



COLORADO STATE UNIVERSITY
EXTENSION

How To Evaluate

How to Evaluate: Methods

Formal evaluations can take on different forms such as focus groups, interviews, and surveys. Because each of these methods can be delivered before, during, or after a program, your choice of method largely depends on “what” and “who” you are evaluating. Resource-intensive focus groups and interviews allow you to collect richer sets of qualitative data, whereas surveys allow you to collect a large quantitative data set relatively quickly and easily.

If using a survey, you have to choose between delivering it in-person, via mail, or online. This decision may be influenced by a number of factors, but one of those factors should be “who” you are evaluating. Seniors, for example may be less inclined to respond to online surveys whereas there is some evidence that the opposite may be true of youth (Olson, 2018). Online surveys tend to be more efficient because you do not have to enter paper survey responses by hand and because online survey platforms like Qualtrics can do much of the analysis and presentation of data for you.

How to Evaluate: Survey Validity and Reliability

With the utility and ease of surveys making them an especially common method of evaluation in Extension, it’s important that you consider validity and reliability when constructing them. **Validity** refers to the extent to which an evaluation actually measures what it intends to measure. A tape measure is valid, for example, because when you compare it to a ruler you can verify that it will measure the length, width, or height of an object according to accepted units of measurement (i.e. feet and inches). **Reliability** refers to the extent to which an evaluation measures information consistently. The tape measure is reliable because it will measure the length, width, or height of an object the same way every time (Michigan State University, 2018).

Using a sustainable energy example, asking survey respondents whether they applied for a utility rebate may provide reliable responses but this would not be a valid way of measuring implementation of cost-effective solar energy. This is because applications may be rejected, the rebate may have been for a different sustainable energy project, or applicants may not have followed through with a project. Conversely, asking survey respondents whether they installed cost-effective solar energy may be a valid way to collect information, but it would not be reliable unless a clear definition of “cost-effective” is provided.

Validity can be determined by statistical analysis, expert review, or comparison with other data. Reliability is often determined by administering the same survey questions at two different points in time and comparing results, but it can also be determined through statistical analysis. Note that validity and reliability are often determined by sets of survey questions, not just a single survey question. Use of whole surveys that have been peer-reviewed and published (referred to as *validated surveys*) can ensure that your surveys are valid and reliable. Such surveys can be found through 4-H Common Measures, the Journal of Extension, Google Scholar, and other means (Michigan State University, 2018). If no such

survey is appropriate and available, Extension specialists or CSU faculty may be able to help determine the validity and reliability of a survey.

Keep in mind that for many Extension programs, surveys should be *subjectively* valid and reliable but they don't always need to be *statistically* valid and reliable. The level of rigor that you use to determine survey validity and reliability should be a function of resource availability, stakeholder expectations, and utility (whether or not you want to publish results, for example).

How to Evaluate: Writing Survey Questions

In addition to ensuring surveys and the questions within them are both valid and reliable, a number of other considerations come into play when writing survey questions. These include the following:

- **Necessary and useful.** Only include questions that serve a distinct, defensible purpose in your evaluation.
- **Accessibility.** This means that potential respondents with different abilities and attributes should be able to answer the questions. In general, questions should be written at a 5th – 6th grade literacy level (Michigan State University, 2018). Having questions available in other languages that may be relevant to your potential respondents (i.e. Spanish) is also important.
- **Level of specificity.** Questions that are too general may not provide you with useful information and may be difficult to respond to accurately. On the other hand, questions that are too specific may not apply to many respondents.
- **Unbiased.** Questions should not lead a respondent to provide an answer they think you want to hear.
- **Sensitive.** It is important to avoid being too direct when asking questions, especially if you are asking about potentially sensitive topics. Only ask questions to which you think respondents will respond truthfully. When asking a potentially sensitive question, consider using response brackets that cover a range of responses (i.e. income between \$25,000-\$50,000).
- **Not double-barreled.** Avoid posing questions that cover multiple topics or traits simultaneously. For example, do not ask respondents if they agree that “the length and content of the workshop was appropriate” because they may have different opinions about the length vs. the content.

In some cases, you may also wish to clarify the **time parameters** of a question. For example, instead of asking whether someone has installed solar energy, asking whether they installed solar energy *in the last year or since your workshop* may help you better understand the context of their response (Trochim, 2005). As a practical matter, you may also want to find a balance of open-ended and closed-ended (i.e. multiple choice, scales, etc.) questions. While open-ended questions allow for input of any kind to come forward, they require much more work to analyze.

Once you have developed questions, you generally want to order them from easier-to-answer to harder-to-answer. Related to this, you also want to start with more general questions before inserting more specific questions (an approach called “funneling”). This provides some level of comfort to the respondent and may even help jostle information from the respondent as the survey progresses. If you are using an online platform such as Qualtrics, you can use “display logic” to only show more specific questions based on responses to the general questions. When moving from one topic to the next, consider inserting text that explains the transition clearly to the respondent. (“Now we’re going to ask you questions about your experience at the workshop”, for example.) Always keep in mind the Golden Rule when constructing surveys and survey questions: is this survey one that you would feel good about responding to (Trochim, 2005)?

How to Evaluate: Survey Response Types

There are many different ways that survey questions and responses can be framed, including multiple choice, ranked order, and open-ended. Each one of these question/response types allows for different types of quantitative analysis based on levels of measurement referred to as scales.

Nominal scales refer to scales in “name only”. For example, when asking “Did you apply for a utility rebate” you may assign numbers 1-2 to multiple choice responses of Yes or No. The numbers have no meaning but simply serve as identifiers. Coding responses to open-ended questions is another example of using a nominal scale. Nominal scales lend themselves to calculations of frequency and percentage, such as calculating the most common responses and the percentage associated with each response.

Ordinal scales refer to scales that are associated with a certain “order”. When asking respondents to self-identify an increase in knowledge of sustainable energy best practices from 1 (Very low) to 4 (Very high), you are using an ordinal scale. Ordinal scales lend themselves to calculating frequencies, percentages, and median responses such as the median for knowledge increase = 3.1.

Interval scales are similar to ordinal scales but also imply that the interval between all responses is the same. When asking respondents to identify their annual energy savings in dollars, you are using an interval scale. Interval scales lend themselves to calculate everything from the previously mentioned scales plus means and standard deviations. **Ratio** scales are the same as interval scales but include a potential value of true zero (a complete lack of the variable of interest) (Trochim, 2005).

In Extension, we often ask questions about level of increase, level of agreement, or level of satisfaction. While statisticians generally label these as ordinal scales, it is common practice to treat them as quasi-interval. This means that although the differences between Extremely Satisfied, Satisfied, Dissatisfied, and Extremely Dissatisfied may not be exactly the same (for example), you can still calculate means and standard deviations. Note, however, that ***analyzing means and standard deviations from quasi-interval scales may not be acceptable if trying to publish data in an academic journal***. Thinking in advance about what you want from the data and how you plan to use the data will help you choose appropriate response types for your questions.

Survey Response Rates

Surveys can be very useful in Extension work, but surveys with low response rates may be more limited in what they reveal from their data. (They may tell you about your respondents, but you will be less able to generalize conclusions to a broader population (Boyd, 2002).) A response rate is simply the number of responses divided by the number of people who received the survey. If you received 40 responses out of the 100 people to whom you had sent the survey, you would have a 40% response rate.

It is difficult to provide an expectation for survey response rates because they depend on many different factors. However, recent research has shown that online academic surveys are likely to receive higher response rates if they follow these recommendations (Saleh & Bista, 2017):

1. Request the aid of authority figures, known personnel, or organizations to the target population to distribute the survey when possible.
2. Target a population that is more likely to have an interest in the research.
3. Offer an incentive for completing the survey.
4. Make every effort to write a survey that is short and concise.
5. Inform the population of the approximate time it will take to complete the survey. (Qualtrics provides this, or you can time yourself completing the survey.)
6. Whenever possible, reduce the number of or eliminate open-ended survey questions.
7. Assure the participants of the anonymity and confidentiality of their responses.
8. Explain how the collected data will be handled, accessed, and stored (and/or disposed of) after the study is completed.
9. Personalize invitations to participate in the study and make them look professional.
10. Send at least one - but not more than three - reminders to the target population to motivate them to complete the survey.
11. Be aware of the time constraints related to time-of-year for the target population (i.e. farmers are busier during the growing season).