Language Testing in Europe: Time for a New Framework?

Book of Abstracts

University of Antwerp
27 – 29 May 2013

Composed by Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer
Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (editors)


Cover: Nieuwe Media Dienst, University of Antwerp
This volume contains all abstracts from our Second International Conference on Language Testing. The first conference, which was organised in 1997, brought together more than 100 participants from 12 countries around the theme ‘Language Testing and HRM’.

This Second International Conference unites more than 150 practitioners, policymakers and researchers from 26 countries. The theme, ‘Language Testing in Europe: Time for a new Framework?’ arose from an urgent need to respond to concrete issues associated with the use of the Common European Framework of Reference for Languages (CEFR) in language testing.

- **Competence and performance**
  What is the link between ‘can do’ performance statements and areas of linguistic knowledge? To what extent can or should the levels be made more explicit in terms of required vocabulary and grammar?

- **Degree of difficulty of the levels**
  How can we make sure that our examinations are measuring at the CEFR levels we claim they are? What evidence do we have to support our claims?

- **Test purpose**
  Why are we testing? What kind of decisions will be made on the basis of information collected via the test? What will be the consequences of these decisions?

- **Practicality**
  How do we link our tests to the CEFR? How practical, applicable and operational is the CEFR for concrete language testing situations?

Lyle Bachman (Professor Emeritus at the University of California), Etienne Devaux (Screening methodologist at SELOR, the Belgian public personnel selection and certification agency), Jan Hulstijn (Professor at the University of Amsterdam) and Waldemar Martyniuk (Executive Director of the European Centre for Modern Languages of the Council of Europe), will give keynote presentations in order to enhance the discussions from their highly specific areas of expertise.

Glyn Jones has been selected to receive the award for the Selected Plenary Presentation.

In all, 53 paper and poster proposals were accepted. All contributions demonstrate a remarkable variety in background, approach and style. They will provide the foundation for three days of intense debate on language testing and the CEFR.

One of the objectives of this conference is to formulate a clear memorandum of understanding. This document will reflect the opinion of each participant. To this end, we conducted a pre-conference survey (the results of which will be presented on the first day). We have also arranged various types of interaction with the audience, including a voting moment, an online survey, discussion groups, a panel and, most importantly, many coffee and food breaks.

We hope that you will enjoy this volume and that you will keep it as a memorable souvenir of a remarkable milestone in the history of language testing in Europe.

Prof Jozef Colpaert & Prof Mathea Simons
CONFERENCE ORGANISATION

Chair
Jozef Colpaert
jozef.colpaert@ua.ac.be
Institute of Education and Information Sciences – Linguapolis – University of Antwerp

Conference manager
Ann Aerts
ann.aerts@ua.ac.be
Linguapolis – University of Antwerp

Local organising committee
Jozef Colpaert
Mathea Simons
Ann Aerts
Margret Oberhofer

Scientific committee
Jozef Colpaert
Mathea Simons
Margret Oberhofer
Charles Alderson
Lyle Bachman
Jan Hulstijn
CONTENTS

FOREWORD ......................................................................................................................... 5

CONFERENCE ORGANISATION .......................................................................................... 6

CONTENTS ............................................................................................................................. 7

PRACTICAL INFORMATION ................................................................................................. 11

CAMPUS MAP .................................................................................................................... 12

SOCIAL EVENTS ................................................................................................................ 13

PROGRAMME ...................................................................................................................... 14

KEYNOTE SPEAKERS ......................................................................................................... 19

Lyle F. Bachman .................................................................................................................. 21
How do Different Intended Uses and Different Views of Language Impact Language Assessment Practice?

Etienne Devaux ..................................................................................................................... 23
Confronting the CEFR to L2 Certification Purposes: Added Value and Methodological Limitations

Jan H. Hulstijn ...................................................................................................................... 24
Natural Tensions between Theory and Practice in the Common European Framework of Reference for Languages (CEFR)

Waldemar Martyniuk ........................................................................................................... 25
The Council of Europe’s Common European Framework of Reference for Languages (CEFR): a 2013 Summary of Developments

SELECTED PLENARY ....................................................................................................... 27

Glyn Jones .......................................................................................................................... 29
Developing a CEFR-Aligned Test from Scratch: a Case Study

PAPER PRESENTATIONS .................................................................................................. 31

Ene Alas & Suliko Liiv ......................................................................................................... 33
Training Interviewers and Raters for the National Examination in the English Language – the Estonian Experience

Marian Amengu Pizarro & Jesús García Laborda ................................................................. 34
Investigating Teachers’ Opinion on the Feasibility and Intended Washback Effect of a High-stakes Oral English Test in Spain

Bernadette Brouwers .......................................................................................................... 36
The Assessment of Language Competence – Moving Forward

Jasminka Buljan Culej ......................................................................................................... 38
ESLC in Croatia: Relation between Years of Learning the Foreign Language and the CEFR Level Achieved
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>About the Impossibility of Assessing Speaking with Focus both on Form and Communicative Output</td>
<td>40</td>
</tr>
<tr>
<td>Defining, Modeling and Piloting SELF, a New Formative Assessment Test for Foreign Languages</td>
<td>42</td>
</tr>
<tr>
<td>What and How Many Words Do We Need? Critical Considerations when Developing a CEFR Vocabulary List: Size, Depth, and Growth</td>
<td>44</td>
</tr>
<tr>
<td>Can the CEFR Assess University Students in a China-Based British University? A case Study at the University of Nottingham Ningbo China</td>
<td>48</td>
</tr>
<tr>
<td>Interchangeability of Test Results and the CEFR – a Validity Argument Approach</td>
<td>50</td>
</tr>
<tr>
<td>How to Save the Common European Framework?</td>
<td>51</td>
</tr>
<tr>
<td>Can a Language Test Verify the Academic Literacy of University Students and how Does that Relate to Study Success?</td>
<td>52</td>
</tr>
<tr>
<td>CEFR Can-Do Statements as a Means of Self-Assessment: is There a Common Understanding, Regardless of the Student’s Gender and Educational System?</td>
<td>55</td>
</tr>
<tr>
<td>The Influence of the CEFR on Rating Scale Design</td>
<td>57</td>
</tr>
<tr>
<td>Combatting the “Can’t do Mentality”: Expert, Peer &amp; Self-Assessment in a French University Context - The “ELLO” Project (étude longitudinale sur la langue orale)</td>
<td>60</td>
</tr>
<tr>
<td>Designing an Online Italian L2/LS Placement Test in Line with the CEFR Standards. Suggestions for Monitoring Reliability and Ensuring its Validity in the Perspective of an International Application</td>
<td>62</td>
</tr>
<tr>
<td>English Tests for Secondary School Leavers in Slovakia</td>
<td>64</td>
</tr>
<tr>
<td>Can Spanish High School Students Speak English?</td>
<td>65</td>
</tr>
<tr>
<td>Proposals of Ubiquitous Delivery of the Foreign Language Paper of the Spanish Baccalaureate General Test</td>
<td>66</td>
</tr>
<tr>
<td>Testing Language Competences for an Intended Practical Use</td>
<td>68</td>
</tr>
<tr>
<td>Investigating the Construct Underlying the CEFR Phonological Control Scale</td>
<td>71</td>
</tr>
<tr>
<td>Work in Progress: How Useful is the CEFR in Designing the Follow-Up Assessment of Learning Outcomes in Foreign Languages in the Finnish Basic Education?</td>
<td>73</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Language Testing in Europe: Time for a New Framework?</td>
<td>-9-</td>
</tr>
<tr>
<td>Ari Huhta, Riikka Ullakonoja, Lea Niimeni &amp; Eeva-Leena Haapakangas</td>
<td>75</td>
</tr>
<tr>
<td>The Use of the CEFR in Diagnosis</td>
<td></td>
</tr>
<tr>
<td>Ben Knight</td>
<td>76</td>
</tr>
<tr>
<td>The English Profile Project: Researching what the CEFR Means in Terms of Specific English Linguistic Knowledge</td>
<td></td>
</tr>
<tr>
<td>Benjamin Kremmel &amp; Franz Holzknecht</td>
<td>77</td>
</tr>
<tr>
<td>Strengths and Weaknesses of the CEFR in Guiding Test Task Design: What the Can Do’s Can Do and What They Can’t (yet)</td>
<td></td>
</tr>
<tr>
<td>Folkert Kuiken &amp; Ineke Vedder</td>
<td>79</td>
</tr>
<tr>
<td>Functional Adequacy as a Fundamental Component of L2 Proficiency</td>
<td></td>
</tr>
<tr>
<td>Birgit Lijmbach</td>
<td>81</td>
</tr>
<tr>
<td>A New Progress Test Dutch as a Second Language, for Higher Educated Learners</td>
<td></td>
</tr>
<tr>
<td>Kasper Maes</td>
<td>82</td>
</tr>
<tr>
<td>CEFR Grammar: Which Rules at Which Level?</td>
<td></td>
</tr>
<tr>
<td>Margret Oberhofer &amp; Jozef Colpaert</td>
<td>83</td>
</tr>
<tr>
<td>Language for Specific Purposes and the CEFR – the EuroCatering.org Example</td>
<td></td>
</tr>
<tr>
<td>Harold Ormsby</td>
<td>85</td>
</tr>
<tr>
<td>Self-Assessment as a Starting-Point for Useful Communication among Learners, (Prospective) Employers and Teachers: Adapting the European Language Portfolio (ELP) for Use in Specific Real-World Contexts</td>
<td></td>
</tr>
<tr>
<td>Alma Ortiz</td>
<td>88</td>
</tr>
<tr>
<td>Proficiency Exams at CELE-UNAM: Guidelines for Analysis with the Common European Framework</td>
<td></td>
</tr>
<tr>
<td>Tina Rutar Leban, Ana Mlekuž, Karmen Pižorn &amp; Tina Vršnik Perše</td>
<td>91</td>
</tr>
<tr>
<td>The Relation between Foreign Language Achievements of Slovenian students Included in ESLC and their Can-Do Statements</td>
<td></td>
</tr>
<tr>
<td>Cédric Sarré</td>
<td>92</td>
</tr>
<tr>
<td>CLES, a Model Framework for 21st Century Higher Education Language Certification?</td>
<td></td>
</tr>
<tr>
<td>Susan Sheehan</td>
<td>94</td>
</tr>
<tr>
<td>A Core Curriculum Inventory for General English</td>
<td></td>
</tr>
<tr>
<td>Carol Spötti &amp; Kathrin Eberharter</td>
<td>95</td>
</tr>
<tr>
<td>CEFR Performance Descriptors and the Missing Formulae</td>
<td></td>
</tr>
<tr>
<td>María Stathopoulou</td>
<td>97</td>
</tr>
<tr>
<td>Investigating Mediation as Translanguaging Practice in a Testing Context: Towards the Development of Levelled Mediation Descriptors</td>
<td></td>
</tr>
<tr>
<td>Sabine Steemans &amp; Ann Vlasselaers</td>
<td>101</td>
</tr>
<tr>
<td>B2 or Not B2? Mapping the Oral Part of the ITNA to the CEFR</td>
<td></td>
</tr>
<tr>
<td>Martine Swennen</td>
<td>102</td>
</tr>
<tr>
<td>From a Low-Stakes Test to a Higher-Stakes Test</td>
<td></td>
</tr>
<tr>
<td>Claire Tardieu, Monique Reichert &amp; Annick Rivens Mompean</td>
<td>104</td>
</tr>
<tr>
<td>The e-CLES Project: How to Make a Scenario-Based Certificate Valid, Reliable and Fair?</td>
<td></td>
</tr>
<tr>
<td>Jennifer Thewissen</td>
<td>106</td>
</tr>
<tr>
<td>The Criterial Power of Accuracy: a Learner Corpus Approach</td>
<td></td>
</tr>
</tbody>
</table>

www.ua.ac.be/LT-CEFR2013
Ülle Türk & Tõnu Tender
Re-Designing the School-Leaving Foreign Language Examinations in Estonia

Jan Van Maele & Lut Baten
Increasing the Applicability of CEFR Descriptor Scales by Bringing the Context Back into the Framework: Practices from the WebCEF and CEFcult Projects

Jane Vinther
CEFR in a Critical Light

Elena Volodina & Sofie Johansson Kokkinakis
Compiling a Corpus of CEFR-Related Texts

Ying Zheng & John De Jong
Linking to the CEFR: Validation Using Priori and Posteriori Evidence

POSTER PRESENTATIONS

Samar Almoossa
Are IELTS and CEFR Enough Indicator of Students Success in Academic Study?

Khalid Alothman
The Use of Multiple Measures to Detect Academic Vocabulary Improvement

Pilvi Alp, Krista Kerge & Hille Pajupuu
Measuring Lexical Proficiency in L2 Creative Writing

Maisa Martin, Ari Huhta & Riikka Alanen
Using CEFR Scales in a SLA Study on Writing in a Second and Foreign Language

YOUR PROGRAMME AT A GLANCE
PRACTICAL INFORMATION

Hof van Liere
The central building of the University of Antwerp’s City Campus (Prinsstraat 13), better known as Hof van Liere, is part of Antwerp’s cultural and historical heritage. The wonderful 16th-century Hof van Liere and the adjacent historical premises, which were erected by the Society of Jesus in the 17th century, are an oasis of tranquility within the bustling city centre.

Registration and information desk
The conference registration and information desk is located at the second floor of Hof van Liere and is always staffed. You can also send an e-mail to ann.aerts@ua.ac.be or LT-CEFR2013@ua.ac.be

Presentation rooms
At Hof van Liere the conference presentations will be held in four fully equipped conference rooms: Tassis, Elsschot, Gresham (second floor) and Prentenkabinet (first floor). A laptop and a beamer are available in each presentation room.

Internet services
All conference rooms are equipped with wireless internet. Logins and guidelines are available at the registration desk.

Lunches and coffee breaks
Lunches are served at the Agora Caffee, at the corner of Grote Kauwenberg and Vekestraat (Building E).
All coffee breaks will be held at Hof van Liere.
SOCIAL EVENTS

**Monday 27 May, 10.30 h:**
Cultural tour in the city centre

**Monday 27 May, 19.30 h:**
Informal gathering and dinner in the city centre:
Da Giovanni (Pizzeria)
Jan Blomstraat 8
2000 Antwerpen
www.dagiovanni.be

**Tuesday 28 May, 19.15 h:**
Conference dinner at Universiteitsclub
Prinsstraat 13B
2000 Antwerpen

**Wednesday 29 May, 14.00 h:**
Farewell drink at Agora Caffee
Grote Kauwenberg 2
2000 Antwerpen
## PROGRAMME

### Monday 27 May 2013

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00 – 14:00</td>
<td>Registration</td>
</tr>
</tbody>
</table>
| 14:00– 14:30 | Welcome and interactive opening session  
                 (Room: Tassis)  
                 Prof. dr. Jozef Colpaert & Prof. dr. Mathea Simons  
                 Institute of Education and Information Sciences, University of Antwerp  
                 Prof. dr. Joke Denekens  
                 President of the Education Board at University of Antwerp |
| 14:30 – 15:30 | Keynote 1: Lyle F. Bachman  
                  How Do Different Intended Uses and Different Views of Language Impact Language Assessment Practice?  
                  University of California/Department of Applied Linguistics, Los Angeles  
                  (Room: Tassis) |
| 15:30 – 16:00 | Coffee break                                                          |
| 16:00 – 16:55 | Paper session 1  
                 (Room: Prentenkabinet)  
                 Jasminka Buljan Culej  
                 Dan Frost, Jean O’Donnell |
|             | Paper session 2  
                 (Room: Tassis)  
                 Ene Alas, Suliko Liiv  
                 Kris Buyse |
|             | Paper session 3  
                 (Room: Elsschot)  
                 Susan Sheehan  
                 Jane Vinther |
## Tuesday 28 May 2013

<table>
<thead>
<tr>
<th>Time</th>
<th>Session 4</th>
<th>Session 5</th>
<th>Session 6</th>
<th>Session 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 – 10:25</td>
<td><strong>Paper session 4</strong></td>
<td><strong>Paper session 5</strong></td>
<td><strong>Paper session 6</strong></td>
<td><strong>Paper session 7</strong></td>
</tr>
<tr>
<td></td>
<td>(Room: Prentenkabinet)</td>
<td>(Room: Tassis)</td>
<td>(Room: Elsschot)</td>
<td>(Room: Gresham)</td>
</tr>
<tr>
<td></td>
<td>Ari Huhta, Riikka Ullakonoja, Lea Nieminen, Eeva-Leena Haapakangas</td>
<td>Giovanna Comerio</td>
<td>Kasper Maes</td>
<td>Tina Rutar Leban, Ana Mlekuž, Keram Pižorn, Tina Vršnik Perše,</td>
</tr>
<tr>
<td></td>
<td>Bart Deygers, Koen Van Gorp</td>
<td>Cedric Sarré</td>
<td>Ying Zheng, John De Jong</td>
<td>Katrijn Denies, Rianne Janssen</td>
</tr>
<tr>
<td></td>
<td>Michael Corrigan</td>
<td>Jesús García Laborda, Mary Frances Litzler, Marian Amengual Pizarro</td>
<td>Sabine Steemans, Ann Vlasselaers</td>
<td>Ben Knight</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Session 4</th>
<th>Session 5</th>
<th>Session 6</th>
<th>Session 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:30 – 11:05</td>
<td><strong>Keynote 3 : Etienne Devaux</strong></td>
<td>Confronting the CEFR to L2 Certification Purposes: Added Value and Methodological Limitations</td>
<td>SELOR (Belgian public personnel selection and certification agency)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Room: Tassis)</td>
<td>(Room: Tassis)</td>
<td>(Room: Tassis)</td>
<td></td>
</tr>
<tr>
<td>11:05 – 11:30</td>
<td><strong>Coffee break</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Session</td>
<td>Room</td>
<td>Speakers/Participants</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>11:35 – 12:30</td>
<td><strong>Paper session 8</strong></td>
<td>Prentenkabinet</td>
<td>Folkert Kuiken, Ineke Vedder, Alma Ortiz</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Paper session 9</strong></td>
<td>Tassis</td>
<td>Daniela Forapani, Martine Swennen</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Poster session</strong></td>
<td>Elsschot</td>
<td>Khalid Alothman, Pilvi Alp, Krista Kerger, Hille Pajupuu, Maisa Martin, Ari Huhta, Riikka Alanen, Samar Almoossa</td>
<td></td>
</tr>
<tr>
<td>12:30 – 13:55</td>
<td><strong>Lunch</strong></td>
<td>Agora Caffee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:00 – 14:30</td>
<td><strong>Selected paper as plenary session: Glyn Jones</strong></td>
<td>Tassis</td>
<td>Developing a CEFR-Aligned Test from Scratch: a Case Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pearson Assessment and Information</td>
<td></td>
</tr>
<tr>
<td>14:35 – 16:00</td>
<td><strong>Paper session 10</strong></td>
<td>Prentenkabinet</td>
<td>Luke Harding, Maria Stathopoulou</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Paper session 11</strong></td>
<td>Tassis</td>
<td>John De Jong, Benjamin Kremmel, Franz Holzknecht, Bernadette Brouwers</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Paper session 12</strong></td>
<td>Elsschot</td>
<td>Margret Oberhofer, Jozef Colpaert, Marian Amengual Pizarro, Jesús García Laborda, Mario García, Marcin Jaźwiec</td>
<td></td>
</tr>
<tr>
<td>16:00 – 16:30</td>
<td><strong>Coffee break</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:30 – 17:15</td>
<td><strong>Reflection and discussion</strong></td>
<td>Tassis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

www.ua.ac.be/LT-CEFR2013
### Paper session 13
(Room: Prentenkabinet)
Jesús García Laborda, Mary Frances Litzler, Teresa Magal Royo, Nuria Otero de Juan
Birgit Lijmbach

### Paper session 14
(Room: Tassìs)
Claire Tardieu, Monique Reichert, Annick Rivens Mompean
Cristiana Cervini, Monica Masperi, Marie Jouannaud, Francesca Scanu

### Paper session 15
(Room: Elsschot)
Yu-Hua Chen, Shaida Mohammadi, Veronica Benigno
Harold Ormsby

---

**Conference dinner**
(Universiteitsclub)

---

**Wednesday 29 May 2013**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 09:00 – 09:45 | **Keynote 4 – Jan Hulstijn**  
Natural Tensions between Theory and Practice in the Common European Framework of Reference for Languages (CEFR)  
Universiteit Amsterdam  
(Room: Tassìs) |
| 09:50 – 10:45 | **Paper session 16**  
(Room: Prentenkabinet)  
Elena Volodina, Sofie Johansson Kokkinakis  
Jennifer Thewissen  
**Paper session 17**  
(Room: Tassìs)  
Carol Spöttl, Kathrin Eberharter  
Jan Van Maele, Lut Baten  
**Paper session 18**  
(Room: Elsschot)  
Zdenka Gadošová, Andrea Billíková  
Ülle Türk, Tõnu Tender  
**Paper session 19**  
(Room: Gresham)  
Raili Hildén, Marita Härmälä  
Lieve De Wachter, Jordi Heeren |
| 10:45 – 11:15 | **Coffee break** |
| 11:15 – 12:20 | **Post-reflection: Time for a New Framework?**  
(Room: Tassìs) |
| 12:30 | **Lunch and farewell drink**  
(Room: Agora Caffee) |
How do Different Intended Uses and Different Views of Language Impact Language Assessment Practice?

Abstract

The use for which an assessment is intended is generally regarded as the most important consideration in its design and development. Similarly, defining the construct (the area, component, or aspect of language ability) we want to measure is widely considered to be a critical decision in the process of developing a language assessment. In practice, language assessments are used for a wide range of uses, or decisions, and historically, test developers have drawn on a variety of theoretical frameworks of language to define the construct to be measured.

For any particular test, but especially for large-scale, high-stakes tests, the test developers and test users are required, by current professional standards, to provide evidence to support their claims about the intended interpretations of test scores and about the intended uses of these interpretations to make decisions. However, in an increasingly global “market” of language testing, there is increasing pressure, on both test developers and test users, to find ways of linking different tests to a common conceptual framework of language use. What is at issue is that these tests may have been developed for very different uses, for different populations of test takers, and may be informed by very different views of the construct to be measured. In this environment, the demand for portability and transferability of interpretations often overrides fundamental concerns for reliability, validity, and fairness.

The primary purpose of “linking” different tests to each other or to a common standard is to enable test users to interpret and use the results of the two tests in the same way. Using two tests “in the same way” requires that the two tests measure similar constructs, that the decisions to be made are similar, and that the consequences of these decisions are similar. In my view, many current linking activities do not provide adequate justification for claims about these basic similarities. Given the pervasiveness of such claims and practice, I think it is imperative for us, as a profession, to address some very fundamental issues about the nature and justification of “linking” different language tests to a common standard.

In this presentation I will begin with a brief overview of the different language frameworks that have informed large-scale language tests in the past half century. I will then use an assessment use argument (Bachman & Palmer, 2010) to analyze the ways in which two different approaches to defining language differ in terms of the claims they make about score-based interpretations. I will then discuss the different uses for which tests based on these two ways of defining language might be most appropriate. Finally, I will return to the issue of the difficulty of “linking” tests based on different types of language frameworks and intended for different kinds of decisions to a common framework of language.

Lyle F. Bachman
University of California, Los Angeles, The United States
lfb@humnet.ucla.edu
References

Etienne Devaux

SELR, Brussel, Belgium
etienne.devaux@selor.be

Confronting the CEFR to L2 Certification Purposes: Added Value and Methodological Limitations

Abstract

As an institutional personnel selection and certification agency, SELOR has a tradition of L2 certification in the Belgian context. Assessing the receptive and productive second language skills of public personnel belongs to our core missions.

In the past few years, we have adapted our tests and assessment methods to the real-life professional context and chose the CEFR as our reference. We worked with an academic experts panel for theoretical issues and involved experienced raters to operationalize decisions. We undertook the revision of the existing methodology and chose a validation model. We gathered in-depth information (qualitative feedback and quantitative studies) to identify satisfactory areas and aspects amenable to improvement. Our work revolved around the main axes of this model: the ‘context validity’ and ‘theory-based validity’ axes helped us revise our expectations and test specifications; we used the ‘scoring validity’ axis to revise our assessment tools and the rater training components. All new methods and contents were duly pretested and validated.

The discussion will focus on the added value and limitations of the CEFR, which we thought might be of interest to other organizations. We found in the CEFR useful common concepts to work in a multilingual context. We experienced that the CEFR is not a self-contained ready-to-use framework and that organizations need specialized partners to use it sensibly. We noticed that the construct choices we made in terms of competence and domains could jeopardize the assumed comparability of CEFR proficiency levels. We observed that the CEFR provides limited input about assessment methods and heterogeneous descriptors and that developing assessment tools requires great care and considerable investments. Using the CEFR for validation purposes and finding data fit for an external criterion validation may also be challenging. This discussion could open perspectives for further studies.
Abstract

The CEFR (Council of Europe, 2011) represents a brave and moderately successful attempt to cater for the interests of stakeholders in the field of language education (learners, curriculum planners, schools, teachers, employers, local and national authorities), while taking account of theories of language use and language acquisition and the empirical research supporting these theories. However, in all walks of human life, there is a natural tension between practice and theory, simply because they serve different purposes. It is therefore impossible to base any framework of language teaching and assessment completely on theory and research in the language sciences. This does not mean, however, that language-acquisition theory and research are irrelevant for the CEFR. In this presentation, I will present a model of language proficiency in native and non-native speakers, and some hypotheses derived from it (Hulstijn, 2011; Hulstijn, in progress), proposed as both sufficiently plausible and sufficiently implausible to deserve to be empirically falsified. The model, embedded in a usage-based approach to language acquisition, distinguishes between basic language cognition (BLC) and higher (or extended) language cognition (HLC). BLC is the language, used in the aural/oral modes, which all native speakers have in common; HLC concerns all other language knowledge and use. Despite the fact that this model, like most models and theories in science, does not render the observed phenomena correctly, I will propose that there is room to use it cautiously in discussions on the CEFR. The main danger currently threatening the CEFR is the diversity of interpretations of its levels, leading to unwanted differences in assessment practices within and across languages. To combat this threat, I will propose that the Vocabulary Range scale (CEFR, 2001, p. 112) be additionally defined in terms of vocabulary-size numbers and that vocabulary tests form part of language-proficiency exams.

References

The Council of Europe’s Common European Framework of Reference for Languages (CEFR): a 2013 Summary of Developments

Abstract

The Council of Europe’s Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR) was developed by a Council of Europe international working party between 1993 and 1996 with a view to promote transparency and coherence in language learning and teaching in Europe. After a pilot scheme, it was officially published in English and French in 2001, proclaimed the European Year of Languages by the Council of Europe jointly with the European Union. The CEFR has since been translated into additional 38 languages and quickly turned to be one of the most influential publications of the last decade in the field of language learning, teaching and especially language testing in Europe and elsewhere. In a steadily growing number of countries, the CEFR has become a powerful instrument for shaping language education policies. The task of relating language policies, language curricula, teacher education and training, textbook and course design and content, examinations and certification systems to the CEFR has been undertaken by a considerable number of public and private stakeholders in Europe and beyond.

Since its publication in 2001, the CEFR has grown to become a core element of an extensive set of materials, a toolkit for different target groups. In my contribution, I intend to summarise the developments around the CEFR by referring to the many related recommendations, tools and instruments that have been developed by the Council of Europe over the last years and made available to the users of the Framework – as a starting point for the discussion on what may be next to happen for the CEFR, in Europe and beyond.
Abstract

This paper is concerned with the genesis of a high stakes test of General English. The test was developed with the intention that each of its six levels should be aligned to the corresponding level of the CEFR, A1 to C2. Procedures recommended for the “Specification” stages of alignment in the Council of Europe’s Manual (Council of Europe, 2009) were followed. The drafting of test specifications was informed by CEFR descriptors, as was the formulation of scoring criteria for speaking and writing. However, various challenges were encountered in this process. Most notably:

- As has been pointed out by Alderson et al (2004) among others, the descriptors are far from complete in their coverage, from one level to the next, of language activities. The higher levels (C1 and C2) are especially underspecified in this respect.
- Even where descriptors do suggest appropriate assessment tasks (e.g “Can understand short simple letters” – Reading Correspondence at A2) they do not help with design decisions such as the linguistic features of texts to be used for comprehension testing, or the criteria for assessing productive tasks – decisions which are often critical in setting the level of difficulty of the task.

The presenter will relate how the test development team endeavoured to meet these challenges by:

- Formulating item writer guidelines according to a schema which aims to specify how CEFR descriptors apply to critical aspects of task design, such as grammatical complexity or choice of distractors
- Applying a similar schema to the formulation of scoring rubrics

References


Language Testing in Europe: Time for a New Framework?
Training Interviewers and Raters for the National Examination in the English Language – the Estonian Experience

Ene Alas & Suliko Liiv
Tallinn University, Tallinn, Estonia
enealas@tlu.ee - liiv@tlu.ee

Abstract
As of 2014, the national examination in the English language in Estonia will attempt to measure students’ language proficiency within two levels on the CEFR scale (B1 and B2) instead of focusing on only B2, as has been the practice so far. The blueprint of the new examination is expected to be available in spring 2013. Changing the concept of the national examination has meant, among other things, altering all sections of the national examination, developing a new interviewer script for the speaking section, designing new marking scales for the subjectively marked sections – writing and speaking – and training interviewers to use the interviewer script as well as training raters to work with the new marking scales. The presentation will briefly concentrate on the challenges posed by the development of the marking scales and relating them to CEFR. The speakers’ main focus, however, is another aspect of quality control - the practical aspect of training teachers of English in Estonia, who act as interviewers and raters within the framework of the national examination in the English language, to use the script and the marking scales reliably. A proposal will be made for a training sequence to reach that end.

References

Investigating Teachers’ Opinion on the Feasibility and Intended Washback Effect of a High-stakes Oral English Test in Spain

Abstract

Research into the influence of tests on teaching and learning referred to as ‘washback’ or ‘backwash’ in the education literature has been extensive over the past decades. The majority of washback studies have emphasised the negative consequences of tests, especially high-stakes examinations, on different areas of the curriculum. More recently, however, high-stakes tests (used for making important decisions which affect people’s futures) have been employed to reform instruction and achieve positive washback (Weir, 1990; Spolsky, 1996; Norris, 2009). Indeed, many different countries in the world have introduced various types of high-stake tests with the aim of improving education and support good practice (Alderson and Wall, 1993; Cheng, 2004; Qi, 2007). The future inclusion of an oral English sub-test in the Spanish University Admission Examination (academic year 2013-2014) is seen as an attempt to improve the level of spoken English among Spanish undergraduates and promote positive washback.

This study investigates the opinion of 13 secondary teachers (out of a total of 15) who participated in the implementation of the pilot oral test conducted in Majorca (Balearic Islands). The teachers evaluated a total of 175 secondary students in May 2012. Results, collected from a questionnaire, show that teachers hold positive views on the organization, structure and design of the new oral test. Furthermore, the majority of them believe that the oral test will affect teachers’ methodology and increase the amount of time devoted to the practice of students’ oral skills in class (Amengual, 2009), although some concerns are raised over gains obtained due to coaching for the examination. Findings also reveal teachers’ concerns associated with the need to receive some training courses in the use of the rating scales to ensure rater inter-reliability. Finally, most teachers question the feasibility of developing this test due to the current economic situation of the country.

References


Bernadette Brouwers
Australian Council for Educational Research, Melbourne, Australia
bernadette.brouwers@acer.edu.au

The Assessment of Language Competence – Moving Forward

Abstract

Australian schools do not participate in national benchmark testing for second languages. In 1990, the Australian Council for Educational Research (ACER), in collaboration with the Australian Bicentennial Multicultural Foundation and the University of Melbourne’s Language Testing Research Centre (LTRC), developed the Assessment of Language Competence (ALC) tests in listening and reading comprehension at three levels and in six languages (Chinese, French, German, Italian, Japanese and Modern Greek). The multiple choice tests are offered to schools annually.

The ALC is widely used by government and independent primary and secondary schools across all Australian states and territories as well as in New Zealand and into SE Asia and the Pacific. The test is not mandated and is used by schools for various purposes. These include motivating students (through the awarding of ALC certificates), formally comparing individual school performance with other schools (via the ALC School Report) and internal monitoring of language programs at the school level (through analysis of ALC School Report data within and across languages).

ALC tests are developed by language specific experts contracted to ACER. Test specifications were originally developed in consultation with the LTRC and formed the basis of guidelines for writing panels. Tests were extensively trialed and piloted and detailed teacher feedback informed the review of early testing.

Item writers attend a training workshop and are guided by a detailed Writers’ Manual. Analysis of the ALC achievement descriptors forms part of the training workshop. These are reviewed annually to ensure they reflect curriculum and assessment developments at national and international levels. Descriptors draw on contexts, topics, communicative functions, text-types and item intents originally informed by the Australian Language Levels Guidelines.

This paper will provide an overview of the ALC item development process and will explore a process for more formally benchmarking and linking the ALC to the CEFR.

References


Jasminka Buljan Culej
National Centre for External Evaluation of Education, Zagreb, Croatia
jbculej@ncvvo.hr

ESLC in Croatia: Relation between Years of Learning the Foreign Language and the CEFR Level Achieved

Abstract

Foreign language teaching in Croatia starts in the first grade of primary education. Learning one foreign language is compulsory for all students from the first grade. In the fourth grade, students have the opportunity to choose a second foreign language as an optional subject.

ESLC testing was conducted in primary schools on a representative sample of eighth grade students at ISCED level 2. Student sampling was done by SurveyLang experts on the basis of the list of all Croatian eighth graders currently learning foreign languages. During the administration of the ESLC, a total of 1,109 (49.6%) students were tested in English, and 1,126 (50.4%) students were tested in the second target language, German.

The majority of participating students (80%) had been learning English for five to eight years, while approximately 18% of students had been learning the first target language since kindergarten.

Students tested in the first target language (English) generally achieve good results in Reading, which was expected. Level A2 and higher, which is considered to be attainable after eight years of foreign language learning, is achieved by 54% of students. Second target language achievements show that 72% of students achieve level A1 and higher in Reading.

When comparing achievements in Listening, the results for the first target language show that 71% of students achieve level A2 and higher. The results in second target language Listening show that 76% of students achieve A1 and higher, while expectations were not met by 24% of students.

In the first target language Writing, 75% of students reached A2 and higher and twenty-five percent of students score below the expected minimum level, out of which 5% achieve pre-A1. Results for second target language Writing show that only 13% of students do not achieve A1 and higher.

References


**Abstract**

“Assessing speaking is not impossible, but difficult (…): teachers often focus narrowly on the development of grammatically accurate speech which may conflict with a learner’s desire to communicate and be understood” (Luoma 2004). Nevertheless, in the last two decades, with communicative and task based approaches as mainstream language teaching methodologies, assessments is focusing on authentic communicative contexts in which not only knowledge, but also skills and attitudes are required (Keeves 1994; Parrondo Rodríguez 2004). A growing number of publications is offering us models and patterns in order to simulate this authentic reality in the classroom (such as Cabré and Gómez de Enterría 2006 for Spanish), but without inquiring into the criteria, scales and templates suitable for a flexible evaluation of performance assessment of language competence in this communicative context in combination with a focus on form, apart from the templates based on and limited to the Common Framework of European Reference for Languages (CEFR(L)), which in our analysis will turn out to be too rigid when assessing both communicative output and focus on form (see also CITO/SLO 2010), as well as regarding the combination of formative and summative evaluation and in a analytic/synthetic way, two other criteria in current assessment, apart from validity, reliability and transparency (Dochy and Gijbels 2010).

Therefore, after presenting briefly the state of the art on performance assessment and the objectives of current language teaching, we will evaluate a corpus of assessment templates with respect to the abovementioned criteria.

Based on the data of our analysis, we will finally propose flexible criteria and templates for a communicative assessment of oral language skills, allowing us to adapt the assessment to the demands of any language course without losing sight of the five criteria. So yes, assessing speaking with focus both on form and communicative output is possible.

**References**


Defining, Modeling and Piloting SELF, a New Formative Assessment Test for Foreign Languages

Abstract

Designing a new foreign language test requires defining what we mean by language and language use (Bachman, 1990). The construct that we use is the cornerstone guiding us when we create original test items and when we design the general architecture of the test. What are the characteristics of the intended test-takers? What do we take 'communicative competence' to be? How can we translate all of this into test items that will be administered online? The Common European Framework of Reference for languages (henceforward CEFR) can guide us but cannot tell us how to anchor our items to the descriptors of each of the skill levels while staying true to task-based and action-oriented approaches to language teaching, and to the role of the learner as a social actor (ALTE, 2011; Weir, 2004).

The purpose of this talk is to describe two aspects of this work in progress:
1. how our team's linguistic, discursive and contextual choices were guided by the principle of situational and interactional authenticity on the one hand, and by the attempt to integrate competences on the other;
2. what the preliminary stages of the piloting process tell us about our students' socio-biographical characteristics and the validity of our test.

SELF (Système d’Evaluation en Langues à visée formative) will eventually cover three language skills (listening, reading and writing1), but the first stage of its development focused on listening, because of the high correlation we observed between oral comprehension level and success in foreign language tests. Developing a test to be used in institutional settings implies a series of inevitable constraints due, on the one hand, to the possible wash-back effects on learning and on teaching models and, on the other, to the high number of test-takers taking the test at the same time.

References


1 It might be useful to indicate that exercises should be self-corrective hence we would focus on "limited production" (i.e.: short answers, discourse completion task, etc.).


What and How Many Words Do We Need? Critical Considerations when Developing a CEFR Vocabulary List: Size, Depth, and Growth

Abstract

Since the launch of the CEFR, there have been various attempts to develop a vocabulary list aligned to it. However, little effort has been made to translate the theoretical frameworks or empirical findings of vocabulary research into the development of such listings. For example, recent studies have concluded that knowledge of 8,000–9,000 word families is necessary for reading authentic English materials and perhaps 5,000–7,000 families for oral communication (Nation, 2006; Schmitt, 2008). In comparison, most of the available pedagogical vocabulary listings contain a much smaller repertoire. The ‘multi-facetness’ of vocabulary knowledge (Read, 2000) should also be acknowledged when developing a CEFR-aligned syllabus. There is a whole gamut of ‘knowing’ a word, from recognizing only one context-dependent sense to frequently using it with a variety of meanings. To address the gap between vocabulary research and alignment with the CEFR, this paper critically reviews the construct of vocabulary as a language ability and focuses the discussion on three main dimensions, i.e. vocabulary size, depth and growth.

To establish the relationship between vocabulary knowledge and the CEFR, we will demonstrate an innovative approach using information from language testing statistics and L1 corpus frequency, taking into account the above three dimensions. The rationale is that L1 corpora offer a comprehensive list of words from authentic text produced by native speakers, while L2 learners’ performance in a language test, specifically their responses to different item types that assess various facets of vocabulary, can be used to gauge the extent to which they know the words in different contexts. Preliminary findings from a pilot study will be reported, and the results confirm the complexity and multi-facetness of vocabulary as a construct. The pros and cons of using various sources to inform vocabulary syllabus design and the implications for CEFR alignment will also be discussed.

References


Giovanna Comerio  
The University of Nottingham Ningbo China, Ningbo, China  
giovanna.comerio@nottingham.edu.cn

Can the CEFR Assess University Students in a China-Based British University? A case Study at the University of Nottingham Ningbo China

Abstract

The University of Nottingham Ningbo China is an English-medium university that, within the Language Centre, offers its students the opportunity to study an additional foreign language (French, German, Japanese, Spanish) up to the level B1. Despite the introduction of the ELP and gradual alignment of the curriculum to the CEFR, when it was time to align the examinations set up by our University to the CEFR some issues arose. The format of our exam papers is consistent with the programme developed and focus on the intellectual and transferable skills involved. The marking criteria are designed for assessing not only the language, but also the intellectual ability of the students to organize their discourse.

This study is a first step of a reflection on how and to what extent CEFR criteria can be integrated in or replace the HE marking criteria and what kind of changes in the our pedagogy this would involve. As practitioners, this study is a comparison of marks obtained by students taking oral, written, listening and reading test according to the CEFR (DELF A1) and our university marking criteria.

This project would not be possible without the practical help and the reflections of Magali Kerbellec, to whom I am deeply grateful. I wish also to thank Filippo Gilardi for his support.

References


Interchangeability of Test Results and the CEFR – a Validity Argument Approach

Abstract

Descriptors of the CEFR are commonly provided to test users to assist them in interpreting the test results. However, when comparing the results of different tests targeted at the same CEFR level, users may be tempted to treat them as interchangeable because the descriptors given are identical. Issues of test-CEFR alignment notwithstanding, the CEFR alone does not support such interpretations, as, according to the Manual on Relating Examinations to the CEFR (Council of Europe, 2009:4), ‘two examinations may both be “at B2 level” and yet differ considerably’. How, then, can we determine whether the results of two exams at the same CEFR level may be used interchangeably? Consideration of this question requires an appropriate methodological approach which takes the intended use of results into account, as they may be interchangeable for some purposes and not others. Although a number of methodologies to appraise the link between tests have been put forward (e.g. Mislevy, 1992; Linn, 1993; Kolen and Brennan, 2004), none are entirely satisfactory. An approach based on validity arguments (e.g. Kane, 2012) will be outlined in this presentation, along with the results of a small study investigating the comparability of two exams of different foreign languages: Cambridge English First (FCE) and Certificato di conoscenza della lingua italiana 3 (CELI 3), an exam of Italian. This talk will be of interest to those seeking to link their tests to the CEFR and test users who must interpret test results for a range of different tests.

References


How to Save the Common European Framework?

Abstract

Since its launch in 2001 several authors have decried the shortcomings of the Common European Framework of Reference for Languages, shortcomings include its bias towards speaking skills at intermediate levels, its lack of precision with regard to the linguistic means to realize the language acts it describes at the various levels it defines.

Early criticism addressed the absence of psychometric means to ascertain the linkage of teaching material and assessment instruments to its descriptive system. Although the Manual for linking examination to the CEFR aims to remedy this shortcoming, its own definition (“to help the providers of examinations to develop, apply and report transparent, practical procedures in a cumulative process of continuing improvement in order to situate their examination(s) in relation to the Common European Framework”) by its density and opacity defies its purpose. Furthermore, in spite of its volume and numerous appendices the Manual never touches upon the question of how to exactly define “being at a level” in measurement terms.

Furthermore the limited number of levels defined in the Framework has led to the misconception that the population of language learners can adequately be described in terms of those levels and has led to stretching existing teaching material and assessment instruments to cover all six levels and to a simplistic models such as 200 guided learning hours per level as suggested on the website of a major language test provider. The experience with the earlier work of the Council of Europe with the Threshold level should have forewarned the language teaching and assessment community that in actual practice most learners need several years to go from one level to the next.

This presentation offers an approach illustrated by ensuing data-driven research results that can remedy these shortcomings and save the Common European Framework rather than replacing it.
Lieve De Wachter & Jordi Heeren
Katholieke Universiteit Leuven, Leuven, Belgium
lieve.dewachter@ilt.kuleuven.be

Can a Language Test Verify the Academic Literacy of University Students and how Does that Relate to Study Success?

Abstract

The TaalVaST language test, developed at the Leuven Language Institute (University of Leuven) has been taken by more than 9000 first year students since September 2010. Its growing popularity shows that a well constructed and piloted language test can replace other screening methods that take up a lot of time and resources. The test has proved to be an efficient tool to, on the one hand, give students an early ‘warning signal’ during their academic education and on the other hand function as a starting point for language tutoring tailored to the needs of a defined target audience.

The construct of the language test has been proved to be statistically valid and reliable. A pilot version of the test was used to determine the construct validity and reliability by means of the simple-item discrimination method, the calculation of the point-biserial correlation coefficient, item facility and the Kuder-Richardson 20 formula. By analyzing students’ secondary education and language use at home the group-differential validity of the test has been determined. Moreover, a study by Huyghe and Marx (2011) found a significant correlation between the language test scores and students’ academic success. That confirms other important studies in the field of language testing (Van Dyk 2010, McNamara 1996) and proves the concurrent validity of the test (Davies 1990, 23-24).

Constructing the test, we deliberately did not start from the CEFR, but used a needs analysis as a basis for the construction of our test items. Essentially, the reading and writing requirements for first year university students were taken as a basis for the items and were only incorporated into the language test after they had proven to be valid. It is demonstrated that the development of a language test is a continuous process of designing, testing and revising.

References


Van Dyk, T. (2010). Konstitutiewe voorwaardes vir die ontwerp van ’n toets van akademiese geletterdheid. Universiteit Stellenbosch. [PhD]


CEFR Can-Do Statements as a Means of Self-Assessment: is There a Common Understanding, Regardless of the Student’s Gender and Educational System?

Abstract

The Common European Framework of Reference (CEFR) contains a set of concrete descriptive scales, which, ideally, should allow for unambiguous communication regarding stages in language learning. Yet there is still need for empirical research on how much room for interpretation is left by the descriptors. The aim of this paper is to investigate whether the use of CEFR can-do statements as a means of self-evaluation shows a gender bias and a country or region-related bias, which is commonly found with other measures of academic self-concept.

The study is based on data from the European Survey on Language Competences (ESLC). The ESLC was carried out in spring of 2011 by order of the European Commission. It was the first survey to provide information on students’ second language competences that can be compared across fourteen countries, five languages and three skills (reading, listening and writing). As part of the survey’s extensive background questionnaire, students were asked to assess their own competences using four can-do statements per skill that were all taken or adapted from the CEFR descriptor scales. The current paper deals with findings resulting from a comparison between the students’ measured CEFR-level on the one hand, and their can-do self-evaluation on the other hand. Students were labeled as ‘overestimating’ their own level, ‘underestimating’ it or having a ‘correct’ self-concept for the tested skill.

Multinomial logistic regression analyses on data from over 40000 students revealed that, when they assess their own skills by means of the can-do statements, the students’ odds of over- or underestimating their competences correlate significantly with their gender and educational system (i.e., country or region). In other words, the can-do statements are currently subject to bias: however concrete the statements may already seem, they are still perceived in a way that reflects the students’ personal or national norms rather than fixed criteria.

References


The Influence of the CEFR on Rating Scale Design

Abstract

The Certificate Dutch as a Foreign Language (CNaVT) offers Dutch task-based language exams for 6 different profiles which have been determined by an extensive needs analysis (Van Avermaet & Gysen, 2006). The task content is co-determined by a pool of subject specialists around the world who verify the authenticity and representativeness of each task and check the items for cultural bias.

For the past years the CNaVT’s rating scale has been dichotomous and analytical. Even though this scale has a proven reliability and usability, it was decided to reshape it into a model that would better reconcile the CNaVT’s philosophy with its stakeholders’ needs: i.e. a clearer alignment with both the CEFR and domain experts’ judgements of language performance (Jacoby & McNamara, 1999).

Redesigning the scale has proven to be an extensive undertaking which touches upon all aspects of language testing. Indeed, altering a dichotomous model into a polytomous band rating scale, which merges performance driven exemplifications (Weigle, 2007) with measurement driven descriptors is an operation so all-encompassing that it necessitates rethinking the entire testing process. Simultaneously, working closely with the CEFR has forced the rating scale developers to critically examine the level descriptors so as to operationalize them in a usable rating scale without neglecting known pitfalls such as validity reduction (Lumley, 2002) and a lack of concreteness (Fulcher, 2010).

This presentation focuses on the role of the CEFR in the rating scale redevelopment process, on its strengths, but also on its shortcomings which prevent it from being a readymade assessment tool. The presentation will include data resulting from the development and validation process. This includes focus groups with subject specialists, stimulated recall interviews with raters as well as qualitative test analyses (i.e. inter and intra rater reliability, correlation coefficients etc).

References


Dan Frost & Jean O’Donnell
Université de Savoie, Chambéry, France
dan.frost@univ-savoie.fr - jean.o-donnell@univ-savoie.fr

Combatting the “Can’t do Mentality”: Expert, Peer & Self-Assessment in a French University Context
The ”ELLO” Project (étude longitudinale sur la langue orale)

Abstract

The CEFR was adopted by the French National Ministry of Education and Research (MENR) in 2005. Its relevance to the French context (Goullier 2007) and its applicability (Taylor 2011, Petit 2007, Luoma 2004) have been discussed by several authors. We will analyse the practical aspects of using the CEFR as a tool in a longitudinal cohort research project involving the assessment of English spoken production among Applied Foreign Languages undergraduates. As practitioners, we were struck by the lack of accurate information, specifically in relation to spoken production, to describe the levels of our students before, during and at the end of their three-year degree. Obtaining a language profile for our students at regular intervals would enable us to design and evaluate our courses more effectively. Consequently, as researchers we started a study in 2011 which involves collecting and assessing several samples of spoken data for each student from first to third year. Three types of assessment are currently being carried out: "expert assessment", “peer assessment” and “self-assessment”.

This paper explains the reasons for choosing the CEFR as our basic measurement tool after having considered other possibilities. We discuss the advantages and disadvantages encountered so far according to the type of evaluation that is being carried out. For example does experts’ use of the scale coincide? How do peers perceive and apply the scale? Are self-assessments using the scale reliable? What features of the scale are experts satisfied/comfortable with? Finally there were attempts to make constructive suggestions in the light of our preliminary findings.

References


Sockett, G. (2011). From the cultural hegemony of English to online informal learning: Cluster frequency as an indicator of relevance in authentic documents, ASp, 60, 5-20.


Abstract

The placement tests are conceived to measure the knowledge of language that students have, in order to enter them into the corresponding course and guarantee a standard level of ability within the class. These tests might be based on the syllabus adopted by the host institution or on unrelated material.

Given that the University of Parma language courses are structured in line with the CEFR, the online test developed by the Language Centre relates to the L2 Italian Syllabus (Lo Duca, 2009), which divides linguistic content (tasks and texts, linguistic functions, form, structure and meaning) into the six levels A1-C2. This paper sets out the guidelines that item writers/editors must refer to while designing the test, including: a) a list of the technical specifications (number of sections required in the test, number of items per section, types of abilities that have to be tested, features of the linguistic elements that must be included), b) a checklist to ensure that items and CEFR levels correspond and c) various indications to assess the reliability of the test and ensure its overall validity. As the online placement developed in Parma has been used both by the University of Zagreb (Croatia) and the University Falun (Sweden), the international impact of this model will also be investigated.

References


Forapani, D. (2011). The tapestry of a placement test. Research into designing an Italian placement test for Erasmus students (who are native speakers of Romance languages), Tuttitalia, 40, 3-6


Zdenka Gadušová & Andrea Billíková
Constantine the Philosopher University, Nitra, Slovakia
zgadusova@ukf.sk - abillikova@ukf.sk

English Tests for Secondary School Leavers in Slovakia

Abstract

Language policy in Slovakia follows the adopted European trends in language teaching and learning – any secondary school leaver should be proficient in two foreign languages apart from their mother tongue. Thus, all secondary school leavers in Slovakia are obliged to take, among other “Maturita exam” subjects, a final exam also in a foreign language either at B1 or B2 levels according to the CEFR. In our paper, we will present the concept of the final exam in English language taken as one of the “Maturita” subjects. We will describe its components and testing techniques. Specific attention will be paid to the process of students’ preparation for the oral part of the exam, which secondary school leavers usually undergo either within English classes at school or on their own. The recent survey and research findings led us to create tailor-made study material for Slovak “Maturita” leavers that is based on reflective learning. During the presentation the material will not only be discussed (its aims and tasks) but also a video-recording showing the way students should be trained for the oral exam, will be demonstrated.

References


< The School Act > <2008> < Act Nº 245/2008> (SK).
Can Spanish High School Students Speak English?

Abstract

Spain will go through a number of educational changes in the next six months because a new educational reform will be implemented from the academic year 2013-2014. Similarly to the “No Child Left Behind Act”, it is thought that assessment will determine the specific support to certain educational programs and school. Assessment is considered to have a great impact in the initiatives for improvement in the overall underachieving educational system (PISA 2012, OCDE indicators 2012, European Survey of Language Competence). When planning the most significant proposals to implement the educational system, the researchers of the OPENPAU project considered the importance of implementing large scale assessments for languages with limited resources. Since computer labs in high schools are neither secure nor updated and do not usually have enough posts, proposals for alternative solutions have been suggested. In that sense, ubiquitous alternatives provide options to increase the number of test candidates taking the test simultaneously. We also considered the demands of teachers and educational boards in test design. Accordingly, we considered that test takers should be able to be tested in speaking, listening, reading and writing. As a consequence, the informatics team considered two main proposals: 1) The design of a specific test to implement speaking exams as the OPI or the speaking sections of TOEFL; 2) to use ubiquitous m-design that can be used in a number of devices such as i-pads, mobiles or tablet PC. The test validity followed the validation standards by Weir (2005). The presentation describes the technological and practical aspects of delivery and application. The paper also proposes a research agenda including different applications from the Baccalaureate General Test.

References


Proposals of Ubiquitous Delivery of the Foreign Language Paper of the Spanish Baccalaureate General Test

Abstract

Spain will go through a number of educational changes in the next six months because a new educational reform will be implemented from the academic year 2013-2014. Similarly to the “No Child Left Behind Act”, it is thought that assessment will determine the specific support to certain educational programs and school. Assessment is considered to have a great impact in the initiatives for improvement in the overall underachieving educational system (PISA 2012, OCDE indicators 2012, European Survey of Language Competence). When planning the most significant proposals to implement the educational system, the researchers of the OPENPAU project considered the importance of implementing large scale assessments for languages with limited resources. Since computer labs in high schools are neither secure nor updated and do not usually have enough posts, proposals for alternative solutions have been suggested. In that sense, ubiquitous alternatives provide options to increase the number of test candidates taking the test simultaneously. We also considered the demands of teachers and educational boards in test design. Accordingly, we considered that test takers should be able to be tested in speaking, listening, reading and writing. As a consequence, the informatics team considered two main proposals: 1) The design of a specific test to implement speaking exams as the OPI or the speaking sections of TOEFL; 2) to use ubiquitous m-design that can be used in a number of devices such as i-pads, mobiles or tablet PC. The test validity followed the validation standards by Weir (2005). The presentation describes the technological and practical aspects of delivery and application. The paper also proposes a research agenda including different applications from the Baccalaureate General Test.

References


Abstract

The paper presents previous conclusions drawn up from a research conducted in the context of the European project Promacolt (www.promacolt.eu), and extends them to the field of language testing based on adapted descriptors of the CEFR.

Promacolt project passed recommendations on how to approach the target users of a foreign language from the perspective of the language course designer/ provider. One of the lessons learnt is the key role played by the course commander, i.e. the organisation which makes possible the actual course (whether a public education organisation with the mission to do so, a private company buying the course for its employees, or a public/private sponsor, for instance).

According to our findings what a test can actually measure must correlate with the expectations of the language course commander, in particular when it is not a language education organisation itself. Often, there is a gap between one and the other because there is a lack of specification in the first phase of the diagnosis of the intended purpose of the language course. In those cases, we can find a satisfactory performance according to the test, but an inadequate/incomplete set of competences gained by the learner for the practical purpose expected by the course commander.

The paper illustrates how the spheres of sociolinguistics, psycholinguistics, cultural aspects and pragmatics can be integrated in language testing, by enriching language competence descriptors based on CEFR when language is intended for a practical use.

References


Investigating the Construct Underlying the CEFR Phonological Control Scale

Abstract

Pronunciation is a difficult skill to master when learning a second language. It is also a challenging skill to assess, with relatively little research on how pronunciation ability develops, and conflicting views on what the goals of pronunciation instruction should be. The “phonological control” scale of the CEFR represents one attempt at describing pronunciation development. On face-value, the scale presents a useful set of descriptive statements, however, as Harding (2012) has argued, the phonological control scale appears to reflect a view of pronunciation development which conflates increasing ease of understanding with a decreasing level of “foreign accent”; a view which is at odds with the research literature (see Munro, 2008). To date there has been little research on the phonological control scale’s validity and usefulness, so any claims that the scale is flawed remain speculative.

The aim of this study was to conduct an empirical investigation of the relationship between ratings of pronunciation ability made using the CEFR phonological control scale and separate measures of comprehensibility (ease of understanding) and strength of accent (the degree to which an accent sounds native-like). 44 non-native speakers of English from a range of first-language backgrounds provided speech samples using a common picture description task. Nine raters then listened to these speech samples and provided ratings on scales designed to measure comprehensibility and strength of accent. The same raters also evaluated the speech samples against the CEFR phonological control performance descriptors, and provided written and oral feedback on their experience of applying these descriptors.

Quantitative and qualitative data will provide evidence from which to draw conclusions about the theory of pronunciation implicit in the phonological control scale. Implications will also be drawn for the usefulness of the scale as a guide both for describing pronunciation development, and for the teaching and assessing pronunciation.

References


Raili Hildén & Marita Härmälä
Finnish National Board of Education, Helsinki, Finland
raili.hilden@oph.fi - marita.harmala@oph.fi

Work in Progress: How Useful is the CEFR in Designing the Follow-Up Assessment of Learning Outcomes in Foreign Languages in the Finnish Basic Education?

Abstract

Our presentation will discuss an on-going national assessment on foreign language proficiency administered by the Finnish National Board of Education. Based on the validity theory of criterion-referenced assessment, we explore how scale descriptors serve in designing the tasks to be used for assessing Finnish 9th graders' foreign language proficiency at the final phase of the compulsory basic education. Drawing on the results of related Finnish research projects, we discuss the strengths and challenges of using the CEFR scales and their national application for assessing learning outcomes in general, and for the purposes of the on-going assessment in particular.

Validity of an assessment is a holistic process, starting from defining the test's desirable consequences and ending at conclusions to be made of the results. In the Finnish context, assessments are implemented to evaluate the attainment of the goals set for the language education at school level. Ultimately, the information gathered through the assessments is used for the further development of the National Core Curriculum (2004) for language education. The major instruments for gaining information are language tasks and questionnaires.

Language tasks used in the national assessment have been specified in relation to the domains and goals of the curriculum, which, in turn, reflect the real life language use of the 9th graders. In addition to the appropriateness for the target population, cultural and ethical issues need to be taken into account in the process of task design (McNamara & Roever 2006; Bachman 2010). In the presentation, we demonstrate the task designing process all the way from choosing relevant proficiency levels and language use situations to compiling the final version of the tests in English, French, German, Russian and Swedish languages.

References


www.ua.ac.be/LT-CEFR2013

http://www.oph.fi/english/sources_of_information/core_curricula_and_qualification_requirements/basic_education
The Use of the CEFR in Diagnosis

Abstract

This presentation reports on how the CEFR has been utilized in a research project focusing on the diagnosis of second or foreign language (SFL) proficiency. The talk first gives an account of the role of the CEFR in the design of the study and, secondly, reports on findings that shed light on the relationship between certain areas of linguistic knowledge and reading and writing at different CEFR levels.

The project is an international 4-year (2010-2013) study into the diagnosis of SFL reading and writing. It seeks to identify the cognitive, affective and linguistic features which predict a learner's strengths and weaknesses in those areas by studying several hundred SFL learners both cross-sectionally and longitudinally. The project aims at deepening our understanding of L2 development and of the factors that affect it, and will formulate hypotheses for further work on SFL diagnosis. The project relates to the activities of the European SLATE (Second Language Acquisition and Language Testing in Europe) research network (see www.slate.eu.org)

We give an account of the ways in which the CEFR influenced the design of the study. First, we describe the selection of CEFR-related reading tests such as DIALANG and the Pearson Test of English General for data-gathering purposes. Secondly, we report on the success of applying CEFR-related scales for rating writing performances in SFL and L1. In this way, it was possible to arrive at CEFR-referenced estimates of the informants' reading and writing ability in two languages.

We also report on findings about the relationship between CEFR levels and linguistic aspects of performance. For example, we provide a characterization of the CEFR levels in reading and writing English in terms of a vocabulary profile. Finally, we reflect on the usefulness and limitations of the CEFR for SFL diagnosis.

References


The English Profile Project: Researching what the CEFR Means in Terms of Specific English Linguistic Knowledge

Abstract

The English Profile Project has supported research projects into what the CEFR levels mean explicitly for English, the Reference Level Descriptions, particularly in terms of grammar and vocabulary. This presentation will share some of those findings, particularly those from research based on the Cambridge Learner Corpus. It will also share some of our work into linking English Profile to curriculum development. Although this presentation relates to research, it is more focused on its significance for teaching and learning than a formal research paper.
Benjamin Kremmel & Franz Holzknecht
University of Innsbruck, Innsbruck, Austria
benjamin.kremmel@uibk.ac.at - fholzknecht@gmail.com

Strengths and Weaknesses of the CEFR in Guiding Test Task Design: What the Can Do’s Can Do and What They Can’t (yet)

Abstract

This practice-related paper will report on opportunities and limitations the CEFR has provided in setting up a national school-leaving exam for modern languages in a European country. While the framework has been particularly helpful in drawing up competence-based test specifications for the traditional language skills, test developers are often faced with shortcomings of the CEFR when it comes to the implementation and operationalization of the level descriptors on a practical level. Although the framework states that task support (e.g. instructions), text characteristics (e.g. text type, discourse structure, presentation, length, relevance and linguistic complexity) and the type of response required can affect the difficulty of comprehension tasks (Council of Europe, 2001, p. 164f), the CEFR lacks specificity when it comes to translating these features into levels on the illustrative scales. On what basis is it then that test developers decide whether a text for a reading comprehension task is at B2 level, whether the response format used is suitable for A2, or whether a writing task will elicit a B2 performance rather than a B1 performance? Such challenges become particularly prevalent when it comes to the design of tasks for the productive skills and to ensuring fairness and equal difficulty for all candidates taking parallel versions of a test. The paper will present advantages and drawbacks experienced throughout five years of working with the CEFR in large-scale assessment and will suggest that the framework can and should be refined for assessment purposes.

References


Abstract

In recent studies in SLA often general measures for assessing linguistic performance, such as complexity, accuracy and fluency (Housen, Kuiken & Vedder 2012). Few studies, however, report on the functional adequacy of the L2 output, considered as being crucial by some authors (De Jong et al. 2012; Kuiken, Vedder & Gilabert 2010, Pallotti 2009). Until now, there has been no unanimity, as to how functional adequacy is to be defined or assessed and by which features it is determined (Iwashita et al. 2008). While functional adequacy is sometimes interpreted as socio-pragmatic appropriateness (McNamara & Roever 2007), in other cases it is operationalized as communicative effectiveness (Upshur & Turner 1995) or successful task completion (De Jong et al. 2012).

The primary focus of the present paper is on the assessment of L2 writing. It is argued that the assessment of linguistic performance in L2 is not possible without taking into account the functional dimension of L2 production, as defined in the CEFR. The analysis is based on the written output of 32 learners of Dutch and 39 learners of Italian at CEFR level B1, who were submitted to two argumentative writing tasks. Functional adequacy was rated by experienced raters on a 6-point Likert scale. During a subsequent panel discussion, the raters were asked to verbalize the reasons underlying their decisions to assign a text to a particular rating level.

In the paper the following questions will be discussed (1) How can functional adequacy as a construct be defined and measured? (2) What are the features of functional adequacy that raters consider to be crucial? (3) How can differences between low proficient and high proficient L2 learners be described in terms of functional adequacy?

References


A New Progress Test Dutch as a Second Language, for Higher Educated Learners

Abstract

I will present a progress test at levels A2 and B1, also on behalf of three colleagues from the Language Centres of three other universities. The test, which assesses the four skills individually, was developed in the past year as there appeared to be a need for such tests, tailored to the target group of university language centres and suitable to be used as a progress test.

I will explain the design of the test, the relationship with the CEFR, the development and pilot phases, and discuss a number of issues we were confronted with and the solutions we found for them. The project has resulted in two tests, one at A2 and one at B1 level, which will be published digitally by Uitgeverij Boom. These tests can be purchased by institutions so that they can set and assess them themselves. Each of the four skills can be individually assessed.

I will demonstrate a few components of the test and the assessment procedure, and of course there will be time for questions, comments and discussion.

This presentation will be held in Dutch.
Kasper Maes
Radboud in'to Languages, Nijmegen, The Netherlands
k.maes@let.ru.nl

CEFR Grammar: Which Rules at Which Level?

Abstract

In 2011, a series of grammar books were published in the Netherlands (Van Dale Uitgevers, 2011). These publications were the result of a project, in which the grammar of six different languages (Dutch, English, German, French, Italian and Spanish) was classified according to the CEFR. This project was a response to the need for concretization of the CEFR level descriptors. First, existing publications regarding the CEFR and grammar were consulted, such as the Profile Deutsch for German and the British Council EQUALS study and Pearson PET General Handbook for Teachers for English. Secondly, the CEFR level descriptors were interpreted for indications of the kind of grammar that should be familiar at a particular CEFR level. A third guideline in the project was the complexity and frequency of particular grammatical structures in sample texts. Based on these three guidelines and the experience of the authors, the grammar rules were classified according to the CEFR. In addition, the language in which the rule itself was explained (easier wording, combining at higher levels) and the difficulty and content of the examples (more difficult vocabulary, longer sentences, different language contexts) were also adapted to the CEFR levels. In this presentation, an overview of the project and its results will be given and some of the choices that were made and the difficulties that were faced will be discussed.

References


www.eurocatering.org is a web-based language learning tool in 12 languages designed for trainees, students and workers in the Hotel and Catering industry to improve their oral language skills. EuroCatering helps the envisaged target groups acquire the basic specific vocabulary and the communicative competences needed to function efficiently in a kitchen, restaurant or hotel abroad by providing learning materials and instructional support. The first part of the presentation discusses the challenges related to a language course for specific purposes to a certain level of the Common European Framework of Reference for Languages (CEFR), however, the CEFR does not refer to languages for specific purposes (Alderson, 2007; Krumm, 2007; Komorowska, 2012). We will also discuss the Reference Level Descriptions (RLDs), which is a new generation of reference descriptions currently available in ten languages.

The second part of the presentation highlights the EuroCatering Language Portfolio as a way to provide a self-assessment tool for students of this vocationally-oriented language learning course.

The presentation closes with consideration of the practicality of the CEFR and the RLDs for developers of an online language course for specific purposes.

References


Self-Assessment as a Starting-Point for Useful Communication among Learners, (Prospective) Employers and Teachers: Adapting the European Language Portfolio (ELP) for Use in Specific Real-World Contexts

Abstract

People often say they are learning an additional language for employment purposes. Indeed, many employers want to have or find employees who are proficient in (a) certain language(s), notably English. However, most often language teachers deal with prospective employees who cannot know who their employer(s) will be. Textbook writers, in turn, bridge gaps by taking generic employment contexts as their guides. Furthermore, employers very, very seldom communicate directly and usefully with teachers or (not surprisingly) with prospective employees about language proficiency expectancies, demands, needs etc.

The CEFR and the subsequent ELP have the goal of improving "the quality of communication among Europeans of different language and cultural backgrounds ... because better communication leads to freer mobility and more direct contact, which in turn leads to better understanding and closer co-operation." (CEFR, pp. xi-xii) Improving communication among learners, (prospective) employers and teachers is, we believe, essential to making this universally applicable goal achievable.

The long-range objective of this design-development project is to create a multidimensional assessment scaffolding (array, matrix, framework) in/on which specific-purpose can-do statements can be located, using categories, classifications and criteria that are found (perceived) and make sense in the real world, and that can be modified as the real world changes.

The project's current reality is preparing Mexican adults with English as L1-L2 or FL to take the International Legal English Certificate (ILEC) examination and/or who have professional needs and interests related to International Legal English. Examples to explain the underlying idea and to show how the current scaffolding works in practice will come from self-assessments used before, during and after exam preparation, and in the creation and updating of individuals’ ELP-like language autobiography and language-experience dossier. One hopes to find other language teachers who would like to do similar projects in their own languages and realities.

References

Short, introductory videos:

Language Testing in Europe: Time for a New Framework?


**Broadly informative webpages:**


**Articles and books:**


Proficiency Exams at CELE-UNAM: Guidelines for Analysis with the Common European Framework

Abstract

The international impact of the CEFR in assessing foreign languages is evident; the emphasis in certifying language proficiency for different purposes has been the cornerstone for exam users. Not only in Europe, but also around the world, diverse stakeholders were interested in analyzing the possibilities of adapting referents for their own academic or commercial purposes.

The Center of Foreign Languages of the National Autonomous University of México (CELE-UNAM) where university students certify their proficiency level in a foreign language has always being at the forefront of the teaching and testing trends. The Center has studied, analyzed, and questioned different approaches and tendencies to adapt or discard them to its students’ needs.

This paper shares the results in the critical analysis of the parameters and referents published in the CEFR in order to confront them and establish the corresponding relationships with the exams produced in CELE.

The rigorous analysis included CEFR documents, ALTE information and CELE’s own Framework of Reference; the proficiency exams used in CELE (English and French versions) were also analyzed in order to establish the guidelines for comparison, design and construction of new exams. The projects’ products include information formats of exams, an operational definition of ‘dominio’ (the term commonly used to refer to the proficiency level of exam takers) and information for exam designers and exam users at all levels.

References


The Relation between Foreign Language Achievements of Slovenian students Included in ESLC and their Can-Do Statements

Abstract

The paper presents the results of the European Survey on Language Competences (ESLC) along with the links between ‘can do’ performance statements based on CEFR and areas of linguistic skills (listening, reading, writing) in Slovenia compared to other educational systems included in the survey. The survey was established to provide participating countries with comparative data on foreign language competence and insights into good practice in language learning. Sixteen European educational systems took part in the survey. Students (in the last year of lower secondary education ISCED2 or the second year of upper secondary education ISCED3) were tested in the two most widely taught foreign languages in their country chosen from the five tested languages: English, French, German, Italian and Spanish. Each sampled student was tested in one language only. The language tests covered Listening, Reading and Writing Language Skills and were based on CEFR. Each student was assessed in two of these three skills. Each student also completed a Questionnaire about his home and school environment regarding foreign languages. A representative sample of 1,500 students was chosen for each foreign language in each educational system. Altogether 53,000 students were tested together with 5,000 foreign language teachers and 2,500 school principals (In Slovenia 290 teachers and 163 school principals).

Students responded to 16 ‘can-do’ statements, providing a self-evaluation of their competence in the tested language. The statements were taken directly or adapted from the descriptor used in the CEFR to illustrate the levels. Statements were chosen to be relevant to the target population. The paper presents the results of the ‘can-do’ statements, moreover it verifies the relationship between can-do self-ratings of students and their achievement at each of the tested skills in Slovenia and compares the results with other included educational systems.

References


http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf
Cédric Sarré
Université Paris-Sorbonne, Paris, France
cedric.sarre@paris-sorbonne.fr

CLES, a Model Framework for 21st Century Higher Education Language Certification?

Abstract

This paper aims at presenting the French Higher Education Language Certificate (CLES – Certificat de compétences en Langues de l’Enseignement Supérieur), a task-based language assessment created in 2000 by the Ministry for Higher Education and based on the CEFR. It is currently available in 10 different languages and at three different levels (CLES 1, CLES 2 and CLES 3) corresponding to levels B1, B2 and C1 of the CEFR. The format of the test is that of a scenario which puts test takers in a realistic situation with a specific mission to complete, the completion of the mission requiring students to read texts, listen to documents, write and speak/interact. In addition to the fact that CLES is available to university students for free, it also has a number of advantages over other better-known private language tests which will be presented.

After a brief presentation of the context in which CLES was developed, the theoretical framework, test specifications and national organisation of the certification will be examined. Then, we will discuss the validity, reliability and feasibility of the test, along with aspects that could potentially be improved, some of which as a direct consequence of the CEFR descriptors. Finally, the positive washback effect of the test on language teaching at university level – which far outweighs any of the issues raised – will be dealt with.

References


A Core Curriculum Inventory for General English

Abstract

What should I teach and assess from A1? When should I introduce the present perfect and when should I test it? How can I improve classroom assessment? The British Council and EAQUALS joined together to answer these questions by creating a core curriculum inventory for CEFR levels A1 to C1 for English. It includes grammar, lexis, functions and notions and topic areas. Alongside the curriculum a series of tasks have been developed for teachers to use for classroom assessment.

My talk discusses the British Council/EAQUALS (The European Association for Quality Language Services) core curriculum inventory. I will begin with a description of the curriculum and the methodology used to create it. Then, I will describe the status of the curriculum and how it can be used by English language practitioners. My talk will conclude with recommendations for practice and areas for further research. Below, I give an outline of each stage of the talk.

The core curriculum inventory represents the core of English language taught at Common European Framework of Reference (CEFR) levels A1 to C1. It includes grammar, lexis, discourse markers, functions and notions.

A number of sources were drawn on, including an analysis of the language implied by CEFR descriptors, an analysis of content common to CEFR-based language school syllabi and popular coursebooks, and a teacher survey.

Each language point appears at the level(s) at which it is considered of most relevance to the learner in the class room. The inventory is a documentation of current best practice that can be used in conjunction with databases of learner language like the forthcoming English Profile. The core curriculum will assist teachers in planning classroom assessment. It has been used to inform the development of language tests.

References

Abstract

The linguistic model underlying the CEFR is a functional notional approach additionally employing descriptive theory to scale what a language learner/user can do with a language and with which assessors can assess performance and competence levels. However, the CEFR ‘can do’ performance statements refer to conspicuous searching for words (General linguistic range), incorrect word choice (Vocabulary control) or lexical gaps (Vocabulary range), implying an approach that focuses on isolated items of language and regards grammar and vocabulary as dichotomous entities rather than an approach to language description that has been the subject of more recent studies in the field. At present, descriptors beyond the basic user level (isolated phrases A1 VR; memorised phrases and formulae A2 GLR) lack any reference to multiword items (Moon 1997), formulaic language (Wray, 2002) or formulaic sequences (Schmitt 2004). The ubiquitous nature of these lexical items has been estimated to cover over 52% of written English and over 58% of spoken language (Erman & Warren, 2000) yet the performance statements as they stand do not cater for this frequency and prevalence which Wray views as a dynamic response to the demands of language use. She argues that, “recognizing the role of formulaicity is fundamental to the understanding of the freedom and constraints of language as a formal and functional system,” (2002.5).

This practice-related paper will outline how CEFR performance descriptors have been used to support the task development of a national school leaving exam across the first and second foreign languages. It will describe how and where the descriptors have been constructive in guiding training and task development across three languages and outline attempts and the rationale behind incorporating formulaicity into both test construct and task development. Results from trials and live data will be presented.

References


Language Testing in Europe: Time for a New Framework?

Maria Stathopoulou
University of Athens, Athens, Greece
mastathop@enl.uoa.gr

Investigating Mediation as Translanguaging Practice in a Testing Context: Towards the Development of Levelled Mediation Descriptors

Abstract

This paper reports on findings of a longitudinal research project exploring the complex nature of interlinguistic mediation—a communicative undertaking which entails purposeful selection of information by the mediator from a source text in one language and relaying this information into another language, with the intention of bridging the communication gap between interlocutors. Although in today’s multilingual contexts, it is essential for individuals to have acquired the skills and strategies that will enable them to use two or more languages in a parallel fashion (an ability foreseen by the Common European Framework of Reference for Languages), mediation as translanguaging practice has not received much attention, probably for reasons related to the monolingual paradigm in mainstream language teaching and testing. Given that the CEFR provides no benchmarked illustrative descriptors relevant to mediation, this research has set out to investigate what counts as successful mediation. Specifically, by drawing data from the KPG English Corpus, which comprises collections of written texts (scripts) produced by users of English who have sat for the Greek national standardized foreign language exams (known as KPG)—the only examination system in Europe which assesses test-takers’ mediation ability—the study identifies successful mediation strategies in scripts of different proficiency levels from different KPG writing test papers over a period of six (6) years. The paper actually presents an inductively and empirically derived Inventory of successful mediation strategies which may contribute to the creation of standardized measures and clear benchmarks for reliable assessment of mediation competence thus complementing the CEFR.

References


Language Testing in Europe: Time for a New Framework?


Abstract

At present, no serious language test can avoid mapping its results to the Common European Framework of Reference for Languages (CEFR). This mapping process initially appears to be straightforward, but all those involved in test development will recognize the many problems that it entails in practice. On the basis of our experiences developing the ITNA (Interuniversity Language Test of Dutch as a Foreign Language), we provide a number of closer insights into the problems faced by test developers and how these might be overcome.

We focus primarily on the speaking part of the ITNA: after briefly introducing the speaking task instructions, we discuss the parameters used to score the candidates’ speaking performance. These parameters are based on the CEFR’s qualitative aspects of spoken language use and the can-do descriptors included in its illustrative scales. We highlight the vagueness of these descriptors and describe our attempts to operationalize them for examiners. We also look at the ‘Grammatical accuracy’ and ‘Phonological control’ scales, comparing the different degrees of proficiency required to achieve each CEFR level. In addition, we discuss the difficulty of establishing a pass/fail threshold, especially with regard to partial mastery, and explain how an experiment with native speakers helped us to establish anchor points for the speaking task.

The second part of the session aims to make these problems and their possible solutions more tangible by placing them within a real working context: we provide audience members with a sample task (in English) and invite them to discover for themselves how theory and practice intersect.
From a Low-Stakes Test to a Higher-Stakes Test

Abstract

In order to put our students in the correct course, whose levels correspond to the CEFR levels, our department uses its own Placement Test (a cloze test with two short essay questions). From experience we know that the results are a fair reflection of the participants’ language skills needed for our courses.

This test is a so-called low stakes test, e.g. there is no pass/fail cut-off point. However, English as the language of instruction is taking a more important place than before and we receive an increasing number of questions from students asking about an assessment of their level of English, for instance because they want to enter a Master’s degree course where English is the language of communication. This means that the stakes are becoming higher and the question is whether we can use our test to determine if a student’s English is good enough to participate in such a programme.

If we want to make higher-stakes claims about our test, we need to ensure that the test is accurate and reliable, irrespective of the test-takers’ background. We are now at the beginning of the project.

References


The e-CLES Project: How to Make a Scenario-Based Certificate Valid, Reliable and Fair?

Abstract

The Higher Education Language Skills Certification (French CLES) is a scenario-based language certification accredited by the French Ministry for Higher Education which is linked to the Common European Framework of Reference for Languages (CEFR). It allows students to obtain a certification to testify their skills in 11 different languages and at three different levels: CLES 1, 2, 3 corresponding respectively to levels B1, B2, C1 of the CEFR. CLES enables five language skills to be assessed: listening, reading, writing, speaking and interacting. As such it belongs to the fourth generation in the history of testing: the action-oriented integrative sociolinguistic approach (cf Spolsky 1981, Reichert, 2011, Tardieu 2013).

This presentation will deal with a submitted Franco-Luxembourgian project aiming at framing the conditions for a valid, reliable and fair CLES. The project will notably involve improving the linking to the CEFR levels: The Dutch CEF Grid¹ as well as the CEF-ESTIM grid² and the expertise in standard setting gained through the Socrates EBAFLS project³ will be used. Key reference materials such as the SurveyLang (Cito, 2011) reports and The Manual (A.L.T.E/Council of Europe, 2011) will be solicited. The expertise of the CLES teams of testers shall be improved through the use of new criterial grids and specific tester trainings. A German and English C-Test (Raatz & Klein-Braley, 1982, Reichert, 2011), as well as additional comprehensive language tests will be used to verify the criterion validity of the CLES. Both reliability and validity of the CLES will be verified before and after the adaptation of the procedures linking the CLES to the CEFR. This will eventually allow the documentation of adequate approaches and methods helping to enhance the validity of language tests such as the CLES with regard to the CEFR.

References


¹ http://www.lancs.ac.uk/fss/projects/grid/
² http://cefestim.ecml.at
³ http://www.cito.com/research_and_development/participation_international_research/ebafls.aspx


The Criterial Power of Accuracy: a Learner Corpus Approach

Abstract

Although its value is undeniable, the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) has been criticised on a number of issues, some fair (e.g. its strong reliance on teacher intuition), some rather unfair (e.g. the underspecificity of its descriptors, which was deliberate and unavoidable). These caveats have been argued to constitute rather “shaky ground” (Husltijn, 2007) for such a major educational document.

To address these issues, Husltijn (2007: 665-666) rightly claims that there is an “urgent need to test empirically the implications of the CEFR using real L2 learners rather than teachers”. We take heed of this need and suggest a learner corpus approach to further specifying the CEFR descriptors for linguistic competence (grammatical accuracy, vocabulary control, orthographic control and punctuation) for learners of L2 English.

The learner data used in this study consist of 223 argumentative learner scripts taken from the International Corpus of Learner English (Granger et al., 2009) (c. 150,000 tokens in total). Each text was submitted to a rigorous rating procedure and assigned to a specific CEFR score, ranging from B1 to C2. Simultaneously to the rating procedure, each script was annotated for errors, i.e. error tagged, following a 40-plus error taxonomy (Dagneaux et al., 2008). Having a CEFR score and an error profile per text meant that it was possible to capture the developmental path followed by each error type in terms of progress, stabilisation and regression (Thewissen, 2013). Carrying out this developmental learner corpus analysis has shed light on a number of error types which constitute potential “criterial features” (Hawkins & Filipović, 2012) for a given CEFR level. Additionally, our analysis of accuracy development raises a number of questions concerning the actual validity of a six-level proficiency scale as proposed by the CEFR.

References


Abstract

The Estonian National Curriculum stipulates that all students have to study at least two foreign languages. The choice is usually made from among the four most popular foreign languages: English, Russian, German and French. Until 2014, the foreign language examination was not compulsory though it was the most popular choice with most of the school-leavers opting for English. The examinations developed by the National Examination and Qualification Centre were supposed to be at level B2 roughly. The introduction of a new National Curriculum in 2012 has led to three important changes: 1) from 2014 onwards a school-leaving examination in a foreign language will be one of the three compulsory school-leaving examinations (the other two being the Estonian language and maths); 2) as students can study foreign languages at either B1 or B2 level, they must have a choice as to which of the levels they want to be tested on; 3) the passmark for the national examinations will be abolished – all school-leavers must take the three compulsory examinations, but they can graduate irrespective of the number of points they gain. These changes have meant that the whole examination system will be overhauled. In the case of foreign languages, two separate examination systems will be introduced. No local examinations in Russian, German and French will be developed and students will be expected to take international examinations. However, as the number of students taking English is large, there will still be a locally designed examination in English. A decision has been made to introduce a bi-level examination at levels B1 and B2. The presentation will discuss the reasoning behind these decisions and will look at the measures to be taken to ensure that the locally designed English examination will indeed be measuring students’ English language competence at the designated levels.

References


Abstract

Citing its deliberate context-free stance, the CEFR presents itself as a framework that is useful for all forms of subjective assessment in self-directed learning (Council of Europe, 2001; North, 2008). At the same time the CEFR authors contend that the framework must remain translatable to each and every relevant context. However, the lack of definition in the CEFR descriptors has proven to present serious challenges even to item writers (De Jong & Jones, 2010). In order to safeguard the usefulness of the CEFR for autonomous learning, we must hence find more effective ways of making descriptors readily accessible.

In this paper the authors will introduce and illustrate three practices for increasing the applicability of descriptor scales by bringing the context back into the framework: (1) including illustrative samples that embody distinctive features of different levels of the scales; (2) adding task-specific descriptors alongside the more abstract official ones; and (3) eliciting feedback in the form of annotations that relate assessments to features of the performance proper.

These practices were originally tested in two European Commission-supported projects, WebCEF (2006-2009) and CEFcult (2009-2011), which centered on developing online assessment platforms for oral skills in relation to the CEFR and INCA scales (INCA, 2004; Baten et al, 2013; Van Maele et al, 2013). Illustrations will be provided for two target use domains: the online job screening interview (for business students in Leuven, Belgium) and the oral presentation of doctoral proposals (for researchers in a program of the Flemish Interuniversity Council and a university in Cuba). To demonstrate the transferability of the presented practices, reference will be made to work-in-progress.

The authors will conclude by suggesting that bringing the context back into the framework is a way of honoring the original vision of the CEFR as an on-going exercise in social moderation.

References


De Jong, J. & Jones, G. (April, 2010). Getting the levels right: deriving item writer guidelines from CEFR descriptors. In 32nd Language Testing Research Colloquium, conducted at the University of Cambridge, Cambridge.


Jane Vinther
University of Southern Denmark, Kolding, Denmark
jvinther@sdu.dk

CEFR in a Critical Light

Abstract

This presentation takes a critical stance in relation to the CEFR and the ‘can do’ statements. The principle behind the framework implies that the ‘can-do’ statements are unitarily understandable and can be interpreted in only one way which will be the same for everyone in every European country. This is probably a fallacy. The political dimension of the CEFR is equally debatable. The convergence towards one system for all disregards national differences and different traditions in educational policy and philosophy. The ecological validity of such a system appears tenuous.

The contemplation of a straightforward set of connectors between a given can-do statement and a corresponding testing framework will lead to a myriad of decisions, all of which will lead to cutting corners and some degree of simplification. The very broad and general statements function at a meta level, and the linguistic expressions or components which will warrant a statement such as “I can with most situations likely to arise whilst travelling in an area where the language is spoken” seem opaque. Furthermore, what does it say about a person’s proficiency in relation to the following statement: “I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible”. Which of these two statements express the highest level of proficiency? It depends among other things on what the criteria are and how the statements are interpreted. Additionally, the difficulty is to find appropriate linguistic elements to be tested for. None of the statements lend themselves to obvious linguistic measurements.

This paper presents data which will throw light on the difficulty of interpreting the ‘can-do’ statements at face value, and how difficult even linguists find it to place the statements in a hierarchy.

References


**Abstract**

This paper reports on initial efforts to compile a corpus of course book texts used for teaching CEFR-based courses of Swedish to adult immigrants. The research agenda behind compiling such a corpus comprises the study of normative “input” texts that can reveal a number of facts about what is being taught in terms of explicit grammar, receptive vocabulary, text and sentence readability; as well as build insights into linguistic characteristics of normative texts which can help anticipate learner performance in terms of active vocabulary, grammatical competence, etc. in classroom and testing settings.

The CEFR “can-do” statements are known to offer flexibility in interpreting them for different languages and target groups. However, they are nonspecific and therefore it is difficult to associate different kinds of competences and levels of accuracy learners need in order to perform the communicative tasks with the different CEFR levels. To address this problem a systematic study needs to be performed for each individual language, both for “input” normative texts and “output” learner-produced texts. In this project we take the first step to collect and study normative texts for Swedish.

The article describes the process of corpus compilation, annotation scheme of CEFR-relevant parameters, and methods proposed for text analysis, namely statistic and empiric methods, as well as techniques coming from computational linguistics/machine learning.

**References**


Åström, M. (2012). Språkporten 1, 2, 3 Textbok med webbdel: Svenska som andraspråk. Studentlitteratur AB.

Ying Zheng & John De Jong
Pearson, London, United Kingdom
ying.zheng@pearson.com - john.dejong@pearson.com

Linking to the CEFR: Validation Using Priori and Posteriori Evidence

Abstract

Linking tests to international standards such as the CEFR is a way of establishing criterion-referenced validity. As is widely acknowledged, validation is a continuous process of quality monitoring. In addition to a posteriori validity evidence, a priori validity evidence - such as test design decisions and the evidence that supports these decisions - also makes a significant contribution to the establishment of validity (Schilling, 2004).

This paper reports on how CEFR scales are operationalized in practice in the course of developing an international English test. Measures to link the test to the CEFR have been studied at different stages of test development. The measures include activities that incorporate the use of CEFR scales in item writing, rating scale developing and human rater training.

A posteriori statistical evidence has been collected from both field tests and live tests. Field test data were used to establish the extent to which scores from this test can be linked to the CEFR, which involved both a test taker-centred approach and an item-centred approach. For the test taker-centred approach, test taker responses on five items from three item types were used: Writing essay (one item), Oral description of an image (two items) and Oral summary of a lecture (two items). These responses were rated on the relevant CEFR scales for writing and speaking by two human raters, independently of the ratings produced to score the test. For the item-centred approach, item writers were required to indicate the most appropriate CEFR level of ability for each item. These estimates were compared with the average item difficulty obtained from field tests.

Furthermore, this paper also reports on the ongoing item seeding process, whereby new test items are seeded in live tests and, following analysis of the results, benchmarked to CEFR-referenced item difficulties.

References


POSTER PRESENTATIONS
Are IELTS and CEFR Enough Indicator of Students Success in Academic Study?

Abstract

Language tests and the level achieved in the CEFR are considered to be the two main gatekeepers for international students who aim to study in an English-speaking university. However, gaining the required scores or passing from the A1 institutional level to the C1 level, instead of focusing on what they need to acquire for their academic study, has become the main aim for many students. Turner (2004, p.97) argues that "students seem to want to ‘train’ to reach the appropriate entrance level score or band rather than to engage with the language as an essential, and integral, part of the learning process of their subject of study."

The present study investigates students' perceptions of their experiences with language institutes, preparation for IELTS tests and how these two endeavors affected their academic study in UK universities. It also explores the relationship between the level these students achieved in the institutions according to the CEFR reference, the IELTS band score they attained and their academic performance. The study was based on a questionnaire that was distributed to 173 Saudi participants and on interviews undertaken with six of them.

The results revealed that participants’ main concerns at every stage of their English language study was gaining the required score on the IELTS test, which led to giving more attention to the test and not preparing for their academic studies. Also, it was revealed that some of the participants in the study finished the advanced level (C1) in the institutions, yet they could not achieve higher than a score of 5-5.5 on the IELTS tests, which indicates the (B2) level in the CEFR. This study suggests that Practitioners should take IELTS band scores and CEFR "can do" statements with caution so students will clearly understand that what they achieved in term of scores or levels is not an indication of their readiness for academic study.

References


The Use of Multiple Measures to Detect Academic Vocabulary Improvement

Abstract

The words contained in the academic word list (AWL) are nowadays seen as essential for EFL learners wishing to study at English speaking universities, particularly when taken together with the most commonly occurring 2000 English words (Nation 2006). However, current practice in many pre-sessional Academic English courses does not fully support the explicit teaching of the AWL within the curriculum, though research suggests that direct teaching leads to higher target word pickup rates than implicit teaching (Sonbul and Schmitt 2009). The study reported here aimed to investigate student learning of academic vocabulary in typical pre-sessional and Academic English courses in the in both native/non-native contexts through multiple measures to assess the vocabulary gains.

A total of 103 students studying on two UK pre-sessional programmes, and 77 students studying on an academic English course based in a non-native context participated in the study. Unlike much previous work, this study used multiple measures to assess three levels of vocabulary knowledge: meaning recognition, written form recall and recall of different forms of the AWL (derivatives). In addition to pre and post testing of learner vocabulary, both learners and tutors were surveyed in order to investigate which learning/teaching method contributed to high pickup rates. In the study I will present the results of the analyses of the data undertaken which show interesting variations in vocabulary gain across a number of different sub-groups. The implications of these findings for teaching, learning and assessment in an academic context will be presented and discussed.

References


Measuring Lexical Proficiency in L2 Creative Writing

Abstract

In order to learn how vocabulary changes with progress in language proficiency and which characteristics of the lexical proficiency should be taken into consideration when assessing writing, we studied creative writings collected at official exams of Estonian as a second language (L2) on levels A2, B1, B2 and C1. For each level, we took for analysis 16 works from examinees who spoke Russian as their mother tongue and who had passed the writing assessment with a score of at least 70%. Vocabulary range was measured by comparing the words used with the frequency dictionary, resulting in a lexical frequency profile (LFP). Lexical diversity was measured via Guiraud’s index (G) and the diversity of the sophisticated vocabulary via Advanced Guiraud (AG). The lexical sophistication (LS) and lexical density of the texts (LD) were calculated, as well. A two-sample t test was used to find the distinguishing characteristics of proficiency levels. For each level, all of the characteristics were correlated with the task scores. The outcome showed that LFP is similar as per level—that is, the writings consisted predominantly of frequent tokens and types. G was distinctive for all levels, increasing with language proficiency, but a positive correlation with score was established for B1 writings, only. AG was significantly different between A2/B1, A2/C1, B1/C1, B2/C1, respectively, but not between B2/B1 and B2/A2. AG correlated with scores only for A2 and C1 levels. LS differentiated between A2/C1, correlating also with the scores for these two. LD differentiated between A2/B1 and A2/B2, but did so without correlating with the scores. Assessors noted different lexical aspects in writings. Assessment guidelines must be complemented so that they take into account the features that distinguish between levels and so that they provide additional support for the assessors.

References


Abstract

The presentation is based on two projects that examine the linguistic (particularly grammatical and lexical) features and combinations of features that characterize different Common European Framework (CEFR) levels. The first project was a cross-sectional study of writing in English as a foreign language and Finnish as a second language in Finland (2007-09). The informants were 450 learners, aged 12 to 15, who completed four different functional writing tasks. The second project (2010-13) is a longitudinal study of the development of writing in English, Finnish and Swedish as L2 in Finland. A total of 550 learners, of different ages, complete the same or similar writing tasks each year for three years. The CEFR played a key role in the overall design of these studies as learners’ writing performances were rated against the CEFR scales and against the more fine-tuned Finnish national curriculum scale (NC) based on the CEFR.

We present the results of multifaceted Rasch analyses on the CEFR and NC scales to show how they worked for rating purposes; this is particularly interesting as there are few published studies on the use of unmodified CEFR scales for rating purposes. Research into the quality of the Finnish NC scale is almost completely lacking.

As an example of a qualitative analysis of the two scales we present the findings based on 4300 comments written on the rating forms by the raters when assessing the learners’ texts. We focus on the use of the term ‘simple’ in the CEFR scales and compare the raters’ interpretations of the term with the contexts it is used in the CEFR scales to determine what the dimensions of simple in the CEFR are and in what way they are connected to the issues in the complexity discussion within the study of second language development.

References


Cohmetrix: http://cohmetrix.memphis.edu/

Direkt profil: http://dl.acm.org/citation.cfm?id=1609838


### YOUR PROGRAMME AT A GLANCE

#### Monday 27 May 2013

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00-14:00</td>
<td>Registration at Hof Van Liere</td>
<td></td>
</tr>
<tr>
<td>14:00-14:30</td>
<td>Welcome and interactive opening session</td>
<td>Tassis</td>
</tr>
<tr>
<td>14:30-15:30</td>
<td><strong>Keynote 1: Lyle F. Bachman</strong> How Do Different Intended Uses and Different Views of Language Impact Language Assessment Practice?</td>
<td>Tassis</td>
</tr>
<tr>
<td>15:30-16:00</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>16:00-16:55</td>
<td>I will attend: <strong>Paper sessions</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>17:00-17:45</td>
<td><strong>Keynote 2: Waldemar Martyniuk</strong> The Council of Europe’s Common European Framework of Reference for Languages (CEFR): a 2013 Summary of Developments</td>
<td>Tassis</td>
</tr>
<tr>
<td>17:45-18:30</td>
<td>Presentation of results from the pre-conference survey and welcome drink</td>
<td>Tassis</td>
</tr>
<tr>
<td>19:30</td>
<td>Social activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dinner in the city centre (optional)</td>
<td></td>
</tr>
</tbody>
</table>

#### Tuesday 28 May 2013

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-10:25</td>
<td>I will attend: <strong>Paper sessions</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>10:30-11:05</td>
<td><strong>Keynote 3: Etienne Devaux</strong> Confronting the CEFR to L2 Certification Purposes: Added Value and Methodological Limitations</td>
<td>Tassis</td>
</tr>
<tr>
<td>11:05-11:35</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>11:35-12:30</td>
<td>I will attend: <strong>Paper sessions or poster session</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>12:30-13:55</td>
<td>Lunch</td>
<td>Agora Caffee</td>
</tr>
<tr>
<td>14:00-14:30</td>
<td><strong>Selected paper as plenary session: Glyn Jones</strong> Developing a CEFR-Aligned Test from Scratch: a Case Study</td>
<td>Tassis</td>
</tr>
</tbody>
</table>
### Paper sessions

14:35-16:00  
I will attend:  
1.  
2.  
3.

16:00-16:30  
Coffee break

16:30-17:15  
**Reflection and discussion**  
Tassis

17:20-18:15  
I will attend:  
1.  
2.

19:15  
Conference dinner  
Universiteits-club

---

**Wednesday 29 May 2013**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Room</th>
</tr>
</thead>
</table>
| 09:00-09:45 | **Keynote 4 – Jan H. Hulstijn**  
Natural Tensions between Theory and Practice in the Common European Framework of Reference for Languages (CEFR) | Tassis                |
| 09:50-10:45 | **Paper sessions**  
I will attend:  
1.  
2.  |
| 10:45-11:15 | Coffee break                                                                             |                       |
| 11:15-12:15 | **Post-reflection: Time for a New Framework?**                                            | Tassis                |
| 12:30   | Lunch and farewell drink                                                                  | Agora Caffee          |