



MONITORING & EVALUATION RESOURCES

Bureau of Educational and Cultural Affairs
Evaluation Division

TIPS FOR DEVELOPING SURVEY INSTRUMENTS AND QUESTIONNAIRES



Evaluation Division
BUREAU OF EDUCATIONAL AND CULTURAL AFFAIRS

The Bureau of Educational and Cultural Affairs' (ECA)'s Evaluation Division has been at the forefront of the Department of State's (DOS) monitoring and evaluation (M&E) efforts since 1999. Throughout its 20 years, the Evaluation Division has built a robust M&E system to ensure that ECA program staff and senior leadership benefit from timely performance data that can be utilized for evidence-based decision-making. The Evaluation Division's priority is to support the ECA Bureau's commitment to meeting and exceeding its programmatic goals by providing the data necessary to drive evidence-based decision-making throughout the Bureau.

For a complete listing of ongoing evaluation projects, an archive of completed reports, and resources for conducting evaluations, visit the ECA Evaluation Division website: <https://eca.state.gov/impact/eca-evaluation-division>

If you would like additional information or have any questions, please contact us at ECAevaluation@state.gov

Contents

- INTRODUCTION..... 1
- IMPORTANT DEFINITIONS..... 1
 - INSTRUMENT TYPES & USES 1
 - QUESTION TYPES 1
 - MEASUREMENT CONCEPTS 2
 - MISCELLANEOUS 2
- DEVELOPING THE INSTRUMENT 3
 - PLAN YOUR SURVEY CAREFULLY 3
 - CREATE YOUR INSTRUMENT SPECIFIC TO THE RESPONDENT GROUP 3
 - WRITE CLEAR INSTRUCTIONS FOR YOUR SURVEY 4
 - CONSIDER THE LENGTH..... 4
- DEVELOPING THE SURVEY QUESTIONS..... 4
 - GENERATE QUESTIONS BASED ON THE PURPOSE OF THE DATA..... 4
 - WRITE UNBIASED QUESTIONS 5
 - ASK THE MOST IMPORTANT QUESTIONS AT THE BEGINNING OF THE SURVEY 5
 - MAKE QUESTIONS SPECIFIC 6
 - ASK ONLY ONE QUESTION AT A TIME..... 6
 - USING SCALED QUESTIONS 6
 - RATING SCALES 6
 - LIKERT SCALES 7
 - MATCHING THE SCALE TO THE NATURE OF THE QUESTION..... 8
 - LEAVE OPEN-ENDED QUESTIONS FOR THE END 9
- FINAL STEPS 9
 - CONSIDER THE TIMING 9
 - REVIEW 9
 - TEST YOUR INSTRUMENT 10

INTRODUCTION

This manual was created to assist you, the grantee, to design surveys that help get the information needed from program participants and other stakeholders to improve programming and demonstrate results. Surveys that are well-designed lead to data that are valid and reliable – which helps you to highlight successes, and determine if and where changes are needed for improvements.

We designed this manual to help you to think about the programs being designed, what outcomes are expected, and whether expected changes occur as a result of the activities implemented. This manual provides some definitions related to surveys and research methods, as well as tips to develop strong surveys and easy-to-understand survey questions. However, it is not possible to provide examples that cover the breadth of programs or activities, nor is it exhaustive of the types of surveys and analysis available. Thus we encourage you to reach out to your Bureau of Educational and Cultural Affairs point of contact if you need further assistance.

IMPORTANT DEFINITIONS

INSTRUMENT TYPES & USES

A **Pre-Test/Post-Test** is used when you would like to observe whether a desired change occurred as a result of your efforts. In a pre-test/post-test, the same instrument is administered both prior to an activity and after. Answers are compared to identify and report any changes. A pre-test/post-test can also help to identify the level of change that occurred. If the desired level of change did not occur, then perhaps the activity can be strengthened. Or rather, if the pre-test determines that a high level of knowledge already exists, then program managers can change activities to accommodate a higher level of proficiency. The pre-test/post-test is perhaps the most reliable method to determine whether a change occurred as a result of your efforts.

Focus Groups are small, deliberately chosen groups of people that are interviewed together to observe reactions to the topics you propose. During a focus group, it is important to observe the responses of all participants as there are usually members that are more outspoken than others. Focus groups are a great way to estimate how others, who have not or will not participate in your activities, will react to the topics proposed to the selected group. Of course, in addition to the observations, the information provided by the focus group participants will also be useful.

Surveys are sets of standardized questions (questionnaire) that are administered to selected individuals or groups of individuals. Surveys can be administered on paper, as an interview, online, or via telephone. You will choose the most appropriate type of survey depending on availability of respondents, security in the area, and/or time constraints.

QUESTION TYPES

Closed-Ended Questions are those that list pre-set answers for respondents. They include multiple-choice or “yes/no” answers. Closed-ended questions are easy to analyze but do not offer additional, anecdotal but important findings. Closed-ended questions include all relevant and possible answers and must be mutually exclusive (there cannot be more than one choice that answers the question).

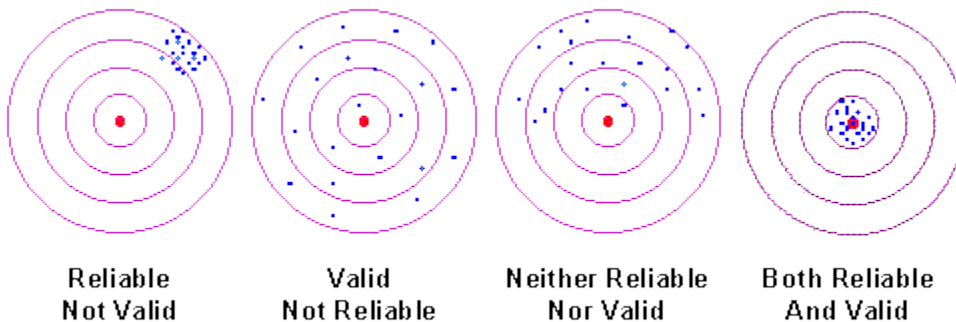
Open-Ended Questions are those that do not have a pre-defined set of answer options. Open-ended questions are used when you want respondents' opinions, additional perspective, or information not captured in other question types. Responses to open-ended questions may inform answer choices for closed-ended questions in future surveys. Examples of open-ended questions include: "Please list any additional activities that you think should be included in future training sessions," or "what was one topic discussed during the training that had the most meaning for you and why?"

Scales are a social science research technique used to measure the qualitative aspects of the group of people you need information from. They associate qualitative attributes to quantitative metrics.

Likert Scales are a type of scale that asks respondents to indicate the level that they agree or disagree (generally, from 'strongly agree' to 'strongly disagree') about a statement. Likert scales generally include five-point, seven-point, or nine-point responses. Notice that they always have an odd number of answer options, which is to make sure the respondent is provided with a neutral response.

MEASUREMENT CONCEPTS

Validity refers to whether or not your survey (or other form of measure) is actually measuring what it is supposed to. **Reliability** is when your survey (or survey question) produces consistent results when used to measure the same thing over and over. These are depicted graphically below. The goal is for your surveys (and all forms of measure) to have both validity AND reliability.



Bias is an unfair preferences or dislike of something. Different aspects of the respondents' experiences or the way a survey is administered could skew the results of a survey. You will generally find bias in a survey in the way a question is worded – we have provided examples of bias (and how to avoid it) below. Careful survey design and planning can help to avoid bias.

MISCELLANEOUS

Data or Information Source Data are information that will help you make decisions about programming and will demonstrate success toward indicators.

Respondents are the individuals taking the survey; the members of the group of people you would like information from, usually the beneficiaries of a program.

Questionnaires are research instruments (surveys) consisting of a series of questions for the purpose of collecting information about a particular subject.

DEVELOPING THE INSTRUMENT

PLAN YOUR SURVEY CAREFULLY

Before you begin to write your survey, you should make a plan for who you need to survey, the questions you want to ask, who will use the information from your survey, how you plan to distribute your survey, and when you need the results by. While not required, a survey design matrix can be helpful in planning the creation of a survey. Once completed, a design matrix summarizes virtually all of the survey design decisions in an easy format.

A survey design matrix is below (each section is described in *italics* along with an example in the third row). Any program stakeholder can utilize this matrix to help them make decisions about what questions are the most appropriate to ask given their program activities.

	Indicator data/question to be answered	Information Source	Sampling	Data Collection (Mode)	Data Collection (When)	Survey Questions	Survey Testing
<i>Description</i>	<i>This will most likely come from the indicator you are reporting against or perhaps a question your team or the client wants to know.</i>	<i>Who can answer the question?</i>	<i>Will we be able to ask everyone to answer the question? If not, how will we choose?</i>	<i>Face-to-face? Through email? The internet? Phone?</i>	<i>After the exchange ends? After a workshop session is completed?</i>	<i>What question(s) will your survey ask to answer the question?</i>	<i>What information did pre-testing the survey reveal? Do we need to clarify wording? Were any questions not answerable?</i>
<i>Example</i>	What percentage of participants improved their perception of Americans during their exchange?	Exchange program participants	We should be able to locate all program participants through email addresses	Will implement through email	Survey will be conducted immediately after the exchange ends	1. How did your perception of Americans change during the exchange: improve? Stay the same? Diminish?	Conducted a pre-test with 5 of 100 participants and they suggested we make a specific change to the language to make the survey more clear

CREATE YOUR INSTRUMENT SPECIFIC TO THE RESPONDENT GROUP

When developing your survey, tailor it to the target group who will be completing it. If you are surveying multiple groups (i.e., American or foreign participants; youth or adults; academic or professional exchange participants), be sure to develop survey instruments for each, as respondents from each group

might not be able to answer questions written for other groups. For example, youth participants will be able to answer questions based on their knowledge gained and experiences from their exchange experiences; however, they will not be able to answer questions that are related to specific activities from a professionally-focused exchange program.

WRITE CLEAR INSTRUCTIONS FOR YOUR SURVEY

- Overall instructions at the beginning of your survey should include:
 - The survey's purpose
 - Who will use the information
 - How the information will be used
 - Whether the responses will be anonymous
 - The approximate time the survey will take to complete
 - Assure respondents of anonymity and confidentiality (if applicable)
- When you have different types of questions it is best to break them up by type and provide instructions for how the respondent should answer the questions at the beginning of each section. Below are example definitions for different types of survey questions:
 - Please select the extent to which you agree or disagree with the statements below (Likert scale type questions).
 - Please select one answer (multiple-choice questions, please select only one answer).
 - Please check boxes for all answers that apply (you may select more than one answer).
- Provide definitions for any terms or concepts that may not be easily understood by all respondents.

CONSIDER THE LENGTH

Consider how many questions you ask and the length of time required to answer all of the questions. Also consider how much time your survey population realistically has available to complete the survey. Try to create a survey that balances the time that respondents would be willing to take to complete the survey and the information required for reporting and analysis. When thinking about the length of the survey, ask yourself: Are you asking so many questions that the respondent will become tired of answering? Asking too many questions creates a risk of a lower response rate and less data.

It is also important to make sure you ask the correct questions and enough questions to report on your indicators and provide the necessary data from the program. The survey design matrix can help you identify the number of survey questions needed to address all information requirements and/or indicators to be answered.

DEVELOPING THE SURVEY QUESTIONS

GENERATE QUESTIONS BASED ON THE PURPOSE OF THE DATA

Survey questions should be created so they are easy for respondents to understand and answer, and also provide the data required for the intended users of the results. When writing survey questions you should think about how you will use the information you obtained from the survey. Ask yourself:

- What questions do you need answered and why?
- Who will use this data and information?

If you do not clarify your objective in the survey it is possible to ask questions that are unnecessary and bog down your respondent. Create questions based on what you answered above. Doing so will yield the most useful data for the intended users of the results.

Case Example

- Problem: Low participation of women in STEM fields
- Program Objective: To expose high school age girls to coding and mentoring to encourage them to pursue STEM fields in University and beyond.
- Indicator: The indicator should describe the type of information that you need to determine success of your program. For instance:
 - Percent of participants who pursue STEM fields upon entering University.
- Survey Questions: You will design a survey instrument(s) with questions that can help you determine if your program has increased the number of females in high school will study in the STEM field.
 - To answer the indicator above, the survey should then ask: “If you are attending University upon completing high school, will you be pursuing a STEM field?” The respondent can then answer “yes” or “no” to this question.

WRITE UNBIASED QUESTIONS

The way that a question is written may influence how people respond, so the questions should be worded in an objective or neutral way - survey questions should never seem to advocate a particular answer (i.e., give a possible response to the respondent by putting an idea in their head or that leads a respondent to a conclusion different from what they would have honestly stated).

Biased Example:

My exchange program was excellent.

Strongly Agree Agree Neutral Disagree Strongly Disagree

The use of the word “excellent” is not neutral and may bias respondents in a way that makes them think more positively about an answer. A different way to ask this question to minimize potential bias would be below:

Unbiased Example:

“Overall, how would you rate your exchange experience?”

Excellent Very Good Good Fair Poor

ASK THE MOST IMPORTANT QUESTIONS AT THE BEGINNING OF THE SURVEY

Some respondents may begin a survey but not finish it. To make sure you collect the most important data, ask the most important questions at the beginning. For example, if you need to know whether a beneficiary increased her advocacy activities after participating in your exchange program, you would ask one or two questions about whether she increased her advocacy activities as a result of the program within

the first few questions. Less important questions might relate to future activities that your organization can offer to alumni – those questions should go at the end of the survey because they are not essential indicator data for reporting or program improvement.

MAKE QUESTIONS SPECIFIC

Be as specific as possible when asking questions in order to avoid confusing respondents. For example, a question included in a post-survey for a civic engagement-themed exchange program such as, “*Do you participate in civic activities and programs?*” is general and it will not likely provide you with the information needed for indicator data and program improvement.

However, you can obtain more useful information by asking more specific questions such as: “*How many civic groups do you participate in?*” or “*How many times per week/month do you participate in civic activities?*” These questions yield more precise, objective answers. Notice that these are separate questions and address only one topic.

ASK ONLY ONE QUESTION AT A TIME

Avoid asking multiple questions in one question. This can confuse respondents; increasing the amount of time spent answering one question may leave the respondent not wanting to continue with the survey or provide you with unclear responses or information that is difficult to analyze. Asking only one question at a time allows the respondent to provide you with valid and useful data.

For example, asking “*How many civic groups do you participate in and what is your role?*” can be confusing for the respondent and she is likely to respond to only one of the questions – resulting in incomplete data. By dividing this question into two parts, “*How many civic groups do you participate in?*” and “*What is your role when participating in civic groups?*” you provide the respondent with clear questions that she can answer individually, yielding more complete data.

USING SCALED QUESTIONS

There are several different types of scaled questions used in monitoring and evaluation and social science research. For the purposes of ease and appropriateness, this manual focuses on scaling responses and Likert scales, the two types of scales you are most likely to use. In general, scales are used to measure personality traits, behaviors or perceptions. They are a good way to determine respondents’ attitudes about a particular topic. When used correctly and without bias, scales are an excellent technique to ask beneficiaries to compare issues and provide feedback about their perceptions.

RATING SCALES

When developing scales, it is important to keep in mind that you create the scale and assign a definition to the value. Although all similar, there are three different types of rating scales:

- Numeric rating scales, where respondents are asked to rate a topic based on a set of numbers;
- Graphic rating scales rate, which look at behaviors or performance (i.e., leadership, teamwork, performance); and
- Descriptive graphic rating scale, where respondents are asked to place a mark along a line that depicts one extreme to the other. This type of rating scale allows the respondent to establish their own criteria-based on opinion – they are not restricted to values previously set by you.

When presenting respondents with a scale, the scale and the meaning of each option should be included in the question.

Example of a numeric rating scale

Telling respondents, "On a scale of 0-5, rate your proficiency" does not offer adequate direction and leaves room for differing interpretation across respondents and survey analysts. In this case, additional direction may include: 0- no experience/proficiency ..., 1-fundamental awareness (basic knowledge) ..., 2-novice proficiency..., 3-intermediate proficiency..., 4-advanced proficiency..., 5-expert proficiency. Respondents will then know exactly what each number means and you will receive more accurate responses.

Example of a Graphic rating scale

Please rate your leadership skills.

- 1- Poor
- 2- Adequate
- 3- Fair
- 4- Good
- 5- Excellent

Example of a Descriptive graphic rating scale

Mark on the line below your effectiveness as a leader:

1	2	3	4	5
Not effective at all				Very effective

LIKERT SCALES

Likert Scales offer a range of answer options — from one opposing perspective to another using either five, seven or nine options. For example, two opposing perspectives on a Likert scale might range from, "completely agree" to "completely disagree." Likert scales also need to include a neutral midpoint, "neither agree nor disagree" for respondents that do not hold a positive or negative opinion on a particular topic.

Example of one type of Likert Scale answer choices:

- 1—Strongly disagree**
- 2—Disagree**
- 3—Neither agree nor disagree (neutral)**
- 4—Agree**
- 5—Strongly agree**

Likert scales allow you to uncover degrees of opinion that can make a difference in understanding the feedback you are getting and can help you pinpoint the areas where you might want to improve your service or product. For example, if you ask an exchange participant whether or not they gained leadership skills, you might receive a response indicating the degree to which they gained those particular skills, providing you with valuable information as to how to prepare for the next cohort. Likert scales are not appropriate to use for measuring actual change in knowledge; only perceptions about change in knowledge.

MATCHING THE SCALE TO THE NATURE OF THE QUESTION

Sometimes, a survey will incorrectly use the same response scale types for all questions. Doing this may result in confused respondents and invalid data. For instance, a common error is using a five-point Likert scale that represents agreement, but framing the question in terms of quality, frequency, or amount. Matching the nature of the questions and response scales ensures that the questions are asked appropriately.

Example : I regularly use my leadership skills.
Strongly Agree Neutral Disagree Strongly
Agree Disagree

Comments: This is a question about frequency and cannot be answered with the traditional Likert-type scale because the definition of “regularly” differs from person to person. **For questions dealing with frequency, it would be better to use a scale that has the options: “always”, “often”, “sometimes”, “never”.**

It would be more appropriate to ask in the following way:

Please rate the extent to which you agree with the following question:
How frequently do you use your leadership skills?
Always Often Sometimes Never

Example 2: The pace of the workshop at the post-program orientation was satisfactory.
Strongly Agree Neutral Disagree Strongly
Agree Disagree

Comments: This is a question about preference and cannot be answered with the traditional Likert-type scale because if “disagree” is selected, you will not understand whether the respondent thought that the pace was too fast or too slow.

It would be more appropriate to ask in the following way:

The pace of the workshop was:
Too fast Too slow Just right

Another option would be to keep the question “the pace of the course was satisfactory” and change the response options to be “yes/no” and if the respondent selects “no” they will be able to explain why in an open-ended question.

Example 3: I sometimes find it difficult to see things from another person’s point of view
Strongly Agree Neutral Disagree Strongly
Agree Disagree

Comments: First, the use of the word “sometimes” is not a useful part of the question and will make your data analysis more difficult. If someone answers that they ‘strongly agree’, then that means it is sometimes difficult for them to see things from another person’s point of view; alternatively, if they answer ‘strong disagree’ then this means the same thing – that they sometimes don’t find it difficult to do so.

It would be more appropriate to ask in the following way (with either a multiple choice or fill-in-the-blank response):

TEST YOUR INSTRUMENT

Before administering a survey, take these actions to identify gaps and problem areas:

- Test the survey internally – take the survey yourself and ask some colleagues to take the survey. When doing so, keep these questions in mind:
 - Are the instructions clear?
 - Are any questions confusing?
 - Are any words confusing?
 - Are all jargon words defined?
 - Would you want or be willing to complete this survey?
 - Is there room for interpretation in any questions or answer choices?
- Test the survey with a member of the group that you will be surveying. Ask the tester to complete the survey keeping in mind the same questions above.
 - If you are unable to test the survey with someone from that specific group, consider asking another individual from that group’s demographic. For example, if you are testing participants that participated in a leadership training program, you might ask a recent alumnus of that program to test the survey.
- If your survey has to be translated, it is a good idea to ask the translator to also take the survey. This ensures that the survey vernacular translates to other languages. They can provide feedback to adjust questions that might need to be altered for either language or cultural differences.

SOURCES

Brinkerhoff, Robert O. The Success Case Method. Berrett-Koehler Publishers, Inc., 2002

Kirkpatrick, James D., Kayser Kirkpatrick, Wendy. Kirkpatrick's Four Levels of Training Evaluation. Association for Talent Development, October 2016.

Taylor-Powell, Ellen and Renner, Marcus. "Collecting Evaluation Data: End of Session Questionnaires" University of Wisconsin-Extension – Cooperative Extension. G3658 – 11. Program Development and Evaluation. (October, 2009)

Taylor-Powell, Ellen and Hermann, Carol. "Collecting Evaluation Data: Surveys" University of Wisconsin-Extension – Cooperative Extension. G3658 – 10. Program Development and Evaluation. (May, 2000)

Taylor-Powell, Ellen. "Sampling" University of Wisconsin-Extension – Cooperative Extension. G3658 – 3. Program Development and Evaluation. (May, 1998)

Wholey, Joseph, Hatry, Harry, and Newcomer, Kathryn. Handbook of Practical Program Evaluation, 3rd edition. Josey-Bass, November 2010.

"Web Center for Social Research Methods". Website:
<https://socialresearchmethods.net/kb/relandval.php>. William M.K. Trochim; 2006.