RELANG

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

Item Construction: Receptive Skills

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

Introduction: Task and item

- A test is composed of a certain number of *tasks*.

- The tasks (particularly with receptive skills) are composed of *instructions, texts* and candidates' *answers* based on *items* of different types, assessed by using a *marking grid/rating scale.*

- Test developers must have a clear idea about:
  - The aim of the task
  - The reason why it has been chosen
  - The skills it refers to

Quality criteria

- Relevance
- Level
- Specificity
- Objectivity
- Acceptability

- Transparency
- Efficiency
- Correct Language-use
- Lay-out

These criteria contribute to the validity and reliability of a test.

Relevance

- Is the item addressing the intended knowledge or skill?
- Does the item <u>not</u> test other knowledge and abilities than the intended ones (e.g. general intelligence, reading in a listening test, grammar in a reading test)?

<u>Advice</u>

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

(CEFR-)Level

- Are text and question indicative of the intended (CEFR-) level?

- Does the question distinguish those who know from those who do not know?

- 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 or 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 the use of extreme words (e.g. always, never)
- Avoid visual characteristics that give unintended information (e.g. long options)
- Be consistent in the ordering of options (alphabetical, chronological, size)
- Do not copy sentences from books

Objectivity

- Do different experts agree on the 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 the introduction give sufficient information?
- Is it clear what a candidate must do (instructions)?
- Is the item not a catch question?
- Is the question not extremely difficult or extremely easy?

Advice

- Use instructions for item writers

Transparency

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

Advice
- Use clear instructions for candidate
- Use clear terminology in line with syllabus and related tests
- Indicate maximum score for an item
- Use item formats or types candidates have been acquainted with

Efficiency

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

<u>Advice</u>

- Limit stimulus material
- Avoid unnecessary scanning of a text by students

Language-use

- Is the language used compatible with the candidates' level of language competence?
- Are interrogative words used consistently?

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

Lay-out

- Does the lay-out support the 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)
- Generally highly 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
- The introduction is easy to read
- The introduction is concise

- The stem *focuses* on a topic
- The stem preferably is phrased as a question
- The stem avoids unnecessary searching by the candidate for relevant information

Characteristics of options in MC-questions

- They are plausible and they exclude each other

- They grammatically link up with the stem

- They are put in a logical or an alphabetical order

- They do <u>not</u> repeat elements in text

- They do <u>not</u> contain words like "always" or "never"

- They do <u>not</u> stand out because of length, formulation, echo

- They do <u>not</u> include clues to items in the same test

- They are (preferably) formulated in a positive way

Construction of MC-distractors

- Distractors must be plausible for those who have not studied the material

- Distractors must represent essential differences

- Distractors should exclude each other (and the correct answer)

- Distractors should be in more or less the same format or phrasing as the correct answer

Matching

- *Matching* is a test task type which involves bringing together elements from two separate lists.

- One kind of matching test consists of selecting the correct phrase to complete each of a number of unfinished sentences.

- A type used in tests of reading comprehension involves choosing from a list something like a holiday or a book to suit a person whose particular requirements are described.

Constructing Open-ended Tests and 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 the question time-consuming

- Time consuming for both candidate and marker

Format of Open-ended Tests

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

- Scoring instructions

Cloze

- *Cloze* is a type of gap-filling task in which complete words/phrases are deleted from a text.

- In a traditional cloze task every *n*th word is deleted.

- Other gap-filling tasks:
  - short phrases are deleted from a text
  - the item writer chooses the words to be deleted (rational cloze)

- Candidates may have to
  - fill in the missing words (open cloze)
  - choose from a set of options (multiple choice or banked cloze).

- Marking of open cloze may require
  - exact word (only the word deleted from the original text is taken as the correct response)
  - acceptable word (a list of acceptable responses is given to markers)

Gap-filling

- *Gap-filling* is any type of item which requires the candidate to insert some written material - letters, numbers, single words, phrases, sentences or paragraphs – into spaces in a text. The response may be supplied by the learner or selected from a et of options.

Sentence transformation

- *Sentence transformation (phrase transformation)* is an item type in which a complete sentence or phrase is given as a prompt, followed by the first one or two words of a second sentence/phrase which expresses the content of the first in a different grammatical form. For example, the first sentence may be active, and the candidate's task is to present the identical content in passive form.

Unclear Open-ended Questions

- Avoid tasks that are not concrete enough:
    - Try to explain why……
    - Can you tell in which way….
    - Can you think why …

- Avoid formulations that may not produce the required response
    - Will the proposed measure result in a positive or negative change

Scoring instructions

The aim of scoring instructions is to increase objectivity in the process of marking and scoring.

Scoring instructions consist of three parts:

- General scoring instructions
- Marking scheme
- Scoring rules

General scoring instructions

They inform the marker/rater how to perform the marking process.

They may tell the marker/rater

- not to write in the candidate's work

- how to prevent *rater effects*

- how many points to give in cases like:

  - answers that are correct but unlike the examples

  - repeated errors

  - computational errors

  - superfluous parts

Scoring instructions

A marking scheme contains model answers and a scoring system, such as :

- maximum number of points per item

- the correct answer + variations, or one or more examples of correct answers (maximum points)

- partially correct answers or aspects + points

Relating scores to answers

Basic principle:

- one point for each relevant element in an answer.

However:

- make sure the weight per item or per cluster of items is relative to its importance or in accordance with the test grid.

Marking schemes: some issues

- What is (not) to be marked: bad drawings / spelling mistakes / bad handwriting

- Knowledge about the candidate (> anonymity)

- Influence of preceding candidates (> mark by question)