**Checklists for test development and assessment**

⦁ **Checking the quality of whole test**

Checking both the presence of the item and its quality.

**Rating**: 3 = very satisfactory. 2 = satisfactory. 1 = unsatisfactory. 0 = item inexistent. NA = not applicable.

|  |  |  |
| --- | --- | --- |
| **Rating** | **N°** | **Item** |
|  |  | Is the test acceptable societally, institutionally, and in the eyes of teachers, parents, and administrators? |
|  |  | To what extent does the test measure what it is intended to measure and provide useful, accurate results? (Validity) |
|  |  | Does the test measure student performance accurately and consistently? (Reliability) |
|  |  | Is the test reasonable and practicable in terms of time and resources? (Manageability) |
|  |  | Are the materials and tasks authentic, meaningful and bias free? (Authenticity and Fairness) |
|  |  | Does the test match the test specifications—text length, questions number, etc.? (Alignment with Specs) |
|  |  | Does the test content match the syllabus content—vocab, gram, writing, etc.? (Alignment with Syllabus) |
|  |  | Is the test appropriate to the overall abilities of the test takers? Does it measure what they *know* and *can do*? |
|  |  | Does the test assess what has been *taught* and *learned* effectively? |
|  |  | Are the suggested materials and tasks accessible to the average students? Is the difficulty of the item appropriate for them? (Accessibility) |
|  |  | Do the tasks avoid tricks (trapping testees) and trivia (unimportant information)? |
|  |  | Is there variety of themes in the Reading, Language, and Writing tasks? (Theme variety) |
|  |  | Are test takers familiar with the topics suggested? (Topic familiarity) |
|  |  | Are knowledge demands minimised and upsetting themes avoided in the test materials? |
|  |  | Are the test items written according to rules of construction for the type(s) chosen? |
|  |  | Does the test involve a balance between knowledge, skills, and attitudes/values? |
|  |  | Are the instructions clear, adequate, and concise? Is language simple, suitable for the students’ reading level, and familiar? (Language clarity and familiarity) |
|  |  | Are test takers familiar with the question types used? (Question familiarity) |
|  |  | Are the test items, as a group, free of overlapping and free of clues within and among them? |
|  |  | Is there a clear objective for each item (task/question)? Is each item format correctly matched to the purpose and content of the item? |
|  |  | Are all correct and possible answers/responses included in the key? |
|  |  | Is the scale for each task/item indicated on the test paper? |
|  |  | Is the test well laid out and legible? Are the items spaced so that they can be read, answered, and scored with the least amount of difficulty? Is the formatting consistent? (Presentation and layout) |
|  |  | Has the test been revised for content validity and language accuracy and appropriacy according to guidelines? (Self / Peer review) |

⦁ **Checking the READING COMPREHENSION Component**

Checking both the presence of the item and its quality.

**Rating**: 3 = very satisfactory. 2 = satisfactory. 1 = unsatisfactory. 0 = item nonexistent. NA = not applicable

|  |  |  |
| --- | --- | --- |
| **Rating** | **N°** | **Item** |
| ⯍ **Checking text appropriateness and quality** | | |
|  |  | Length**1**. Is it appropriate? |
|  |  | Text originality and source**2**. Is the text original? Is the source acknowledged? |
|  |  | Text quality. Does the text represent professional-quality writing? |
|  |  | Text adaptation. If it is adapted, has it been made accessible for all students appropriately? |
|  |  | Text content relevance, validity, and suitability. Is the text content interesting, useful, meaningful, and suitable? |
|  |  | Genre**3** familiarity |
|  |  | Type**4** familiarity |
|  |  | Theme relation to syllabus |
|  |  | Topic familiarity |
|  |  | Topicality of issues. Are the text issues recent, up-to-date, and of interest currently? |
|  |  | Readability of the text, measured by computer software |
|  |  | Meaning (single or multiple levels; **ideas**: simple/concrete or complex/abstract; **theme**: obvious or implicit) |
|  |  | Purpose (explicit or implicit; clear or intricate) |
|  |  | Organization (simple or complex; explicit or implicit; conventional or unconventional) |
|  |  | Language style (literal or complex; conversational or unfamiliar) |
|  |  | Vocabulary (clear, contemporary or figurative, academic) |
|  |  | Sentence structure (simple, compound, or complex) |
|  |  | Knowledge Demands (common and everyday, moderate, or extensive) |
|  |  | Match between text and students’ reading skills and cognitive capabilities. |
|  |  | Match between text and students’ prior knowledge, experience, motivation, engagement, and interest. |
|  |  | Match between text and tasks/questions. Is the text appropriate for assessing a range of reading skills? Is it suitable for a good exploitation (generating the required number of items)? |
| ⯍ **Checking the comprehension questions relevance and quality** | | |
|  |  | Tasks/Questions accessibility |
|  |  | Instructions. Are they clear, adequate and in conformity with the recommendations? |
|  |  | Language used**5**. Is it accurate and familiar? |
|  |  | Do the questions avoid focusing on one or two paragraphs? |
|  |  | Do the questions have answers within the text**6**? |
|  |  | Are the questions free of clues within and among them? |
|  |  | Is the number of questions as recommended? (Basic Ed.: 4–5 Qs; Secondary: 6–7) |
|  |  | Variety of tasks/question types |
|  |  | Familiarity with tasks/question types |
|  |  | Variety of question levels. Is there a balance of literal & higher-order items**7**? Are they familiar? |
|  |  | Are items requiring a variety of possible answers avoided? (Except reaction to text)**8** |
|  |  | Do the responses make minimal demands on writing ability? |
|  |  | Completing a paragraph with words from the text**9**: Is the related paragraph number indicated? Are students instructed to write the words without making any changes to them? |
|  |  | ‘Circle the adjectives/adverbs which apply to…’: Are the suggested items familiar to testees?**10** |
|  |  | False/Justify item type—Instruction**11**: “For each of the following statements, pick out one detail from the text showing that it is false”.  Is the number of the paragraph containing the detail indicated? |
|  |  | **Multiple-Choice Items** |
|  |  | Instruction: Tick the “most appropriate” – not “correct” – option. |
|  |  | Is most information put in the stem, without including unnecessary material? |
|  |  | Are all alternatives the same length, parallel and brief, without repeated material? |
|  |  | Are distractors plausible? |
|  |  | **True-False Items** |
|  |  | Instruction: ‘Circle’ – not ‘Write’ – ‘True or False’. |
|  |  | Have you avoided lifting a statement from the text and modifying it slightly to create an item? |
|  |  | Have you worded the statement so carefully that it can be judged clearly true or false? |
|  |  | Is the statement relatively short, without unnecessary material, and positively stated? |
| ⯍ **Checking the scoring criteria and answer key quality** | | |
|  |  | Is the number of expected answers equal to the number of allotted marks? |
|  |  | Are all correct and possible answers/responses included in the key? (Relevant as well as irrelevant answers/responses should be specified.) |
|  |  | Are badly-constructed answers tolerated? (As long as information is not obscured.) |
|  |  | Is the mark scheme indicated on the test paper? |
| ⯍ **Checking the whole component** | | |
|  |  | Appropriateness and quality of text and questions |
|  |  | Item writing conformity to rules and guidelines of the development of the item types used |
|  |  | Manageability in terms of difficulty and time |
|  |  | Avoiding tricks and trivia |
|  |  | Language and instructions clarity, appropriacy, relevance, and familiarity |
|  |  | Presentation and layout appropriateness: formatting, answer space, on same page, etc. |
|  |  | Revision and proofreading |
|  | 🡪 **Overall appropriateness, quality, and construction of the component** | |

**Notes**

1. How many extra words to tolerate beyond the required length? 10% of the required length may be tolerable, especially if the text is not ‘complex’.

The text cannot be taken from extra-curricular, commercial books, previous exams, etc. It should not be an approximation to texts examinees have read in class.

*Genre*: Article, letter, report, speech, review, story, biography, etc.

*Type*: Narrative, expository, persuasive, argumentative, descriptive.

Check to avoid double barrelled, loaded, leading, ambiguous, or biased items; avoid including superfluous information.

Make sure all items can be answered correctly only by understanding the text and referring to it, e.g. without using prior knowledge or common sense. Is all information necessary for comprehension in the passage itself?

1. You can code the questions as asking for either higher-level or lower-level ideas from the text.
2. Note that for the reaction-to-text item type, the student response should be reasonably worthy of the mark allotted, e.g. it should not lift expressions or ideas literally from the text or be trivial, etc.
3. The instructions should specify the policy towards misspelled responses.

▪ The words should be part of the Textbook wordlists; see Appendix 4 for a word list for each level, from Year 6 Basic Ed. to Year 4 Secondary Ed. Bear in mind that for Year 4 Secondary Ed., a number of lessons have been omitted as indicated in Appendix 5.

▪ Care should be taken when marking this item type – if the question asks for two items: for an answer providing: • 2 correct items plus one incorrect item 🡪 1 mark is allotted rather than 2. • 2 correct items plus 2 incorrect items 🡪 no credit. • 1 correct item plus 1 incorrect item 🡪 1 mark. • 1 correct item plus 2 incorrect items 🡪 no credit...

▪ A similar marking issue may arise when marking a test item that asks examinees to pick out from the text, say *three* details and write each detail in one table cell, e.g. causes: (a) ... | (b) ... | (c) ... : for an answer providing: • all 3 correct items written in one cell (e.g. cell (a)) 🡪 full credit (3 marks). • 3 correct items plus one incorrect item all written in one cell 🡪 partial credit (deduce 1 mark from total mark: 2 marks instead of 3). • 3 correct items written in one cell plus one incorrect item written in another cell 🡪 partial credit (deduce 1 mark from total mark: 2 marks instead of 3).

1. The instruction should specify that the detail ranges from one word to one full sentence, but no more, and that it should be written entirely, without using ellipsis (…….).

⦁ **Checking the LANGUAGE Component**

**Rating**: 3 = very satisfactory. 2 = satisfactory. 1 = unsatisfactory. 0 = item absent entirely. NA = not applicable

|  |  |  |
| --- | --- | --- |
| **Rating** | **N°** | **Item** |
| ⯍ **Fill in Task** | | |
|  |  | Do the blanks include a balance of content words and function words? |
|  |  | Do the options include a balance of content words and function words? |
|  |  | Do the options include two plausible distractors? |
|  |  | Do the options appear in the textbook wordlist? |
|  |  | Do the options precede the gapped paragraph, in a box, separated by slashes? |
|  |  | Are the blanks numbered? |
|  |  | Are the blanks represented by lines (\_\_\_) of equal length rather than by dots (......)? |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple and clear—accessible to all students?* |
|  |  | *Is the task material authentic? Is the task contextualized?* |
|  |  | *Is the first sentence of the paragraph integral, providing enough context?* |
|  |  | *Are the items well spaced out (evenly spread)?* |
|  |  | *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there only one key for each item? Are misspelt answers penalized?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, with adequate spaces for answers, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |
| ⯍ **Tense/Form Task** | | |
|  |  | Is the instruction precise? (*Put the bracketed words in the appropriate tense/form.*) |
|  |  | Are the bracketed verbs supplied in the base form? |
|  |  | Are students familiar with the bracketed words? |
|  |  | Are students familiar with the tenses/forms required? |
|  |  | Is there a balance between tense and form in the task? |
|  |  | Are items requiring a double task, e.g. adding a prefix & suffix (able 🡪 disabled), avoided? |
|  |  | Are all the bracketed words in **bold type**? |
|  |  | Are the blanks represented by lines (\_\_\_) of equal length rather than by dots (......)? |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple and clear—accessible to all students?* |
|  |  | *Is the task material authentic? Is the task contextualized?* |
|  |  | *Is the first sentence of the paragraph integral, providing enough context?* |
|  |  | *Are the items well spaced out (evenly spread)?* |
|  |  | *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there only one key for each item? Are misspelt answers penalized?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, with adequate spaces for answers, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |
| ⯍ **Multiple-Choice Task** | | |
|  |  | Is the instruction precise? (*Circle the “correct” option.*) |
|  |  | Do options in the same item share the same part of speech – 3 verbs, 3 preps, etc.? |
|  |  | Does the item test one thing at a time—e.g. tense OR form? |
|  |  | Is there a balance between content words and function words? |
|  |  | Are distractors plausible but clearly incorrect? |
|  |  | Do distractors include only correct forms and vocabulary? |
|  |  | Are all the bracketed options in **bold type**? |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple and clear—accessible to all students?* |
|  |  | *Is the task material authentic? Is the task contextualized?* |
|  |  | *Is the first sentence of the paragraph integral, providing enough context?* |
|  |  | *Are the items well spaced out (evenly spread)?* |
|  |  | *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there only one key for each item? Does the key position vary randomly?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |
| ⯍ **Matching Task** | | |
|  |  | Are there more response options than premises? |
|  |  | Are premises and options rather short in order to reduce the amount of reading? |
|  |  | Are premises longer than options? |
|  |  | Are the premises in the right order? |
|  |  | Are the lists of options and premises related to one central theme, familiar to testees? |
|  |  | Have the task sentences been divided into sentence parts only once?  Are the premises numbered (1, 2, 3...) and the options identified by letters (a, b, c...)? |
|  |  | Are the instructions appropriate? (*Match each sentence part in Column A with its corresponding part in Column B to get a coherent paragraph. The sentence parts in Column A are in the correct order. Use each part in Column B only once. Write your answers in the box.*) *Answers*:1 + ... | 2 + ... | 3 + ... | 4 + ... | 5 + ... |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple and clear—accessible to all students?* |
|  |  | *Is the task material authentic? Is the task contextualized?* |
|  |  | *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there only one key for each item?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |

⦁ **Checking the WRITING Component**

**Rating**: 3 = very satisfactory. 2 = satisfactory. 1 = unsatisfactory. 0 = item absent entirely. NA = not applicable

|  |  |  |
| --- | --- | --- |
| **Free writing** | | |
|  |  | Is the topic syllabus related? |
|  |  | Does the topic consider the educational level and background of the test takers? |
|  |  | Is the task clear and focused? Is it sufficiently narrow/broad for the time available? |
|  |  | Does the prompt clearly specify the communicative situation/context? |
|  |  | Does the prompt clearly specify the purpose to write? |
|  |  | Does the prompt clearly specify the audience? |
|  |  | Does the prompt clearly specify the genre/text type? |
|  |  | Does the prompt clearly specify the response length? |
|  |  | Does the topic interest and motivate test takers to write? Is it suitable for expressing personal attitudes and for creative writing? Does it prevent rote learning? |
|  |  | Does the task stimulate using one’s ideas and reinvesting personal knowledge & language resources? |
|  |  | Is scaffolding avoided? (Helpful notes and hints should not be provided.) |
|  |  | If quotes are used, are they relevant and appropriate? |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple, accurate, and suitable for the students’ reading level?* |
|  |  | *Does the task avoid tricks and trivia?* |
|  |  | *Is the task well laid out, with adequate space for the response, on the same page?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Are the evaluation criteria indicated on the test paper (adherence to task and content adequacy, lexical appropriacy, etc.)?* |
|  |  | *Is there in the answer key a model/sample response or an outline of the major points that should be included? (Clarify what is required in the response as well as what is irrelevant in it.)* |
|  |  | *Have you left wide enough margins for your comments, points, etc.?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |
| **Guided Writing** | | |
| ⯍ **Information transfer** | | |
|  |  | Is the examinee’s task clearly defined? |
|  |  | Is the data provided relevant and appropriate? |
|  |  | Is overloading tables/biography data with many details and notes avoided? (Suggest a reasonable amount of information to be developed in a paragraph written in the specified and provided number of lines.) |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?*  *Is language simple, accurate, and suitable for the students’ reading level?*  *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there a written model/sample response included in the answer key?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, with adequate spaces for answers, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |
| ⯍ **Developing notes into sentences** | | |
|  |  | Are the instructions appropriate and clear? (Develop full sentences from the following notes to get a coherent paragraph; supply the correct grammatical form, including changes in word form and addition of such elements as prepositions, articles, auxiliary verbs, etc.) |
|  |  | Are the notes made up of chunks/series of words rather than single words? |
|  |  | Are the notes in the right order? |
|  |  | Are there slashes that indicate where an addition or change should be made? |
|  |  | *Is the task manageable and appropriate in terms of difficulty and time?* |
|  |  | *Is language simple, accurate, and suitable for the students’ reading level?* |
|  |  | *Is the task material authentic? Is the task contextualized?* |
|  |  | *Does the task avoid tricks and trivia? Is it clue free?* |
|  |  | *Is there a written model/sample response included in the answer key?* |
|  |  | *Is the mark scheme indicated on the test paper?* |
|  |  | *Is the task well laid out, with adequate spaces for answers, on the same page?* |
|  |  | *Has the task been revised and proofread?* |
|  | 🡪 **Overall appropriateness, quality, and construction of the task** | |

**On a scale of 1 to 10, how would you rate the overall quality of the test? 🡪 ....... / 10**