

University of Wisconsin-Extension
Cooperative Extension
Madison, Wisconsin



**Program Development &
Evaluation**

Use Comparison Groups to Strengthen Your Evaluation

**Audio Conference
Friday, November 22, 2002
9:30-11a.m.
Heather H. Boyd**

Use Comparison Groups to Strengthen Your Evaluation

**Audio Conference
Friday, November 22, 2002
9:30-11a.m.
Heather H. Boyd
University of Wisconsin-Extension, Cooperative Extension**

Agenda

- 9:30 Welcome and Roll Call**
- 9:40 Overview of audio conference**
- 9:45 Definition of a comparison group**
- 9:50 When a comparison group can help strengthen your evaluation**
- 9:55 Time for Questions**
- 10:00 Types of comparison groups**
- 10:30 Time for Questions**
- 10:45 Practical Steps you can take in order to use a comparison group**
- 10:55 Steps for being recorded as having participated in the audio conference**



Today, we are going to talk about comparison groups – what they are and why you might want to use them in evaluating your programs.

The ideas presented here are some basic concepts. This audio conference is meant to be a starting point – an introduction to basic concepts. If you are able to complete the worksheet during or after this workshop, you are on your way to using comparison groups.

What is a comparison group?

Generally speaking, a comparison group is a group of people who:

- Have not participated in your program AND
- Are similar to your program participants AND
- Have characteristics that the research base demonstrates that are important relative to the intended outcomes of your programming

Why would I want to use a comparison group in my program evaluation?

Using a comparison group helps you address the question of whether your programming is “making a difference” at its most fundamental level – how your learners are different from people who are similar to your learners but have not experienced your UW-Extension educational programs – on outcomes of interest to your educational programs.

Evaluations that help gather the most credible evidence that a program is making a difference in the lives of its participants include a comparison group.

Example: Mandatory Parenting Course for Divorcing Parents

Let us say that you teach a parenting and communication curriculum that the family court officials in your county declared mandatory for parents who are in the process of divorcing. Your program assumes that improved communication with one’s about-to-be-former spouse about your children helps to make the transition smoother for everyone.



For the sake of this example, the central intended outcome of your programming is to reduce the occasion of post-divorce child custody litigation.

Even though attendance in the program is mandatory, some individuals attend and some do not. One natural comparison group for your program evaluation is the parents who are mandated to attend, but do not. This would be a good comparison group for your program evaluation, if you can show that the participants do not differ from the non-participants in any way that is important relative to your intended outcomes. If the non-attendees differ from the participants in any important way – as could be the case – this is not the best comparison group for you to use. For example, if non-participants do not attend because they are incarcerated, they would probably not be a very useful comparison group. The incarcerated parents might be very different than those who have not interacted with the penal system.

How do I determine the key characteristics that are important to the outcomes that I want to measure? In other words, how do I know what similar characteristics my participants and non-participants need to have?

This is what the research literature is for, and it is related to Extension's niche in community-based education. Principles of your curriculum probably use university-based research – research that describes characteristics of people who experience specific outcomes you intend. If you are not sure what these principles are in your field or for your curriculum, make an effort to find out. Read the literature in your field – both academic journals and the writing of practitioners. Ask others who teach curriculum similar to yours what assumptions they based their curriculum on and why.

I can see that my programming is making a difference. Why would I care if I have a comparison group or not?

Others who scrutinize your programming might find the results of a program evaluation that uses a comparison group to be persuasive, credible evidence for the worth of your program. Moreover, if those others make decisions about funding or otherwise supporting your programs, you probably want to have strong, credible evidence that your programs matter. Keep your Extension education committee and your county board in mind as groups who might be interested in the results of a comparison group study.

You might also want the information when you apply to funding from different agencies, and you yourself might be interested in using the information in marketing your program to other people in your county.

How do I “control for” all the factors that affect the results of my program?

For the most part, you cannot control for all things that could affect the outcomes you are looking for. Even when you do use “control” characteristics in your evaluation, most people acknowledge that the control is theoretical. However, you do need to try to



measure the most important ones – the ones that are linked to the outcomes or results that you intend. If you do not measure the most important ones, someone who knows your content area could easily find this flaw.

What can a program evaluation – whether I use a comparison group or not -- probably NOT do for me?

Program evaluations, even ones that use comparison groups, generally do not:

- *“Prove” that your programs make a difference*
Program evaluation can help provide credible evidence that your programming **could** be making a difference for your participants and for your community.
- *Find that your program is the “cause” and the outcomes you see are the “effects”*
Experimental design might be able to do this for you, but even then, questions can remain, depending on details of the design.
- *Turn your harshest critics who have made up their mind about the worth of your programming into your supporters*
Consider funneling your best energies into repairing and improving personal and working relationships rather than incorporating comparison groups into your evaluations. In some cases, you should both improve relationships and incorporate comparison groups into your evaluation design.

Four Types of Comparison Groups

There are four most useful types of comparison groups for program evaluation:

- Method One – Select participants and non-participants randomly
- Method Two – Use your computer to control for differences with regression analysis
- Method Three – Match participants and non-participants on key traits
- Method Four – Give your learners a pre- and post-test

First, let us focus first on the groups that are most likely to help you gather the most credible information that your program “makes a difference” for your program participants. Then, we’ll discuss practical ways you could use comparison groups for your community education programs.



Method One – Select Participants and Non-Participants Randomly

Design an experiment that uses random assignment to choose who participates in your program and who does not.

This design helps to collect some of the most convincing evidence about the difference a program can make because each person has an equal chance of being selected for either your group of learners or your non-participants. Random selection helps protect against choosing only certain people – such as the ones that really need the most help where you might show the most improvement -- or any other group that could lead to biased results.

Random assignment is easy to do and takes only a few simple steps with a random numbers table. As an alternative, go to www.randomizer.org for an easy to use tool for drawing your potential participants at random.

If using random selection seems inappropriate or uncomfortable to use because you want everyone who is interested to participate in your programs, you have options. You could include those who were not selected to participate in the current offering to participate in an offering that takes place at a later date .

Using this method might be very practical if you want to test something such as the best instructional design and delivery methods for similar curricula. For example, let us say that you have a dairy modernization curriculum that is available as face-to-face instruction as well as on a CD-ROM, and that you want to test which delivery method is best for the learning outcomes you intend. You could randomly assign participants to one delivery method or the other if you knew that they were similar on other characteristics related to the outcome variables in which you are interested.

Method Two – Use Your Computer to Control for Differences with Regression Analysis¹

If you have the appropriate training and software, such as SPSS² or MicroSoft Excel³, this is an easy method. You tell the computer that there are certain variables (the traits of your program participants and non-participants) that you want to treat as “control variables.” Run an analysis where you “control for” the influence of these variables and determine what sort of difference participation in your educational programming makes once these other influences are controlled. You might be familiar with this process if you have ever done a regression analysis.

¹ Regression Analysis attempts to answer, “What values in the dependent variable can we expect given values of the independent variables?” An independent variable is an influence that is thought to be associated with or cause change in a dependent variable. One example could be that your teaching program is an independent variable that is associated with change in the outcome you intend.

² An annual license for this statistical software is available for \$150. You can order this through DoIT software sales on the WISPLAN web page at <http://cf.uwex.edu/ces/wisplan/software/>

³ MicroSoft Excel is available in MicroSoft Office software. Your computer may already have a copy of Excel.



Let us take the example of a group of elected officials in your county who you surveyed about their comprehensive land-use planning process. On the survey, you ask them about the degree of exposure they have had to your programming and you ask them about your intended outcome – informed decision-making – in whatever way you have defined it. From the survey, you know whether or not they have participated in your programming, in what ways and how they have made decisions about land-use planning for the level of government they represent.

For the sake of this example, the research literature tells you that an influence on informed decision-making in land-use planning is someone’s environmental attitudes. Let us also say that the research tells you that the level of government that an elected official represents affects their decisions about land-use planning. Your computer software could help you “control for” the influences of environmental attitudes and level of government represented relative to informed decision-making. Your results would then tell you about the “pure” effect of your educational programming – at least as well as you have measured it in this study.⁴

Method Three – Match Participants and Non-Participants on Key Traits

Match participants and non-participants according to key traits that the literature demonstrates may influence your outcomes of interest (age, sex or education, for example, depending on the program you teach). Then, compare the two groups on your central outcome. Key traits could vary by the learners and the program. That is why you need to be familiar with the literature in your field.

It might seem difficult to find groups to compare to your learners, but if you think creatively, some pre-existing groups may be readily available. Assuming that these groups are not different from your learners in some important ways, some natural groups for a comparison group may include:

- Applicants you accepted to your program but who decided not to participate – if you can provide information that says that the decision to not participate is not related to an important difference in the groups
- Program participants from previous programs
- Program participants from educational programs offered by other agencies
- People in your county who are similar to your program participants but have not participated in either your or similar programs

For example, if you are a 4HYD educator who teaches a communication curriculum about public speaking, you might be able to compare the youth in your communication programs with youth in other public speaking programs.

⁴ Assuming that there are no intervening variables or confounding or spurious relationships, or that you have already accounted for these in your data set.



Practically speaking, Methods Two and Three attempt to simulate Method One. Methods Two and Three are designed to be as much like an experiment as possible, but they do not use random assignment. They are called “quasi-experimental” design and they try to come as close as possible to random assignment by minimizing pre-existing differences between the two groups that are being compared.

This last method is not technically a comparison group that includes people that have not participated in your offerings, but it could help you to make comparisons.

Method Four – Give Your Learners a Pre- and Post-Test

With this method, you are comparing your learners to themselves, which is not using a true comparison group. You are really just comparing your learners “before” and “after” they have experienced your programming by using a “pre-” and “post-” test design.

There are many legitimate critiques of this method – or any other method that examines only one group such as learners. A few of these critiques include:

- The effect of history – How do we know that an external influence did not help to create the change in the intended outcome?
- Maturation – How do we know that the participants would not have changed in this way anyway without experiencing your programming?
- Testing – How do we know that participants have not improved on their outcome measures just from having taken the test before?

Using a pre- and post-test might be the easiest and most practical way for you to evaluate your programming. Just be aware that if you choose it, you may have to contend with these types of critiques.

Timing for pre- and post-tests can occur at almost any interval. What assumptions are you using about when your participants might change in their learning or behavior? Possible examples include testing:

- Immediately before and immediately after a workshop
- A baseline measure one year before a program and a post-test measure one year after
- Before and after an exercise within a workshop



Additional reading:

Wholey, J., Hatry, H., & Newcomer, K. (1994). "Constructing Natural Experiments" in *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass Publishers.

Human Resources Development Canada. *Quasi-Experimental Evaluation*. Retrieved September 19, 2002 from <http://www11.hrdc-drhc.gc.ca/pls/edd/QEE.shtml>

