Training Terminology – Toolkit

Criterion-Referenced Test

Step by Step

1. **Analyze learning content**

   Creating the content hierarchy is one of the most important steps in a course planning process. One of the approaches mainly used to validate a hierarchy is the Bloom's Taxonomy.

   Bloom's cognitive classification consists of six levels as follows:

   1) *Knowledge*: recalling or recognizing facts

   2) *Comprehension*: Understanding relationships and meanings

   3) *Application*: Being able to appropriately apply solutions to familiar situations

   4) *Analysis*: Being able to separate a situation from its constituent part

   5) *Synthesis*: Being able to combine elements to form a whole

   6) *Evaluation*: Having the ability to appraise the value of something – a subjective ability

   The cognitive level needs to be defined before starting to create CRTs.

2. **Create Measurable Learning Objectives**

   See Learning Objectives toolkit.

3. **Create cognitive items**

   A test poses a question and either provides possible answers or distracters (called closed-ended items) or allows for a free-form response (open-ended items).

   - *Closed-ended questions*: They are often used when you are looking for a response that would have a predictable correct answer. Common item types are: multiple-choice, true/false, matching. They are easily scored by a computer and will provide the best chance of establishing an acceptable level of reliability for a test.

unitar
United Nations Institute for Training and Research

- **Open-ended questions**: They are used when you are gathering information or looking for a response that is not easily classified as right or wrong. Common types of open-ended questions are: essay, short answer, fill-in, numeric.

  According to the cognitive level defined for CRTs, certain types of cognitive items are more appropriate for creating the tests.

Level	Item Types
1. **Knowledge**: Recalling or recognizing facts	• True/ false • Matching • Multiple-Choice • Fill-In • Short Answer • Essay
2. **Comprehension**: Understanding relationships and meanings	• True/ false • Matching • Multiple-Choice • Fill-In • Short Answer • Essay
3. **Application**: Being able to appropriately apply solutions to familiar situations	• True/ false • Multiple-Choice • Fill-In • Short Answer • Essay
4. **Analysis**: Being able to separate a situation into its constituent parts	• Multiple-Choice • Short Answer • Essay
5. **Synthesis**: Being able to combine elements to form a whole	• Short Answer • Essay
6. **Evaluation**: Having the ability to appraise the value of something – a subjective ability	• Essay

unitar
United Nations Institute for Training and Research

Guidelines for Writing Test Items

General guidelines

- Monetary references should be applicable to all potential learners (for instance, Euro, USD, etc.)

- Personal information should always be fictional, although references to Social Security, personal IDs, telephone numbers, mail codes, and so forth should be in the format applicable to the job performance assessed by the test.

- Avoid slang, idioms, and abbreviations. (for example, use "that is" instead if "i. e.")

- If you expect a cross-cultural use of a test written in English, plan to write the items at a reading level appropriate to those for whom English is a second language.

- If you use names, select them to reflect the cultural norms of the country where the test will be given.

Guidelines for writing True/ False items

- Use true/ false items in situations where there are only two likely alternative answers, that is, when the content covered by the question is dichotomous.

- Include only one major idea in each item.

- Make sure that the statement can be judged reasonably true or false.

- Keep statements as short and as simply stated as possible.

- Avoid negatives, especially double negatives; highlight negative words such as not, no, none if they are essential.

- Attribute any statement of opinion to its source.

- Randomly distribute both true and false statements.

- Avoid specific determiners (such as always or never) in the statements.

Guidelines for writing Matching items

- Include only homogeneous, closely related content in the lists to be matched.

- Keep the lists of responses short --- five to fifteen entries.

- Arrange the response list in some logical order, for example, chronologically or alphabetically.

- Clearly indicate in the directions the basis on which entries are to be matched.

- Indicate in the directions how often a response can be used; responses should be used more than once to reduce cueing due to the process of elimination.

unitar

United Nations Institute for Training and Research

- Use a larger number of responses than entries to be matched in order to reduce process of elimination cueing.
- Place the list of entries to be matched and the list of responses on the same page.

Guidelines for writing Multiple-Choice items

For writing the stem:

- Write the stem using the simplest and clearest language possible to avoid making the test a measure of reading ability.
- Place as much wording as possible in the stem, rather than in the alternative answers; avoid redundant wording in the alternatives.
- If possible, state the stem in a positive form.
- Highlight negative words (no, not, none) if they are essential.

For writing the distractors:

- Provide three or four alternative answers, in addition to the correct response.
- Make certain you can defend the intended correct answer as clearly the best alternative.
- Make all alternatives grammatically consistent with the stem of the item to avoid cueing the correct answer.
- Vary randomly the position of the correct answer.
- Vary the relative length of the correct answer; don't allow the correct answer to be consistently longer (or shorter) than the distractors.
- Avoid specific determiners (all, always, never) in distractors.
- Use incorrect paraphrases as distractors.
- Use familiar looking or verbatim statements that are incorrect answers to the question as distractors.
- Use true statements that do not answer the question as distractors.
- Use common errors that learners make in developing distractors; anticipate the options that will appeal to the unprepared learner.
- Use irrelevant technical jargon in distractors.
- Avoid the use of "All of the above" as an alternative; learners who recognize two choices as correct will realize that the answer must be "all of the above" without even considering the fourth or fifth alternatives.
- Use "Non of the above" with caution; make sure it is the correct answer about one third to one-fourth of the times it appears.

- Avoid alternatives of the type "both a and b are correct" or "a, b, and c but not d are correct"; such items tend to test a specific ability called syllogistic reasoning as well as the content pertinent to the item.

- Items with different numbers of options can appear on the same test.

- If there is a logical order to options, use it in listing them; for example, if the options are numbers, list them in ascending or descending order.

- Check the items to ensure that the options or answer to one item do not cue learners to the correct answers of other items.

Organizing the distractors:

Pattern 1:
a. correct answer
b. incorrect answer
c. incorrect answer
d. incorrect answer

Pattern 2:
a. correct answer
b. plausible misconception
c. incorrect answer
d. incorrect answer

Pattern 3:
a. correct answer with correct condition (such as, because, since, when, if, etc.)
b. correct answer with incorrect condition
c. incorrect answer with incorrect condition
d. incorrect answer with incorrect condition

Pattern 4:
a. correct answer
b. incorrect answer
c. correct answer with incorrect condition
d. incorrect answer with incorrect condition

Guidelines for writing Fill-In items

- State the item so that only a single, brief answer is likely.

- Use direct questions as much as possible, rather than incomplete statements, as a format.

- If you must use incomplete statements, place the blank at the end of the statement, if possible.

- Provide adequate space for the learner to write the correct answer.

- Keep all blank lines of equal length to avoid cues to the correct answers.

unitar
United Nations Institute for Training and Research

- For numerical answers, indicate the degree of precision required (for example, "to the nearest tenth") and the units in which the answer is to be recorded (for example, "in pounds").

Guidelines for writing Short Answer items

- State the question as clearly and succinctly as possible.

- Be sure that the question can truly be answered in only a few sentences rather than requiring an essay.

- Provide guidance regarding the length of response anticipated (for example, "in 150 to 200 words").

- Provide adequate space for the learner to write the response.

- Indicate whether spelling, punctuation, grammar, word usage, and other elements will be considered in scoring the response.

Guidelines for writing Essay items

- State the question as clearly and succinctly as possible; present a well-focused task to the learner.

- Provide guidance regarding the length of response anticipated (for example, "in 5 or 6 pages").

- Provide estimates of the approximate time to be devoted to each essay question.

- Indicate sufficient space for the learner to write the essay.

- Indicate whether spelling, punctuation, grammar, word usage, and other elements will be considered in scoring the essay.

- Indicate whether organization, transitions, and other structural characteristics will be considered in scoring the essay.

A cookbook for the Subject-matter expert (SME) to determining the test length

Test length determination in seven steps:

1. Have the SMEs identify the number of chapters, units, or modules that need to be assessed.

2. Have the SMEs identify the objectives for each unit.

3. Rate the objectives by criticality.

4. Rate the objectives by domain size.

5. Draw the line.

6. Multiply the criticality by the domain size.

7. Adjust the proportions to fit the time allotted for testing.

unitar
United Nations Institute for Training and Research

4. Create rating instruments

A checklist is created by categorizing the performance or quality of a product into specifics, of which the rater "checks" its presence or absence. Checklists are known to be more reliable because they combine a "yes" or "no" evaluation from the rater with particular behaviors or qualities. A checklist significantly reduces the rater's required degree of subjective judgment. As a result, the level of observation errors is also reduced.

5. Report scores

There are some minimal types of information that would help the organization make decisions surrounding the learner's performance:

- Report the learner's score as mastery or not.

- Describe the test measures that were used to assess the skills or knowledge.

- Indicate what, if any, remediation options are available for learners who do not meet mastery. Typically, remediation means on-the-job supervision and coaching, repeating a course, or individual study of areas that need improvement, followed by retesting.

- Provide the name of the person the manager can call if he or she has questions about the test.

Checklist for content validity of tests

Essential elements to evaluate tests' content validity:

1. Job Analysis:

- A content validity study must include an analysis of the important work behaviors required for successful job performance.

- The analysis must include an assessment of the relative importance of work behaviors and/ or job skills.

- Relevant work products must be considered and built into the test.

- If work behaviors or job skills are not observable, the job analysis should include those aspects of the behaviors that can be observed, as well as the observed work product.

2. For Tests Measuring Knowledge, Skill, or Ability

- The test should measure and be a representative sample of the knowledge, skill, or ability.

- The knowledge, skill, or ability should be used in and be a necessary prerequisite to performance of critical or important work behavior.

unitar
United Nations Institute for Training and Research

- The test should either closely approximate an observable work behavior, or its product should closely approximate an observable work product.

- There must be a defined, well-recognized body of information applicable to the job.

- Knowledge of the information must be a prerequisite to the performance of required work behaviors.

- The test should fairly sample the information that is actually used by the employee on the job, so that the level of difficulty of the test items should correspond to the level of difficulty of the knowledge as used in the work behavior.

3 For Tests Purporting to Sample a Work Behavior or to Provide a Sample of a Work Product

- The manner and setting of the test and its level and complexity should closely approximate the work situation.

- The closer the content and the context of the test are to work samples or work behaviors, the stronger the basis for showing content validity.