Aligning tests to proficiency scales

What can we learn from the CEFR?

ILR Testing Committee
March 17, 2023

Mika Hoffman

Michigan Language Assessment

Goals for today

- Explore CEFR-based materials that support aligning tests to proficiency levels
- Discuss how these might or might not be useful for ILR assessment

Note: focus is on the CEFR descriptors and test alignment from a system and structural perspective, not on concepts of proficiency or how the content of the CEFR does or does not align with ILR

CEFR Overview

- Launched in 2001
- A product of the Council of Europe
- Six levels: A1, A2, B1, B2, C1, C2
- Descriptors for **activities**, **strategies**, and **competences**
- Descriptors are not language-specific
- Activities include reception (listening/reading/signing), production (speaking/writing/signing), interaction, and mediation
- Strategies and competences have their own descriptors

COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT

Purpose & Use

- "Provides a metalanguage for discussing the complexity of language proficiency…and for education policy makers to reflect on learning objectives and outcomes…"
- Not intended to serve "a gate-keeping function of assessment instruments"

COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT

Descriptors

Illustrative descriptors

- Not language-specific
- Can-do statements
- Descriptors have been translated into multiple languages
- Developed centrally

Reference Level Descriptions

- Developed for specific languages
- Detailed and specific content (vocabulary, grammar)
- Usually developed locally
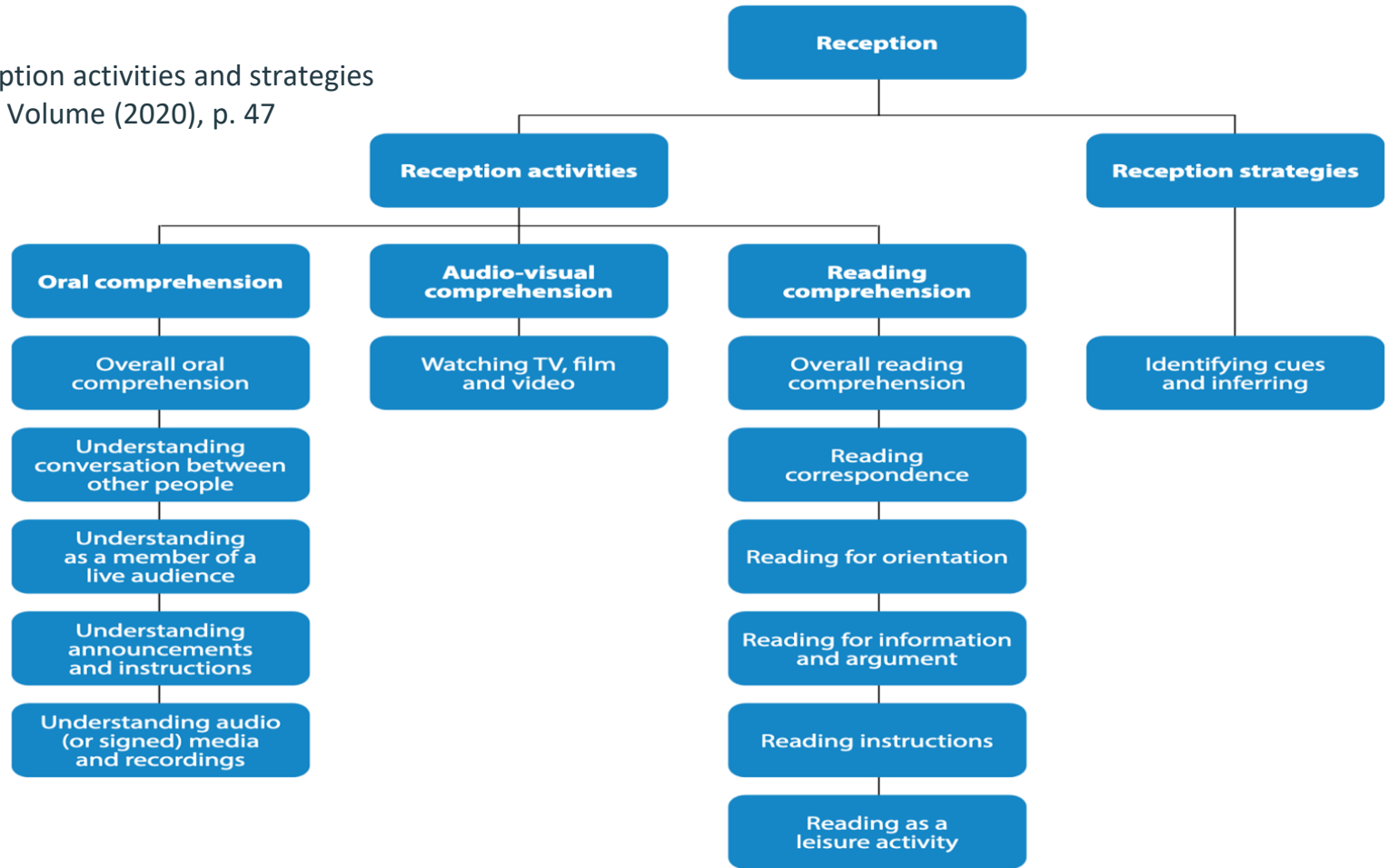- Calibrated to CEFR via centrally determined criteria

COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT

Structure of descriptive scheme
CEFR Companion Volume (2020), p. 32
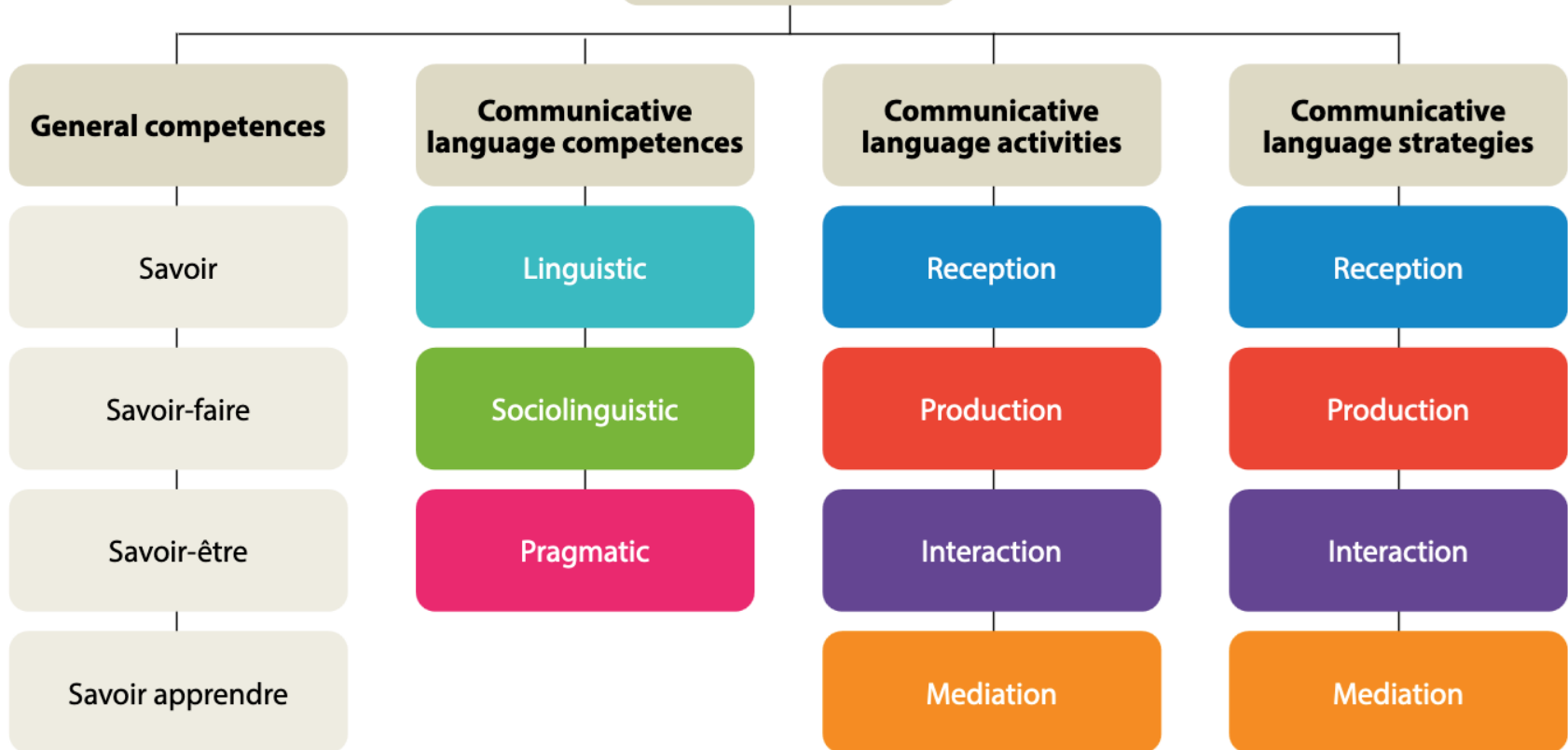
Structure of reception activities and strategies
CEFR Companion Volume (2020), p. 47

	Overall oral comprehension
C2	Can understand with ease virtually any kind of language, whether live or broadcast, delivered at fast natural speed.
C1	Can understand enough to follow extended discourse on abstract and complex topics beyond their own field, though they may need to confirm occasional details, especially if the variety is unfamiliar.
	Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts.
	Can follow extended discourse even when it is not clearly structured and when relationships are only implied and not signalled explicitly.
B2	Can understand standard language or a familiar variety, live or broadcast, on both familiar and unfamiliar topics normally encountered in personal, social, academic or vocational life. Only extreme [auditory/visual] background noise, inadequate discourse structure and/or idiomatic usage influence the ability to understand.
	Can understand the main ideas of propositionally and linguistically complex discourse on both concrete and abstract topics delivered in standard language or a familiar variety, including technical discussions in their field of specialisation.
	Can follow extended discourse and complex lines of argument, provided the topic is reasonably familiar, and the direction of the argument is signposted by explicit markers.
B1	Can understand straightforward factual information about common everyday or job-related topics, identifying both general messages and specific details, provided people articulate clearly in a generally familiar variety.
	Can understand the main points made in clear standard language or a familiar variety on familiar matters regularly encountered at work, school, leisure, etc., including short narratives.
A2	Can understand enough to be able to meet needs of a concrete type, provided people articulate clearly and slowly.
	Can understand phrases and expressions related to areas of most immediate priority (e.g. very basic personal and family information, shopping, local geography, employment), provided people articulate clearly and slowly.
A1	Can follow language which is very slow and carefully articulated, with long pauses for them to assimilate meaning.
	Can recognise concrete information (e.g. places and times) on familiar topics encountered in everyday life, provided it is delivered slowly and clearly.

CEFR Companion Volume (2020), p. 48

COE resources

CEFR (2001) ("The CEFR")

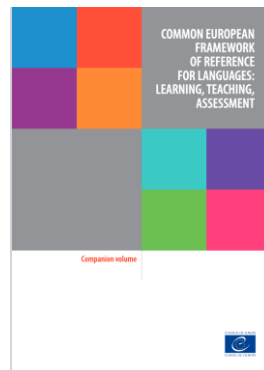https://rm.coe.int/1680459f97

CEFR companion volume (2020)

https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4
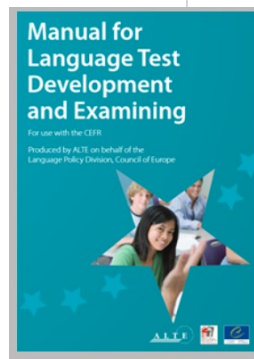
Relating Language Examinations to the CEFR

https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr

Manual for Language Test Developing and Examining

https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2b

Relating Language Examinations to the CEFR

5 stages

- Familiarization
  - Exercises with placing descriptors at the correct level
- Specification
  - Detailed analysis of test tasks and their alignment to CEFR
- Standardization training/benchmarking
  - Illustrative examples
  - "A text does not have a level. The most that can be said about a text is that it is appropriate for inclusion in a test aimed at a particular level." (p. 49)
- Standard-setting
  - Multiple methods considered
  - Little discussion of the information provided by items at specific levels toward the overall cut score
- Empirical validation

Guidelines and Standards

https://www.alte.org/Materials

- Minimum standards
- Guide for item writing
- Code of Practice
- Principles of Good Practice
- Multiple checklists related to the minimum standards
- Can-do statements

http://www.ealta.eu.org/guidelines.htm

- Guidelines for Good Practice

ALTE Q-Mark

- Audit for quality according to the ALTE Minimum Standards
- Focus on the validity argument
- Takes into account purpose and context

Relevance for ILR

Relevant

- Materials specifically addressing alignment to a multilevel proficiency scale
- Grids and checklists
- Sensitivity to different assessment contexts

Not so relevant

- Many more descriptors for CEFR than for ILR
  - Familiarization activities may not translate well to ILR
- Little discussion of potential mismatch between proficiency level and item/task difficulty
- Many procedures are only relevant to large-scale, high-stakes testing

Discussion

Thank you!