Linking Examinations to the CSE

Barry O’Sullivan
Jamie Dunlea
British Council
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

Professor Barry O’Sullivan

Barry’s recent work includes the development and validation of a new business-to-business language test called APTIS. He is the founding president of the UK Association for 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 MAIA (Kuala Lumpur) and at the University of Luson.

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 non-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.
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
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
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

Curriculum

STANDARDS

Linking Project

Delivery

Teacher training
Text books
Classroom design
Etc.

Assessment

Formative
[developmental]
Summative
[judgemental]
Stands

Textbook X
Textbook Y
Textbook Z

Curriculum X
Curriculum Y
Curriculum Z

Test X
Test Y
Test Z

Common Understanding

Textbook A
Curriculum A
Test A

Common Understanding
## China’s Standards of English

### CSE Scale

<table>
<thead>
<tr>
<th>Number of Descriptors</th>
<th>C3</th>
<th>C2</th>
<th>C1</th>
<th>B3</th>
<th>B2</th>
<th>B1</th>
<th>A3</th>
<th>A2</th>
<th>A1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listening</strong></td>
<td>14</td>
<td>35</td>
<td>61</td>
<td>121</td>
<td>153</td>
<td>142</td>
<td>99</td>
<td>50</td>
<td>29</td>
<td>704</td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td>50</td>
<td>61</td>
<td>106</td>
<td>120</td>
<td>147</td>
<td>149</td>
<td>165</td>
<td>103</td>
<td>19</td>
<td>920</td>
</tr>
<tr>
<td><strong>Speaking</strong></td>
<td>35</td>
<td>101</td>
<td>137</td>
<td>169</td>
<td>116</td>
<td>120</td>
<td>84</td>
<td>71</td>
<td>20</td>
<td>853</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td>11</td>
<td>51</td>
<td>73</td>
<td>125</td>
<td>111</td>
<td>116</td>
<td>104</td>
<td>80</td>
<td>32</td>
<td>703</td>
</tr>
<tr>
<td><strong>Organizational knowledge</strong></td>
<td>12</td>
<td>26</td>
<td>32</td>
<td>54</td>
<td>58</td>
<td>79</td>
<td>80</td>
<td>49</td>
<td>38</td>
<td>428</td>
</tr>
<tr>
<td><strong>Pragmatic knowledge</strong></td>
<td>25</td>
<td>33</td>
<td>43</td>
<td>49</td>
<td>52</td>
<td>50</td>
<td>46</td>
<td>47</td>
<td>41</td>
<td>386</td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td>126</td>
<td>203</td>
<td>199</td>
<td>153</td>
<td>65</td>
<td>2</td>
<td>122</td>
<td>100</td>
<td>179</td>
<td>748</td>
</tr>
<tr>
<td><strong>Interpreting</strong></td>
<td>10</td>
<td>91</td>
<td>110</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>304</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>283</td>
<td>601</td>
<td>761</td>
<td>884</td>
<td>702</td>
<td>658</td>
<td>578</td>
<td>400</td>
<td>179</td>
<td>5046</td>
</tr>
</tbody>
</table>

### Global Scale

- **Listening**: 704
- **Reading**: 920
- **Speaking**: 853
- **Writing**: 703
- **Organizational knowledge**: 428
- **Pragmatic knowledge**: 386
- **Translation**: 748
- **Interpreting**: 304
- **Total**: 5046
THEORETICAL BASIS OF THE CSE

underlying models
Use Oriented Language Ability

Linguistic Knowledge

Comprehension
Listening
Reading
I/T

Mediation

Expression
Speaking
Writing

Translation knowledge
Interpreting knowledge

Planning
Execution
Evaluation/Repair

Pragmatic Production
Persuade
Explain
Convey

Texts
Description
Narration
Exposition
Argumentation
Instruction
social
graphic

Functions
Describe
Narrate
Explain
Argue
Instruct
Interact

Topics

Language Activities
Exploring the Approach

LANGUAGE MODEL

LANGUANGE USE MODEL

LEARNER MODEL

Implyed in contextualisation of CSE in China’s Education System
A Socio-Cognitive CSE

Context of Use

The Test System
- Test Task
  - Performance
  - Parameters
  - Test Administration

The Scoring System
- Theoretical fit
- Accuracy of decisions
- Value of decisions

The Test Taker
- Individual Characteristics
  - Physical—psychological—experiential
- Cognitive
  - Processes
  - Resources

PERFORMANCE

Consequence/Impact

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
A Socio-Cognitive CSE

LANGUAGE USE MODEL

LEARNER MODEL

Implied in contextualisation of CSE in China’s Education System
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.
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
NEEA – CSE – British Council – Cambridge

Working Group

Objectives
Establish the linking process
Pilot the process (using Aptis)
Apply the process to IELTS
Overview of the Process

Review of Test Specifications

CSE Descriptors

Expert Panel estimate of the link

Compare data from research project (CSE and test and teacher)

Comparison of other linking claims

Convincing claim of the levels the tests at and the cut points between levels for that test

China's Standards of English

Aptis

9
8
7
6
5
4
3
2
1
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
  o Communicative activities
  o Strategies
  o 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

Review of Test Specifications
(particularly the construct definition – e.g. document analysis & expert evidence)

CSE Descriptors
(as related to specific tasks/items)

REJECT
or
Proceed to Next Phase
## Content analysis templates

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

Expert Panel

Review all parts of the test and estimate where the cut points between CSE levels fit on the test’s reporting scale

REJECT or Proceed to Next 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

Appropriate Criterion
(e.g. review claims of links to appropriate tests or frameworks from other published studies)

REJECT or Proceed to Decision Phase
Decision Phase

Triangulate Results from the Phases undertaken

REJECT or Accept Link Claim
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

Aptis
- Judgement Round 1
- Analysis & Report
- Discussion
- Judgement Round 2
- Analysis & Report
- Preliminary Decision
- Analysis
- Final Decision

IELTS
- Judgement Round 1
- Analysis & Report
- Discussion
- Judgement Round 2
- Analysis & Report
- Preliminary Decision
- Analysis
- Final Decision
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

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