



RELANG

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

Testing Receptive Skills Reading and Listening

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



Introduction: Tasks and items

- Test composed of a certain number of *tasks*.
- Tasks composed of
 - instructions
 - texts
 - candidates' answers based on different item types,
- Tasks assessed using *marking grid/rating scale*.
- Test developers to have clear idea about:
 - aim of task
 - reason why it has been chosen
 - skills it refers to

Quality criteria

- Relevance
- Level
- Specificity
- Objectivity
- Acceptability
- Transparency
- Efficiency
- Correct Language use
- Lay-out

Criteria contribute to validity and reliability of test.

Relevance

- Does item address intended knowledge or skill?
- Does item not test other knowledge and abilities (e.g. general intelligence, reading in a listening test, grammar in a reading test)?

Advice

- Use test matrix (test grid)
- Relate questions to purpose of the test
- Make items recognizable for candidate

(CEFR-)Level

- Are text and question indicative of intended (CEFR-) level?
- Does question distinguish those who know from those who do not?
- What type of behaviour is addressed (reproduction, production, etc.)?

Advice

- Make experts work in a team
- Assign screeners
- Avoid manipulation of wording to influence level of difficulty
- Avoid unnecessary information
- Use data analysis (through pretesting and afterwards)

Specificity

- Can the item only be answered by those who have learned for the test?
- Are there no unintended hints (test-wiseness)?

Advice

- Avoid use of extreme words (e.g. always, never)
- Avoid visual characteristics giving unintended information (e.g. long options)
- Be consistent in ordering of options (alphabetical, chronological, size)

Objectivity

- Do different experts agree on correct answer?

Advice

- Use written tests
- Organize discussion among experts
- Organise extensive screening
- Avoid bias (gender, culture, experience, attitudes)
- Do not ask candidates for their opinions
- Develop marking schemes

Acceptability

- Does introduction give sufficient information?
- Is it clear what a candidate must do (instructions)?
- Is item not a catch question?
- Is item not extremely difficult or extremely easy?

Advice

- Use instructions for item writers

Transparency

- Does candidate know how many answers, details or arguments are expected?
- Does item relate to what candidates expect (syllabus, preparatory work)?
- Do candidates know maximum score for each item?
- Is item format known to the candidates?

Advice

- Use clear instructions for candidates
- Use clear terminology in line with syllabus and related tests
- Indicate maximum score for item
- Use item formats or types candidates are familiar with

Efficiency

- Is all relevant information presented in the most efficient way (time needed, introductory information, length or complexity of answer)?

Advice

- Limit stimulus material
- Avoid unnecessary scanning of text by candidates

Language use

- Is language used compatible with candidates' level of language competence?

Advice

- Use short sentences
- Use a limited number of standardised words for questions: *How, When, Why etc.*
- Do not use unnecessary negative or extreme formulations

Lay-out

- Does lay-out support test?

Advice

- Make use of one lay-out format
- Make sure questions and parts of questions are clearly identifiable
- Use clear numbering of questions (1, 2, 3, ... Not: 1a, 1b, 2a ...)
- Use symbols according to rules / conventions
- Make sure all pictures, tables etc. are graphically correct
- Make sure references to pictures, tables, etc. are correct

Constructing Multiple-choice items

Advantages

- Easy to score
- Highly objective
- Low writing load (efficient)
- Higher reliability
- Broad content domain sampling
- Possibility of mechanical scoring
- Possibility of different versions

Disadvantages

- More difficult to test higher and creative skills & competencies
- Poor face validity
- Construction time-consuming and expensive
- No possibility for partially correct answers
- Negative backwash effect on teaching
- Demands reading skills
- Risk of guessing

Characteristics of introduction and stem

- The introduction
 - only contains relevant information
 - easy to read
 - concise
- The stem
 - *focuses* on a topic
 - preferably phrased as a question
 - avoids unnecessary searching by the candidate for relevant information

Characteristics of MC options

- plausible and they exclude each other
- grammatically link up with the stem
- put in a logical or an alphabetical order
- do not repeat elements in text
- do not contain words like “always” or “never”
- do not stand out because of length, formulation, echo
- do not include clues to items in the same test
- (preferably) formulated in a positive way

Characteristics of MC distractors

- plausible for those who have not studied the text
- represent essential differences
- exclude each other (and the correct answer)
- in more or less the same format or phrasing as the correct answer

Matching

- bringing together elements from two separate lists
- popular in tests of receptive skills
 - selecting correct phrase to complete unfinished sentences
 - choosing from a list (e.g. a book) to suit particular requirements

Constructing Open-ended Items

Advantages

- Suitable for productive and creative skills & competencies
- Possibility for partially correct answers
- Possibility for several correct answers (!)
- High acceptability

Disadvantages

- Difficult to achieve high reliability in interpretation of correct answers
- Scoring can be inefficient
- Clarity of formulation of question time-consuming
- Time consuming for both candidate and marker

Format of Open-ended Tests

- Test paper (questions, tasks or assignments)
 - Short answer
 - Gap-filling or completion
 - Long answer
 - Essay

- Scoring instructions

Gap-filling

- *Cloze*
 - complete words/phrases deleted from text
 - every *n*th word deleted.
- Other gap-filling tasks:
 - short phrases deleted from text
 - item writer chooses the words to be deleted (rational cloze)
- Candidates may have to
 - fill in missing words (open cloze)
 - choose from a set of options (multiple choice/banked cloze).
- Marking of open cloze may require
 - exact match
 - list of acceptable responses for markers