4th Annual EALTA Conference, Sitges, 15-17 June 2007

Defining good practice in relating exams to the CEFR

Spiros Papageorgiou
Department of Linguistics and English Language
Lancaster University

Overview

- EALTA Good Practice Guidelines
- Define good practice in relating exams to the CEFR
- Report on a PhD study using data from the Trinity College London CEFR project

EALTA Guidelines for Good Practice

LINKAGE TO THE COMMON EUROPEAN FRAMEWORK

- What evidence is there of the quality of the process followed to link tests and examinations to the Common European Framework?
- Have the procedures recommended in the Manual and the Reference Supplement been applied appropriately?
- Is there a publicly available report on the linking process?

Defining good practice

- What constitutes ‘evidence’ of the quality of the process?
- What constitutes ‘proper application’ of the procedures recommended in the Manual?
- What kind of information should a publicly available report contain?

Evidence

- Procedural evidence (Kaftandjieva, 2004: 20)
- Good practice entails explanation of the following:

1. *Why linking to the CEFR?*
2. *Are the judges really experts in using the CEFR?*
3. *How relevant is the test content to the CEFR?*
4. *How do candidates perform in relation to the CEFR?*
5. *What happens in relation to other tests claiming the same CEFR level?*
6. *How about the quality of the test itself?*
7. *What should a report contain?*

CEFR linking process in the Manual

- Does the Manual offer guidance for providing evidence?

1.	*Why linking to the CEFR?*	*Ch. 2*
2.	*Are the judges really experts in using the CEFR?*	*Ch. 3*
3.	*How relevant is the test content to the CEFR?*	*Ch. 4*
4.	*How do candidates perform in relation to the CEFR?*	*Ch. 5*
5.	*What happens in relation to other tests claiming the same CEFR level?*	*Ch. 6*
6.	*How about the quality of the test itself?*	*Ch. 6*
7.	*What should a report contain?*	*Ch. 7*

Why linking to the CEFR?

- Acceptance of qualifications, more useful for users
- Meaningful way of reporting test scores

Trinity report:

1. *'wide acceptance that Trinity qualifications could have after being related to the CEFR'*
2. *'the project aims at explaining to test users what a score means when taking a Trinity exam'*

Who is qualified to make judgements?

- ‘many people, including so-called experts, claim a knowledge of the CEFR, where in reality they are only familiar with part of it’ (Alderson, 2005: 66)

- Trinity judges were posted a hard copy of the CEFR and were asked to study the whole volume.
- Instructed to pay attention to Chapters 3, 4, 5, 7
- Explained the purpose of the forthcoming Familiarisation activities

Examining familiarity with the CEFR levels in the Trinity project

- Familiarisation as a separate stage & part of Specification and Standardisation
- Sorting descriptors into levels
- Intra- and inter-judge consistency
- Agreement with the CEFR
- Judges' scaling of the CEFR descriptors
- Both CTT and IRT

Evidence of in-depth understanding of the CEFR levels

- Rank order correlations: only indicative of judges' general understanding of progression from one level to another

Scales	Spearman correlations		
	Mean	Min	Max
Speaking1	0.911	0.871	0.928
Speaking 2	0.958	0.913	0.985

- Descriptive statistics revealed lower scores

Scales	N	Mean	Min	Max	SD
Speaking 1	30	16.5	12	20	2.78
Speaking 2	30	20.67	12	27	4.29

Familiarisation claim

- The validity of the claim is strengthened if:

1. *Judges read the whole CEFR volume, not just Tables 1 and 2*
2. *Familiarisation is repeated throughout a linking project*
3. *The judges discuss which levels they cannot distinguish*

Effect of training: mixed results

- Descriptive statistics

Scales	Date	N	Mean	Min	Max	SD
Speaking	06/09/2005	30	16.5	12	20	2.78
Speaking	07/09/2005	30	20.67	12	27	4.29
Speaking	22/11/2005	30	19.4	12	28	5.13
Speaking	28/02/2006	30	18.82	12	26	4.49

- Rasch occasion measurement report for Speaking- differences in difficulty

Statistics	Results
Reliability	.83
Fixed (all same) chi-square	24.5
d.f.	3
significance (probability)	.00

Test content description in relation to the CEFR

C2	Gr12	Gr12	Gr12	
C1	Gr10 Gr11	G10 G11	G10 G11	G10 G11 G12
B2.2	G9 Gr10	Gr8 Gr9	Gr9	Gr9
B2.1	Gr7 Gr8 Gr9	Gr7 Gr8	Gr7 Gr8	Gr7 Gr8
B1.2	Gr6			
B1.1	Gr5 Gr6	Gr5 Gr6	Gr5 Gr6	Gr5 Gr6
A2.2	Gr4	Gr4	Gr4	Gr4
A2.1	Gr3	Gr2 Gr3	G3	Gr2 G3
A1	Gr1 Gr2	Gr2	Gr2	Gr2
Below A1		Gr1	Gr1	Gr1
	Overall Impression	Spoken Interaction	Spoken Production	Listening Compr.

Unrealistic expectations for test content in relation to the CEFR

- Has no difficulty in understanding any kind of spoken language, whether live or broadcast, delivered at fast native speed (C2 Listening)
- *'the C2 descriptor is completely unrealistic' (Trinity judge)*

- Can we really design a listening test at C2?

Specification claim

- The validity of the claim is strengthened if:

1. *Scales and descriptors are explicitly stated; simply referring to the level is not enough*
2. *Test specifications are aligned to the CEFR levels*
3. *The effect of group dynamics is examined*
4. *Bias of insiders is considered*

Bias of insiders

- Insights from a qualitative study into how judges interact during the Specification

046 Lora	look at the B1 under the global scale on page 4
047 Tim	yes
047 Lora	yes B1
048 Tim	it pretty well fits [.]

Performance in relation to the CEFR

- Systematic investigation into the factors influencing participants' ratings will contribute to the validity evidence of the cutscore (McGinty, 2005)
- Focus group discussions with the Trinity judges to investigate factors affecting decision-making during the Standardisation stage

Standardisation claim

- The validity of the claim is strengthened if the following are considered:

1. *Is the performance of the borderline pass candidate at the intended level?*
2. *When compensatory composite scoring is used, is a candidate at the intended level even if he/she failed some components?*

Empirical validation

- Internal validation: requirement for initiating the CEFR linking, not simply part of the linking process.
- External validation: Exam providers' cooperation is needed to compare performance on tests claiming the same level.

Impact of the CEFR linking

- Trinity report acknowledges positive impact:

  'The project may also be used by Trinity as an impetus to further research and on-going improvements to the examinations and examiner standardisation'

Further research

- What is the impact (if any) of CEFR linking on:

1. *Test quality*
2. *Test use*
3. *Teaching and learning practices*
4. *Curricula*
5. *Textbooks*

- How can positive impact be distinguished from negative impact?

Contact details

Spiros Papageorgiou
Dept of Linguistics & English Language
Lancaster University
LA1 4YT
UK
s.papageorgiou@lancaster.ac.uk