The Bell project: data collection

Constant Leung  
Kings College London

Michael Evans  
Cambridge Faculty of Education

Neil Jones  
Consultant, Cambridge English

Yongcan Liu  
Cambridge Faculty of Education
Comparative Judgment

• Comparative judgment is one way of generating data for an Item Response Theory (IRT) analysis, so first it is necessary to introduce IRT.

• IRT provides test scales which have the features of true measurement (think of a thermometer):
  – Judge different tests on the same scale;
  – Work with meaningful units;
  – Enable comparison, meaningful interpretation of points on scale

• (Unfortunately IRT is hardly used in operational educational testing in the UK).
Item bank links all levels

Tests at appropriate level

Measurement scale

Standards consistently applied

Intermediate

Preliminary

Breakthrough

Learners located on scale
IRT – The Rasch model
Comparative judgment

Can-do statements from different sources
Rank statements

Compare pairs of samples: better or worse?

Rank whole set
Automatically set standards

CEFR A2

CEFR B1
Construct a measurement scale
CIEP multilingual benchmarking of Speaking
Sevres June 2008

![Graph showing rankings and ratings for different languages: German, English, Spanish, French, and Italian. The graph plots rankings on the y-axis against ratings on the x-axis, with distinct markers for each language level (A1, A2, B1, B2).]
No More Marking

- [https://nomoremarking.com](https://nomoremarking.com)
- This is the website we are using for the study.
- The name shows CJ being promoted as a superior alternative to marking – because relative judgment is more accurate than absolute judgment.
- Do visit the site:
  - take the *colours test* to prove that CJ works
  - Look at the Ofqual study, showing big differences in the standards of maths exams
- The site makes it easy for anybody to participate as a judge.
Retell short narrative stories through pictures

Speech is still often fragmented, but carries meaning.
Retell short narrative stories through pictures.

Can understand questions and instructions addressed carefully and slowly to him/her and follow short, simple directions.

Which is the higher performance level?
Can communicate immediate, concrete matters in longer utterances, with some cohesion, e.g. Plant die because chloroplasts get no sun, no glucose.

Can express how he/she feels in simple terms, and express thanks.
Repeat words, simple phases

Answer questions with one or two words (e.g., “Where is Sonia?”)
Retell short narrative stories through pictures

Speech is still often fragmented, but carries meaning.
Can communicate immediate, concrete matters in longer utterances, with some cohesion, e.g. Plant die because chloroplasts get no sun, no glucose.

Can understand questions and instructions addressed carefully and slowly to him/her and follow short, simple directions.
Answer questions with one or two words (e.g., “Where is Sonia?”)

Can express how he/she feels in simple terms, and express thanks
Answer questions with one or two words (e.g., “Where is Sonia?”)

Speech is still often fragmented, but carries meaning.
<table>
<thead>
<tr>
<th>Source</th>
<th>Skill</th>
<th>Level</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASSEA</td>
<td>Speaking</td>
<td>Step 2</td>
<td>Speech is still often fragmented, but carries meaning.</td>
</tr>
<tr>
<td>NASSEA</td>
<td>Speaking</td>
<td>Step 3</td>
<td>Can communicate immediate, concrete matters in longer utterances, with some cohesion, e.g. Plant die because chloroplasts get no sun, no glucose.</td>
</tr>
<tr>
<td>CEFR</td>
<td>Spoken interaction</td>
<td>A1</td>
<td>Can understand questions and instructions addressed carefully and slowly to him/her and follow short, simple directions.</td>
</tr>
<tr>
<td>CEFR</td>
<td>Spoken interaction</td>
<td>A2</td>
<td>Can express how he/she feels in simple terms, and express thanks</td>
</tr>
<tr>
<td>WIDA</td>
<td>SPEAKING</td>
<td>Level 2 Beginning</td>
<td>Answer questions with one or two words (e.g., “Where is Sonia?”)</td>
</tr>
<tr>
<td>WIDA</td>
<td>SPEAKING</td>
<td>Level 3 Developing</td>
<td>Retell short narrative stories through pictures</td>
</tr>
</tbody>
</table>
Goals of empirical study

• Drawing on teachers’ experience:
• Empirical equating of ‘levels’ referenced from different fields of work
• Enabling better shared understanding of levels.
• Verify performance of can-do statements (i.e. do users understand them in the same way?)
• Construction of scales with the potential to link different levels of assessment (classroom, external, self) into a common framework:
• Relate a wide range of evidence.
Thank you for your attention!

• Constant Leung:  constant.leung@kcl.ac.uk
• Neil Jones:  neilfjones@ntlworld.com
• Michael Evans:  mje1000@cam.ac.uk
• Yongcan Liu:  yl258@cam.ac.uk