Extending CEFR descriptors for an analytic rating scale of academic writing at C1 to C2 – a stop-gap measure?

“THE COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES – HOW DO WE DEAL WITH ITS GAPS?”

6th Bremen Symposium on language learning and teaching at universities

Dr Helen Heaney
Dr Armin Berger
Department of English
University of Vienna
Preview

- Brief overview of the rating scales project
- Our stop-gap measures for grammar
- The validity gap
- Validation of the speaking scales
- Future work
The Austrian context

- 2000: Standards for foreign languages laid down in curricula based on CEFR
- 2008: School Education Act amended; pilot projects for school-leaving exam in English & French
- 2009: Educational Standards for 14 year olds
- 2014/15: Centralized, standardized school-leaving exams nationwide

BMB (2004), Innsbruck University (2011), LTC (2009a)
Overview of the rating scale project (2006-2010)

Extending CEFR descriptors for an analytic rating scale of academic writing at C1 to C2 – a stop-gap measure?

Bremen 24.02.2017
Initial decisions

- Five levels for each dimension
- Descriptors to be formulated for levels 1, 3 & 4
- Anchor descriptors to be extracted from the CEFR:
  - Top grade (=1): C2
  - Bare pass (=4): C1
- Extended descriptors for C2/C1 to be produced in initial rating procedure
Brainstorming for grammar

- Agreement, tense, verb complementation
- Pronouns, articles, prepositions
- Word order
- Punctuation
- Range of structures
- Appropriacy of construction – register
- (Over)complexity
- Ease of interpretation; accessibility/strain on reader
First round

• Four prompts
• Three performances selected (excellent, pass, fail)
• One dimension rated by each university, with rationale (2-3 sentences) for each rating
Grammar: C2

- Maintains consistent grammatical control of complex language, even while attention is otherwise engaged [...]. [C2: GRAMMATICAL ACCURACY:114] Shows confident handling of constructions appropriate to the given genre; well formed sentences connected into meaningful subtexts showing awareness of focusing devices such as cleft-sentence, passive constructions, etc.; observing principle of end-weight, theme-rheme structuring; hardly any errors of agreement, tense, number, word order, articles, prepositions, use of modal verbs, etc.; meaning never obscured.

LTC (2009b), Heaney (2013)
Grammar: C1

- Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.  

Uses a limited number of complex constructions correctly (such as sentence types deviating from SVO pattern, modals, adverbials, etc.) Accurate but unambitious use of grammatical structures. Grammatical inaccuracy in the text does not impede overall understanding.
Second round

• Are the scales & descriptors applicable to the various prompts?
• What modifications, if any, do we wish to make to the scales?
• How can we formulate the descriptors between C2 and C1?

Third round

• Benchmarking
Final version: Grammar C2

Maintains consistent grammatical control of complex language, even when dealing with cognitively demanding content [while attention is otherwise engaged]. [C2: GRAMMATICAL ACCURACY:114]

Shows confident handling of a wide range of constructions and punctuation appropriate to the given genre; well formed sentences [...] showing awareness of focusing devices, passive constructions, etc.; observing principle of end-weight; no errors of agreement, tense (choice & form), number, word order (including marked word order like cleft sentences, inversion, extraposition, existential there), articles, prepositions, expressions of modality, negation, relative clauses, adjectives/adverbs, conditional structures, conjunctions, comparative constructions, pronoun reference within sentences. No dangling participles. Meaning never obscured.
Final version: Grammar C1+

Consistently maintains a high degree of grammatical accuracy; errors are rare and insignificant [difficult to spot].

Hardly any errors of agreement, tense, number, word order (including marked word order like cleft sentences, inversion, extraposition, existential there), articles, expressions of modality, negation, relative clauses, adjectives/adverbs, conditional structures, conjunctions, comparative constructions, pronoun reference within sentences.
Final version: Grammar C1

Consistently maintains a high degree of grammatical accuracy although occasional errors may occur. [C1: GRAMMATICAL ACCURACY:114]

Uses a limited number of complex constructions correctly (such as sentence types deviating from SVO pattern, modals, adverbials, etc.) Accurate but unambitious use of grammatical structures. Grammatical inaccuracy in the text does not impede overall understanding. Occasional errors of agreement, tense, number, word order, articles, expressions of modality, negation, etc.
Dealing with the gaps in grammar

• Intuitive list of complex grammar

Shows confident handling of a wide range of constructions and punctuation appropriate to the given genre; well formed sentences [...] showing awareness of focusing devices, passive constructions, etc.; observing principle of end-weight; no errors of agreement, tense (choice & form), number, word order (including marked word order like cleft sentences, inversion, extraposition, existential there), articles, prepositions, expressions of modality, negation, relative clauses, adjectives/adverbs, conditional structures, conjunctions, comparative constructions, pronoun reference within sentences. No dangling participles. Meaning never obscured.
More recent research

- Ongoing consultation process on an extended version of the CEFR’s illustrative descriptors (CE 2016-2017)
- Core Inventory for General English (North, Ortega & Sheehan 2011)
- English Profile
Core Inventory for General English – C1

• Practical inventory of language points relevant for teaching at different levels
  ◦ Analysis of language implied by CEFR descriptors
  ◦ Analysis of syllabuses
  ◦ Analysis of popular course books
  ◦ Teacher surveys

North, Ortega & Sheehan (2011)
Core Inventory for General English – C1

- Grammar covered at C1
  - Futures (revision)
  - Inversion with negative adverbials
  - Mixed conditionals in past, present and future
  - Modals in the past
  - Narrative tenses for experience, incl. passive
  - Passive forms, all
  - Phrasal verbs, especially splitting
  - Wish/if only regrets
English Profile

• Reference level descriptors for national and regional languages
• Cambridge learner corpus
• Criterial features “characteristic and indicative” of L2 proficiency at a particular level
• Transitional features
  ◦ Positive linguistic features
  ◦ Negative linguistic features

Hawkins & Filipovic (2012)
English Profile: C1

- Higher mean length of utterance
- Various raising constructions with specific verbs, i.e. *believe, find, assumed, discovered*
- Double genitive
- Modal *might* (deontic)
- Reduction in errors of noun agreement, derivation of determiner, missing quantifier, missing preposition
English Profile: C2

- Higher mean length of utterance
- Various raising constructions with specific verbs, i.e. *declare, remember, presumed*
- Reduction in errors of noun agreement, noun countability, replace quantifier, replace adverb
The validity gap

- Progression remains intuitive
  - “… descriptions of expected outcomes, or impressionistic etchings of what proficiency might look like as one moves through hypothetical points or levels on a developmental continuum” [our emphasis] (Clark 1985)

- The ‘systematic approach’ (North 2014) lacks definiteness, e.g.
  - Level 1: Consistently produces flawless paragraphs.
  - Level 3: Most paragraphs are well developed.
  - Level 4: Can produce well-developed paragraphs.
# Validation of the ELTT speaking scales

<table>
<thead>
<tr>
<th></th>
<th>Descriptor Sorting</th>
<th>Descriptor Calibration</th>
<th>Descriptor-Performance matching</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>21 university teachers</td>
<td>16 students (trial group)</td>
<td>78 students</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 university teachers</td>
<td></td>
</tr>
<tr>
<td><strong>Instruments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td>sorting task</td>
<td>2 task prompts</td>
<td>153 video performances</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rating sheets</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rater guidelines</td>
<td></td>
</tr>
<tr>
<td><strong>Procedures</strong></td>
<td>descriptor sorting</td>
<td>descriptor scaling</td>
<td>task development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mock exams</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>performance matching</td>
</tr>
<tr>
<td><strong>Analyses</strong></td>
<td>correlations</td>
<td>multi-faceted Rasch</td>
<td>multi-faceted Rasch</td>
</tr>
</tbody>
</table>
Descriptor sorting and calibration

- 174 descriptor units and sorting task
- 21 university language teachers
- Correlations
- Multi-faceted Rasch analysis

<table>
<thead>
<tr>
<th>Sorting task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>highly proficient</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

I. Oral presentation and spoken interaction

A. Lexico-grammatical resources and fluency

| 1. Shows great flexibility formulating ideas | 1 | 2 | 3 |
| 2. Can formulate ideas in differing linguistic forms | 1 | 2 | 3 |
| 3. Can convey finer shades of meaning precisely | 1 | 2 | 3 |
| 4. Can use appropriate register | 1 | 2 | 3 |
Descriptor-performance matching

- 174 descriptor units
- 153 videotaped student performances
- 8 university language teachers
- 21,909 valid data points
- Multi-faceted Rasch analysis

Rating sheet

<table>
<thead>
<tr>
<th>clearly above the student's level cannot perform like this</th>
<th>slightly above the student's level may perform like this in favourable circumstances</th>
<th>reflects the student's level adequately can perform like this</th>
<th>slightly below the student's level may perform like this even in more difficult circumstances</th>
<th>clearly below the student's level can perform better than this</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

A. Lexico-grammatical resources and fluency

1. Shows great flexibility formulating ideas
2. Can formulate ideas in differing linguistic forms
3. Can convey finer shades of meaning precisely
4. Can use appropriate register
5. 

1 | 2 | 3
Descriptor evaluation

- Quality criteria
  - Soundness of the calibration
  - Model fit
  - Consistency across methods
  - Congruence with original intention

<table>
<thead>
<tr>
<th></th>
<th>excellent</th>
<th></th>
<th>good</th>
<th></th>
<th></th>
<th>adequate</th>
<th></th>
<th>poor</th>
<th></th>
<th>problematic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>LGF</td>
<td>7</td>
<td>23.33</td>
<td>8</td>
<td>29.67</td>
<td>6</td>
<td>20.00</td>
<td>6</td>
<td>20.00</td>
<td>3</td>
<td>10.00</td>
<td></td>
</tr>
<tr>
<td>PVI</td>
<td>6</td>
<td>26.09</td>
<td>4</td>
<td>17.39</td>
<td>6</td>
<td>26.09</td>
<td>5</td>
<td>21.74</td>
<td>2</td>
<td>8.70</td>
<td></td>
</tr>
<tr>
<td>PSCW</td>
<td>7</td>
<td>33.33</td>
<td>3</td>
<td>14.29</td>
<td>4</td>
<td>19.05</td>
<td>6</td>
<td>28.57</td>
<td>1</td>
<td>4.76</td>
<td></td>
</tr>
<tr>
<td>PGSP</td>
<td>9</td>
<td>52.94</td>
<td>2</td>
<td>11.76</td>
<td>2</td>
<td>11.76</td>
<td>4</td>
<td>23.52</td>
<td>0</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ICRW</td>
<td>10</td>
<td>28.57</td>
<td>1</td>
<td>2.86</td>
<td>10</td>
<td>28.57</td>
<td>14</td>
<td>40.00</td>
<td>0</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>IINH</td>
<td>12</td>
<td>25.00</td>
<td>17</td>
<td>35.42</td>
<td>10</td>
<td>20.83</td>
<td>9</td>
<td>18.75</td>
<td>0</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>29.31</td>
<td>35</td>
<td>20.11</td>
<td>38</td>
<td>21.84</td>
<td>44</td>
<td>25.29</td>
<td>6</td>
<td>3.45</td>
<td></td>
</tr>
</tbody>
</table>
Lessons learnt

• The project has shown the benefits of the symbiosis between intuitive, qualitative and quantitative scale development methods (cf. Galaczi et al. 2011).
• Evidence-based validation is fundamental.
• Validation methods are complementary rather than interchangeable.
• The utility of scale validation methods should be evaluated in terms of their combined effect.
An expanded model for rating scale development

(based on Galaczi et al. 2011)
Future work

• Revision and/or fine-tuning of descriptors for writing
  – Descriptive scheme of the CEFR
  – Further sources of CEFR-related description for academic writing
  – Systematic performance analysis
  – Corpus analysis (e.g. CAARLLA)

• Validation of the writing scale using both qualitative and quantitative methods

• Investigating the use of the scale in different contexts

• Rater training
Extending CEFR descriptors for an analytic rating scale of academic writing at C1 to C2 – a stop-gap measure?

Yes!
Bibliography


Extending CEFR descriptors for an analytic rating scale of academic writing at C1 to C2 – a stop-gap measure? Bremen 24.02.2017
Bibliography


