Benchmarking EAP Written Assessment with the CEFR

A/Prof Thomas Roche

UECA Assessment Symposium 2019, The University of Queensland, 13-14th July, 2019
Overview

1. Why focus on EAP Written Assessment?
2. What are we assessing? EAP – the construct.
3. Assessment in Australian Higher Ed & ELICOS - the context.
4. Benchmarking – the legislation
5. UECA Benchmarking – the project
   5.2 Activity: the CEFR
5.3 CEFR and Peer Review Portal
   6. Activity: Assessing Writing with the CEFR.
7. Conclusion
1. Why EAP Written Assessment?

1. The Students:
   1.1 over 700,000 international enrolments in education in Australia in 2018
   1.2 approx. 30% of Higher Ed enrolments enter via an ELICOS pathway (often EAP) – responsibility to get it right

2. Writing is a good proxy for overall academic English proficiency:
   2.1 majority of university assessments are written
   2.2 good (best?) predictor of overall language proficiency and in-course performance (Ushioda & Harsch 2011; Roche & Harrington, 2015).
1. Why EAP Written Assessment?

Academic achievement in HE is influenced by a number of factors (Andrade, 2006; Bayliss & Ingram, 2006):

- **Personal factors:** motivation & self-efficacy (Phakiti, Hirsch and Woodrow, 2013)
- **Social factors:** supportive peer network (Evans and Morrison, 2011)

as well as by **factors Direct Entry/EAP programs can directly influence:**

- **English Language Proficiency (ELP)** (Elder, Bright, and Bennett, 2007; Humphreys et al 2012)
- **Cultural factors:** transition from local high-school to global higher-education requires adjustment (Andrade, 2006)
  - understanding of the HE provider’s teaching and assessment practices (Lee and Greene, 2007)
  - digital literacy: ability to access, assess & disseminate information in online mode (Roche, 2017)
  - understanding of institutional policy – notions of text ownership/authorship (Pennycook, 1996; Morrison, 2017)

EAP curriculum (ie. learning outcomes/teaching/learning and assessment) should address these factors in order to produce graduates with the attributes/competencies/skills that will enable them to succeed in HE.
2. What are we assessing – EAP the construct?

A broader view of what language is to be taught and assessed. Moving both from skills (speaking, reading, writing and listening) to include explicit focus on higher-order language skills: academic literacies (paraphrase, synthesis, referencing etc.)

Develop students’ understanding of
Text ownership
Academic identity features

literacy practices
academic integrity: paraphrasing/in-text citation/referencing
employing appropriate register and pragmatic use (e.g. netiquette, report features)

• Academic language requires argumentation, which often involves combining different sources (Cho & Bridgeman, 2012; Cumming, 2013; Hyland, 2006; Pennycook, 1996).
• Story of Ram, the MIT student.
Assessment is the tool by which students are judged to have met unit learning outcomes (ULOs).

- *International White Paper on the future of assessment* (Coates, 2018) flag changes in how HE assessment is produced, implemented, scored, report but there are also changes to legislation on how it is regulated...

### 3. Assessment in Australian Higher Ed & ELICOS

<table>
<thead>
<tr>
<th></th>
<th>Traditional 1990s and before</th>
<th>Stretched 1990s to 2020</th>
<th>Next Generation 2020s and after</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authority</strong></td>
<td>University</td>
<td>University or regulator or testing org (80s)</td>
<td>? Shared</td>
</tr>
<tr>
<td><strong>Production</strong></td>
<td>Solo academics</td>
<td>Academic teams</td>
<td>Co-creation</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>Paper</td>
<td>Paper and online</td>
<td>Online</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Campus</td>
<td>Campus and online</td>
<td>Online</td>
</tr>
<tr>
<td><strong>Implementation</strong></td>
<td>Universities</td>
<td>Unis/Specialists</td>
<td>Specialists</td>
</tr>
<tr>
<td><strong>Scoring</strong></td>
<td>Solo academics</td>
<td>Moderated practice</td>
<td>? Automated</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td>Generic</td>
<td>Contextualised</td>
<td>Customised</td>
</tr>
</tbody>
</table>
4. Benchmarking – the legislation

<table>
<thead>
<tr>
<th>Documentation</th>
<th>Key Points</th>
</tr>
</thead>
</table>
| **ELICOS Standards 2018 (Standard P4.1 c (ii)):**  
Department of Education and Training  
National standards for ELICOS providers and courses: | • Contain a general requirement to ensure that assessment is *valid, reliable, fair and clearly referenced* to criteria.  
• formal measures must be in place to ensure that assessment outcomes are comparable to other criteria used for admission to the tertiary education course of study |
| **Higher Education Standards Framework**  
[5.3.1, 5.3.4; 5.3.7; 1.4.1; 1.4.3; 1.4.4] | • admissions policies and procedures are documented and are designed to ensure students have the English proficiency they need to succeed  
• Focus on testing rather than assessment |
| **TEQSA Guidance Note: External Referencing** | • The ELICOS provider must be able to show that a valid and reliable mechanism (or combination of mechanisms) is in place to independently demonstrate this comparability  
• **Initial focus on testing.**  
• UECA Executive provided feedback to TEQSA on the draft TEQSA Guidance Note on English Language Standards. *Jan 2018 broadened external referencing to benchmarking (CEFR, ISLPR), tracer studies, external testing (IELTS, CAE, PTE).* Woot. |
5. UECA Benchmarking: the project

- *External Referencing of the ELICOS Standards and International Education [ERESIE]* Project was established to facilitate mutual learning among member Centres and enable institutions to validate their policies, processes and assessment standards.

- **UECA 2018-2019 Benchmarking Project** focus on *Standard P4: Assessment of ELICOS students.*

- External reference point identified is Common European Framework of Reference (CEFR):
  - three basic categories at six levels:
    - *Basic user* (A1 & A2), *Independent user* (B1 & B2) and *Proficient user* (C1 & C2).
    - It now defines ‘plus’ levels (e.g. B1+, B2+) to differentiate within these (North, 2014).

- ‘the practice of colleagues providing and receiving feedback on one another’s unit/subject outlines, assessment tasks and marking criteria to ensure that assessment is aligned to intended learning outcomes and includes a calibration process to ensure comparability of achievement standards and an opportunity for professional learning’. (Booth et al., 2015)
5.1 CEFR as an external frame of reference

- Originally developed as a frame of reference for communicating about threshold standards of language across Europe

- Takes real-world language use as a starting point
- Council of Europe has released a *CEFR Companion Volume with New Descriptors (2018)* containing written work descriptors, p.173-235; *NEW* written reports and essays p.77
- **CEFR levels and IELTS correspondence:** not 1:1

<table>
<thead>
<tr>
<th>CEFR</th>
<th>IELTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1+</td>
<td>7.5</td>
</tr>
<tr>
<td>C1</td>
<td>7.0</td>
</tr>
<tr>
<td>B2+</td>
<td>6.5</td>
</tr>
<tr>
<td>B2</td>
<td>6.0</td>
</tr>
<tr>
<td>B1+</td>
<td>5.5</td>
</tr>
<tr>
<td>B1</td>
<td>5.0</td>
</tr>
<tr>
<td>A2</td>
<td>4.5</td>
</tr>
<tr>
<td>A1</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**Diagram:**

- C2 (Produce user) 9
- C1 (Independent user) 8
- B2 (Basic user) 6
- B1 (Intermediate user) 5
- A2 4.5
- A1 4.0
5.2 Activity: The CEFR (5min)

• **Activity One:** Look at CEFR WRITING Descriptors from the CEFR Companion Volume (2min)

• **Q1.1 Discuss with a partner** – can you see any positives to using this tool for assessing writing samples from different university EAP programs?

• **Q1.2 Discuss with a partner** – can you see any issues with the wording of the scale?
5.3 The Project: CEFR issues and advantages

- CEFR is not without critics:

- Initial versions of the CEFR were formulated to be applicable across a range of contexts and situations (North, 2000) – context/genre independent (i.e. underspecified) as such it underrepresents the complexities of academic writing (McNamara, Morton, Storch & Thompson, 2018).


- development and validation of the level descriptors Deygers (2019) problematics

- But the new CV includes descriptors for the ‘C’ levels (Goodier, 2018) and there is consensus amongst some that it can be used as a base to develop a shared understanding of threshold standards (Deygers, Van Gorp, & Demeester, 2018; (Harsch & Martin, 2012) - “socialization into communities of practice” (Kramsch, 2002), Mediation /appropriateness
5.4 The Project: Participants and timeline

20 Participating Institutions

- ACU
- CQU
- Curtin
- Flinders
- Hawthorn (Uni of Melbourne)
- James Cook Uni
- Monash
- QUT
- RMIT Training
- SCU
- Swinburne
- Uni of Adelaide
- UOW College
- UNE
- Newcastle
- USydney
- UTas
- UWA
- VU
- Western Sydney Uni

5.3 The Project: CEFR Issues and advantages

- Phase 1
  - Project Administration: Sign collaboration agreement; Fill in institution context statement & pay invoice; Sign up to the Portal: August-September 2018

- Phase 2

- Phase 3
  - National Exchange Workshop
    - Validation of results prior to workshop
    - Assessment bank and review of assessor’s comments
    - Exchanging and presenting on good practice
    - Calibration of standards
    - Institutional/national recommendations
    - May 2019

- Phase 4
  - Self Review Phase
    - Zoom meetings on Peer Review Portal and assessment training
    - Institutional and disciplinary teams
    - Self review reports
    - Schedule of external review of assessment
    - October, 2018-March, 2019

- Phase 5
  - Draft Interim Report: early Jan 2018
  - Final Report: June 2019
5.5. The Project: *How* and *what* to benchmark?

**Develop shared interpretation** (Nulty, 2017, Sadler, 2013, Bloxham, 2009): participating Centres were sent the new CEFR Companion Volume as pre-reading.

**Identify assessment samples**: For each written in-course, formative assessment item worth 20% or more: three samples each of a Pass and a (just) Fail with the marking rubric; and or each written Exit or Capstone/summative assessment three samples each of up to three Grades/levels with the marking rubrics

**Co-develop the scales and interpretation** (Harsch & Martin, 2012): UECA Committee discussed and agreed on a version of the CEFR scales for use. Three Guiding Documents were produced over 6 months.

**Assess together**. Assessing (judgment) involves both subjective and objective knowledge (Kahneman, Slovic, & Tversky, 1982) therefore even experienced raters will differ in their assessment at times (Harsch & Hartig, 2015).

Employed consensus moderation (Nulty, 2017; Sadler, 2013):
- trial mark the sample of student work against the rubric;
- compare with each other provisionally allocated marks;
- engage in focused discussion about how marks should be allocated;
- reach agreement on an appropriate overall academic achievement standard

**Calibrate**
**7. Conclusion**

- The UECA project took a validated, user-oriented Proficiency Scale containing generalised, abstract descriptions of what learners are likely to be able to do at levels. The CEFR was our external frame of reference.

- Centres broadly welcomed outcomes of the benchmarking: suggestions for improvement and involvement in the project which developed a shared interpretation of CEFR levels.

- **Issues:** Though updated to capture the features of reports and essays, there are still gaps in the scales (e.g. B2+ Grammar and Vocab range) and occasionally ill-matched descriptors for academic language (e.g. C1+ appropriateness “including emotional, allusive and joking usage”).

- **Response:** Develop relevant subscales for common UECA assessment items. Use sector validated samples to develop annotated samples at levels for further rater-training.
Thank you
see you next year at SCU