

ECML Colloquium December 2017:
Ensuring quality in language testing and assessment:
the contribution of the CEFR

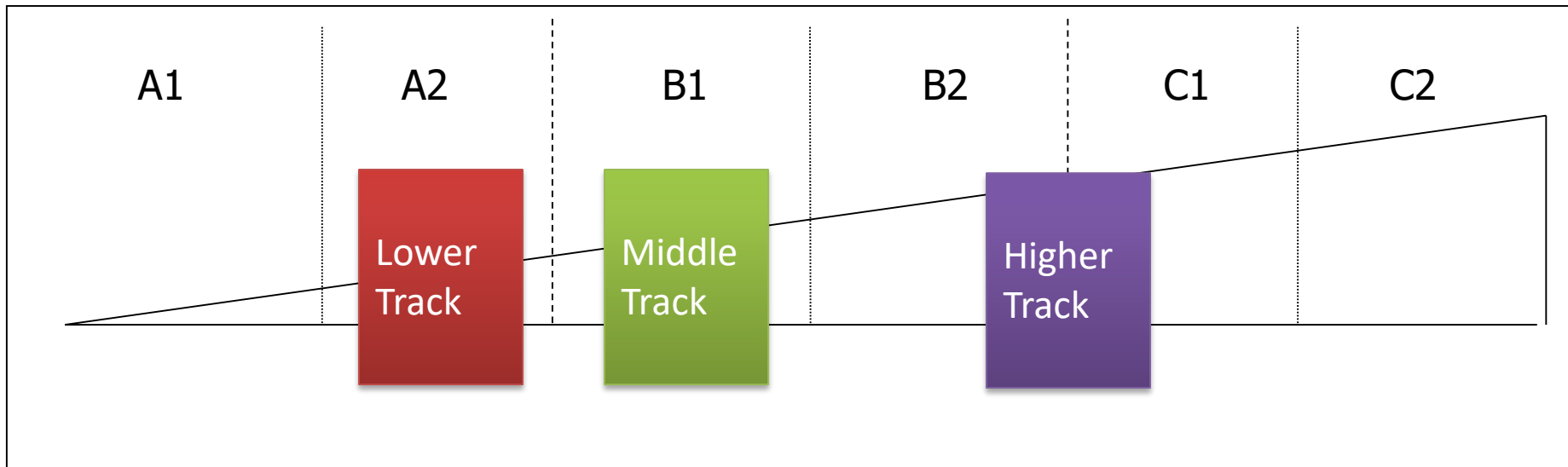
*Standard Setting in the context
of evaluating the German
Educational Standards in EFL*



Educational Standards and the CEFR

- Standards 2003 onwards – output-orientation
- CEFR as ‘model’ for the foreign languages:
Curricula revised after 2001,
Standards based on CEFR
- Differentiation of Standards reflecting
three tiers of school system
 - 2003/2004 Standards for lower/middle track
 - 2012 Standards for higher track
 - Core subjects German, first Foreign Language (E/F),
Maths, Natural Sciences

CEFR and Standards



National Standards: Competence Model

Functional communication competences	
Communicative skills	Application of linguistic resources
<ul style="list-style-type: none">• Listening / visual comprehension• Reading comprehension• Speaking (conversation/ speech)• Writing• Language mediation	<ul style="list-style-type: none">• Vocabulary• Grammar• Pronunciation and intonation• Spelling
Intercultural competences	
<ul style="list-style-type: none">• Socio-cultural orientation knowledge• Awareness of cultural diversity• Behavioural competences in intercultural encounters	
Methodological competences	
<ul style="list-style-type: none">• Text reception (listening-, listening-/ visual- and reading comprehension)• Interaction• Text production (speaking and writing)• Learning strategies• Presentation and use of different media• Awareness and organisation of learning	

Evaluating the Standards



Institut zur Qualitätsentwicklung
im Bildungswesen

- IQB: Institute for Educational Quality Improvement, Humboldt-University Berlin
- Founded 2004 by the KMK – Innovation in Germany
- Commissioned with two strands:
 - standards-based tests to evaluate Educational Standards
 - comparison tests to be used in schools for diagnostic purposes
- Teachers involved in all phases
 - Training as item writers over 3 years
 - Active in the ‚Landesinstitute‘

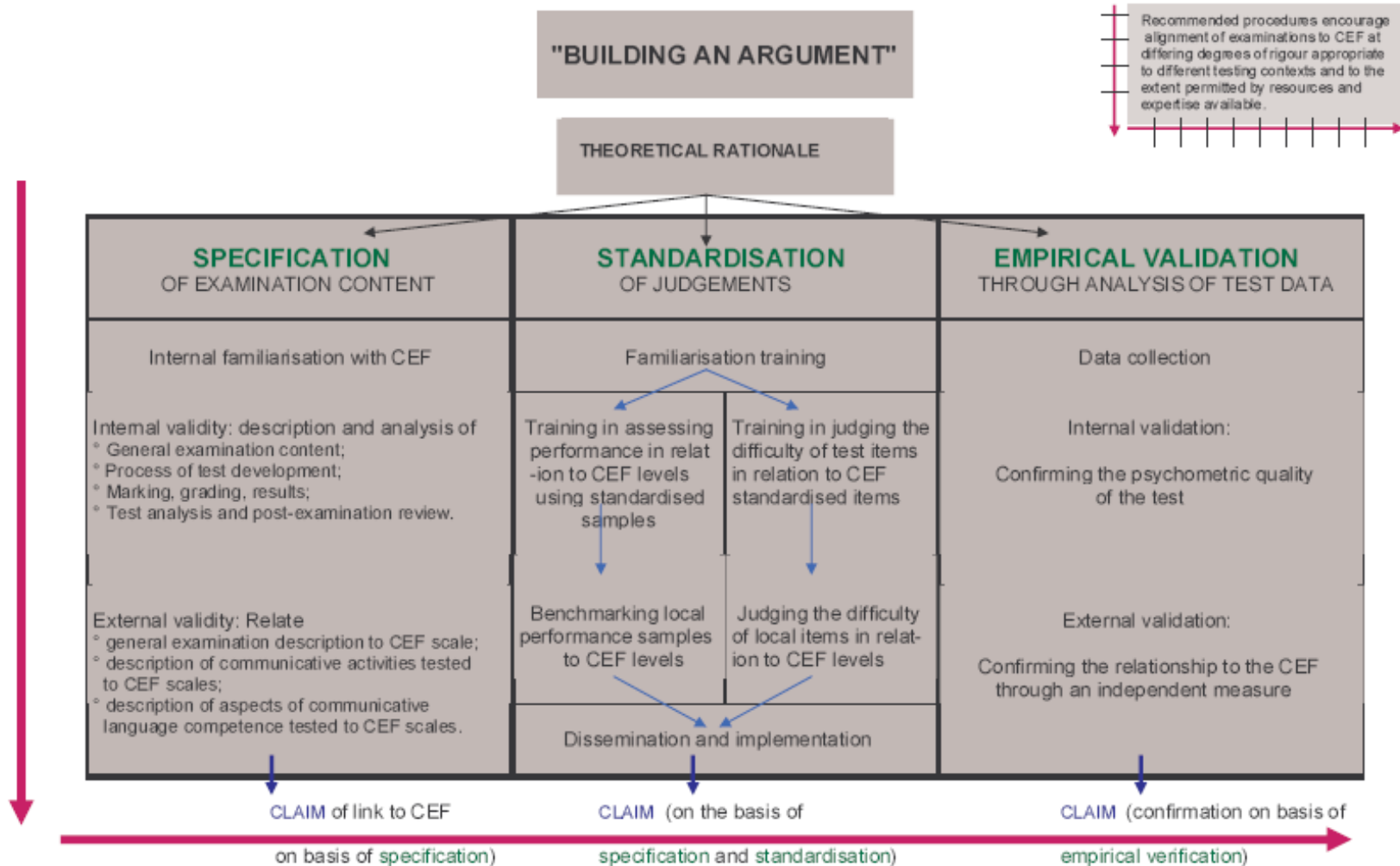
Standards-based Tests for EFL (Rupp et al., 2008)

- Lower and middle track
- Four skills
- Test specs based on CEFR levels A1 to C1
level-specific specs, test items/tasks and rating scales (Harsch & Rupp, 2011; Harsch & Martin, 2012)
- One strand: large-scale monitoring (grades 8/9/10)
Aim: information about population
- Second strand: annual comparison tests (grades 6/8)
Aim: 'diagnostic' information for the classroom
- Both strands linked, formally aligned to CEFR (Rupp et al., 2008; Harsch et al., 2010)
- Advisory Group of international experts

Alignment to the CEFR (Harsch et al. 2010)

Following guidance from Manual (CoE 2009):

FIGURE 1.1: VISUAL REPRESENTATION OF PROCEDURES TO RELATE EXAMINATIONS TO THE CEF



Alignment to the CEFR (Harsch et al. 2010)

Following guidance from Manual (CoE 2009)

During test development

- Familiarisation
- Specification (Dutch Grid, ALTE Grids, item templates)

During Formal Standard Setting Workshop:

- Familiarisation of panellists
- Standardisation for Reading, Listening, Writing

Before / After Workshop

- Empirical Validation

Multi-faceted Rasch Analysis

students	raters	criteria	tasks
	X		
5	X		C1
	X		C1
	X		
4	XX		
	XXX		
	XXX		
3	XXX		B2 B2
	XXXXXXX		
	XXXXXXX		
2	XXXXXXX		
	XXXXXXX		
	XXXXXXXXXX		
	XXXXXXXXXX		
1	XXXXXXXXXX		
	XXXXXXXXXX 6	O	
	XXXXXXXX 1 8 9 10 12	V H	
0	XXXXXXXX 3 5 7 13	G	
	XXXXXXXX 2 4		B1 B1 B1
	XXXXXXXX 11		
-1	XXXXXXXX		A2
	XXXXXX	F	
	XXXXX		
-2	XXX		A2 A1
	XXXX		
	XXXX		A2
-3	XX		
	XX		A2
	XX		
-4	X		A1
	X		
	X		
-5	X		

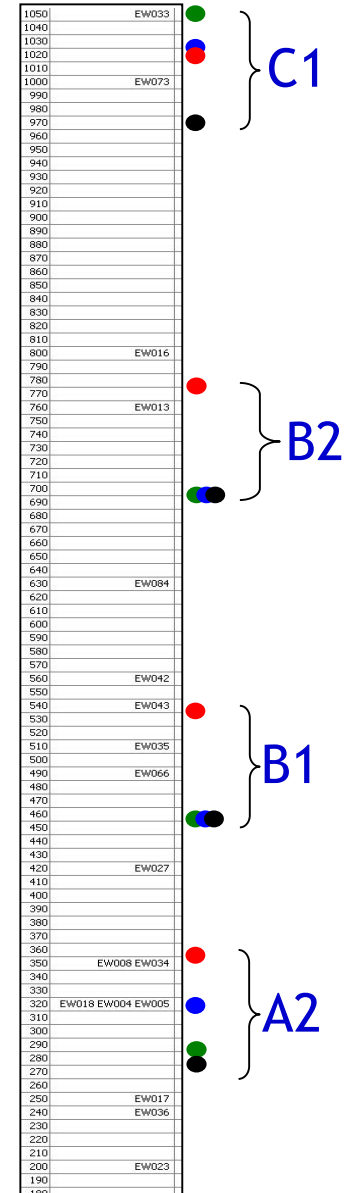
Standard Setting summer 2008



- **Cooperation:**
 - IQB
 - BEAR (Wilson, Berkeley California)
- **Research Project - design:**
 - 2 methods: Bookmark ●● and Criterion Map ●●
 - 2 groups for each method
homogeneous (teachers) ●● / heterogeneous ●●
- **Judgment of panellists:**

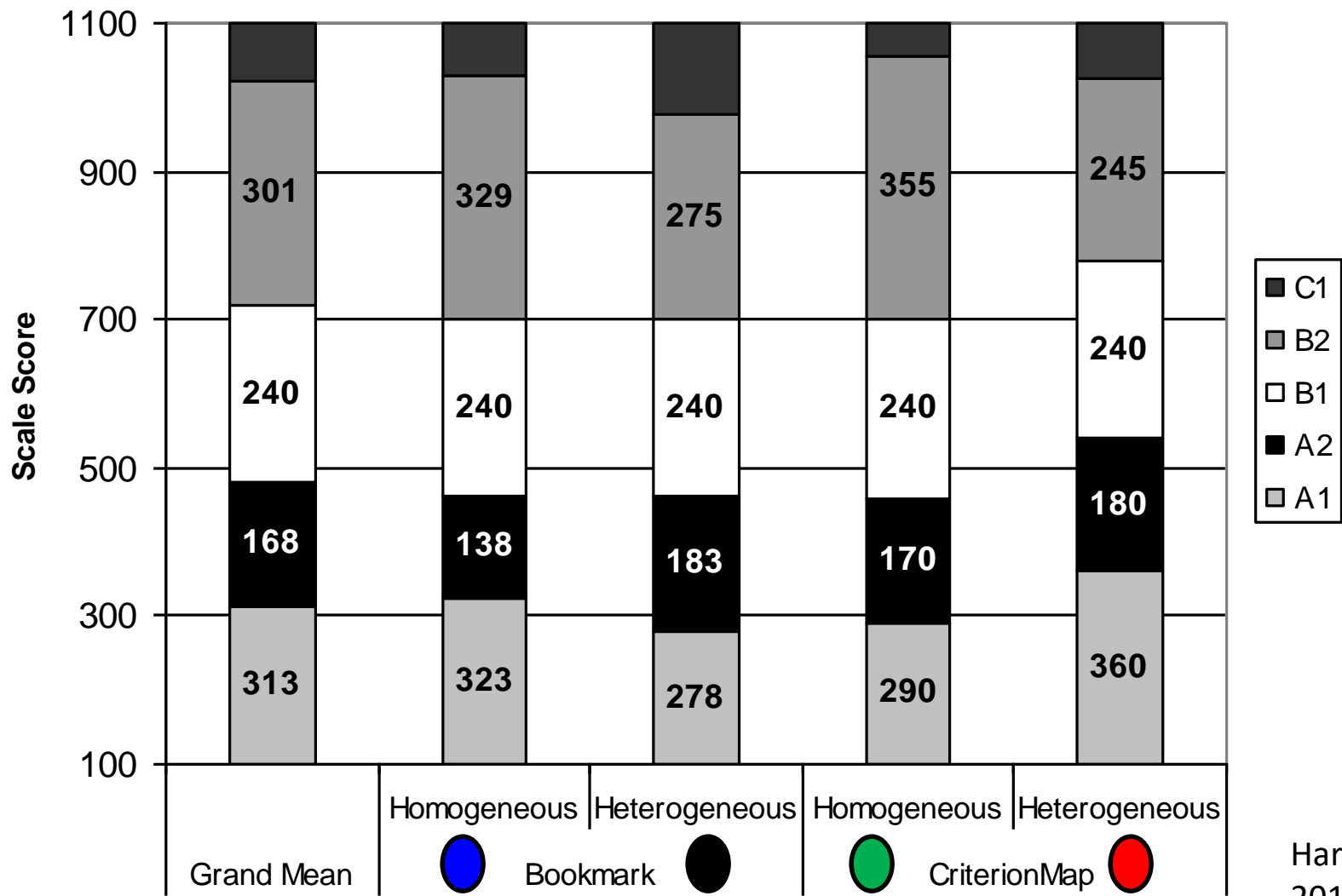
set cut-scores at the task which a learner at the beginning of a CEF level can solve with more than 67% probability
- **Challenge:**

combine recommendations from the 4 groups



Writing Final cut-scores and grand mean

Writing



Conclusions

- Group and method effects
(trends: bookmark more lenient; teacher group in bookmark more lenient for receptive skills)
- => combine cut-scores to grand mean recommendations by averaging across all four panels
- Panels' cut-scores = recommendations, resulting in uneven intervals (proficiency levels)
- => final political decision, also influenced by pragmatic communication needs (defensibility, plausibility for stakeholders):
definition of equidistant proficiency levels

THANK YOU FOR YOUR ATTENTION



Prof. Dr. Claudia Harsch | Research into Language Learning and Teaching | Faculty 10
Director of the Foreign Languages Centre for the Universities in the Land Bremen

www.fb10.uni-bremen.de | www.fremdsprachenzentrum-bremen.de

harsch@uni-bremen.de

References for the Educational Standards

- IQB: www.iqb-hu.berlin.de
- KMK. (2003). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Abschluss*. Darmstadt: Luchterhand.
- KMK. (2004). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Hauptschulabschluss*. Darmstadt: Luchterhand.
- KMK. (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Neuwied: Wolters-Kluwe.
- KMK (2012). *Bildungsstandards für die fortgeführte Fremdsprache (Englisch / Französisch) für die Allgemeine Hochschulreife*.
- All Standards available as pdf at <https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html>, access 28.11.2016.
- IQB 2009: Initial cut-score recommendations, access 28.11.2016
www.iqb.hu-berlin.de/bista/ksm/KSM_SekI_MSA_Eng.pdf
- IQB 2014: Cut-scores 2014, access 28.11.2016
www.iqb.hu-berlin.de/bista/ksm/iKSM_Englisch_Le.pdf

References

- Harsch, C. (2010). Schreibbewertung im Zuge der Normierung der KMK-Bildungsstandards: Der „niveauspezifische Ansatz“ und ausgewählte Schritte zu seiner Validierung. In: Aguado, Vollmer & Schramm (Hrsg.): *Fremdsprachliches Handeln beobachten, messen und evaluieren: Neue methodische Ansätze der Kompetenzforschung und Videographie*. KFU Kolloquium Fremdsprachenunterricht. Frankfurt/Main: Lang, 99-117.
- Harsch, C. & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing* 17, 228–250.
- Harsch, C., Pant, H. A. & Köller, O. (2010). *Calibrating standards-based assessment tasks for English as a first foreign language. Standard-setting procedures in Germany*. Münster: Waxmann.
- Harsch, C. & Rupp, A.A. (2011). Designing and scaling level-specific writing tasks in alignment with the CEFR: a test-centred approach. *Language Assessment Quarterly*, 8(1), 1-34.
- Köller, O., Knigge, M. & Tesch, B. (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.