

Linking Examinations to the CSE

Barry O'Sullivan
Jamie Dunlea
British Council



第三届语言测试与评价国际研讨会
暨第五届英语语言测评“新方向”研讨会：

衔接测评、教学与学习

Plenary

上海
2017年12月2-3日



Assessment Research Group

Research and test expertise

Assessment Research Group

Assessment Advisory Board

Assessment Research Awards and Grants

Research into Reading Grants Scheme

Language Assessment Literacy

Assessment Research Publications

Professor Barry O'Sullivan



Barry's recent work includes the development and validation of a new business to language test called Aptis. He is the founding president of the UK Association of Language Testing and Assessment and holds honorary and visiting chairs at the Universities of Reading and Roehampton in the UK and the University of Technology MARA (Kuala Lumpur) and at the University of Lisbon.

Jamie Dunlea



Jamie works on a range of language test development and validation projects for assessment systems designed and developed by the British Council, as well as collaborating on projects with researchers and organisations internationally. Jamie joined the British Council in 2013, and was previously Chief Researcher at the Eiken Foundation of Japan, a not-for-profit organization which develops and administers EFL examinations in Japan.

He has 25 years of experience working in EFL education, first as a teacher, then in test development and production and assessment research.

<http://www.britishcouncil.org/exam/aptis/research>

Assessment Research Group

Research and test expertise

Assessment Research Group

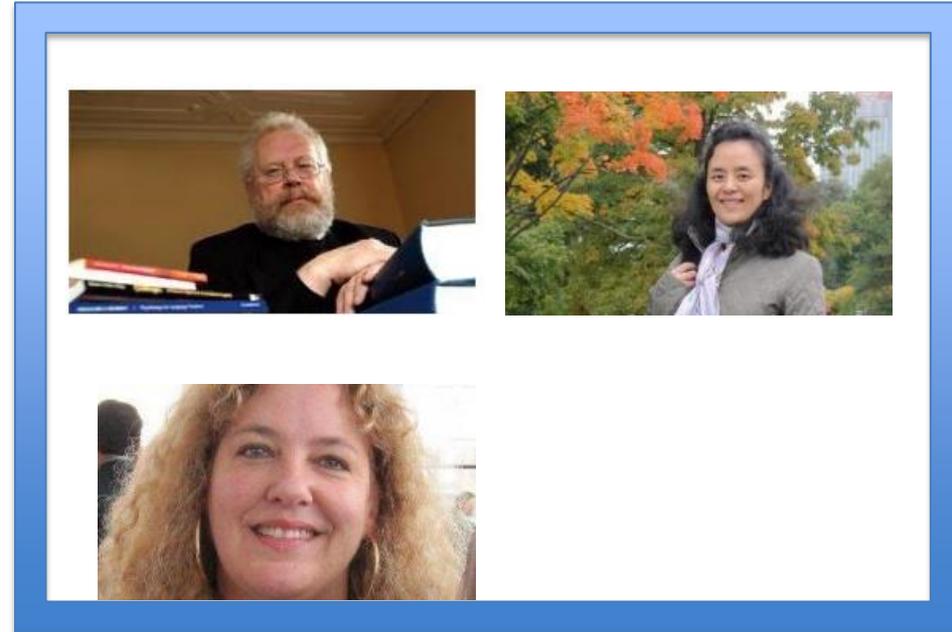
Assessment Advisory Board

Assessment Research Awards and Grants

Research into Reading Grants Scheme

Language Assessment Literacy

Assessment Research Publications



<http://www.britishcouncil.org/exam/aptis/research>

Assessment Research Group



Aptis Test Development Approach

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



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.



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.

Background



教育部考试中心
NATIONAL EDUCATION EXAMINATIONS AUTHORITY

第三届语言测试与评价国际研讨会 暨第五届英语语言测评“新方向”研讨会：

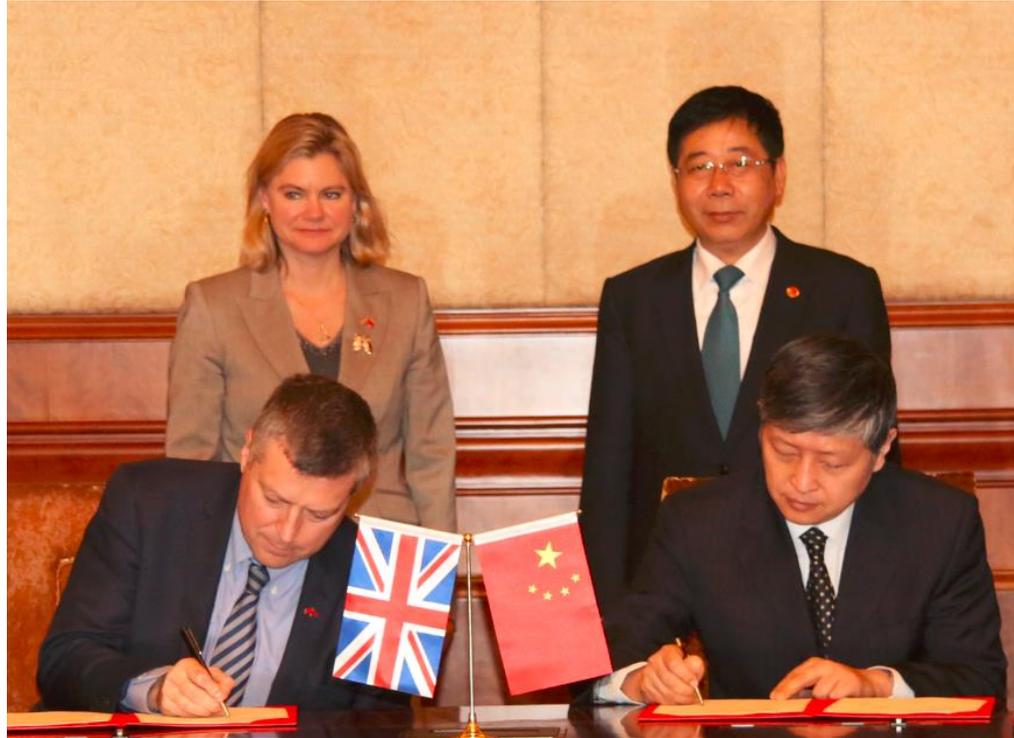
衔接测评、教学与学习

上海



2017年12月2-3日

Background



Action Plan under the UK-China Partners in Education Framework

**Approved by the Ministry of Education of the People's Republic of China and the
Department for Education on behalf of the Government and Devolved
Administrations of the United Kingdom of Great Britain and Northern Ireland**

The Linking Project

MEMORANDUM OF UNDERSTANDING

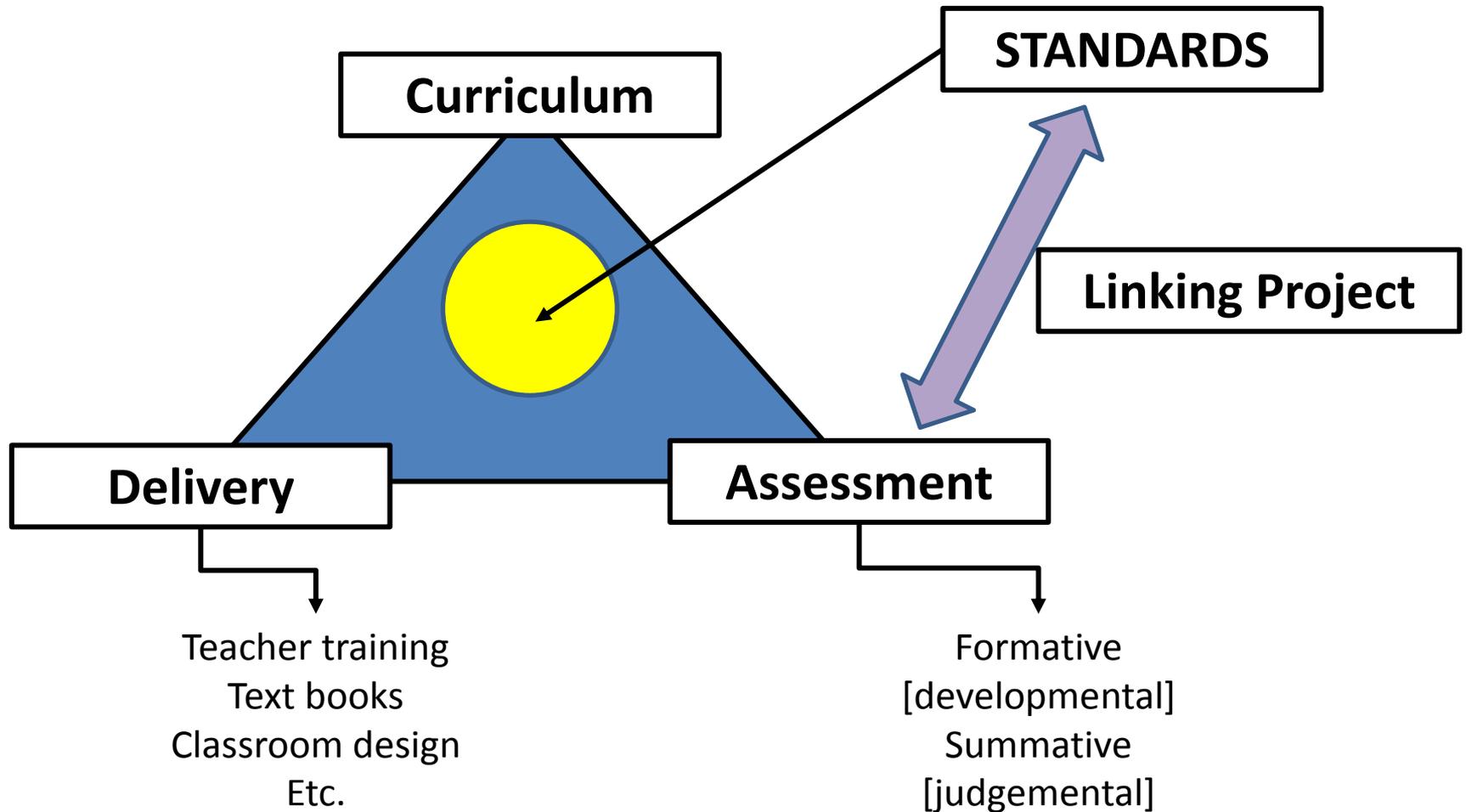
on Conducting the Collaborative Research on Linking UK English Tests to China's Standards of English



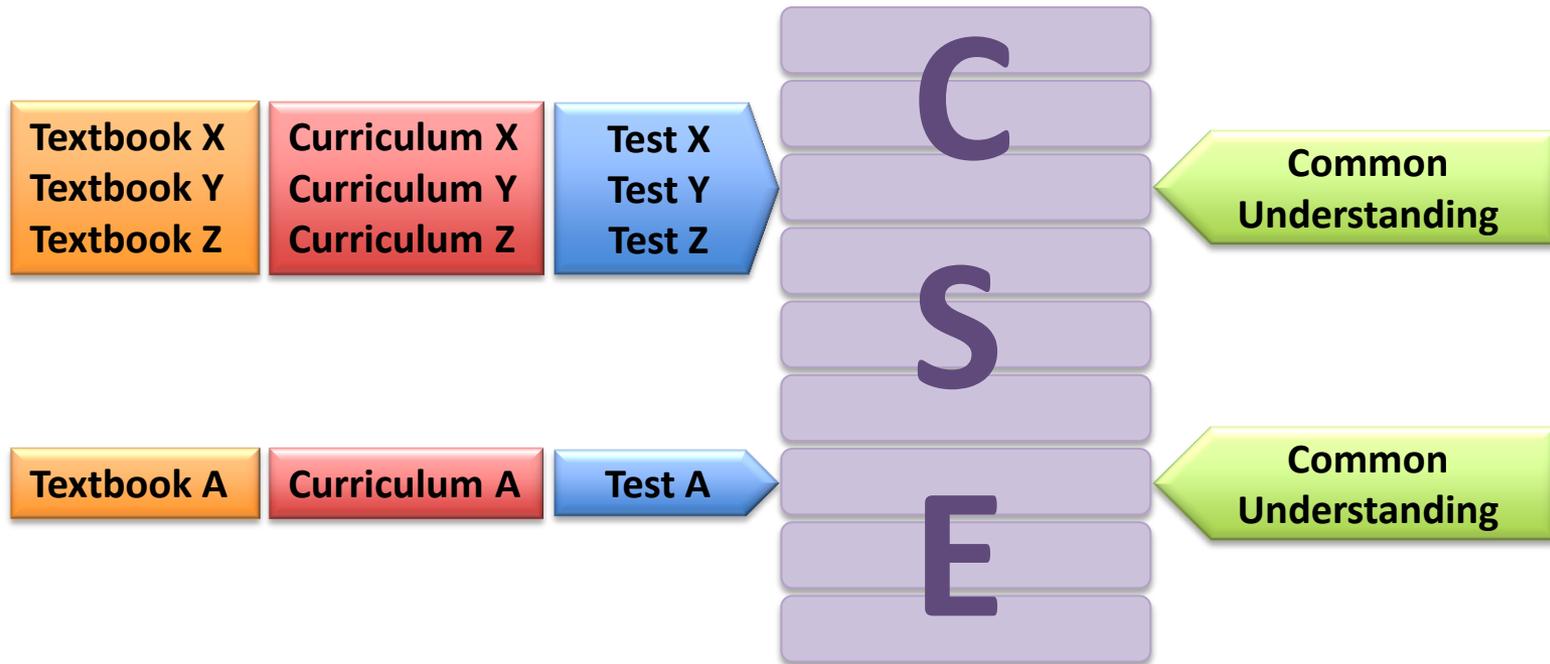
NEEA British Council

Identify an appropriate set of procedures through which test developers should demonstrate a theoretical and empirical link between a specific test and the CSE

The Learning System



Standards



China's Standards of English

C3	能准确、透彻地把握各种形式语言材料，并对其进行分析评价。能自如地与他人进行深入沟通和交流，内涵丰富，表达自然、娴熟、细腻，语言具有个人风格。
C2	能理解不同体裁和题材的复杂的语言材料，领悟其深刻内涵、语言风格或论证特点。能在多种场合熟练使用各种策略进行交流，并就不同话题展开充分的描述、说明、解释或论证，发表自己的意见，评析他人的观点，表达流利、地道。
C1	能理解多种话题和体裁的语言材料，包括一些抽象的或较为复杂的专业材料，批判性地审视、评析材料的内容，准确把握主旨要义，深刻理解隐含意义。能就多种话题进行深入交流或讨论，有效地描述、说明、解释、论证或评析事件或观点，交流顺畅，表达得体。
B3	能理解专业领域内有关一般性话题的语言材料，把握要点及其逻辑关系，批判性地分析材料的内容、隐含的意义、各方的观点或态度。能参与多种话题，包括专业领域内一般性话题的交流和讨论，有效地传递信息，发表见解，并对不同意见进行比较和评析，表达连贯、得体。
B2	能理解不同场合中有关一般性话题的语言材料，抓住主旨，把握重点，明晰事实、观点与佐证的细节，领悟他人的意图、态度。能在不同场合进行交流或讨论，表明观点或态度，能就一般性话题进行准确、有效地描述、说明或阐述，并能建立和维持较为广泛的人际关系。
B1	能理解在一般社交场合中有关熟悉话题的语言材料，抓住主题和主要内容，把握主要事实与观点，清楚他人的意图、态度。能就这些话题进行交流，叙述事件发展、描绘事物状况、介绍相关活动、说明事物要点，或简单论述个人观点，表达清晰、连贯，取得良好的交际效果。
A3	能理解日常生活中一般社交场合中简单的语言材料，获取特定或关键信息，抓住要点，推断他人的意图。能在日常生活或一般社交场合中用简单的语言与他人交流，描述活动、交待事情、说明情况、表达观点等，交流基本连贯、顺畅，并能建立良好的人际关系。
A2	能理解日常生活中常见的、简单的语言材料，获取基本的真实性信息，抓住主要内容。能就熟悉的话题或身边的事物用简单的语言进行交流，陈述信息、叙述事件、描述情况、表达基本的意图等，实现基本的交际目的。
A1	能理解日常生活中十分熟悉的、简单的语言材料，识别相关活动或身边事物的基本信息。能用基本的、十分简短的话语与他人交流，互相问候、相互介绍或描述，交待身边事物的基本信息、表明态度等，建立基本的人际关系，交流时有时需要借用重复、手势或其他语言之外的辅助手段。

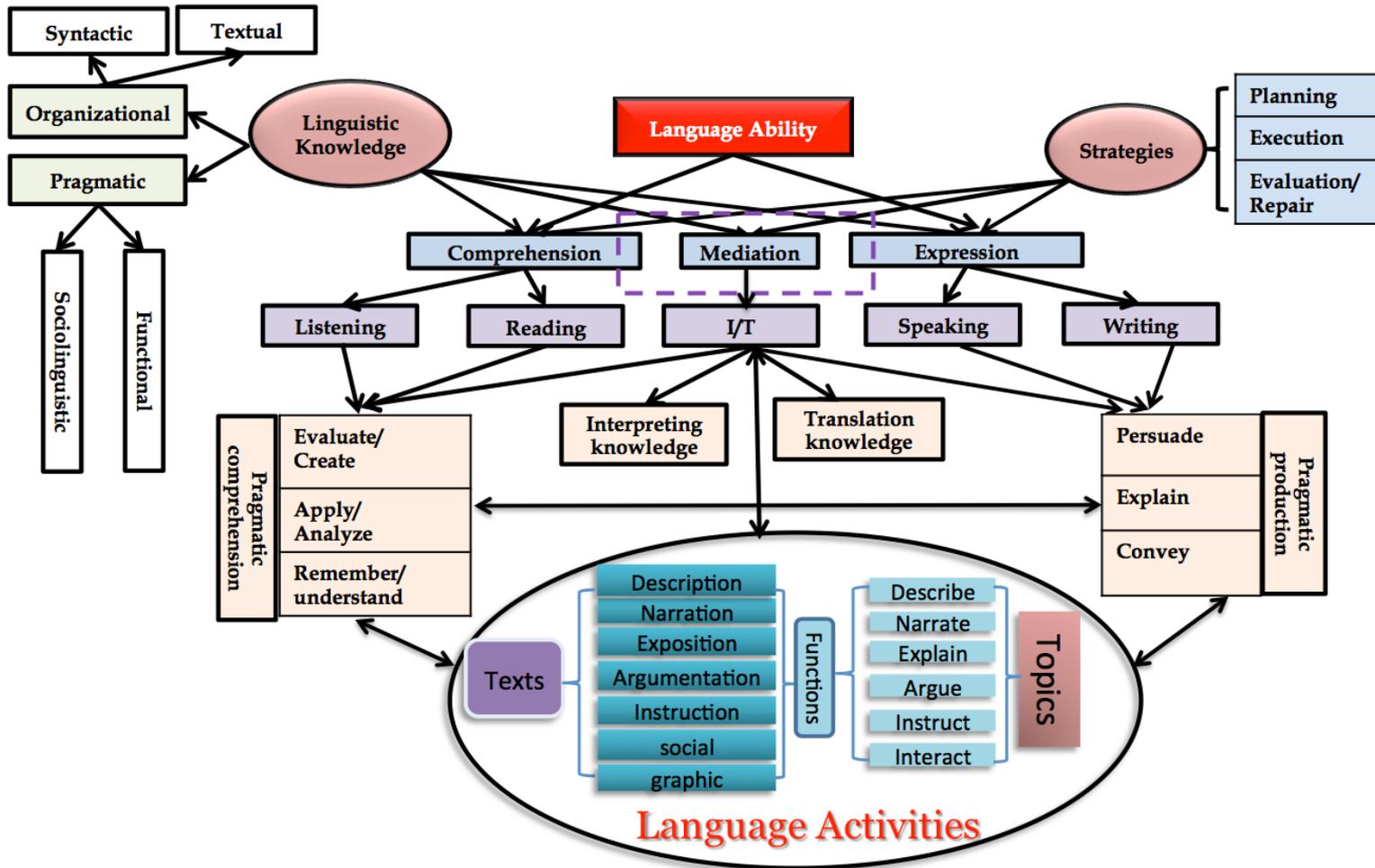
Number of Descriptors										
CSE Scale	C3	C2	C1	B3	B2	B1	A3	A2	A1	Total
Listening	14	35	61	121	153	142	99	50	29	704
Reading	50	61	106	120	147	149	165	103	19	920
Speaking	35	101	137	169	116	120	84	71	20	853
Writing	11	51	73	125	111	116	104	80	32	703
Organizational knowledge	12	26	32	54	58	79	80	49	38	428
Pragmatic knowledge	25	33	43	49	52	50	46	47	41	386
Translation	126	203	199	153	65	2				748
Interpreting	10	91	110	93						304
Total	283	601	761	884	702	658	578	400	179	5046

Global Scale

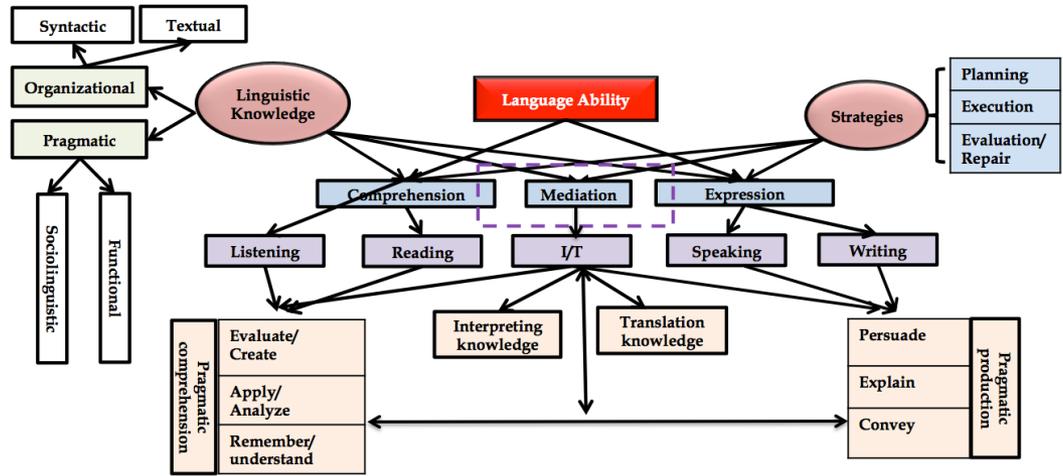
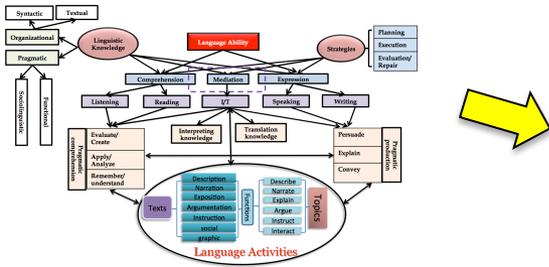
underlying models

THEORETICAL BASIS OF THE CSE

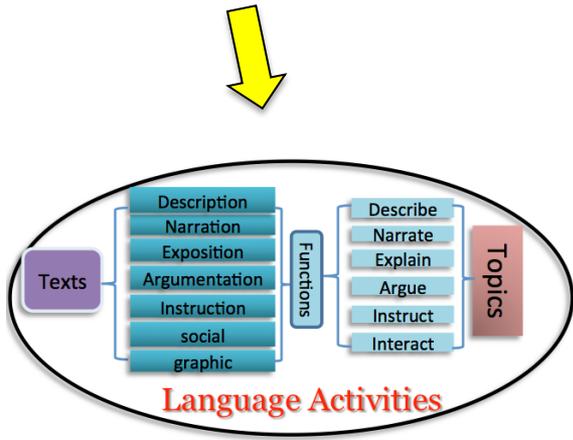
Use Oriented Language Ability



Exploring the Approach



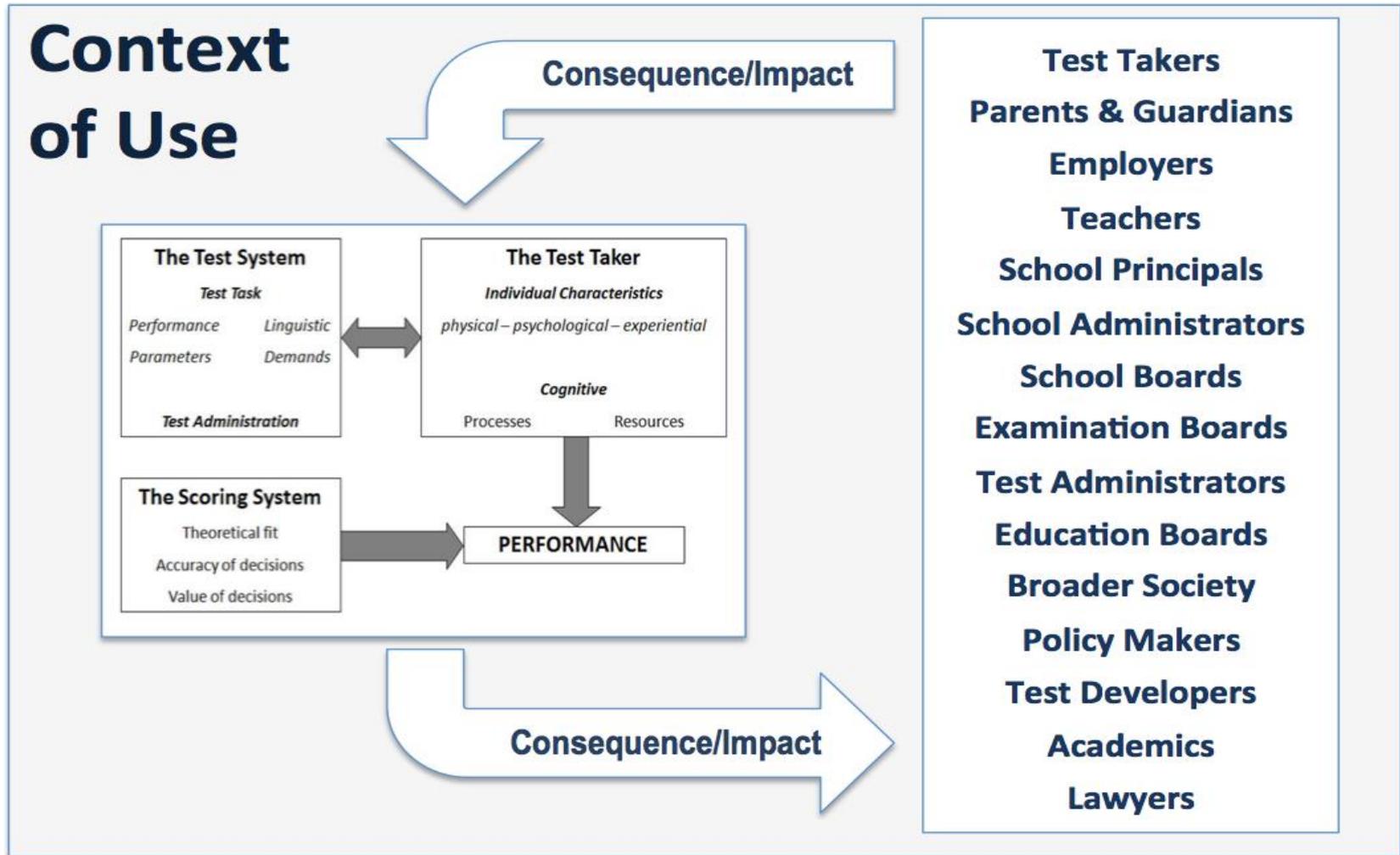
LANGUAGE MODEL



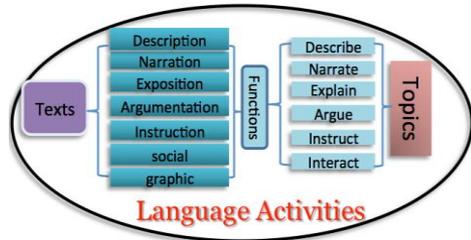
LANGUAGE USE MODEL

LEARNER MODEL
 Implied in contextualisation of CSE
 in China's Education System

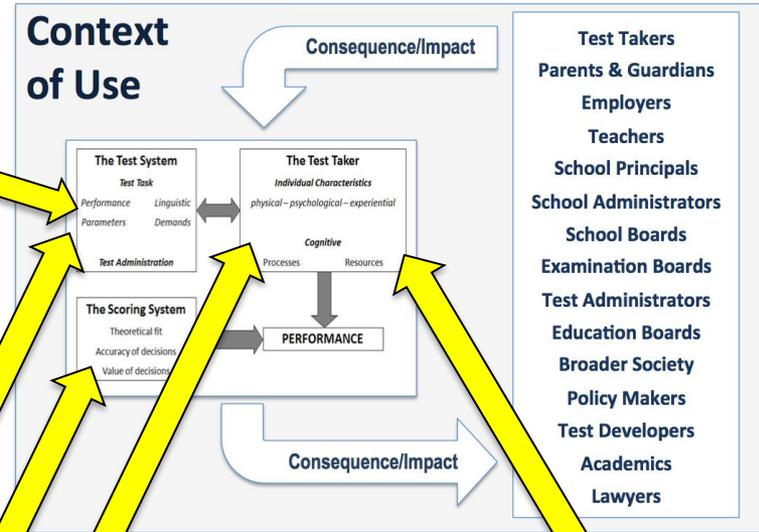
A Socio-Cognitive CSE



A Socio-Cognitive CSE



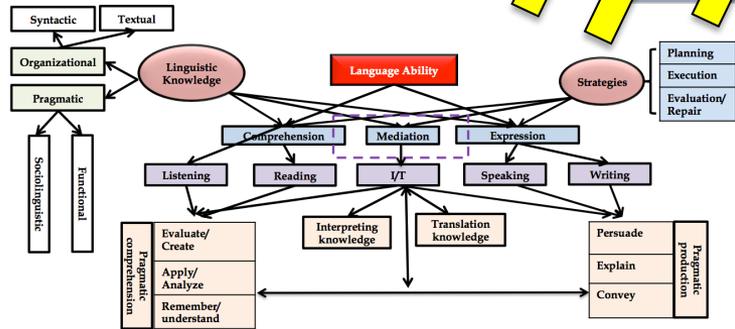
LANGUAGE USE MODEL



- 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

LEARNER MODEL

Implied in contextualisation of CSE in China's Education System



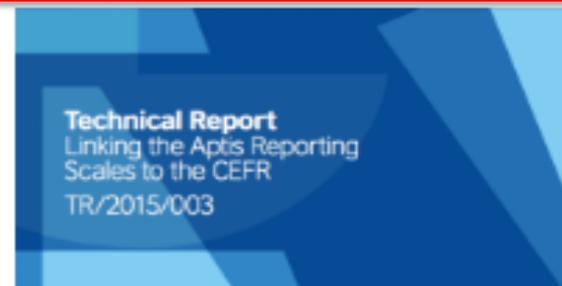
LANGUAGE MODEL

Assessment Research Group



Aptis Test Development Approach

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



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.



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.

the approach

CONCEPTUALISING LINKING

Sources & Types of Evidence

Construct Related The trait/ability being tested

Measurement Related Consistency & accuracy

Criterion Related Comparison with other estimates

Project Description

Steering Group

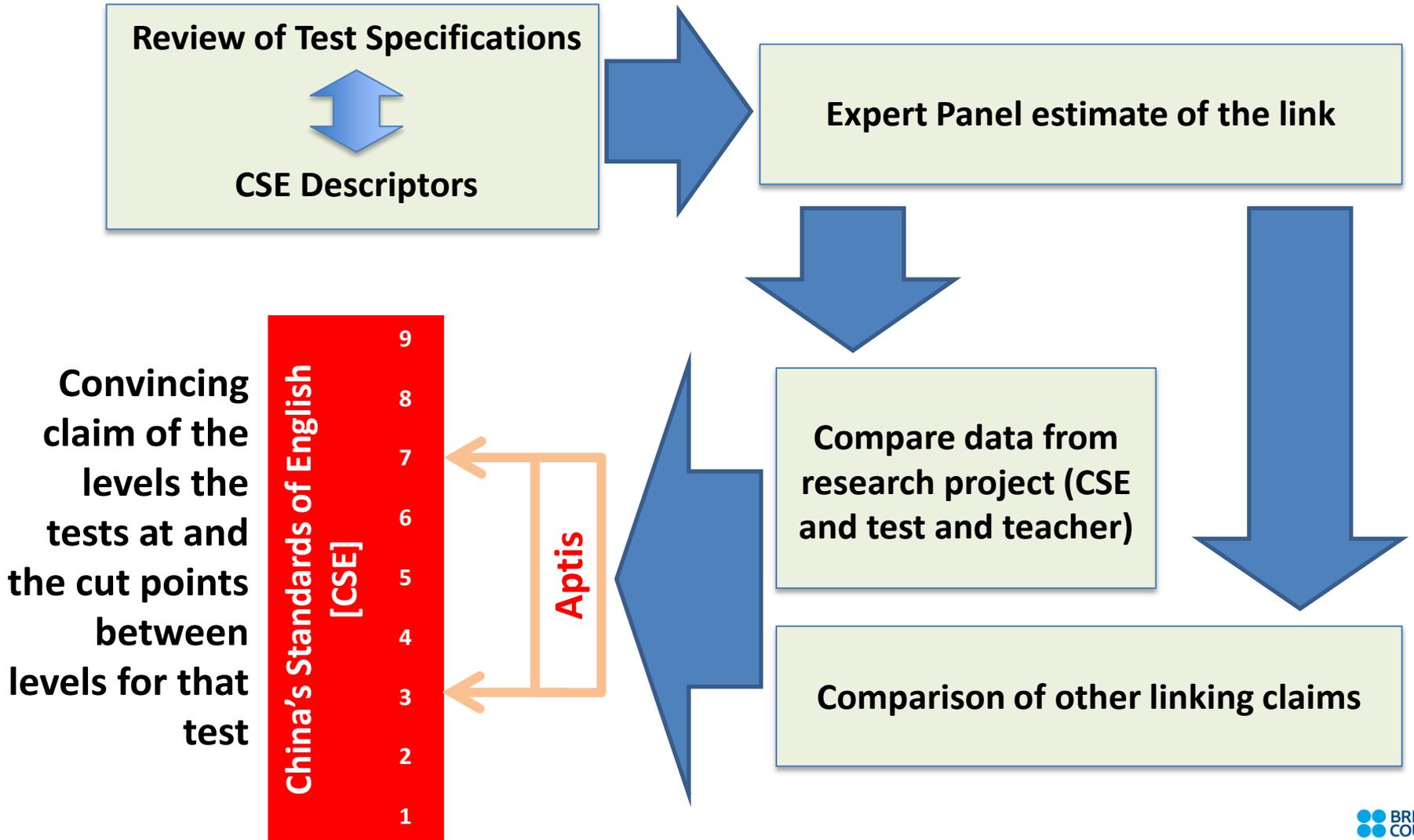
Working Group

NEEA – CSE – British Council – Cambridge

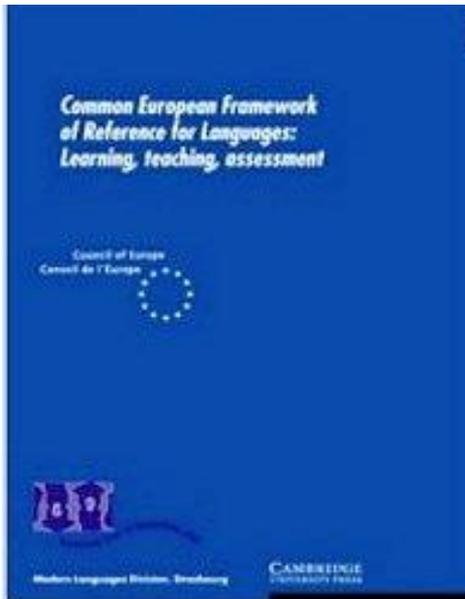
Objectives

Establish the linking process
Pilot the process (using Aptis)
Apply the process to IELTS

Overview of the Process



Drawing on experience



- Published by the Council of Europe in 2001
- “Formal origins of the CEFR date back to 1991” (Morrow,2004)
- 40 years of research in language education in Europe (Morrow,2004; Trim, 2010)
 - *Waystage, Threshold, Vantage*
- Main scaling studies carried out in Switzerland in 1994-1995

3 key goals from the CEFR:

- [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, and accordingly will aid European mobility.*

Linking to what part of the framework

“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

Linking to the CEFR

Manual for linking exams to the CEFR (2009)

❖ Familiarization

- Ensure knowledge of the CEFR

❖ Specification

- Describe what the exam covers and how exam content relates to the CEFR

❖ Standardisation

- Training to ensure a common understanding of the Common reference levels

○ Standard setting

- Systematic setting of cutoff points for CEFR levels on a test

❖ Validation

- The collection and analysis of test data...to provide evidence that both the examination itself and the link to the CEFR are sound

Linking studies in Europe

- Aptis (O'Sullivan, 2015)
- City & Guilds Communicator IESOL Examination (O'Sullivan, 2008)
- Dutch state foreign language examinations (Berger, Kuiper, & Maris, 2009; Noijons & Kuipers, 2010)
- TestDAF (Kecker & Eckes, 2010)
- Trinity College Examinations (Papageorgio, 2007; Papageorgio, 2009)
- TOEFL iBT (Tannenbaum & Wylie, 2008)
- GEPT, Taiwan (Wu & Wu, 2010)
- GEPT, Taiwan (Brunfaut & Harding, 2014)
- EIKEN, Japan (Dunlea & Figueras, 2012)
- EIKEN, Japan (Dunlea, 2016)
- VSTEP, Vietnam (Tran, Nguyen, Dang, Nguyen, Nguyen, Huynh, Do, Nguyen, Davidson)

Assessment Research Group



Aptis Test Development Approach

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



Linking the Aptis Reporting Scales to the CEFR

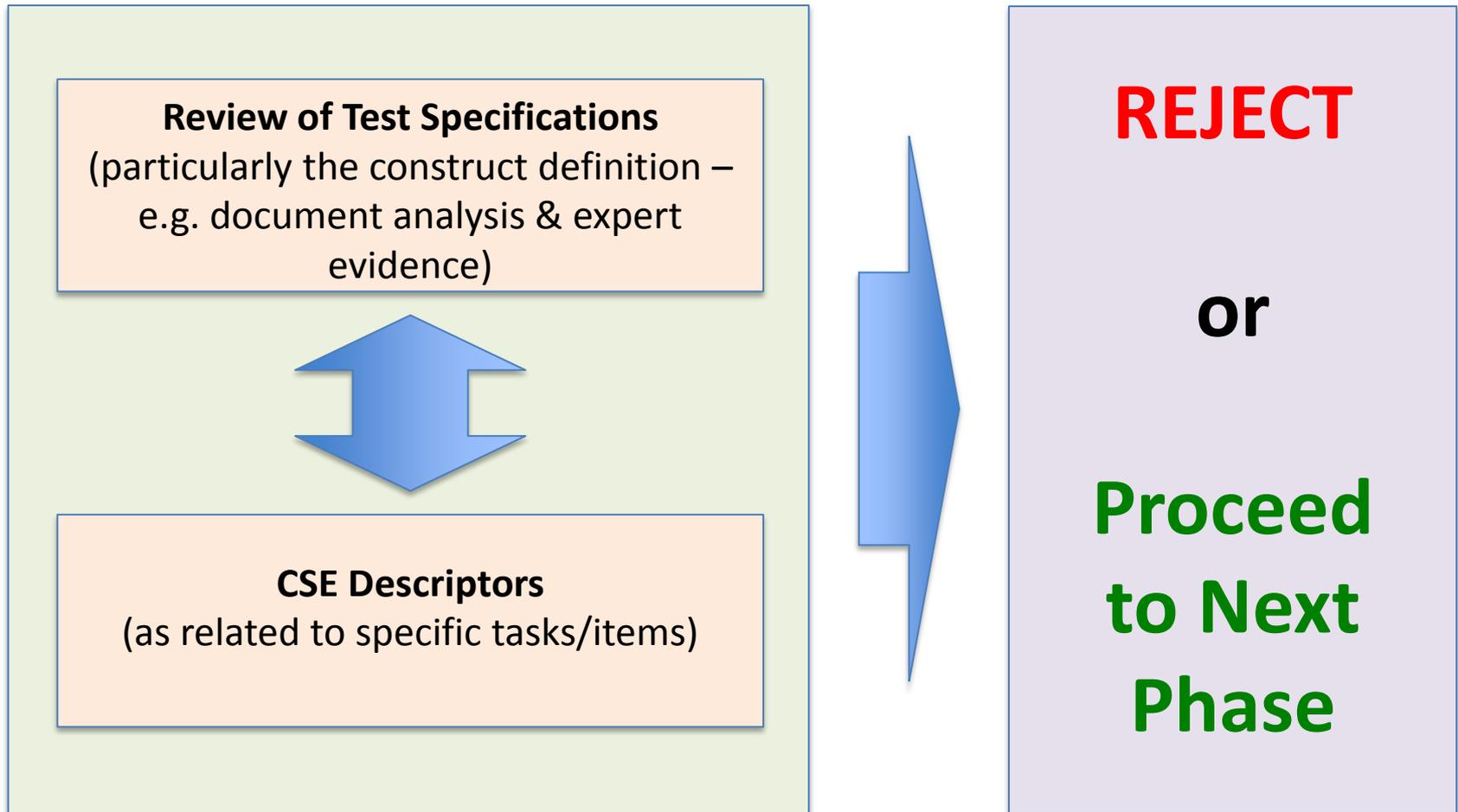
The Aptis development project marked a new era for the British Council, even though it had been involved in a number of test development projects in the past.



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.

Construct Appropriateness



Content analysis templates

Categories Reading	Task 1	(Task 1) Item 1	(Task 1) Item 2	(Task 1) Item 3	(Task 1) Item 4	(Task 1) Item 5
	CONSENSUS	CONSENSUS	CONSENSUS	CONSENSUS	CONSENSUS	CONSENSUS
Features of the TASK	Features of the TASK	Features of the TASK	Features of the TASK	Features of the TASK	Features of the TASK	Features of the TASK
Skill focus	sentence comprehension, lexis					
Task Level (CEFR)	A1					
Response format	Multiple choice gap fill					
Items per task	5					
Cognitive processing 1	Careful reading: local					
Cognitive processing 2	Establishing propositional meaning (cl./sent. level)					
Content knowledge	1 (General)					
Cultural specificity	1 (Neutral)					
Features of the Input Text	Features of the Input Text	Features of the Input Text	Features of the Input Text	Features of the Input Text	Features of the Input Text	Features of the Input Text
Domain	Personal					
Discourse mode	Descriptive					
Nature of information	Only concrete					
Topic	Daily life					
Text genre	Personal letters / e-mail					
Presentation	Verbal (written)					
Features of the Response	Features of the Response	Features of the Response	Features of the Response	Features of the Response	Features of the Response	Features of the Response
Key information		Within Sentences				
Operation		Main idea / conclusions				
Question presentation		Verbal (written)				
Option Presentation		Verbal (written)				

Assessment Research Group



Aptis Test Development Approach

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



Linking the Aptis Reporting Scales to the CEFR

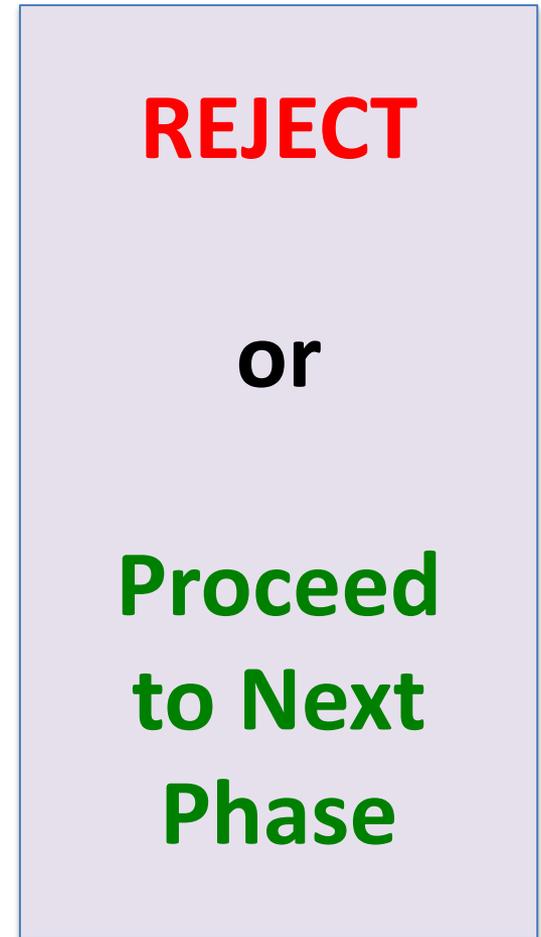
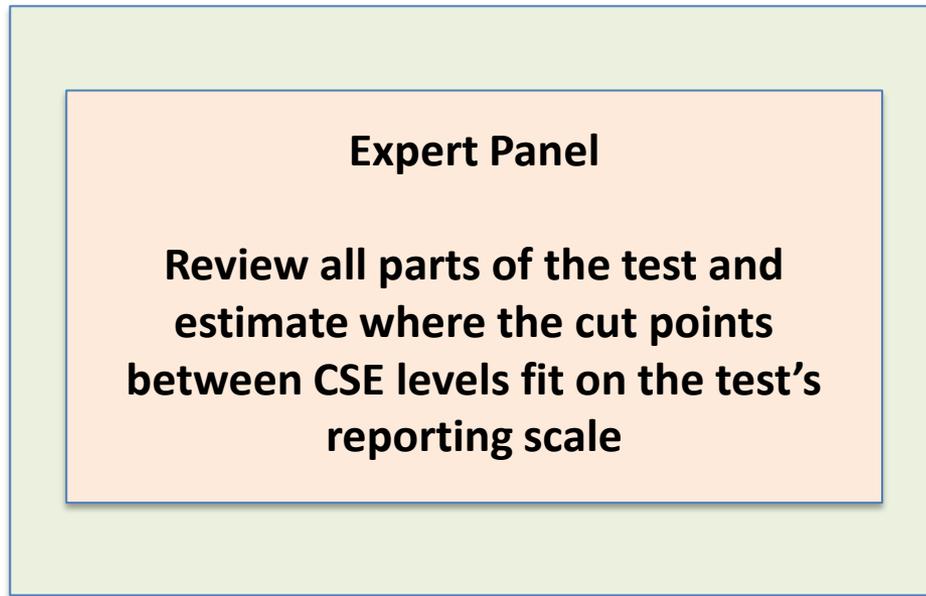
The Aptis development project marked a new era for the British Council, even though it had been involved in a number of test development projects in the past.



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.

Expert Panel Phase



Measurement Phase

Appropriate Comparisons (e.g. teacher estimations of CSE level compared to test results and/or scaling data from other studies and scales)

and/or

Compare Scaling Data
(e.g. CSE scaling compared to CEFR scaling or other framework where appropriate)

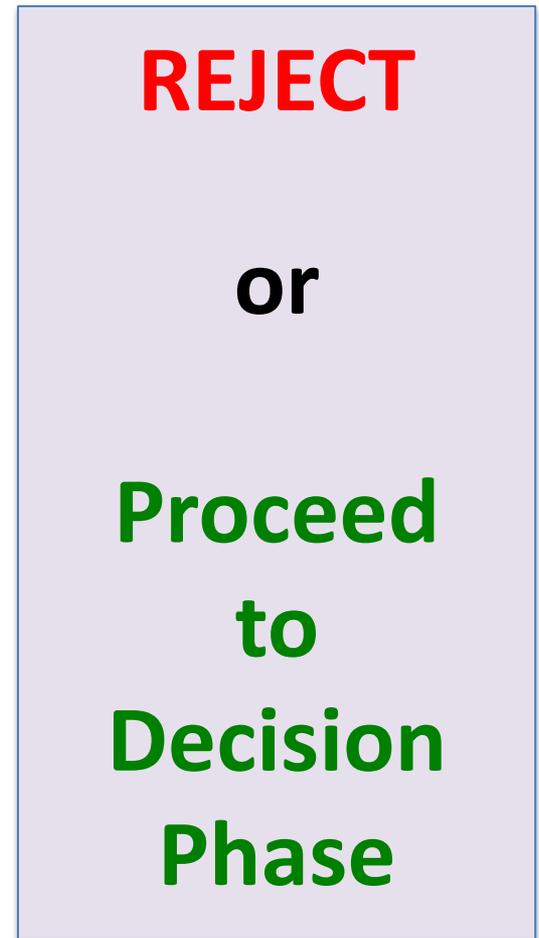
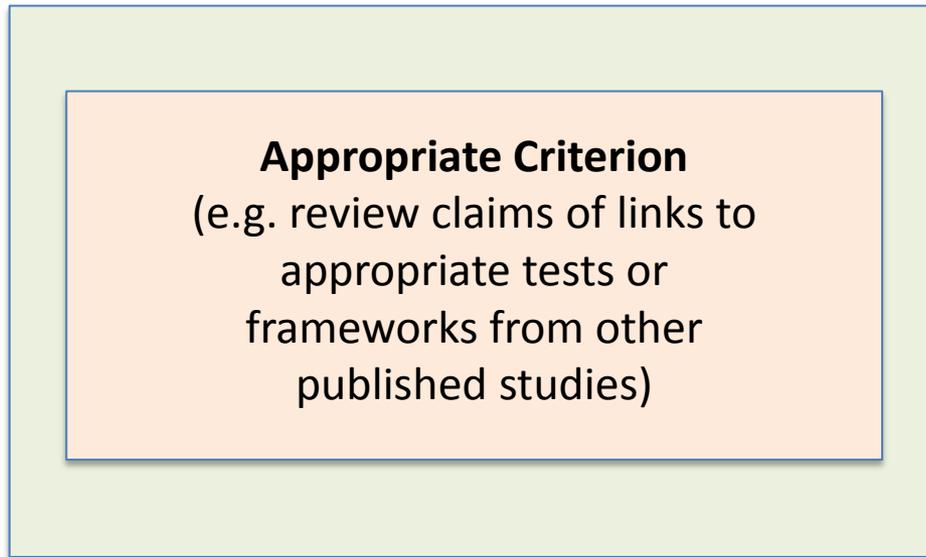


REJECT

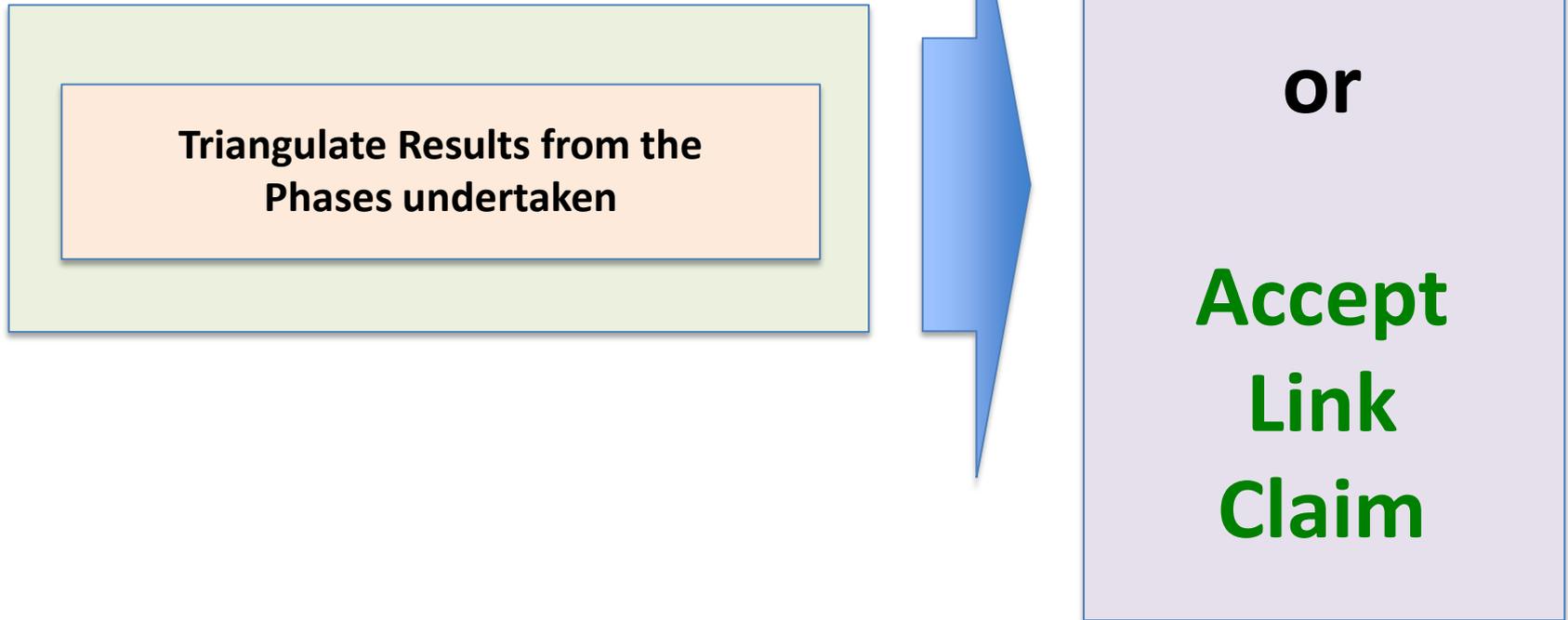
or

**Continue
with
Criterion
Phase**

Criterion Phase



Decision Phase



Linking to the CEFR: standard setting

- The proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance. (Cizek, 1993)
- The results “are seldom, if ever, purely statistical, psychometric, impartial, apolitical, or ideologically neutral activities.” (Cizek & Bunch, 2007)

Linking to the CEFR: standard setting

Pilot phase: Aptis and IELTS Listening

- Panel: university educators + specialists with testing expertise and knowledge of the tests
- Pre-standard setting session familiarization activities
 - Familiarization with CSE listening scales using self study preparation booklet before the event
- Standard setting session for Listening
 - 3 day event
 - Day 1: training and focus on CSE levels and standard setting training
 - Day 2: standard setting for Aptis listening
 - Day 3: standard setting for IELTS listening

Linking to the CEFR: standard setting

- The proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance. (Cizek, 1993)
- The results “are seldom, if ever, purely statistical, psychometric, impartial, apolitical, or ideologically neutral activities.” (Cizek & Bunch, 2007)

Pilot phase for listening: some Findings

Expert Panel

16 members

Researchers

Educators

Testing experts

Leaders

Working Group

members

Judgement Round 1

Analysis & Report

Discussion

Judgement Round 2

Analysis & Report

Preliminary Decision

Analysis

Final Decision

R
E
V
I
E
W

Judgement Round 1

Analysis & Report

Discussion

Judgement Round 2

Analysis & Report

Preliminary Decision

Analysis

Final Decision

Aptis

IELTS

Pilot phase for listening: some Findings

- Expert Panel**
- Overwhelmingly positive about experience
 - Similarly confident in outcomes

- Preliminary Analysis**
- Aptis listening ranges from CSE 3 to 7
 - IELTS Listening ranges from CSE 4 to 8

Next steps

- Replicate similar standard setting panels for
 - Reading
 - Writing
 - Speaking
- Collect student test data and teacher judgments for those students (ongoing)
- Collate construct definition and all standard setting data into coherent, comprehensive validity arguments as evidence of any claimed links between the exams and CSE
- Create a manual outlining the methodology to demonstrate best practice in linking tests

Conclusion

Critical

```
graph TD; A[Critical] --> B[The Construct Phase must be included]; A --> C[The process must be fully followed];
```

**The Construct Phase
must be included**

**Otherwise any later claim is
meaningless**

**The process must be
fully followed**

**Otherwise linking projects &
claims cannot be compared
or evaluated**

Thank You