



Reflections on using the Draft Manual: The Cambridge ESOL FCE case study

H. Khalifa, A. Salamoura
& A. French



Why First Certificate of English Examination?

- Candidature
- Stability
- Mid point in ESOL's proficiency scale
- Updated (for launch in Dec 08)
- Coverage of testing focuses



Familiarisation & Standardisation Procedures – our experience

- **Format:** face-to-face workshop
- **Duration:** a day and a half
- **Participants:** internal and external



Familiarisation & Standardisation Procedures – our experience

▪ **Materials:**

- those recommended by the Manual as perceived relevant to Cambridge ESOL's context
- Data collection instruments:
 - Participants' feedback questionnaire
 - Audio-recordings of discussions
 - Participants' responses to activities

▪ **Data Analysis:** qualitative & quantitative



Workshop aimed at achieving:

- a common understanding
- a broad awareness
- a shared knowledge
- a participant-led action plan
- data collection for research purposes



Pre-workshop tasks

- Self-assessment of FL ability:
 - participants use Table 1-global scale (based on activity C)
- Descriptor-sorting activity:
 - participants use own experience within the teaching and testing field but not CEFR material
- Relevant readings
 - e.g. Taylor & Jones (2006) Cambridge ESOL exams and the Common European Framework of Reference (CEFR), *Research Notes* 24, 1-5



Face-to-face Workshop Activities: Familiarisation

- Presentations about:
 - the origins, aims and nature of the CEFR
 - its relevance for language assessment
 - its relevance & implications for participants as professional language testers working with Cambridge ESOL
- Descriptor-sorting: using a variety of CEFR scales across 4 skills (activities D & E)



Face-to-face Workshop Activities: Standardisation

- Focus on training stage of standardisation procedures
 - Rating speaking and writing performances
 - Rating reading and listening tasks
- Scope: focus on CEFR B2 and its adjacent levels



Post-workshop tasks

Based on gained KSA, participants were asked to:

- revisit pre-workshop descriptor-sorting tasks to enable comparisons & evaluation of face-to-face familiarisation activities,
- develop action plans for cascading gained KSA,
- complete feedback forms on the workshop experience



Reflections on Familiarisation & Standardisation based on:

**workshop experience, participants'
feedback, workshop leader & project
coordinator's observations**



Usefulness of introductory activity C

Question Asked:

How effective is the Self-assessment activity in

- familiarising participants with the CEFR Global scale
- highlighting key words that signify criterial performance differences between adjacent CEFR levels?



Participants' responses

- On a 4 point scale, 2/9 found it **very effective** & 7/9 found it **effective**.

“Reflecting on my own language performance made the CEFR ‘real’, rather than just a series of statements. It also highlighted for me the issue of the ‘spiky profile’ which many students have...”

“This activity clarified for me that the descriptors are typical standards of performance from which individuals will necessarily depart in certain aspects, as they have different strengths and weaknesses.”



Usefulness of qualitative analysis activities D & E

Question Asked:

How effective are the Descriptor-sorting activities in familiarising participants with skill specific CEFR descriptors across levels?



Participants' responses

- On a 4 point scale, 4/9 found it **very effective** & 5/9 found it **effective**.

“It got me thinking about which abilities we associate with which levels and to what extent it’s possible to differentiate easily between adjacent levels. Honing in on individual words like ‘complexity’ was thought provoking.”

“We talked about level overlap as well and how this affected our decision making i.e. ‘the most able B1 / least able B2 candidate’. How can we differentiate!”

“I found some of the language used in the descriptors made the level obvious e.g. ‘familiar contexts = B1’, ‘most general topics = B2’, ‘flexibly and effectively = C1’... How helpful is this? Do we use it to reinforce what we already think, or use it as a basis to allocate a level? If for the latter, then I don’t think it’s helpful because it’s neither exclusive nor definitive.”



Usefulness of qualitative analysis activities D & E

Question Asked:

What were some of the difficulties encountered when classifying the CEFR descriptors into levels?



Participants' responses

Difficulties arose from

- ambiguity in the CEFR descriptor
 - *e.g. “can narrate a story” (what kind of story, simple, complex, of medium complexity?), the meaning of ‘style’*
- trying to relate CEFR's classification with real-life experience
 - *e.g. “only extreme noise” affecting our ability to understand is classified as B2 whereas ‘extreme noise’ affects everyone’s ability to understand at whatever level*



Usefulness of time management table

- (30 mins) Brief presentation of CEF
 - not adequate unless participants already fairly familiar with CEFR
 - we felt at least 45' needed, positive feedback from participants
- (45 mins) Introductory activity and discussion
 - Adequate
- (105 mins) Qualitative activity, discussion and conclusions
 - Adequate for the suggested 40 descriptors

Rating Performances & Tasks

Question Asked:

How effective are the suggested activities in training participants to rate standardised performances & calibrated tasks across the CEFR B1-C1 levels?

Participants' responses:

Made no difference	Made some difference	Effective	Very Effective
1/9	1/9	3/9	3/9



Few Quotes

“I found this activity very interesting having spent the morning immersed in the descriptors... In most cases we agreed with the bench marks, but were wary of giving B2 to the 4 samples we found had been awarded C1... Looking again at the CEFR descriptors we changed our minds about 2 of them. Overall, though, we felt reassured that FCE / B2 / CEFR descriptors did not conflict with one another, but actually bore each other out.”

“I would have preferred less evaluation and more discussion of one’s evaluation with one’s fellow assessor.”

“It made us think carefully about the scales and the tasks we were given to rate but it felt as though the different [sample providers] had different interpretations of the ratings so we did get some of these ratings wrong.”



Reflections on non-Manual mandated activities



1. Background reading activity

Question Asked:

How useful have you found reading “independent” but CEFR-relevant material as an introduction to the CEFR and the workshop activities?



Participants' responses

- On a 4 point scale, 3/9 found it **very useful** & 6/9 found it **useful**.

“It alerted me to the main issues underpinning the CEFR levels and made me realise their potential impact on FCE. ”

“Some of this was information I had come across already. It did, however, revisit the wider picture but as the workshop was fairly closely based on looking at the levels, the article did not impact directly on it.”

“Students are becoming more CEFR aware and ask which level of exam they should enter for and why. It's useful to have the background knowledge of Cambridge ESOL's history, links and attitude to the CEFR, as well as having another point of reference when saying 'Well, I should advise you to take PET / FCE / CAE because.....' ”



2. Juxtaposing adjacent CEFR levels

Question asked:

Does focus on one CEFR level and its adjacent levels during Familiarisation and Standardisation help participants better understand the key features of the target level?



Participants Responses

On a 4 point scale, 8/9 found it **very useful** and 1/9 found it **useful**.

“Among the most useful things we did: we had some issues with deciding whether a competence at a given level didn’t better describe an adjacent level; there were definitely blurred edges in adjacent levels ...”

“It would have been too abstract to deal with one level by itself. It was helpful to compare both with the adjacent CEFR levels and our experience of PET and CAE.”

“... it’s important to consider the levels immediately above and below you when you are trying to write tasks at a particular level so that you can have a firm grasp on what it is that makes your level different from the other two.”



What do we take from this experience?

- Contextualisation / local validity
- General versus specific dilemma
- Data: why, which, what, & how
- Practicalities: resources, time allocation/constraints, participant selection- who to invite from our network



What do we take from this experience?

- Suggested procedures are good starting point but:
 - complement with activities that suit own context, e.g., readings, active engagement through cascading plans
- Focusing on one target level was useful and time efficient.
- Looking at the adjacent levels helped understanding key features of the selected target level.



What do we take from this experience?

- Perceive the exercise not as one off but as iterative.
 - Clarify/simplify ambiguous terminology in descriptors, e.g., can narrate a story: simple-complex continuum
 - PLAN, PLAN, PLAN



Thanks are due to:

Workshop Leader and participants
for constructive feedback