Annex xx –

Guidelines on Test Development

Source: Making a Perfect Fit: Objectives and Test Items by Celia T. Adriano, PhD, College of Education, University of the Phlippines, Diliman.

Following is a discussion on the different types, their uses and limitations, and tips for item development:

True or False (T/F)	
Description	Used when test items are naturally dichotomous; i.e., statement is absolutely true or absolutely false. Items can be written in two forms: simple (answer consists of only two choices), or compound (two choices plus a conditional completion response: "Example: If False, what is the learning domain referred to?).
Limitations	● There is 50 - 50 chance of getting item correctly simply by guessing. With the high guessing factor, T/F items do not discriminate between learners of varying ability. ● T/F items can often lead test developers to favor testing of trivial knowledge.
Tips 	● Base T/F items on statements that are absolutely true or false, without qualifications or exceptions. ● Write item statements as simply and as clearly as possible. ● Use compound T/F items if you want to eliminate guessing factor. ● Avoid lifting statements directly from presentation, lecture or other materials so that memory alone will permit a correct answer. ● Avoid using negatively stated items. ● Include up to 15% more false items than true items as false items tend to discriminate more highly than true item.

Examples:

	A terminal learning objective refers to what learners should be able to do by the end of a training program.
	The affective learning domain involves knowledge and the development of intellectual skill. If False, write the correct answer: _____
	A terminal objective does not refer to what learners should be able to do by the end of a session. (Negatively stated)
	According to Kirkpatrick, "All training programs should be evaluated at least at Level 1." (Direct lift from presentation)

Fill-in	
Description	Requires learners to answer a question or to finish an incomplete statement by filling in a blank with the correct word or phrase. Minimizes guessing as possible answers are not presented. Thus, learners do not simply recognize the correct answer, but has to recall or create the response.

Limitations	It can be challenging to construct questions in a way that desired response is clearly indicated. If not prepared correctly, fill-in items are difficult to score since more than one answer may have to be considered correct.
Tips 	● Omit only significant words from the statement. ● Do not omit so many words that the intended meaning is lost. ● Make sure there is only one correct response. ● Avoid lifting statements directly from presentation, lecture or other materials. ● Limit the required response to a single word or phrase.

Examples:

	The _____ domain refers to the manner in which we deal with things emotionally, such as feelings, values, motivations, and attitudes.
	The affective domain refers to the manner in which we deal with _____ emotionally, such as feelings, values, motivations, and attitudes.
	The _____ domain refers to the manner in which we deal with things _____, such as _____, values, motivations, and attitudes.

Matching	
Description	Items consist of a brief direction and two related lists of words, symbols, pictures or statements, and learners are required to match the items on the left column (stimuli) with those on the right (responses).
Limitations	It can be challenging to construct items as multiple possible stimuli and responses need to be identified. Some test-savvy learners are good in using elimination process, thus may not discriminate between learners of varying ability.
Tips 	● Use homogenous concepts or related content otherwise matching items become extremely easy by process of elimination. ● Include directions that clearly state the basis for matching the stimuli with the responses. Explain whether or not a response can be used more than once and indicate where to write the answer. ● Arrange the list of responses in some systematic order if possible (e.g., chronological, alphabetical). ● Keep matching items brief, limiting the list of stimuli to below 10. ● Include more responses (but no more than three) than stimuli to help prevent answering through the process of elimination.

Example: Match the following elements of a terminal objective:

	1			2	
	1. A	A. Audience		1. A	A. Audience
	2. B	B. Behavior		2. B	B. Behavior
[?]	3. C	C. Criteria	[?]	3. C	C. Beliefs
	4. D	D. Condition		4. D	D. Condition
		E. Degree			E. Degree
		F. Duration			F. Decimal

Note that the items on the right set are not homogenous. ("Beliefs" and "Decimals" are not related to the other responses.)

Multiple Choice	
Description	Consists of two parts: 1) the stem, which contains the task: question or problem; and 2) response options. Multiple choice test is highly reliable and measures all levels of cognitive ability except the two highest ones, synthesis and evaluation (which are beyond multiple choice formats because they require original responses from learners. Scoring is easy and accurate.
Limitations	It can be difficult and time-consuming to construct multiple choice items. Consequently, some test developers resort to simple recall of facts, thus neglecting the higher levels of cognitive ability. This test type places a high degree of dependence on the developers' writing ability.
Tips 	● When possible, state the stem as a direct question rather than an incomplete statement. ● Present a definite, explicit and singular question or problem in the stem. ● Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word. ● Provide at least four options, with only one correct response. ● The incorrect options or distracters should be homogeneous and plausible. ● Make the options grammatically parallel with each other, and consistent with the stem. ● When possible, present alternatives in some logical order (e.g., chronological, most to least, alphabetical).

Example: What learning domain involves coordination and use of motor-skill areas?

	1		2		3
?	1. Ideomotor 2. Neuromotor 3. Psychomotor 4. Sensorimotor	?	A. Deductive B. Cognitive C. Inductive D. Psychomotor	?	A. Competency B. Psychomotor D. Cognition E. Handling practical problems

Note that the options on set 1 are not homogenous. Aside from not being homogenous, set 2 contains options that are not grammatically parallel.

Short Answer	
Description	Requires learners to create original responses or recall information unassisted (rather than recognize presented information). Short answer questions require responses longer than those for fill-in test items and shorter than those for essay questions. Applicable when learning objectives involve higher levels of cognitive domain (e.g., comprehension, application or analysis)
Limitations	Since this is a subjective type of test, assessing and scoring responses will be more time consuming. Some test developers also sometimes overlook the preparation of scoring rubric when developing the test questions, thus assessment and scoring can become challenging.
Tips 	● State the questions as clearly as possible. ● Provide guidance on the length of expected response. ● Make sure questions can be adequately answered in a few sentences. ● Ensure alignment of questions to specific learning objectives. ● Provide sufficient space for response.

Examples:

?	In 150 to 200 words, what mechanisms can you include in the program design to increase probability of learning transfer to the workplace?
?	In 150 to 200 words, differentiate the features and applications of inductive and deductive learning methodologies.
?	As a training program designer, what can you do to ensure that the agency will get maximum return on training investment?

Note that the third question is rather broad and loaded, and requires discussion on several factors that may not be possible for the learner to cover, especially in a few sentences.

Essay	
Description	An essay test consists of a small number of questions/tasks to which learners are expected to demonstrate their ability to recall factual knowledge, organize this knowledge, and present the knowledge in a logical, integrated response.
Limitations	Like short answer test, the major challenge in an essay test is scoring the answer reliably, especially when assessment and scoring rubric is not clear. It takes more time to score and it uses up more testing time, yet may not provide an objective measure of students' achievement or ability.
Tips 	● Prepare essay items that elicit the type of behavior that needs to be measured, as stated in the learning objective. ● State learners' task in clear and concise manner. ● Ask questions that will elicit responses on which experts could agree that one answer is better than another. ● Indicate for each item a point of value or weight and an estimated time limit for answering.

Examples:

	Briefly explain how the 70-20-10 learning model can be applied when executing the Learning Management Cycle of Analysis-Design-Development-Evaluation. (10 points, 10 minutes)
	How can the application of the 70-20-10 learning model enhance the overall execution of the agency's L&D function?

Note that the first question is more specific and will yield responses where experts can easily agree on. It also indicated the value of the item and the allotted response time. The second question is broad, and can yield a wide range of answers which can be difficult to rate.