RELANG

Relating language examinations to the common European reference levels of language proficiency: *promoting quality assurance in education and facilitating mobility*

Standard Setting

European Centre for Modern Languages and European Commission cooperation on Innovative Methodologies And Assessment In Language Learning

Linking Procedures in the Manual

- Familiarisation with the CEFR
- Linking on the basis of specification of examination content
- Standardization and Benchmarking
- **Standard setting**
- Validation: checking that exam results relate to CEFR levels as intended

Standard Setting in Indirect Tests

- For tests with a numerical score, performance standards have to be set. A performance standard is the boundary (cut-off score) between two levels on the scale reported by a test. A cut-off score of 30 says that a score of 30 or more on the tests grants a level of a particular level (e.g. B1) or higher, while a lower score points to a level lower than the level of the cut-off score (here: B1).

- The process to arrive at a cut-off score is commonly referred to as standard setting. In the case of receptive skills (reading and listening) or underlying competences (grammar, vocabulary), cut-off scores need to be decided upon.

Scores, Grades and Standards

- The decision if a person has reached a given CEFR level is based on grading, not on scoring (marking).
- Therefore a score must be transferred to a grading scale.
- Transformation of scores to grades is based on a cut-off score on a test.
- A cut-off score is the border between the lowest acceptable score on a test to be assigned the relevant CEFR-level and the highest score on that test to fail that level.

How to arrive at Standards?

- Group decisions (panel)
- Group is familiar with CEFR
- Test content has been specified in terms of the CEFR
- Standard setting procedures have been 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 student behaviour
- Effects of decisions
- Documentation is needed to judge validity of the procedure

Basket Method

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

- In other words: what is the minimum CEFR level that is required to give a correct response to each task or question in a test?

- Panellists to put each item in a "basket" corresponding to one of the CEFR levels relevant to the test in question.

- Candidates at higher levels should also be able to give the correct response.

- A correct response cannot reasonably be required at lower levels.

Conversion Judgements to Cut-off Scores

- Suppose that for a 50 item test, two items are placed in Basket A1, seven in Basket A2 and 12 in Basket B1, then it follows that according to this panel member, 2+7+12 = 21 items should be responded to correctly by any one who is at Level B1 or higher. **This number, the minimum requirement, is interpreted as the cut-off score.**

- If a test aims at Level B1, it is not necessary to provide baskets for all levels. The three highest ones could be labelled as "B1", "B2" and "higher than B2".

Modified Angoff Method

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

- *Borderline person at B1* has the ability to be labelled as B1, but only to such an extent that the slightest decrease in that ability would no longer make him a B1 person.

- Task for panel: keep this *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 Tucker-Angoff method

Percentage chance correct for borderline person

	Rater 1	Rater 2	…	Rater 15	
Item 1	25%	32%	…	35%	
Item 2	48%	55%	…	45%	
Item 3	33%	38%	…	28%	
…	…	…	…	…	
Item 49	21%	30%	…	35%	
Item 50	72%	80%	…	90%	
Average	**65%**	**72%**	**…**	**78%**	**Standard 75%**