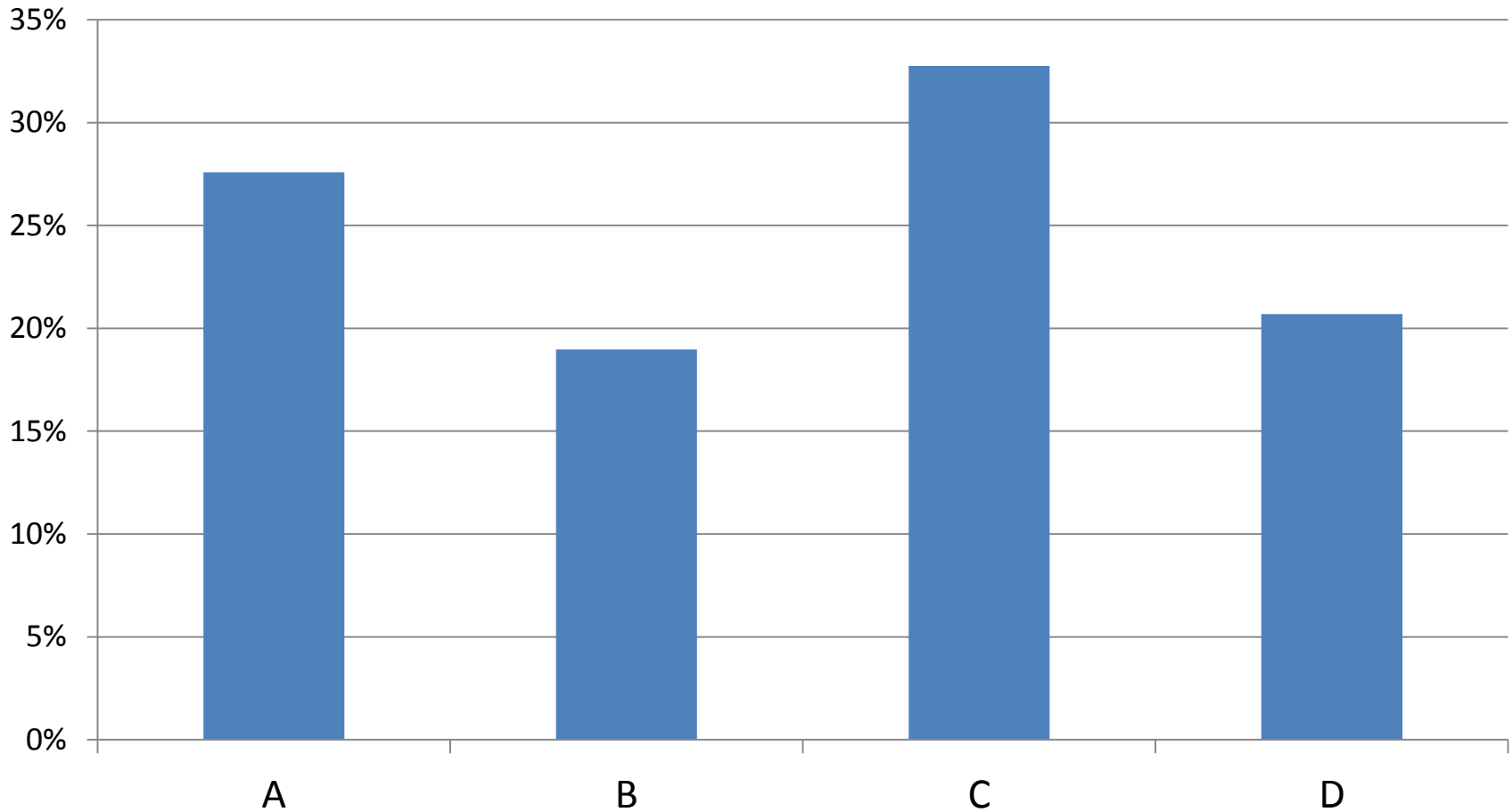
# Guiding principles

- We need standards, benchmarks, proficiency scales to provide a clear framework to describe the language ability we want students to achieve
- The CEFR is our starting point

# Familiarity with the CEFR

- A) I had read the CEFR and was familiar with its aims and contents, including the Common Reference Levels.
- B) I was familiar with the aims of the CEFR, but had not studied it in detail.
- C) I had heard of the CEFR but was not familiar with its aims or contents.
- D) I had not heard of the CEFR.

# Familiarity with the CEFR: workshop group



# The CEFR is NOT....

## ❖ A test, or tests



➤ *the CEFR is useful for developing tests, and setting learning goals, but the CEFR does not say what kind of tests are should be used*

## ❖ Just a proficiency scale or can-do statements

➤ It is a 9-chapter, 290-page book

## ❖ A set of rules or fixed ways of teaching or testing

# Is it useful: Goals of the CEFR

CEFR LEVEL
C2
C1
B2
B1
A2
A1

## 3 key goals from the blue book:

*[Provide] a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe.*

*Define] levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis.*

*[Facilitate] the mutual recognition of qualifications gained in different learning contexts*

# Familiarity with the CEFR

“At the heart of the CEF are the  
Common Reference levels.”  
(Morrow, 2004)

C2

Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

C1

Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

B2

Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

B1

Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

A2

Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.

A1

Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

C2

Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

C1

Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

B2

Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

B1

Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

A2

Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.

A1

Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.



# Familiarity with the CEFR

Provides a principled basis for evaluating the claims of test developers (and for test developers to evaluate their own claims) from both quantitative and qualitative perspectives

CEFR LEVEL
C2
C1
B2
B1
A2
A1



Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

# Familiarity with the CEFR

The Global Scale is “just the tip of the iceberg” (Morrow, 2004)

How many *illustrative scales* are there?

❖ Descriptors grouped in **54 scales**

- Communicative activities
- Strategies
- Communicative language competences

# Using the CEFR: Self-assessment grid

**The CEFR is not just about testing.  
It can be used for setting learning goals, discussing levels with students, designing learning tasks, and self-assessment**

# Is it useful?

To use the CEFR as basis for task development (test tasks and learning tasks) , we need to add detail, and fill in the gaps.

*The CEFR can be a springboard to task and test development*

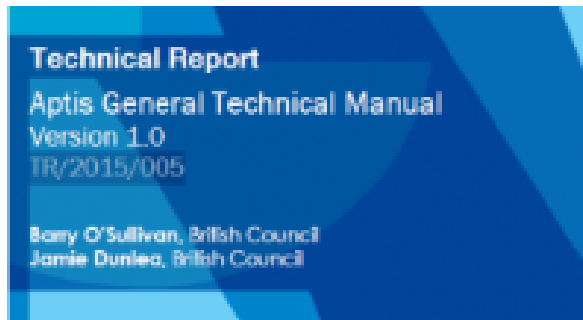


# Test specifications

*The chief tool of language test development is a test specification, which a test is a generative blueprint from which test items or tasks can be produced. A well-written test specification (or “spec”) can generate many equivalent test tasks.*

(Lynch & Davidson, 2002)

# Case studies: where to find detailed specs



## [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.



## Test of English for Academic Purposes

<http://www.eiken.or.jp/teap/group/report.html>

<https://www.beds.ac.uk/crella/projects/teap>

# Socio-cognitive model

## What is validity?

Does the test measure what we want it to measure?

CONTEXT VALIDITY

COGNITIVE VALIDITY

Are the scores from the test accurate, reliable, meaningful?

SCORING VALIDITY

Are the scores useful for test users to make decisions?

CONSEQUENTIAL VALIDITY

CRITERION-RELATED VALIDITY

# Socio-cognitive model: O'Sullivan 2017

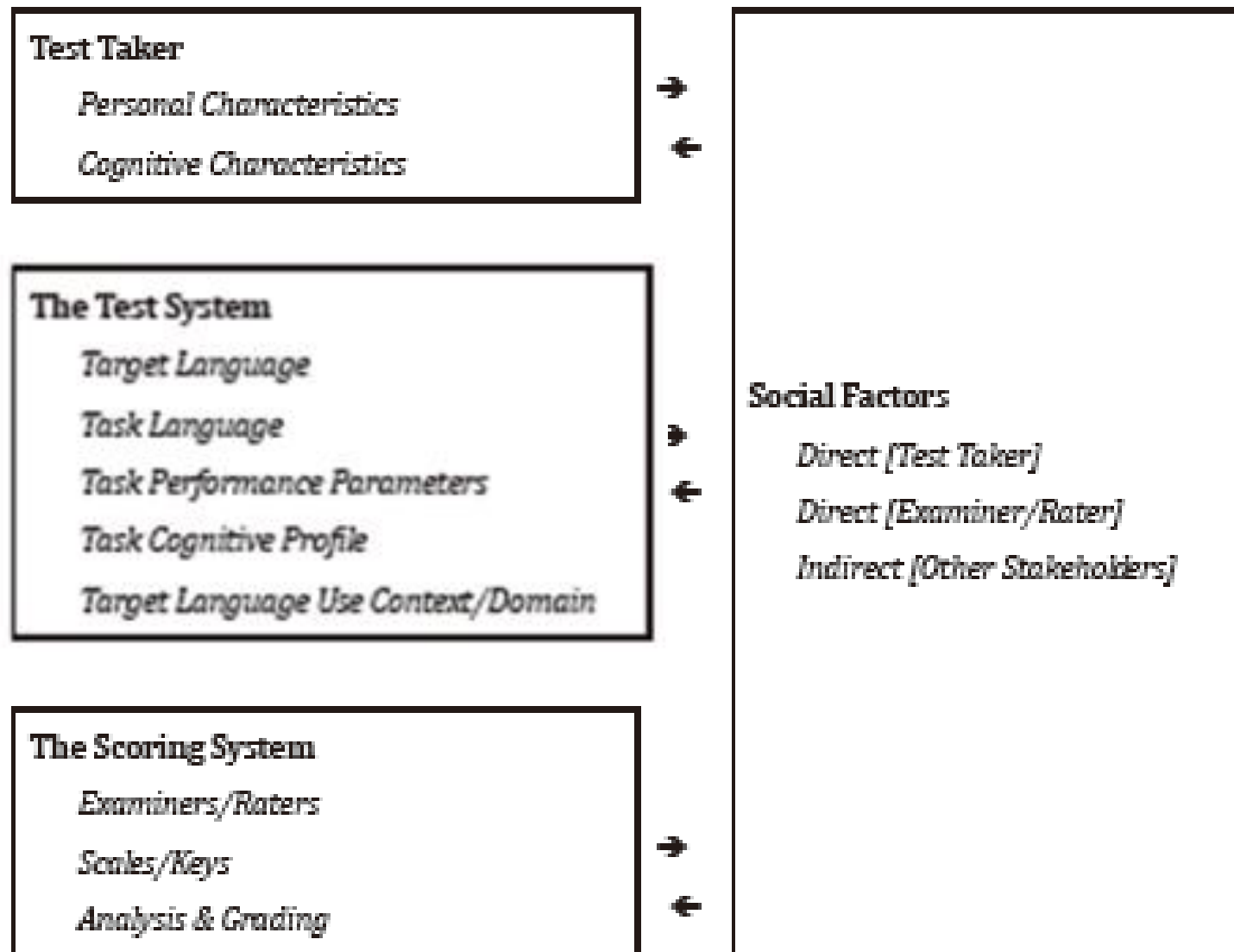


Figure 6: Simplified test system model



# Socio-cognitive model: O'Sullivan 2017

## The Test System

*Target Language*

*Task Language*

*Task Performance Parameters*

*Task Cognitive Profile*

*Target Language Use Context/Domain*

# Filling the gaps in the CEFR

OVERALL READING COMPREHENSION	
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. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.
B1	Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.
	Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language
A2	Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.
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.

# Filling the gaps: a test task

Choose one word from the list for each gap. The first one is done for you.

Dear Morgan,

Thank you for a wonderful weekend. I had a really  time with you and

Becky. Your wife is a good cook and she  a very nice dinner.

I am writing this note in my hotel room and I can  the park from my

window. My plane leaves tomorrow and I will take a taxi to the airport .

breakfast. I hope you and Becky will come and  with me in Rome next

summer. I am feeling a little tired now and I  to have a sleep.

Thanks again and see you soon,

James

# Filling the gaps: a test spec

Test	Aptis General	Component	Reading		Task	Multiple Choice Gap-Fill
<b>Features of the Task</b>						
<b>Skill focus</b>	<del>Reading comprehension up to the sentence level</del>					
<b>Task Level</b>	A1	A2	B1	B2	C1	C2
<b>task description</b>	Multiple-choice gap fill. A short text of 6 sentences is presented. Each sentence contains one gap. Test takers choose the best option from a pull-down menu for each gap to complete the sentence. The first sentence is an example with the gap completed. Each gap can be filled by reading within the sentence.					
<b>Cognitive processing Goal setting</b>	<b>Expeditious reading: local</b> (scan/search for specifics)			<b>Careful reading: local</b> <b>(understanding sentence)</b>		
	<b>Expeditious reading: global</b> (skim for gist/search for key ideas/detail)			<b>Careful reading: global</b> (comprehend main idea(s)/overall text(s))		
<b>Cognitive processing Levels of reading</b>	<b>Word recognition</b>					
	<b>Lexical access</b>					
	<b>Syntactic parsing</b>					
	<b>Establishing propositional meaning (cl./sent. level)</b>					
	<b>Inferencing</b>					
	<b>Building a mental model</b>					
	<b>Creating a text level representation (disc. structure)</b>					
<b>Creating an intertextual representation (multi-text)</b>						

# Filling the gaps: a test spec

Features of the Input Text												
<b>Words</b>	40-50 words (including target words for gaps)											
<b>Domain</b>	Public			Occupational			Educational			Personal		
<b>Discourse mode</b>	Descriptive			Narrative			Expository		Argumentative		Instructive	
<b>Content knowledge</b>	General								Specific			
<b>Cultural specificity</b>	Neutral								Specific			
<b>Nature of information</b>	Only concrete			Mostly concrete			Fairly abstract			Mainly abstract		
<b>Lexical Level</b>	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10		
<b>Text genre</b>	E-mails, letters, notes, postcards											
Features of the Response												
<b>Target</b>	<b>Length</b>	1 word	<b>Lexical</b>	K1	<b>Part Speech</b>	<b>of</b>	Noun, verb, adjective					
<b>Distractors</b>	<b>Length</b>	1 word	<b>Lexical</b>	K1	<b>Part Speech</b>	<b>of</b>	Noun, verb, adjective					
<b>Key</b>	Within sentence			Across sentences			Across paragraphs					

# Task specs: Where to start?

## CEFR: Overall Reading Comprehension

**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. Has a broad active reading vocabulary, but may experience some difficulty with low frequency idioms.

## CEFR: Reading for Information and Argument

**B2** Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

# Test task based on spec

## READING



Read the passage quickly. Choose a heading for each numbered paragraph (1-7) from the drop-down box. There is one more heading than you need.

1	<input type="text"/>
2	<input type="text"/>
3	<input type="text"/>
4	<input type="text"/>
5	<input type="text"/>
6	<input type="text"/>
7	<input type="text"/>

### Bone Wars

In the summer of 1868 a train carrying a group of American scientists made its way through the western frontier state of Wyoming. On board was O.C. Marsh, an expert in geology and the first person in the country to hold the position of professor of palaeontology at the University of Yale. Like his fellow passengers, Marsh was impressed by the enormous landscapes of dry rock, and he knew that the ancient stones must hold evidence of prehistoric life. It was during this journey that he made a decision that was to have a lasting impact not only on his own professional career but on the American scientific community.

1. In 1800 the French naturalist Georges Cuvier identified a fossil [old bone] as the remains of a small flying reptile. This was the first recorded example of a species that later became known as the dinosaur. Although these creatures no longer existed, Cuvier showed that they could be studied through an examination of fossil records, buried and preserved in rock. So the science of palaeontology – the study of prehistoric life – began.

2. Over the next two decades some spectacular finds were made by English



# Task specs: an example

Test	Aptis General	Component	Reading	Task	Matching headings to text	
<b>Features of the Task</b>						
<b>Skill focus</b>	<i>Expeditious global reading of longer text, integrating propositions across a longer text into a discourse-level representation.</i>					
<b>Task Level</b>	<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>
<b>task description</b>	<i>Matching headings to paragraphs within a longer text. Candidates read through a longer text consisting of 7 paragraphs, identifying the best heading for each paragraph from a bank of 8 options.</i>					
<b>Cognitive processing Goal setting</b>	<b>Expeditious reading: local</b> (scan/search for specifics)			<b>Careful reading: local</b> (understanding sentence)		
	<b>Expeditious reading: global</b> (skim for gist/search for key ideas/detail)			<b>Careful reading: global</b> (comprehend main idea(s)/overall text(s))		
<b>Cognitive processing Levels of reading</b>	<b>Word recognition</b>					
	<b>Lexical access</b>					
	<b>Syntactic parsing</b>					
	<b>Establishing propositional meaning</b> (cl./sent. level)					
	<b>Inferencing</b>					
	<b>Building a mental model</b>					
	<b>Creating a text level representation</b> (disc. structure)					
<b>Creating an intertextual representation</b> (multi-text)						



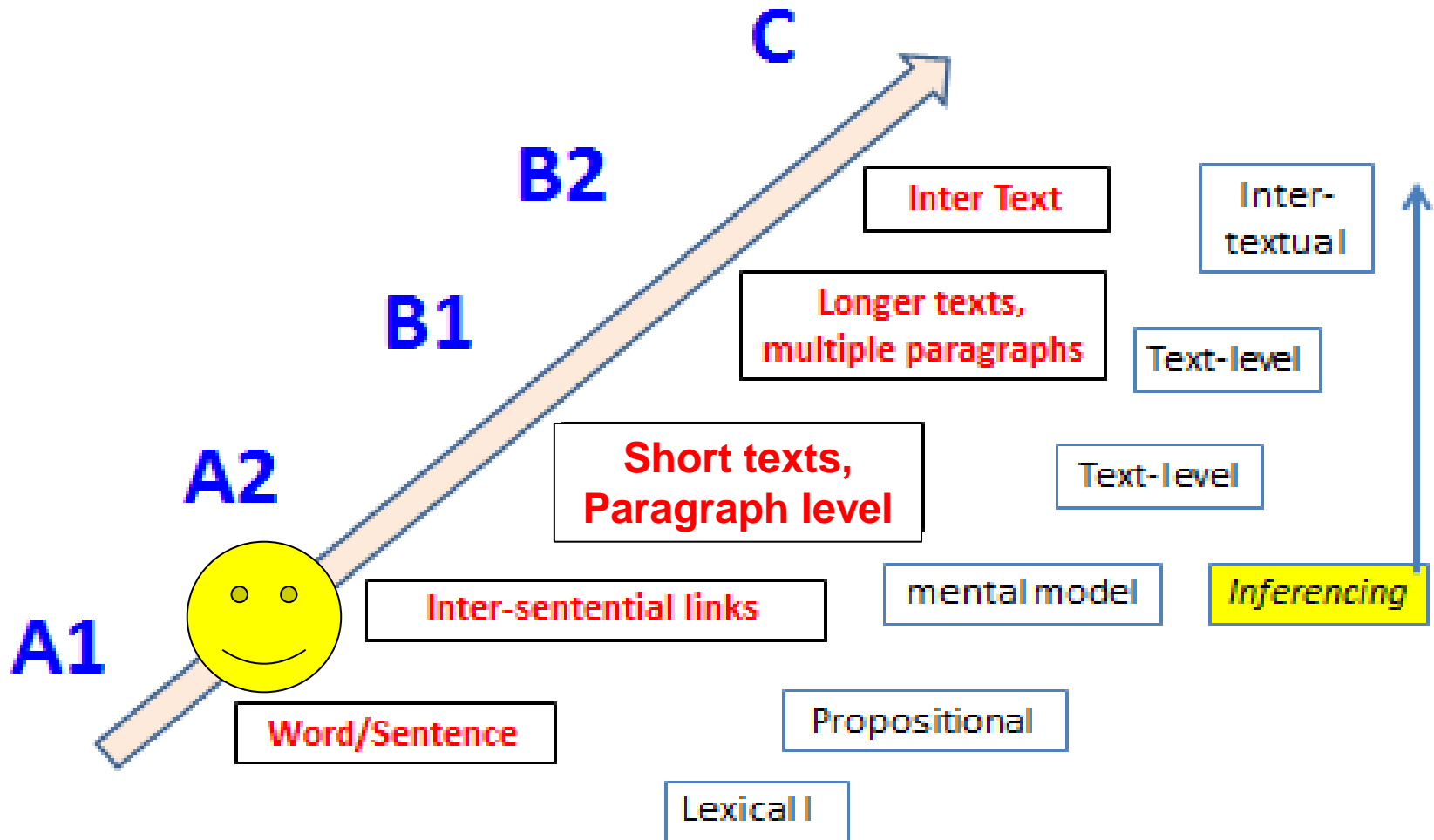
# Task specs: an example

Features of the Input Text										
Words	<i>700-750 words</i>									
Domain	Public	Occupational			Educational			Personal		
Discourse mode	Descriptive	Narrative	Expository		Argumentative			Instructive		
Content knowledge	General							Specific		
Cultural specificity	Neutral							Specific		
Nature information	Only concrete		Mostly concrete			Fairly abstract		Mainly abstract		
Lexical Level	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
Readability	<i>Flesch-Kincaid Grade Level 9-12</i>									
Grammar	<i>A1-B2 Exponents</i>			Average sentence length				<i>18-20 words</i>		
Text genre	<i>Magazines, newspapers, instructional materials (such as extracts from undergraduate textbooks describing important events and ideas, etc).</i>									

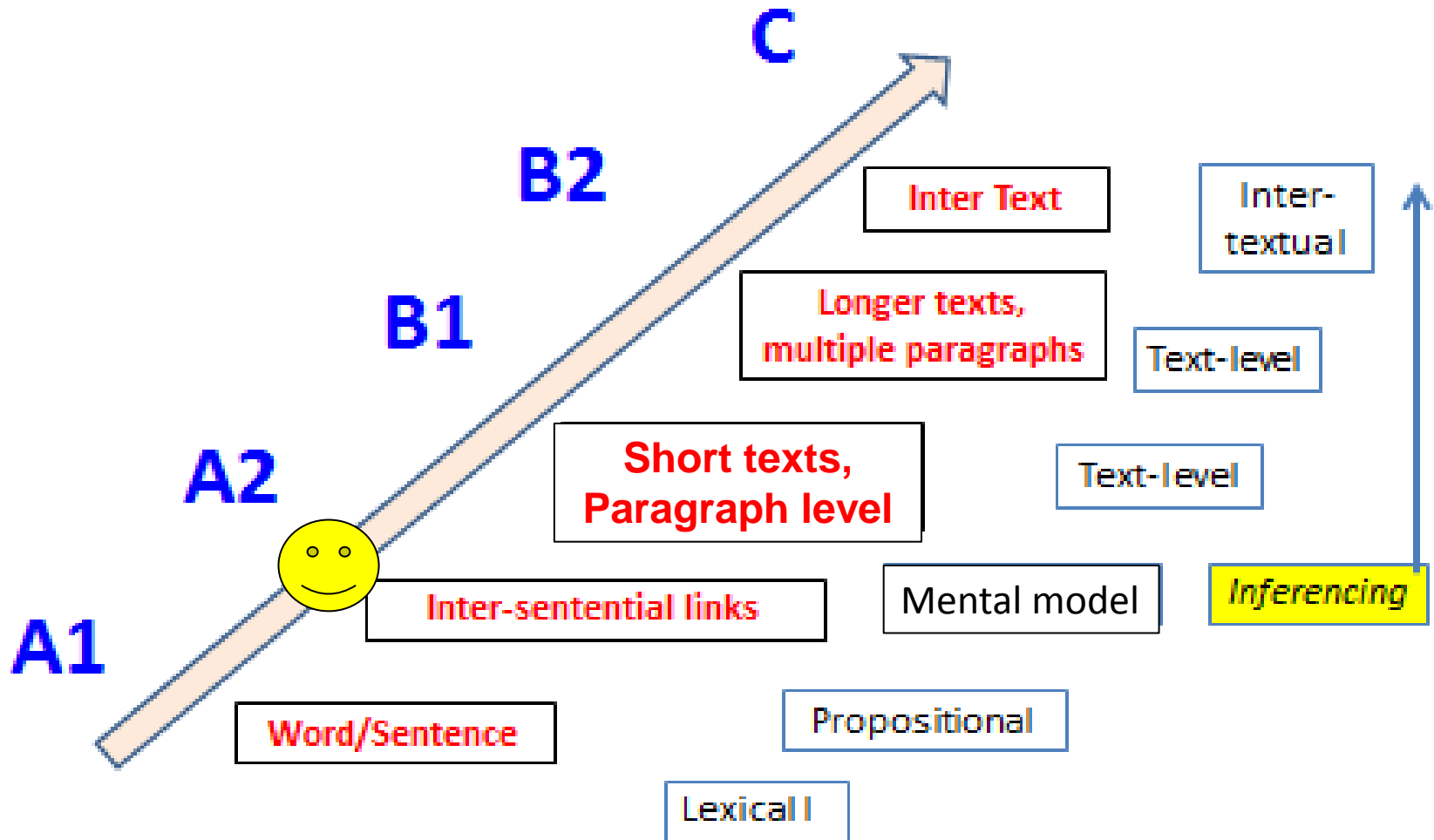
# Aptis Reading test spec

Skill focus	Lvl	Task description	Cognitive processes
Sentence level meaning	<b>A1</b>	A short text with 5 gaps. Filling each gap only requires comprehension of the sentence containing the gap. Text-level comprehension is not required.	<ul style="list-style-type: none"> <li>Careful local reading</li> <li>Syntactic parsing</li> <li>Understanding propositional meaning</li> </ul>
Inter-sentence cohesion	<b>A2</b>	Reorder jumbled sentences to form a cohesive text	<ul style="list-style-type: none"> <li>Careful global reading</li> <li>Inferencing</li> <li>Building a mental model</li> </ul>
Text-level comprehension of short texts	<b>B1</b>	A short text with 7 gaps. Requires comprehension of text across sentences.	<ul style="list-style-type: none"> <li>Careful global reading</li> <li>Building a mental model</li> </ul>
Integrating macro-propositions and understanding important ideas in longer texts	<b>B2</b>	Matching the most appropriate heading to paragraphs. Requires integration of micro- and macro-propositions within and across paragraphs, and comprehension of discourse structure of more complex and abstract texts.	<ul style="list-style-type: none"> <li>Expeditious global reading</li> <li>Creating a text level representation</li> </ul>

# Operationalizing the Model



# Operationalizing the Model



*Thank you!*

*Jamie Dunlea*

*Senior Researcher*

*Assessment Research Group, British Council*

*Jamie.dunlea@britishcouncil.org*