University of Wisconsin-Extension
Cooperative Extension
Program Development and Evaluation

Evaluation *Quick Tips*

Evaluation Designs Common in Measuring Outcomes

#36|2010

Evaluation design indicates when data collection will occur, which people or groups will be studied, how they will be selected, and what comparisons, if any, will occur. When measuring outcomes, the evaluation design provides the level of confidence you can have related to how much change has occurred and how much of the change is due to the program. Sometimes, *evaluation design* is confused with *evaluation method*. *Evaluation design* is concerned with from whom and when data will be collected. *Evaluation method* refers to how data is collected: survey, focus group, interview, observation, etc.

1. AFTER ONLY
   This is a one-time assessment conducted after the program is completed. It is a common design but, without a comparison or other data collection, you cannot be sure if the program had an effect on participants or whether other factors had an effect.

   Example: At the end of the Winter Leadership Camp, a questionnaire is given to participants to measure what leadership skills participants thought they gained and how they might use those skills in the future.

2. RETROSPECTIVE POST
   Conducted at one point in time, at the end of an event, this design asks participants to reference two points in time thus, providing a comparison. Participants are asked to report, their opinion, knowledge, attitude, skills, or behavior at the end of the program <u>and</u> to recall that same aspect prior to the program. This design is used when you are asking participants about their *perceptions* of their knowledge, skills, attitudes, behaviors, etc.

   Example: At the end of a Winter Leadership Camp, a questionnaire is given to participants. The questions are structured to get responses from participants about their leadership skills at two points in time – at the end of the program and before the program started. Results from the 'end of program' are compared to 'prior to the program' to determine the extent to which leadership skills changed.

3. BEFORE AND AFTER
   Data are collected at two points in time, providing a comparison to determine the extent of change resulting from the program. Program participants or situations are looked at before the program and then again after the program. For example, program participants take a pre- and then post-test or you might observe behaviors before and after the program. This design is used when you are measuring actual skills and knowledge, such as through a test or observations. While the difference between Time 1 and Time 2 should be the effect of the program, remember that often other things can happen to participants (and influence change) than just their participation in the program.

   Example: At the start of the Winter Leadership Camp, participants are given a questionnaire to *test* their knowledge of leadership concepts and attributes. At the end of camp, participants are given the same questionnaire. The scores obtained before the camp are compared to the scores on the post questionnaires to determine extent of knowledge change.

4. CASE STUDY
   The case study design uses multiple sources of information and multiple methods to provide an in-depth and comprehensive understanding of the 'case.' The case may be the program, one participant, or one site, for example. Its strength lies in its comprehensiveness and exploration of reasons for observed effects relative to the particular case.

   Example: One participant of the Winter Leadership Camp is selected (and agrees) to serve as a case study. The youth completes the Before-After questionnaire, keeps a journal and is interviewed about her camp experiences and leadership skill development. Her parents are interviewed before and after the camp about her leadership skills. Camp counselors are interviewed as well. All the data from the multiple sources are summarized to describe the extent and nature of change resulting from camp for a participant.

5. MIXED METHOD
A mixed method design involves the use of several data collection methods and the blending of qualitative and quantitative methods; at least, one qualitative method (designed to collect words) and one quantitative method (designed to collect numbers). Its purpose is to understand more comprehensively and to gain a more complete and full portrait of the program.

Example: A structured survey of all youth attending the Winter leadership Camp is conducted to assess leadership knowledge, skills and behaviors and a series of 12 in-depth interviews are conducted with participating youth and families to capture the major effects of attending camp.

Evaluation designs can be **strengthened by:**

- **Adding points in time.** Any one-time data collection is limited to that snapshot in time. You can strengthen these studies, by collecting data at other times. For example, you might add a one-month follow-up data collection to the after only design. Or, you might collect data on skill development *during* the program, possibly at several points over the course of a program.
- **Combining multiple methods of data collection.** For, example, you might use a questionnaire and also do some observations. Or, you might conduct interviews with participants as well as have them write a journal. This is moving towards a *mixed method* design.
- **Using multiple sources of information.** For example, you might collect information from youth participants and volunteers working with the youth. Or, you might solicit information from parents and teachers as well as youth.
- **Using comparisons (people, groups, sites)**. Adding a comparison of one or more groups, individuals, or sites can strengthen your design. Comparison groups refer to groups that are not selected at random but are from the same population. When they are selected at random, they are called *control groups.* The purpose of a comparison group is to add assurance that the program (the intervention), not something else, produced the observed effects. It is essential that the comparison be ***very similar*** to the program group.

  Consider the following possibilities as comparisons:

  - Between program participants and nonparticipants. Example: You collect data from youth participating in the 4-H animal science program and youth with animals but who do not participate in 4-H.
  - Between different participants (or different groups of participants) experiencing different types of program delivery. Example: In a youth financial literacy program, youth in two sections of the 8th grade participate in a simulation for learning financial management and the other two 8th grade sections receive non-experiential learning. At the end of the program, both groups receive the same retrospective questionnaire. Changes in the two sets of scores are compared to see what difference the two types of program delivery made in developing financial management skills of youth.
  - Between sites where the program operates and sites without program intervention. Example: Seven of the 14 4-H clubs receive a new type of leadership development course. You use the same evaluation procedure for all clubs and compare the results to see what difference the new leadership development course makes.

References:

Sabatelli, R.M., Anderson, S.A., 2005. *Assessing Outcomes in Child and Youth Programs: A Practical Handbook.* Revised Edition. University of Connecticut, School of Family Studies, Center for Applied Research, September 2005.

Weiss, Carol H., 1998. *Evaluation.* 2nd edition. Upper Saddle River, NJ: Prentice Hall.