Cultural Orientation Assessment Toolkit: Frequently Asked Questions about Assessment and Evaluation

This document provides guidance on frequently asked questions about evaluating cultural orientation (CO) programs provided to refugees in the United States. It offers suggestions on monitoring and evaluating CO programs in general, with a particular emphasis on developing and using assessments of participant learning.

In this FAQ, CO assessment refers to the evaluation and documentation of what refugees learned in orientation; in other words, to what degree participants can demonstrate that they have learned new concepts and skills.

All suggestions in this FAQ describe best practices in assessment and evaluation, but should not be interpreted as mandates for any U.S. resettlement agency. Decisions about the implementation of the Model CO Assessment or the use of any other monitoring or evaluation activities should be based on the needs and requirements of local and national agencies and their funders.

Model CO Assessment Connection

Some sections of the FAQ refer to the Model CO Assessment that was developed by the Cultural Orientation Resource (COR) Center at the Center for Applied Linguistics (CAL) in February 2014. The Model CO Assessment is one tool that agencies can use to assess the effectiveness of their CO program and is aligned with the Reception and Placement (R&P) CO Objectives and Indicators. Guidance on how to administer that assessment was provided in a separate instructions document; this FAQ provides additional information on the use of assessment data.

The contents of this document were developed by the Center for Applied Linguistics under an agreement financed by the United States Department of State Bureau of Population, Refugees, and Migration.

Contents

How can the Objectives and Indicators (O&I) be used?

The R&P O&I document has a number of uses. First, it **outlines the content that should be covered** in R&P CO. CAL's online R&P Orientation Curriculum is aligned to the O&I. CO providers may wish to align their existing curricula to the O&I in order to ensure that all objectives are covered and to help staff find lessons and activities that address particular objectives.

Second, the O&I can also help you define the areas in which you should **tailor curricula to your client population**. Once your curriculum is aligned to the O&I, you can go through the O&I to prioritize objectives that are particularly important for your clients (or particular groups of clients).

Third, the O&I may be used by **administrators, funders and other stakeholders as part of program evaluation** to ascertain whether refugees are being taught the information that they need to be successful. Assessments aligned to the O&I are one way to investigate whether CO was implemented as intended (see also What kind of data collection instruments can we use to assess CO learning?).

What is the difference between monitoring and evaluation?

Monitoring is a routine process of documenting program activities, usually for the purpose of ensuring that programs are doing what they intended to do. Data collected usually indicates that something was done or not done, how much was done, and who was involved.

Evaluation uses a variety of data to make judgments about the quality of the program. Sometimes, this includes whether the program was effective in producing the intended outcomes (this is called impact evaluation). Evaluation data goes beyond "what" and "how much" questions to ask why and how the program worked.

Monitoring and evaluation work together in that evaluations often rely on data collected as part of monitoring, monitoring results may inspire evaluation questions, and evaluation findings may result in new categories of information to monitor.

For example, an agency or affiliate might *monitor* the amount of time that CO instructors spend discussing housing and whether they have discussed each content objective from that section of the Objectives & Indicators. Later, they might *evaluate* whether participants understood the information. An *impact evaluation* might examine whether having brief or extended exposure to housing topics in CO helped refugees procure and maintain appropriate housing.

For more information on monitoring and evaluation, see these online resources:
Linking Monitoring and Evaluation to Impact Evaluation (InterAction)
Monitoring & Evaluation: Some Tools, Methods & Approaches (World Bank)

What is the difference between outputs, short-term outcomes, and long-term outcomes?

Program evaluation may measure different aspects of what happens as a result of a program:

- Outputs - measures of what the program did
- Short-term outcomes - what happens to participants right after the program
- Long-term outcomes - what happens to participants over the course of months or years

In the case of CO, outputs are the amount of CO delivered and to whom; short-term outcomes are changes in participants' knowledge, behavior, and attitudes; and long-term outcomes are changes in families, communities, and systems over a period of time.

CO assessment measures short-term outcomes, namely, whether what was taught in CO was learned, understood, and retained. This is important because if the information is not learned in the first place, we cannot say that CO has any impact on refugee resettlement. In the short term, we can say CO was effective if participants are able to demonstrate that they understood and remember the concepts learned.

Additionally, CO may predict long-term outcomes, since the program theory underlying CO is that what is learned in CO helps refugees resettle more successfully than if they did not have CO. Long-term outcomes may be measured by such indicators as economic self-sufficiency, strong social connections, and maintenance of personal safety. Whether CO contributes to long-term outcomes would be the subject of impact evaluation or research.

The long-term impact of CO would be hard to evaluate because so many variables intervene, such as participants' knowledge and beliefs (which may cause them to choose not to accept guidance provided in CO), access to community members who can reinforce or undermine CO messages, opportunities to practice skills learned in CO, and even luck, since the most well-informed CO participant may do everything right and still fall on hard times.

What kind of data collection instruments can we use to assess CO learning?

The matrix below classifies data collection instruments along two dimensions--formality and directness:

- **Formal measures** ask the same questions the same way to all participants; **informal measures** are based on on-the-spot decisions about what information to gather or use non-standardized procedures.
- **Direct measures** allow participants to demonstrate knowledge or skills directly tied to what they learned in CO; **indirect measures** include participant reflections on their learning and assessments of how well concepts and skills learned in CO are applied outside the classroom.

> *Model CO Assessment Connection*
> The data collection instrument developed for the Model CO Assessment is a formal, direct data collection instrument. However, you could use other data collection instruments to assess CO learning.

	Direct	*Indirect*
Formal	**Assessment** (oral or written questions meant to elicit knowledge) **Skills test** (watching participant do a task that was learned)	**Self-assessment or satisfaction survey** (oral or written assertion of how much the client feels he or she has learned) **Long-term impact** (causal connection between CO participation and indicators of integration)
Informal	**Asking questions** (checking for understanding or listening to responses during discussion) **Skills observation** (watching participant do a task that was learned but not using formal assessment)	**Observation of behavior** (whether participants use what they learned in everyday life) **Peer interview** (asking other people about the participant's knowledge and skills)

About these types of data collection instruments:

- **Assessment.** See What is the best format for a CO assessment?
- **Skills test.** See What is the best format for a CO assessment?
- **Self-assessment or satisfaction survey.** Refugee self-assessments can be a powerful pedagogical tool. Having CO participants reflect on how much they know before, during, and after CO helps them synthesize knowledge and feel a sense of accomplishment. Drawbacks include participants relying on their memory and impressions of how much they knew prior to CO, lack of a frame of reference for what participants believe is the "right" amount of learning, and response bias (participants providing what they believe is the desired answer rather than the true one).
- **Long-term impact.** See What is the difference between outputs, short-term outcomes, and long-term outcomes?

- **Asking questions** and **skills observation**. Very similar to formal assessments but without the use of a formal instrument such as a test or checklist. These may be reported by two methods:
    - "Portfolio" approach (collection of formative assessment data, e.g., in-class responses to questions, written activities, log of participants' skill demonstrations)
    - "Report card" approach (instructors summarize their perception of participant learning)
- **Observation of behavior.** The participant's case worker may be able to report on such behavior as the participant's use of community resources, the types of questions the participant asks outside of CO, the beliefs and attitudes of the participant, etc.
- **Peer interview**. This involves asking people who know the participant (friends, family, co-workers, employers) about their perceptions of to what degree the participant demonstrates knowledge and skills learned in CO in their everyday life.

What is the best format for a CO assessment?

All types of learner assessments have benefits and drawbacks. For example, assessments that are administered one-on-one and allow participants to provide open-ended responses provide the most rich and useful information about what has been learned, but they also are time-consuming to administer and analyze. True/false questions are easy to administer, but the high probability of guessing correctly (50%) means that the results may not be reliable or valid. Multiple choice questions are harder to get right by guessing, but they are hard to deliver orally because it can be hard to remember the questions and answer choices.

Other types of question formats that can be used in CO assessment include having the refugee point to or circle written items, sort or match pictures, demonstrate a skill, or use a rating scale (e.g., strongly agree, agree, disagree, strongly disagree; always, sometimes, rarely, never).

(See also What is the connection between the purpose of an assessment and how it is developed?)

Model CO Assessment Connection

The format chosen for the Model CO Assessment is an orally-administered, one-on-one assessment. The assessment design includes oral prompts based on participants' responses, so that assessors can follow up on partially correct answers. A group-administered or written assessment would not allow for prompts or follow-up questions.

In the Model CO Assessment, most of the items are open-ended questions requiring a response of a few words or sentences. The open-ended format of the questions allows participants maximal opportunities to demonstrate what they know by allowing for a wide range of correct responses.

What are the steps involved in creating a CO assessment?

Decide on the topics you want to ask about. Use the R&P Objectives and Indicators and your CO curriculum to select the content to assess. The format of the assessment and other variables such as how much time you have may also affect the topics you choose.

> *Model CO Assessment Connection*
> The Model CO Assessment is one tool that can be used to assess what refugees learned in CO. This question provides guidance for those agencies that wish to develop other assessments to use for this purpose.

Develop questions. Good questions to use on a CO assessment have the following characteristics:

- They measure something important and meaningful that relates to adapting to life in the U.S.
- They pertain to content that is emphasized and practiced during CO
- They are applicable across refugee populations (age, sex, education level)
- They do not reference specific program requirements that might change over time or vary across participants
- They would not evoke an emotional reaction in the participant

Align the assessment to your setting, curriculum, and client population. Check that your assessment questions and tasks do the following:

- Use the same terms and vocabulary that were introduced during CO
- Use locally-relevant names (e.g., "Metro" for public transportation in Washington, DC)
- Are translated accurately and appropriately (where applicable)

Pilot test. Try out the assessment with former clients who can provide feedback on whether each question is clear. Using refugees who have been in the U.S. for a year or two can be helpful because they remember what it was like to be in CO but have sufficient perspective to be able to help reframe or reword confusing questions.

Get feedback. You want to include as many types of stakeholders in CO assessment development as possible right from the beginning, because they will have different perspectives based on their job and how they plan to use the data (see How might different stakeholders use CO assessment data?). In addition to seeking feedback on the content of the assessment (what is assessed and how questions are worded), you should seek the following types of feedback:

- From case managers and others who will administer the assessment: Feedback on the instructions and format of the assessment and on any computer or paper-based systems for entering data
- From native speakers of the language: Feedback on the wording of translations and cultural sensitivity/bias of any questions
- From administrators and funders: Feedback on the way that data are recorded, analyzed, and reported

What is the connection between the purpose of an assessment and how it is developed?

There are many purposes for assessment that are based on the questions that stakeholders have about their program. These include:

- Program evaluation (How do participants do in general, over time?)
- End-of-lesson or end-of-unit test (Did participants understand all the concepts in a unit or lesson?)
- Curriculum development (Are there topics that participants have more difficulty understanding than others?)

Most assessments can be valid for more than one purpose, but you need to think about how you want to use the resulting data while you are thinking about the format and content of the assessment.

If you want to…	*Do this…*
Demonstrate that all participants are able to demonstrate basic skills or achieve a minimal level of competence	Create as simple an assessment as possible that covers only basic skills
Be sure that participants have an opportunity to demonstrate the full extent of their understanding	Administer the assessment one-on-one so that follow-up prompts can be asked and misunderstandings about the meaning of the question can be clarified
Compare the scores of a variety of participant groups	Be sure that the assessment is administered the same way for everyone (in the same time frame post-arrival, either orally or in writing, one-on-one or in a group) or present findings disaggregated by these variables
Be able to identify the characteristics of beginning, intermediate, and advanced learners	Create a variety of easy, medium, and hard questions to ensure that there will be a wide range of scores among participants
See if participants increase their knowledge and understanding over time	Give the same assessment at two time points, and design the assessment to be relatively difficult at the time of the first assessment so that there is room for scores to increase
Find out which topics participants have the most trouble with	Ask a number of questions on each topic that you are interested in (at least 3-5, more if the topic is complex)

Model CO Assessment Connection

The purposes of the Model CO Assessment are:

- To ensure all refugees are able to demonstrate basic competencies
- To be able to identify groups of refugees that are not able to demonstrate basic competencies (if any)

The Model CO Assessment may be used as a pre-test and post-test to see if participants increase their knowledge and understanding over time, or it may be used at the end of the R&P period to show that refugees can demonstrate basic competencies. A post-only assessment would allow you to suggest that refugees have the knowledge and skills they should acquire during the R&P period (whether from CO or otherwise), but a pre/post implementation may allow you to suggest that CO contributed to participants' increase in knowledge.

The open-ended, oral interview format was chosen for the Model CO Assessment because

- It provides participants maximum opportunities to demonstrate what they know
- It elicits correct answers through prompts and partial credit
- It allows for a range of correct answers depending on individual or local circumstances
- It is an easy-to-understand format for participants who may be unfamiliar with Western testing practices

What ethical guidelines should we follow for involving refugees in learner CO assessment, monitoring, and evaluation?

At present, collecting data from refugees for the purpose of evaluating a CO program is exempt from formal oversight from an Institutional Review Board (IRB), although data collected for *research* purposes does require IRB approval.

Whether you use an IRB or not, ethical guidelines like the Standards for Educational and Psychological Testing and The Program Evaluation Standards should be followed. Ethics in testing and evaluation generally concern the following:

- Protecting participants from distress caused by uncomfortable testing conditions, overly difficult or offensive test items, or fatigue
- Ensuring participants' privacy
- Informing participants about the purposes and consequences of the test or survey

In particular, consider IRB and ethics guidelines for protecting the rights and welfare of "vulnerable groups." Refugees may feel obliged to participate in testing due to their dependence on the resettlement agency, their lack of experience with participants' rights in test settings, and/or cultural attitudes toward obeying authority. Although agencies are not required to obtain informed consent from participants on routine assessments used for monitoring and evaluation, agencies should review their assessment procedures to consider what rights refugees have to opt out of testing (or what the ramifications might be if they do), how to inform participants of their rights (including the right to continue to receive services whether they participate or not), and whether procedures respect culturally-diverse perspectives on autonomy and human rights.

Should we use the same assessment for everyone or have different assessments for different groups?

If you are creating an assessment for program monitoring and evaluation purposes, using the same assessment for all CO participants means that you can compare scores across participants and across time. It would be hard to compare a group's average score on one assessment with another group's average score on a different assessment without using complex statistics.

Model CO Assessment Connection

Because the Model CO Assessment focuses on minimal competence and basic skills and was designed to maximize the use of non-verbal or short-answer responses, it should be feasible to use this assessment with a variety of refugees (including those with more/less education or more/less prior knowledge of CO concepts). It is likely that many refugees will score toward the high end of the scale, since the assessment focuses on those concepts that we expect refugees to have learned by the end of the R&P period. For the purposes of this assessment, finding few differences in score between groups (e.g., ethnic, linguistic, gender) and having most or all refugees score toward the top of the scale would be a positive outcome. (See also What is the connection between the purpose of an assessment and how it is developed?)

What do our CO assessment data mean?

Data from assessments of CO learning (such as the Model CO Assessment) tell us whether instruction was effective in the short-term, meaning that participants are able to demonstrate that they understood and remember the concepts learned. Research on the connection between short-term learning and long-term success in resettlement would have to be done to tell us what scores on CO assessment mean in relation to the larger goals of refugee resettlement programs. (See also What is the difference between outputs, short-term outcomes, and long-term outcomes?)

Given the type of information taught in CO, it is possible that people who score low on a CO assessment may be able to access knowledge and skills gained from CO in real-life encounters. As we have all experienced, sometimes abstract knowledge we learned in school does not make sense until we have a chance to apply it in real life. Because of this, low CO assessment outcomes might not be sufficient evidence to determine that CO was ineffective for some participants.

Another complicating factor in interpreting CO assessment data is the fact that CO is not the only source of information that refugees have. To know for sure that CO was responsible for new learning, a researcher would have to compare the outcomes of those who attended CO versus those who did not and carefully account for what other sources of knowledge each group was exposed to. Nevertheless, results from Canadian and Australian post-CO surveys as well as CAL project activities (refugee focus groups, pre/post-CO assessments) indicate that, broadly speaking, CO has the desired effect of providing useful information and setting appropriate expectations.

Finally, interpreting assessment data involves more than finding average scores. In order for scores to be useful for program improvement, you also need to know answers to questions that help contextualize your findings and provide guidance on how to make changes to improve outcomes. Such questions might include:

- Who is being assessed and what knowledge did they have before participating in CO?
- What kind of instruction did participants have?
- What kind of training do instructors have? What is the quality of the materials that they use?

(See also What kind of data could we collect to evaluate the quality of our program beside learner assessment?)

What are some comparisons we might want to make with our CO assessment data?

The following are some ways to disaggregate your CO assessment outcome data to look at how different groups compare to each other:

- Older and younger refugees
- Men and women
- Refugees from different ethnic/linguistic backgrounds
- Refugees with prior education/literacy and those without
- Refugees with U.S. ties and free cases

You also might want to compare outcomes of refugees who had different kinds of CO (e.g., one-on-one or in classes, or comparing previous cohorts to a cohort that experiences a new curriculum). If contextual factors (client population, curriculum or instructional approach, resources available to support CO) or assessment tools change, you will need to account for these changes in explaining why scores have increased or decreased.

How might different stakeholders use CO assessment data?

CO assessment data may be used for a variety of monitoring, evaluation, and research purposes. Primarily, stakeholders want to know if the programs they run are effective. These stakeholders may use assessment data to suggest changes to the program, curriculum, or instruction.

> Tip: Be ready to help stakeholders understand your data by tracking information on your population and your program. (See What kind of data could we collect to evaluate the quality of our program beside learner assessment?)

At the local level, CO instructors and other agency staff can use CO assessment information to make changes to instruction. For example, they might find they need to improve the effectiveness of a particular activity or to choose certain topics to emphasize in greater depth for particular client populations. They might also use CO assessment data to make decisions about staff training.

Assessment information may also be useful to share with local partners and community organizations to help them refine their services or instructional delivery.

Researchers may also find the data useful to better understand refugees' experiences or to suggest how CO affects long-term resettlement.

> *Model CO Assessment Connection*
>
> The intended uses of the Model CO Assessment are
>
> - To demonstrate the effectiveness of CO in the short-term (understanding and retention of concepts) or to indicate a need for overall program improvement
> - To point to areas for further investigation (topics, populations, program implementation models)

Finally, assessment data may be useful for program advocacy. Showing a funder that program outcomes have improved over time or that outcomes correlate to long-term success may help encourage them to provide additional resources to continue or expand the program.

How do we compile our data and share our findings from CO assessment?

Early in your assessment planning process, think about how different stakeholders might use your assessment and evaluation data (see How might different stakeholders use CO assessment data?). The most important consideration for how to compile data and share findings is what is required from your funders. National headquarters may request information to facilitate their monitoring and evaluation processes and to use in their reports (aggregating findings across affiliates).

> Tip: Use backwards planning to think about what information you will want to convey, what information you will need to collect in order to convey that information, and how and when to collect that information accurately and efficiently.

Programs such as Microsoft Excel or Microsoft Access may be used to capture CO assessment data. If your agency has a custom application for tracking refugee information, you may want to request that fields be added to enter assessment findings.

Whether you are looking at CO assessment data or results from other monitoring and evaluation activities, the typical format for sharing assessment findings is a written report. Some organizations also use more dynamic presentation formats such as PowerPoint or websites. A digital dashboard, which can be created with Microsoft Excel or with specialized software like Tableau or iDashboards, is an interactive tool that lets users visualize data in different ways.

What kind of data could we collect to evaluate the quality of our program beside learner assessment?

In order to get a full picture of any program, evaluators look at factors along the full continuum of the program's logic model, including inputs, processes, outputs, and outcomes (see also What is the difference between outputs, short-term outcomes, and long-term outcomes?)

In the list below, the information described under *inputs* and *processes* may provide context to your findings regarding the outputs and outcomes of your program. You may also want to look at *processes* in order to investigate possible areas in which to improve program implementation.

Inputs:

- Demographic and language background of refugees (sources: WRAPS biographical data, historical/ethnographic data)
- CO learning needs (sources: feedback from overseas CO providers, client survey or interview, feedback from community based organizations)
- Resources available for CO (sources: budget, personnel reports, lists of available instructional materials)

Processes:

- Quality of the curriculum and alignment with O&I (sources: curriculum alignment rubric, instructor surveys, classroom observations, lesson plans)
- Quality of instruction (sources: classroom observations, instructor and supervisor interviews or surveys, client satisfaction surveys)
- Quality of program management (sources: administrative reports, recruitment and enrollment procedures and reports, logistical information on program implementation, processes used for feedback and continuous improvement of the program)
- Quality of staff and professional development (sources: training agenda, staff qualifications, instructor and supervisor interviews or surveys)

Outputs and Outcomes:

- Reach of CO program (sources: number of hours of CO provided per participant, number of participants)
- Learning outcomes (see What kind of data collection instruments can we use to assess CO learning?)
- Client satisfaction (sources: participant surveys or interviews, feedback from community-based organizations)
- Impact evaluation (sources: research on CO participation and long-term outcomes in resettlement and integration)

Glossary

Aggregated: Data findings that are combined in order to look at large-scale trends (for example, a national headquarters might aggregate findings to look at CO assessment results across 30 affiliates). See also **disaggregated**.

Client population: The sub-groups of refugees who participate in CO. Groups that may have distinct needs include older refugees, unaccompanied minors, urban and rural refugees, men and women, refugees with low levels of education or English, and refugees from different ethnic or linguistic backgrounds.

CO: Cultural Orientation

CO assessment: The documentation of what refugees learned in orientation; in other words, to what degree participants can demonstrate that they have learned certain concepts and skills. The Model CO Assessment is one example but other formal or informal measures may be used.

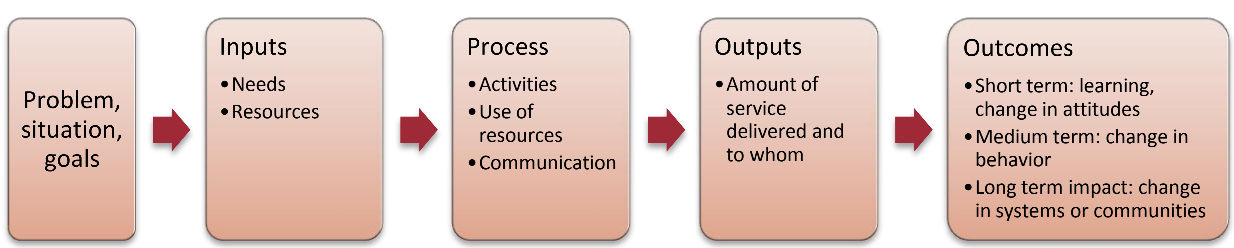
Disaggregated: Data findings that are broken down to be summarized by sub-group (such as comparing average scores for males and females). See also **aggregated**.

Evaluation: The use of a variety of data to make judgments about the quality of the program. Sometimes, this includes whether the program was effective in producing the intended outcomes (this is called impact evaluation). Evaluation data goes beyond “what” and “how much” questions to ask why and how the program worked.

Formative assessment: Testing or measurement that takes place in order to find out what concepts need to be re-taught or emphasized, or—at the program level—how to improve instruction or the CO program.

Impact evaluation: A specific type of program evaluation that investigates to what degree the program caused something to happen for individuals, populations, or systems.

Logic model: A logic model is a visual representation of how the program works. It usually includes inputs, processes, outputs, and outcomes. See also **program theory**.



Monitoring: A routine process of documenting program activities, usually for the purpose of ensuring that programs are doing what they intended to do. Data collected usually indicate that something was done or not done, how much was done, and who was involved.

Program theory: Also called program logic, theory of change, or causal model. This is the description of how your program addresses a particular problem or situation and what outcomes are expected as a result. For example, a simplistic program theory underlying CO is that arriving refugees are overwhelmed with new information and experiences; classes or one-on-one orientation helps them understand their new environment; and as a result of orientation, refugees will develop knowledge, skills, and attitudes that help them develop realistic expectations, confidence, and the ability to get along in their new environment. See also **logic model**.

Reliable: An assessment (or a question on an assessment) is reliable when you would get the same findings if you repeated the assessment multiple times. Things that contribute to whether an assessment or question is reliable include whether questions are asked the same way and under the same circumstances in each administration of the assessment and consistency in scoring guidelines.

Stakeholders: People or institutions that have an interest in a program, such as agency headquarters, funders, refugees, receiving communities, and instructors.

Summative assessment: Testing or measurement that is done to make judgments about whether CO was effective.

Valid: An assessment (or a question on an assessment) is valid when it is measuring what you think it should be measuring. Things that contribute to whether an assessment or question is valid include whether participants understand the questions, whether the format of the assessment is appropriate for the population, whether the question is closely related to the underlying construct of interest, and whether there is sufficient evidence to make inferences about what the assessment results mean.