Writing multiple choice questions for CO learner assessments
Center for Applied Linguistics, December 2014

Overarching guideline: We are not testing a participant's ability to take written tests, so we want to do as much as possible to make the questions easy to answer and to avoid tricks or distractions in the way the questions are organized and worded. Having well-designed questions gives you confidence that people who get them right really know the answer and people who get them wrong really do not.

Directions: For populations that are less familiar with multiple-choice tests, consider including directions and an example of how to answer the questions, such as the following:

For each question, circle the ONE answer that is correct. Here is an example:
Which of the following describes the sun?
   a.   It is dark.
   b.   It is bright.
   c.   It is cold.
   d.   It is green.

You may have participants circle the answer, circle the letter next to the answer, fill in a bubble (if you can use a scannable form), or put an X or checkmark next to the right answer. Any of these procedures is fine; it is best to pick a procedure that is already familiar to participants. You may also want to suggest they use a pencil so that they can erase their answer if they change their mind.

Length: There is no standard length of a multiple-choice assessment that is preferable in all circumstances. In general, a longer assessment with multiple questions per topic is more *reliable* (all people with the same amount of knowledge would get the same score each time they took the test) and *valid* (each question and the assessment overall really measure what you think they are measuring). But good information can be gathered from a shorter assessment if it is well written. A comprehensive CO assessment would probably need at least 10-15 questions to cover a variety of critical topics.

Customization: You may wish to give a different assessment to different participants based on their characteristics (e.g., education level) or based on what was taught in CO. However, remember that you can only compare scores across the same assessment given in the same way. It is technically possible to determine if two different assessments are at the same level of difficulty but the process is extremely complicated.

Order: Be sure to start the assessment with one or two easy questions to build participants' confidence. Also, check that the correct answer shows up in the A, B, C, or D spot relatively randomly.

Guideline	Select questions that are important, relevant, and can be answered consistently.
Rationale	Because there is limited time available for assessment, ensure that all items address concepts that are important to refugee resettlement and cultural orientation. In order to make statements about the usefulness of CO, items should also address concepts that are covered in CO (preferably not concepts that participants may be able to guess correctly based on their general knowledge). Further, in order to have an assessment that can be used across refugee populations, the concept underlying the question should apply to all or nearly all refugees (based on education level, whether they are parents, age, etc.). Similarly, questions should not reference specific program requirements that might change over time or vary across participants (e.g., asking exactly how much cash assistance they will receive).

Guideline	Limit the use of true/false or yes/no questions.
Rationale	In the first overseas CO assessment pilot, we found that an entire true/false assessment would not do a very good job of distinguishing between people who understood CO content and people who do not, which is the purpose of this type of assessment. But true/false or yes/no questions, when used along with other types, might be a good choice for certain items. If you do use yes/no or true/false questions, be sure to follow the following guidelines:

- Items must be clearly and always true or false; do not include items that are subject to interpretation or opinion, or could be true or false depending on circumstances.
- Do not use negatively phrased items (e.g., "Refugees are <u>not</u> met at the airport by the resettlement agency").
- Avoid true/false questions when there are multiple interpretations of why the person got it right or wrong. For example, it is relatively clear what the participant is thinking when they mark true or false to a question like "In America, you should bribe police officers to get out of trouble." However, it is less clear what is the source of misunderstanding if a participant incorrectly answers a question like "It takes a long time to adjust to life in the United States"—possible alternatives are that it takes no time at all, it takes a short time, it never happens at all, or it is impossible to know so the answer must be false.
- Make about half of the items true and half false (does not have to be exactly half).
- Put all true/false questions together so you can provide instructions before participants answer them.

Guideline	Limit the use of tricky multiple-choice question types ("Which of the following does NOT…", select all that apply, answer choices that include "all/none of the above").
Rationale	Negatively phrased items can be tricky for test takers, who may miss the word(s) in the sentence that make it negative (e.g., *not*, *except*). If using a negatively phrased item, be sure to put the negative words in capitals, bold, underline, or whatever mechanism is possible in the translated language to help participants see those words. Questions that ask participants to *select all that apply* are also more difficult because participants may be distracted by trying to figure out how many choices in a question could be correct. For this type of item, consider a cluster of true/false questions, for example:

Which of the following will you receive from the resettlement agency?
T   F   Clothing
T   F   Food
T   F   A car
(etc.)

Finally, *all of the above* and *none of the above* might be tricky for participants if they are not familiar with that type of alternative in a multiple-choice question. It could also limit the effectiveness of an item due to participants being able to select the right answer based on partial knowledge (see references for more details). |
| *Guideline* | Try to avoid questions that might evoke an emotional reaction in the participant. |
| *Rationale* | This not only makes good intuitive sense, but it is also one of the ethical responsibilities of test-makers under the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 2014). It's appropriate to discuss the negative consequences of some resettlement choices and to address concerns during CO sessions, but it is important to avoid worrisome distractions during assessment. Some strategies for addressing this include wording questions in the abstract as much as possible (e.g., something that might happen to *a refugee* rather than something that might happen to *you* or *your family*); alternating more positive and hopeful messages that emphasize refugee choices with questions about negative consequences; and piloting your questions ahead of time with volunteers. |

Guideline	When having an assessment translated, consider whether questions should be worded in first, second, or third person.
Rationale	Since the questions on a CO assessment refer to things that will happen to refugees, they might be worded in different ways. For example, if a question asked what happens at the destination airport, an option might read: • I will be met at the airport (first person) • You will be met at the airport (second person) • He or she will be met at the airport (third person) Many languages other than English also have an impersonal pronoun which refers to a person of indeterminate gender (e.g., *on* in French). Consider which form is most appropriate based on the language into which the assessment is being translated, thinking about cultural norms of address (e.g., if it might be rude to say "you"), how the person will interpret the question (e.g., if participants might interpret the use of "I" to mean the person writing the assessment), and helping participants to answer based on what is likely to happen to most people, not necessarily what they are personally anticipating.
Guideline	Ensure that the broadness or narrowness of the answer choices will allow you to use data appropriately.
Rationale	Many guides on multiple-choice question-writing suggest that answer alternatives be homogeneous, meaning that they relate closely to each other. In the example below, the question on the left has homogenous alternatives because all four are about cultural adjustment. Because the participant has to think carefully about four plausible, closely-related statements, you can be fairly confident that a person getting this question right really understands the concept. On the other hand, in the question on the right, there is a statement about cultural adjustment, but you can't be confident that participants understand that some things will be the same and some will be different just because they correctly skip over choice (d) (also, they might mark [c] correct before even reading [d]). However, the question on the right allows you to sample a broader range of knowledge about life in the U.S., which is especially important when you have a lot to cover in your assessment. The question on the right might be less *reliable* in a technical sense, but might contribute more to your assessment because it allows you to assess a broader range of topics.

Example	Alternatives are homogeneous	Alternatives are heterogeneous

	Which of the following statements about life in America is correct?
	  a. It will be just like the country you came from, nothing will change
	  b. You will be on your own to learn everything you need to know
	  c. Everything will be easy to get used to
	  d. Many things will be new and different and it will take time to adjust
		Which of the following statements about life in America is correct?
		  a. Affordable apartments are dirty
		  b. You will be discriminated against
		  c. Americans have freedom of speech
		  d. It will be just like the country you came from, nothing will change

Guideline	Use the most simple vocabulary possible, unless a more sophisticated term was taught or used frequently during CO.
Rationale	Ensure that translations use the simplest and most straightforward vocabulary possible to make the test accessible for participants with low levels of education. However, more complex terms (such as *resettlement*, *self-sufficiency*, or *cultural adjustment*) can and should be used in questions when the term was used and emphasized in instruction and the term is important for refugees to know.

Example	*Uses complex term*	*Uses simple vocabulary*
	Who will be your most accurate source about resettlement? a. Friends and family who are already in the U.S. b. An American friend or neighbor **c. Your resettlement agency case manager** d. What you see on television	Who will be your most accurate source about getting used to living in America? a. Friends and family who are already in the U.S. b. An American friend or neighbor **c. Your resettlement agency case manager** d. What you see on television

Guideline	The correct answer should not stand out because of how it is worded.
Rationale	Participants may be distracted by trying to guess the answer based on clues unrelated to the content. For example, if one answer choice is longer or more detailed than the other three, participants might choose that answer because it looks different. Or, participants might avoid one choice that uses "always" or "never" if none of the other choices do. Also be careful of grammatical clues, such as in this example: The best type of fruit is an: a. apple b. banana c. watermelon In this case, the answer must be (a) because it starts with a vowel. This is especially important in languages where there must be grammatical agreement between nouns, verbs, and other types of words.

Guideline	Answer choices should be mutually exclusive.

Example	*Yes*	*No*
	What is your age group? a. 20-29 b. 30-39 c. 40-49 d. 50-59	What is your age group? a. Under 30 b. Under 40 c. Under 50 d. Under 60

Guideline	Answer choices should be plausible.	
Example	*Yes*	*No*
	In America, what is something that only U.S. citizens have the right to do? **a. Vote in elections** b. Own a business c. Drive a car d. Buy a home	In America, what is something that only U.S. citizens have the right to do? **a. Vote in elections** b. Own a business c. Climb trees d. Eat vegetables

Guideline	Answer choices should be in a logical order (when applicable).	
Example	*Yes*	*No*
	At what age do Americans generally retire (stop working at a job)? a. 45 b. 55 **c. 65** d. 75	At what age do Americans generally retire (stop working at a job)? **a. 65** b. 55 c. 75 d. 45

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Salt Lake City, UT: Brigham Young University Testing Services and The Department of Instructional Science. Available: https://testing.byu.edu/handbooks/betteritems.pdf

Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). New York: Macmillan. (See also 11th edition, 2012)