Unwarranted claims about CEF alignment of some international English language tests

John H.A.L. de Jong
June, 2009
Founding Pearson Language Tests

- 2004 Start-up & planning
- 2005 Recruitment
- 2006 October: Start development of PTE Academic
- 2007 First field test
- 2008 Second field test
- 2009 Launch
Test Development - Field Testing

- Total of > 10,400 candidates in 2 Field tests
- Total of 158 countries of birth
- Total of 126 languages spoken at home
- Thousands of items trialled
- Close to 300 human markers trained
- Items rated on up to 7 different traits
- Total of 2.6m human ratings gathered
  - 2m for Speaking
  - 0.6m for writing
Four methods

- Item writer training & writing to CEF
- Item estimates of CEF level in reviewing process
- Inclusion of DIALANG anchors in field test
- Human rating on CEF scale of candidates
Two independent estimates of CEF cut-offs

\[ y = 0.4082x + 0.5057 \]

\[ R^2 = 0.9934 \]
Problem

If we have preliminary mapping of PTE Academic to
- CEF
- TOEFL
- IELTS

AND they already have linking claims to CEF

THEN \textbf{B2 on PTE A} = \textbf{B2 on TOEFL} = \textbf{B2 on IELTS}

etc.
Claims ...

Our exams are linked to the CEFR

Mapping TOEFL iBT® on the Common European Framework Reference (CEFR)
Beware of misunderstandings

Being at a CEF level means:
Being more likely than not be able to perform language activities at the level of efficiency such as captured in the exemplary descriptors provided in the CEF.

It CANNOT be
Being able to perform ALL language activities for a particular level provided in the CEF.

Simply because the CEF only contains exemplary descriptors. For example all descriptors for Mediation are missing.
Beware of misunderstandings

follows two basic dimensions

Quantity

Quality
The quantity development is in fact multidimensional and quality can develop along each of the dimensions.
A quote from the CEF:

(…) all knowledge of a language is **partial**, however much of a ‘mother tongue’ or ‘native language’ it seems to be. (…) In addition, a given individual **never has equal mastery** of the different component parts of the language in question, for example, of oral and written skills, or of comprehension and interpretation compared to production skills.

(CEF, page 169; emphasis added, JdJ).
Profiled development
CEF scaled on IRT model

Chance to perform task

100%
80%
60%
50%
40%
30%
20%
10%
0%

Pearson Language Tests
CEF scaled on IRT model

95% at Level A1

But only 15% at level B2
Fact finding

Then what are the claimed relationships for

- TOEFL
- IELTS

And how do they relate to evidence?
## Claims and (incoherent) evidence

<table>
<thead>
<tr>
<th>Top 50</th>
<th>IELTS U</th>
<th>TOEFL U</th>
<th>IELTS P</th>
<th>TOEFL P</th>
<th>IELTS P-U</th>
<th>TOEFL P-U</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>28</td>
<td>33</td>
<td>29</td>
<td>34</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>100%</td>
<td>55%</td>
<td>65%</td>
<td>57%</td>
<td>67%</td>
<td>47%</td>
<td>57%</td>
</tr>
<tr>
<td>Average</td>
<td>6.8</td>
<td>87</td>
<td>6.8</td>
<td>87</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>stdev</td>
<td>0.5</td>
<td>14</td>
<td>0.3</td>
<td>12</td>
<td>0.3</td>
<td>14</td>
</tr>
<tr>
<td>max</td>
<td>7.0</td>
<td>111</td>
<td>7.0</td>
<td>111</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>min</td>
<td>5.0</td>
<td>57</td>
<td>6.0</td>
<td>61</td>
<td>-0.5</td>
<td>-39</td>
</tr>
</tbody>
</table>

1) There are 51 schools because 2 schools are ex aequo at rank 50.
## Figure 2: Indicative IELTS band scores at CEFR and NQF levels

<table>
<thead>
<tr>
<th>Corresponding NQF Level</th>
<th>Corresponding CEFR Level</th>
<th>IELTS approximate band score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 3</td>
<td>C2</td>
<td>7.5+</td>
</tr>
<tr>
<td>Level 2</td>
<td><strong>C1</strong></td>
<td><strong>6.5/7.0</strong></td>
</tr>
<tr>
<td>Level 1</td>
<td>B2</td>
<td>5.0/5.5/6.0</td>
</tr>
<tr>
<td>Entry 3</td>
<td>B1</td>
<td>3.5/4.0/4.5</td>
</tr>
<tr>
<td>Entry 2</td>
<td>A2</td>
<td>3.0</td>
</tr>
</tbody>
</table>

## Claims ets

<table>
<thead>
<tr>
<th>Test Section</th>
<th>Total Score Scale Range</th>
<th>A1</th>
<th>A2</th>
<th>Minimum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOEFL iBT Total¹</td>
<td>0–120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT Reading</td>
<td>0–30</td>
<td>8</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT Listening</td>
<td>0–30</td>
<td>13</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT Speaking</td>
<td>0–30</td>
<td>8</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>TOEFL iBT Writing</td>
<td>0–30</td>
<td>11</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

¹ TOEFL scores are reported in 1-point increments.
... and (incoherent) evidence

<table>
<thead>
<tr>
<th>Top 50 1) USA</th>
<th>IELTS U</th>
<th>TOEFL U</th>
<th>IELTS P</th>
<th>TOEFL P</th>
<th>IELTS P-U</th>
<th>TOEFL P-U</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>28</td>
<td>33</td>
<td>29</td>
<td>34</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>100%</td>
<td>55%</td>
<td>65%</td>
<td>57%</td>
<td>67%</td>
<td>47%</td>
<td>57%</td>
</tr>
<tr>
<td>Average</td>
<td>6.8</td>
<td>87</td>
<td>6.8</td>
<td>87</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>stdev</td>
<td>0.5</td>
<td>14</td>
<td>0.3</td>
<td>12</td>
<td>0.3</td>
<td>14</td>
</tr>
<tr>
<td>max</td>
<td>7.0</td>
<td>111</td>
<td>7.0</td>
<td>111</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>min</td>
<td>5.0</td>
<td>57</td>
<td>6.0</td>
<td>61</td>
<td>-0.5</td>
<td>-39</td>
</tr>
</tbody>
</table>

1) There are 51 schools because 2 schools are ex aequo at rank 50

=C1 says Cambridge

=ETS says ETS
No different in the UK ...

<table>
<thead>
<tr>
<th>Top 10</th>
<th>IELTS U</th>
<th>TOEFL U</th>
<th>IELTS P</th>
<th>TOEFL P</th>
<th>IELTS P-U</th>
<th>TOEFL P-U</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>90%</td>
<td>100%</td>
<td>90%</td>
<td>100%</td>
<td>90%</td>
</tr>
<tr>
<td>Average</td>
<td>6.5</td>
<td>94</td>
<td>6.7</td>
<td>97</td>
<td>0.3</td>
<td>2</td>
</tr>
<tr>
<td>StDev</td>
<td>0.4</td>
<td>6</td>
<td>0.3</td>
<td>6</td>
<td>0.4</td>
<td>4</td>
</tr>
<tr>
<td>Max</td>
<td>7.0</td>
<td>100</td>
<td>7.0</td>
<td>107</td>
<td>1.0</td>
<td>10</td>
</tr>
<tr>
<td>Min</td>
<td>6.0</td>
<td>85</td>
<td>6.5</td>
<td>89</td>
<td>0.0</td>
<td>0</td>
</tr>
</tbody>
</table>

=C1 says Cambridge

=B2 says ETS
Lynda Taylor (2004) concludes her article:

The current alignment is based upon a growing body of internal research, combined with long established experience of test use within education and society, as well as feedback from a range of test stakeholders regarding the uses of test results for particular purposes. As we grow in our understanding of the relationship between IELTS, other Cambridge ESOL examinations and the CEFR levels, so the frame of reference may need to be revised accordingly.
So is who is dumb?

Distribution of TOEFL iBT and IELTS Candidates

- **A1+A2**: TOEFL iBT - 20%, IELTS - 5%
- **B1**: TOEFL iBT - 40%, IELTS - 40%
- **B2**: TOEFL iBT - 30%, IELTS - 30%
- **C1**: TOEFL iBT - 10%, IELTS - 20%
- **C2**: TOEFL iBT - 0%, IELTS - 10%
ETS: TOEIC ~ TOEFL Concordance

So: TOEFL and TOEIC mapping to CEF corresponds to earlier concordance of TOEFL & TOEIC
Comparing IELTS ~ TOEFL iBT

- C1
- B2
- B1

IELTS Scores vs. TOEFL iBT Scores

CEF Claims
Comparing IELTS ~ TOEFL iBT

- C1: USA Universities (n=332)
- B2: UK CAN AUS (n=197)
Comparing IELTS ~ TOEFL iBT

- CEF Claims
- Empirical (n=217)
- USA Universities (n=332)
- UK CAN AUS (n=197)
So what appears?

Data from
1. USA Universities
2. UK Australian and Canadian Universities
3. International student distributions
4. Equating study run by ETS
5. Equating results using PTE Academic as anchor

All agree to suggest that

\[ \text{B2 on PTE A} = \text{B2 on TOEFL} = \text{C1 on IELTS} \]
### Underestimate explained

06/09/2008 at 23:16:41

<table>
<thead>
<tr>
<th>Common European Framework</th>
<th>Guided Learning Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>approximately 180–200</td>
</tr>
<tr>
<td>B1</td>
<td>approximately 350–400</td>
</tr>
<tr>
<td>B2</td>
<td>approximately 500–600</td>
</tr>
<tr>
<td>C1</td>
<td>approximately 700–800</td>
</tr>
<tr>
<td>C2</td>
<td>approximately 1,000–1,200</td>
</tr>
</tbody>
</table>
Dear John de Jong,

The numbers quoted on this page are an approximate figure for how long it might take to prepare a candidate for each level of the CEFR if they have little/no previous experience. A candidate who was already of C1 standard therefore, may only require a further 200-400 guided learning hours to reach level C2, rather than an additional 1000-1200.

I hope that this is of assistance.

Kind Regards

Terry Knox
ESOL Helpdesk
Cambridge ESOL Customer Services
Underestimate explained
IELTS 8 Very Good User: C2 or B2?

Band 8 Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.

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.

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.
Questions?

john.dejong@pearson.com