RELANG

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

Marking, Rating and Grading

European Centre for Modern Languages and European Commission cooperation on INNOVATIVE METHODOLOGIES AND ASSESSMENT IN LANGUAGE LEARNING

The Process

Three steps:

marking, grading and reporting of results

returned test materials → marking → grading → reporting of results

Marking

- Marking is an activity by which marks are assigned to test responses
  - *marker* (less skilled) v *rater* (trained)
  - *clerical* (human) marking
  - *machine* marking

Rating

- Trained judgement necessary
- A single 'correct answer' cannot be prescribed
- More scope for disagreement between judgements
- A greater danger of inconsistency
  - between raters (inter-rater reliability)
  - in work of individual rater (intra-rater reliability)

Rating Scale

- A set of descriptors describing performances at different levels:

  – it shows which mark/grade each performance level should receive

- A rating scale reduces variation in human judgements.

Types of Rating Scales (1)

- **Holistic vs. analytic**
  - a single mark for performance using a single scale describing each level of performance
  - scales for different criteria; mark given for each
- **Relative vs. absolute**
  - scales worded in relative, evaluative terms (e.g. 'poor', 'adequate', 'good')
  - scales define performance levels in positive, definite terms
- **Checklists**
  - marks based on yes/no judgements on specific requirements

Types of Rating Scales (2)

- **Generic vs. task-specific:**
  - a scale or set of scales for all tasks
  - rating criteria specific to each task
  - a combination of both also possible
- **Comparative vs. absolute judgement:**
  - a scale defined through exemplar performances
  - a rater decides performance's relationship to exemplar(s)
  - a mark is a ranking on a scale

Rating Process

- Raters to have a shared understanding of standard (based on examples of performance)
- Small-scale exams
  - raters arrive at a shared understanding through free and equal discussion
- Large-scale exams
  - a stable and meaningful standard
  - experienced examiners communicate standard to newcomers

Rater Training

- A series of steps from more open discussion towards independent rating:
  - guided discussion of samples: markers come to understand the targeted level
  - independent marking of samples
    - comparison with the pre-assigned mark
    - discussion of reasons for discrepancies
  - independent marking of several samples
    - checking how close markers are to pre-assigned marks

Grading

- Tests reporting results in terms of CEFR levels: grading to be *criterion-referenced*
  - performances evaluated with respect to a fixed, absolute criterion or standard
- Exams may be designed to report over several CEFR levels, or just one
  - passing/failing the level
- Identifying the score which corresponds to achieving a level: *standard setting*