Standard Setting

Linking Procedures in the Manual

- Familiarisation
- Specification
- Standardization and Benchmarking
- Standard setting
- Validation

Standard Setting in Indirect Tests

- For tests with numerical score, performance standards to be set
  - Receptive skills (reading, listening)
  - Underlying competences (grammar, vocabulary)
- Performance standard
  - Boundary (cut-off score) between two levels on the scale
- Process to arrive at cut-off score: **standard setting**

Scores, Grades and Standards

- Decision if a person has reached a given CEFR level is based on grading, not on scoring (marking)
- Score must be transferred to a grading scale
- Transformation of scores to grades based on a cut-off score on a test
- Cut-off score
  - border between lowest acceptable score for relevant CEFR-level and highest score to fail that level

How to arrive at Standards?

- Group decisions (panel)
- Group is familiar with CEFR
- Test content specified in terms of the CEFR
- Standard setting procedures formalized
- Careful selection and training of panel members

General Procedures

- Length: 2-3 days (including familiarization)
- 2-3 rounds with in-between
  - information on panel members' behaviour
  - information on candidate behaviour
- Effects of decisions
- Documentation needed to judge validity of procedure

Basket Method

- Basic question: *At what CEFR level can a test taker already answer the following item correctly?*

  - What is the minimum CEFR level required to give a correct response to this item?

- Panelists to put each item in a "basket" corresponding to one of the relevant CEFR levels

- Candidates at higher levels to give correct response

- Correct response not required at lower levels

Conversion of Judgements to Cut-off Scores

- 50 item test
  - 2 items in Basket A1
  - 7 items in Basket A2
  - 12 items in Basket B1
  - 2+7+12 = 21 items to be answered correctly for Level B1 or higher
  - **cut-off score: 21**
- not always necessary to provide baskets for all levels
  - for B1 test: "lower than B1", "B1" and "higher than B1"

Angoff Method

- Basic concept: *minimally acceptable person/borderline person*

- *Borderline person at B1*
  - has the ability to be labelled B1
  - slightest decrease in ability: no longer B1

- Task for panel: keep *borderline person* in mind for all the judgments

- For each item: What is the **probability** that the *borderline person* gives the correct answer ?

Basic Data in Angoff method

Percentage chance correct for borderline person

	Item 1	Item 2	...	Item 49	Item 50	
Rater 1	25%	32%	...	43%	35%	
Rater 2	48%	55%	...	35%	45%	
Rater 3	33%	38%	...	52%	28%	
...	...	...	...	...	...	
Rater 15	72%	80%	...	48%	90%	
Average	**65%**	**72%**	**...**	**45%**	**78%**	**Standard 75%**