

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

Funded by
the **bell** foundation

KING'S
College
LONDON

 **UNIVERSITY OF
CAMBRIDGE**
Faculty of Education

 **CAMBRIDGE ENGLISH**
Language Assessment
Part of the University of Cambridge

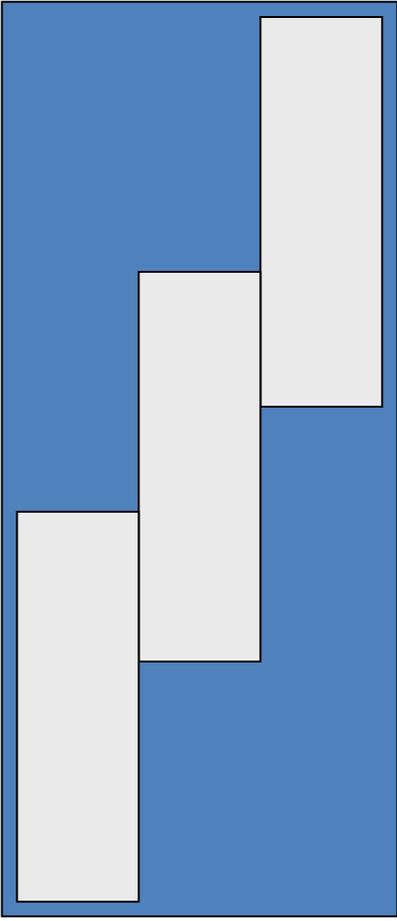
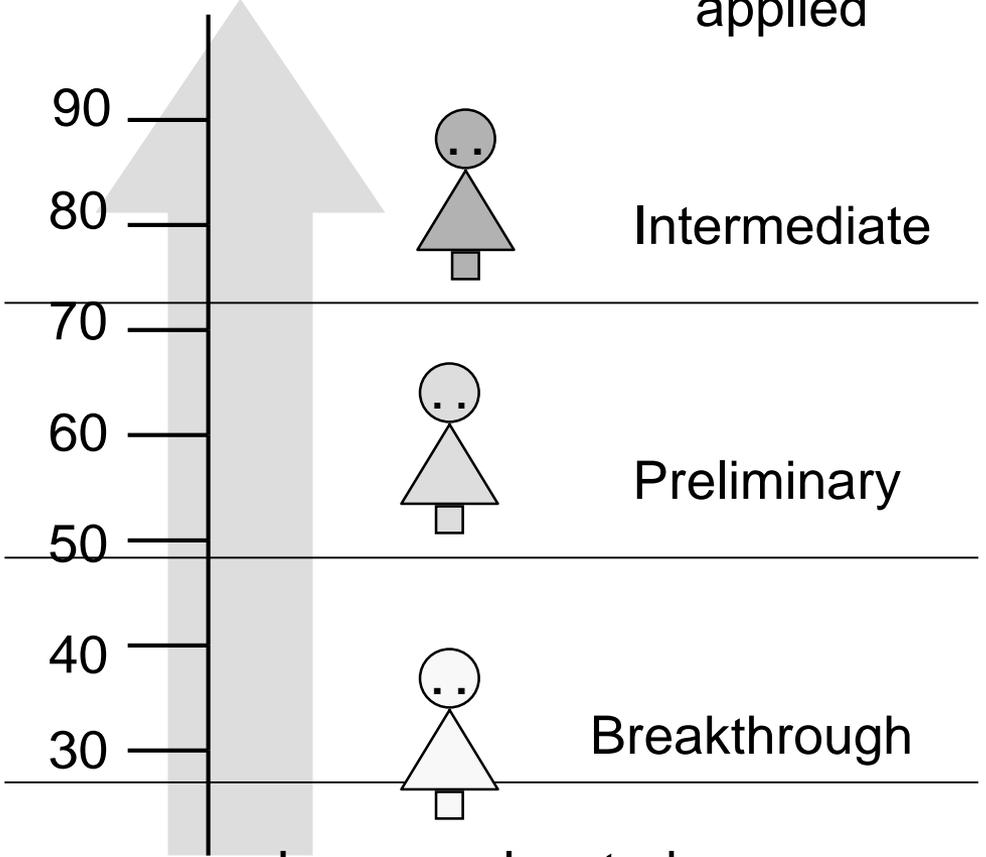
Comparative Judgment

- 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).



Measurement scale

Standards consistently applied



Test 1

Test 2

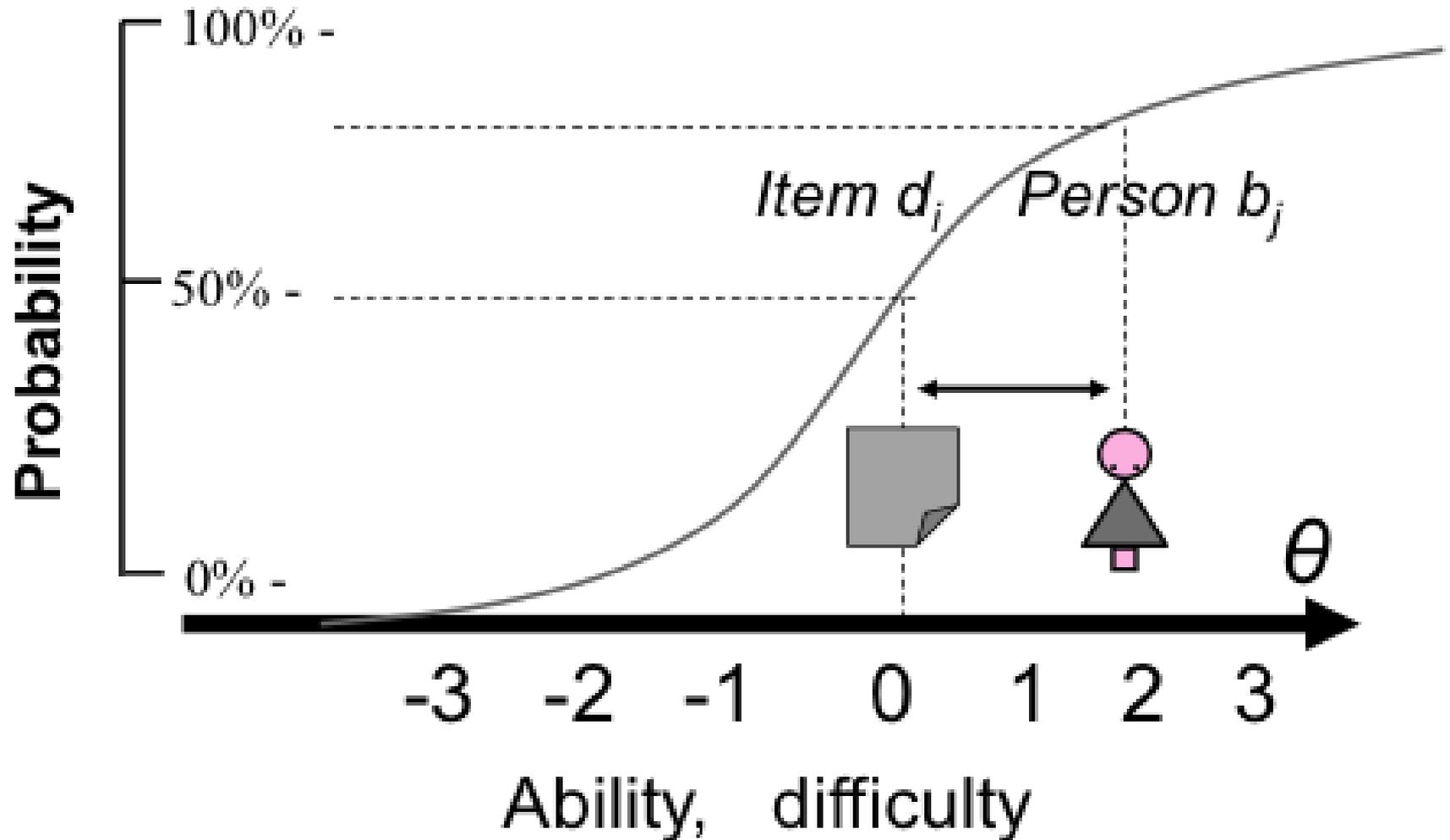
Test 3

Tests at appropriate level

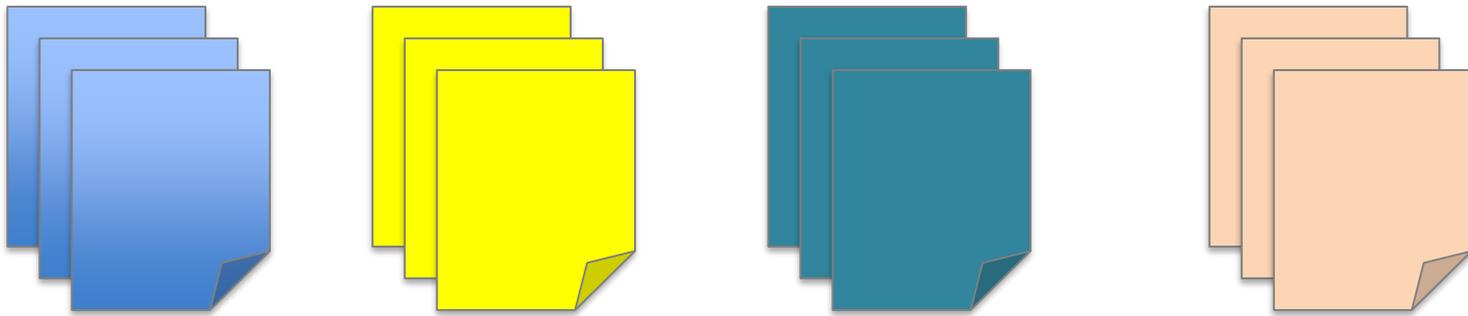
Learners located on scale

Item bank links all levels

IRT – The Rasch model

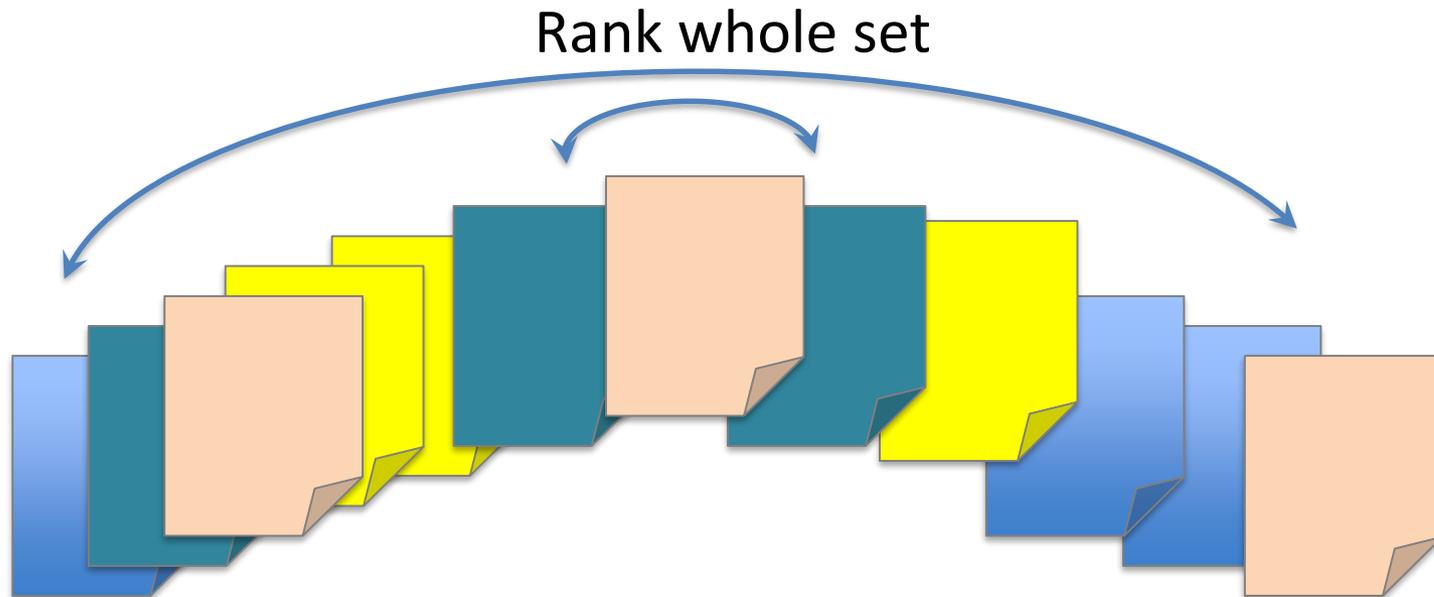


Comparative judgment



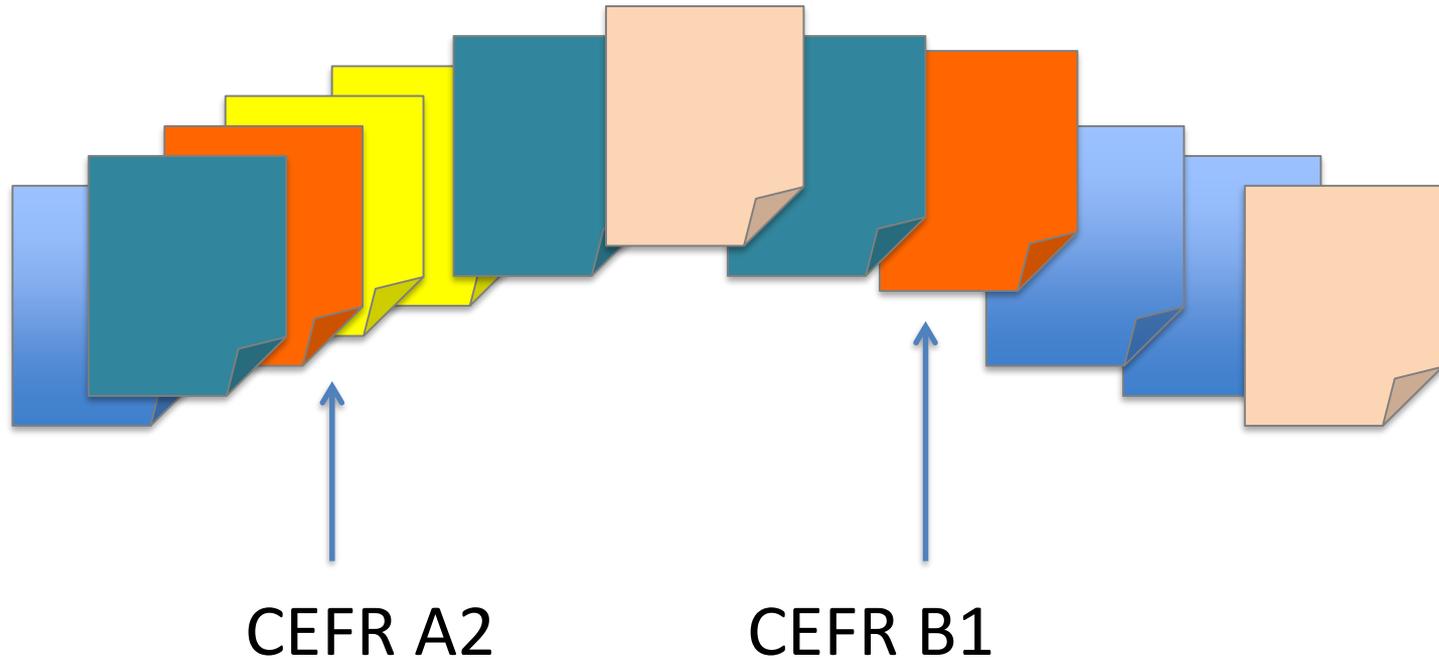
Can-do statements from different sources

Rank statements

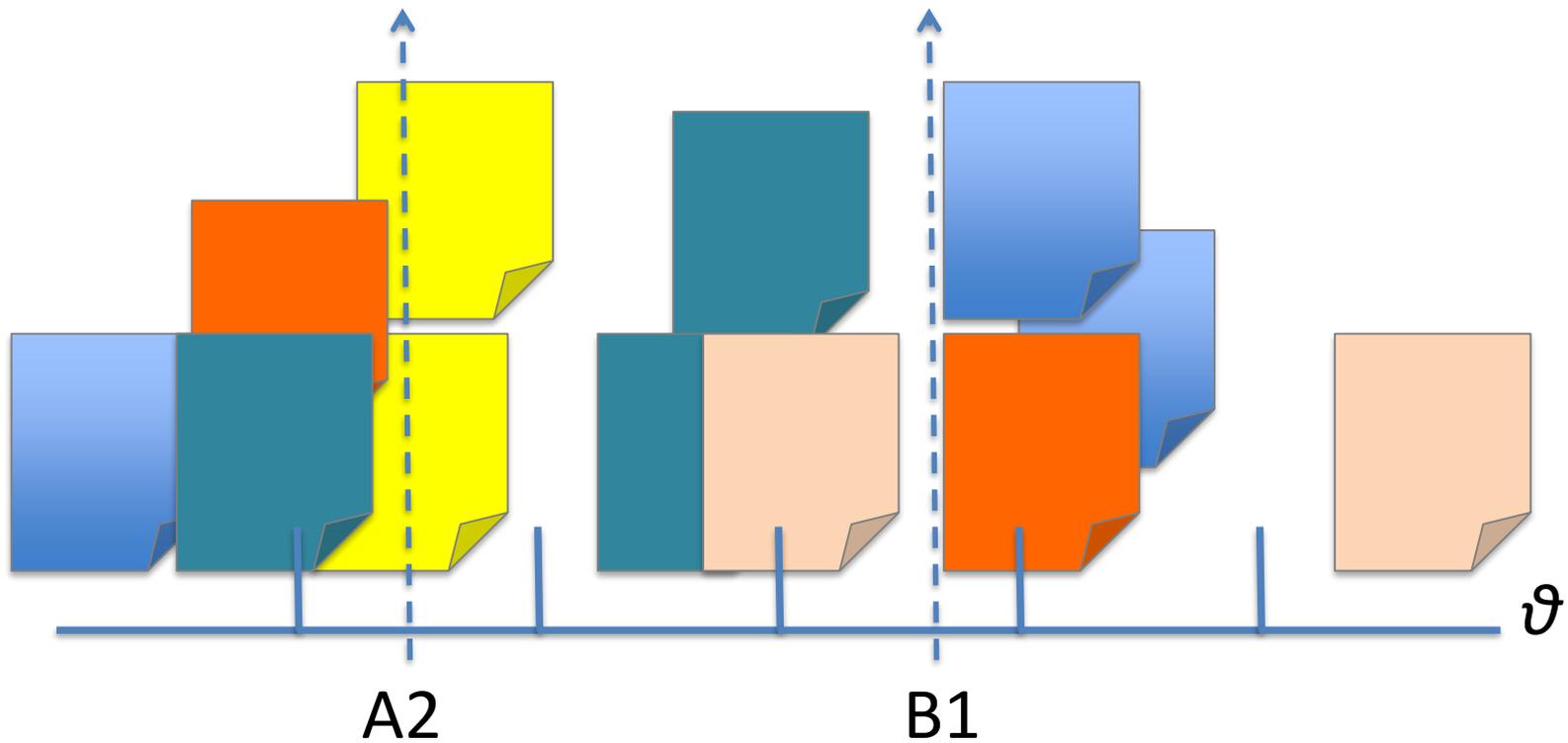


Compare pairs of
samples: better
or worse?

Automatically set standards

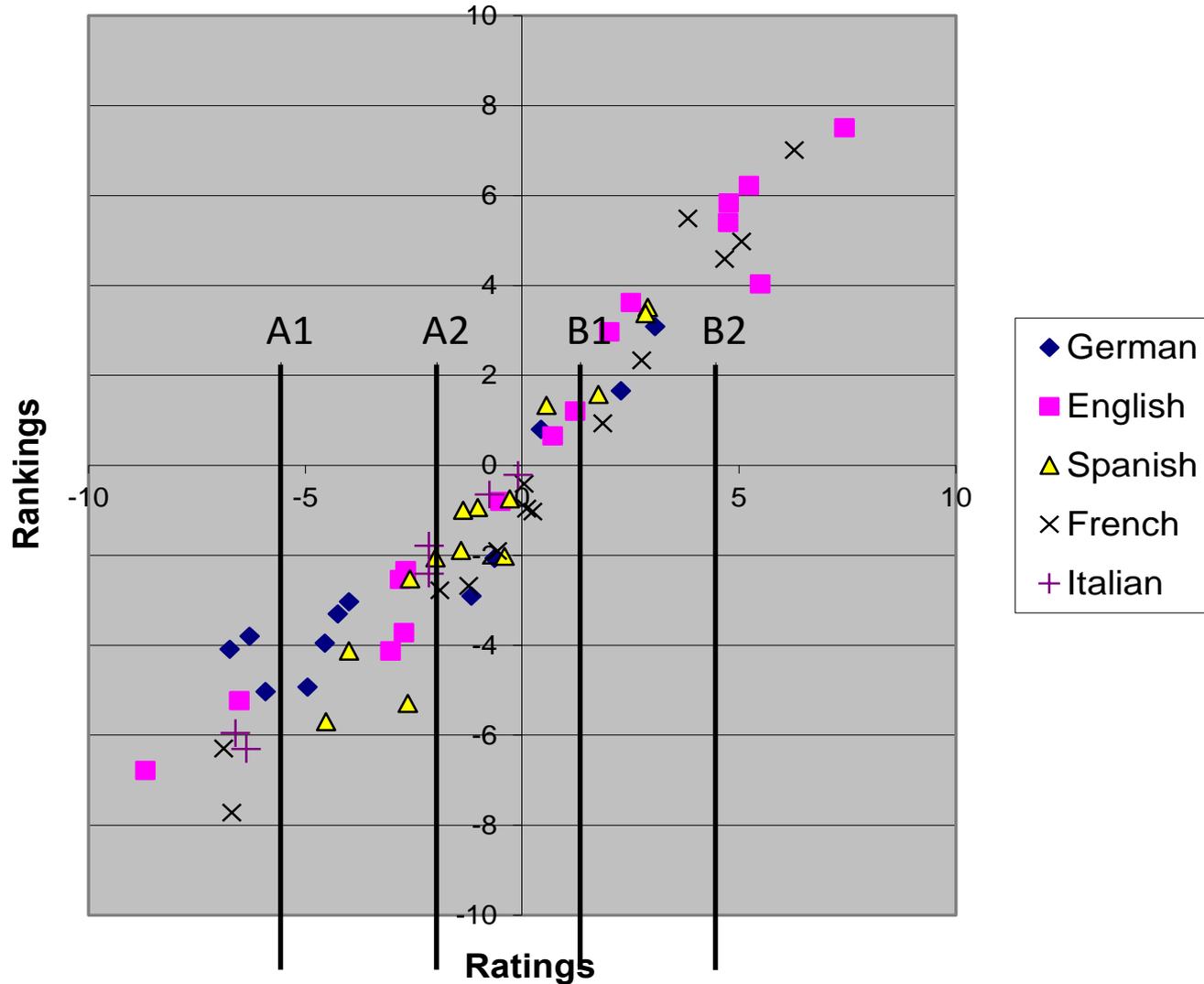


Construct a measurement scale



CIEP multilingual benchmarking of Speaking

Sevres June 2008



No More Marking

- <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.



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



< Left <

Which is the higher performance level?

> Right >



☐ ⬆️ ⬇️ ⬇️ 1 of 1 - + >>

Repeat words, simple phases

☐ ⬆️ ⬇️ ⬇️ 1 of 1 - + >>

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.

Source texts

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

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: neiljones@ntlworld.com
- Michael Evans: mje1000@cam.ac.uk
- Yongcan Liu: yl258@cam.ac.uk