

Data Gathering Designs to Evaluate Educational Programs¹

C.L. Taylor²

A SOUND DESIGN

Ross (1989) states that the results of an educational program may be (1) the effects of the program, (2) the effects of extraneous factors that the educator has no control over, and (3) the effects of the design, such as a faculty sampling procedure. The educator, seeking to determine the effects of the educational program, needs to establish that other factors do not explain its results.

A sound data gathering design makes it very unlikely that uncontrolled influences produce a program's results. For example, a sound design determines if 4-H'ers knowledge of citizenship is the result of the 4-H program or of some other factor or program; a sound design determines if a client's knowledge about energy conservation in the home results from the Learn by Mail series "Energy Efficient Home" or from other factors.

Sound design provides quality control, which is as important in gathering data for evaluation as it is in other aspects of social science research, for it increases confidence in the conclusion.

DATA GATHERING DESIGNS

The following are some common data gathering designs used to evaluate programs:

- Experimental

- Nonrandomized Pre/Post Test
- Time - Trend
- Before and After
- After Only

Experimental Design

The experimental design requires both a **control** and an **experimental group**, with random assignment to each. This is the top of the line in data gathering designs. It is rigorous. It is designed to rule out the effects of everything but exposure to the educational program.

An **education program** is given to the experimental group; **no** program is given to the control group. **Random** assignment—each member equally likely to be placed in the control or experimental group—rules out the possibility that prior differences in groups could account for the difference between groups after exposure to the educational program.

The experimental design **accurately assesses** a program's impact. However, it is **costly** in time and other resources, more costly than most educators can afford. Furthermore, even if a control group is available, there are certain ethical concerns about providing educational programs for certain groups of people and withholding the information from other groups. There are many practices in the natural or physical sciences where research has already

1. This document is Bulletin PE-25, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. Publication date: March 1993.
2. C. L. Taylor, Professor and Extension Specialist, Cooperative Extension Service, Institute of Food and Agricultural Sciences (IFAS), University of Florida, Gainesville FL 32611.

The Institute of Food and Agricultural Sciences is an equal opportunity/affirmative action employer authorized to provide research, educational information and other services only to individuals and institutions that function without regard to race, color, sex, age, handicap, or national origin. For information on obtaining other extension publications, contact your county Cooperative Extension Service office.

Florida Cooperative Extension Service / Institute of Food and Agricultural Sciences / University of Florida / John T. Woeste, Dean

established the results. For example, it is not necessary for educators to attempt to evaluate the yield of a new variety, since the researcher has already established it. It is the educator's role to evaluate the rate of adoption of this new variety.

Non-Randomized Pre\Post Test

Ary (1985) discusses the nonrandomized control group pretest/posttest design, which has applications for many educational program situations. The procedure is rather simple. Participants are not randomly assigned to their groups. Both groups are pre and post tested, but only one group receives the educational program.

Time-Trend Design

The time-trend design has practical application to many educational programs. It is not as rigorous as the experimental design; using it, scientific proof cannot be established.

Time-trend design requires measurement **before**, **during**, and **after** the educational program. Control groups are not used. The measurements taken before the educational program is started become benchmark data and, in effect, the control, since the control groups are not used.

As an example, consider the following program objective:

Individuals will learn basic principles of investment, estate planning, and budgeting to cover expenses and allow for increased savings.

One of the educator's tasks is to evaluate the extent to which savings increased. So the educator establishes, from a representative sample of the entire target group, the amount of money in savings. After conducting the educational program, the educator periodically re-assesses the amount of money in savings and compares it with the original balance.

From the difference between the final assessment, or the final amount of money saved, and the original savings balance, the educator subtracts the **trend**. The trend refers to what would have happened without an educational program; it represents the predicted or expected change resulting from other influences, and so must be subtracted from the final results, the last savings balance. The difference between the final assessment and the original data

(minus the trend) is the result of the educational program.

The time-trend design is appropriate for **long-range** programs, and it requires **no control** groups. However, the extra time necessitated by the repeated data collections permits factors other than the educational program to influence results. These other factors include such changes as clients' maturation, economic or social shifts, and other instabilities of data.

Before and After Design

The before and after design usually refers to measuring or collecting baseline data immediately before a program starts and at the program's end or shortly thereafter. The amount of time that elapses between the pre-data collection and post-data collection is determined by the nature of the educational program. The before and after design has no controls.

Consider an example of use of the before and after design. To assess the effectiveness of a week long in-service training in nutrition for Extension Home Economics, an educator gave a pre and post test. One question revealed the following program results.

- Breast milk contains all the nutrients needed by a baby for at least the first three months of life.

Percent Correct Responses? See Figure 1.

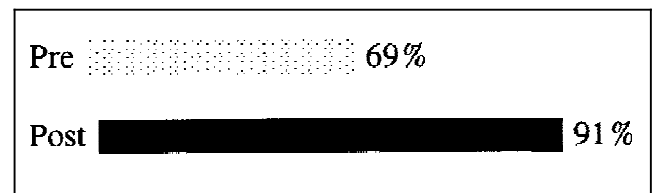


Figure 1. Results of the pre- and post-test.

Consider another example. To assess the impact of an educational program to increase acres of improved pasture, an agriculture agent determines the number of acres before and shortly after the educational program has been conducted (which is probably after establishment season), and compares them.

In this example, there are certain internal validity problems. If acreage of improved pasture had increased, other agencies could have had a major

influence. This may be controlled to some degree in various ways, such as, comparison with the trend and asking a sample of those who improved their pasture who or what influenced them to add new acres.

After Only Design

This design operates with **no baseline** data and **no controls**. It is simple to conduct and rather economical. However, it provides only **weak** conclusions about the impact of the program. Generally it yields data or information suitable **for descriptive** purposes.

A common example of the after only design is the questionnaire handed out at the end of a meeting asking participants' (1) what their reactions were, (2) how much they felt they learned, and (3) how valuable the information will be to them.

A DESIGN SHOULD MEET CERTAIN CRITERIA

Two major criteria are important in an evaluation data gathering design: **internal validity**, which refers to the extent to which the design allows the effects of the educational program to be accurately determined, and **external validity**, which refers to the extent to which the data can be generalized across time, geography, environment, and human involvement. For Extension's purposes, internal validity is of more concern than external validity.

There are various threats to internal validity.

1. **Time:** Other educational programs and other influences occur between pre-test and post-test, and thus provide alternative explanation of the effects.
2. **Maturation:** Processes within the clients or animals produce changes, such as growth or fatigue or expanded capability.

3. **Instability:** Instability refers to fluctuations in sampling of clients, changes in regulations etc.

4. **Testing:** Taking a test may effect the scores on a second test.

The data gathering design is stronger—the evaluation's conclusions more reliable—when it is valid internally.

CONCLUSIONS

What designs are appropriate to gather data for evaluating County Extension activities and programs?

Data gathering designs that require controls seem somewhat impractical to use with most Extension programs, for at least two reasons. First, Extension is usually interested in evaluating annual or long-range programs, as opposed to evaluating a specific method or technique. This is because Extension is not accountable for how the job is done, but for the consequences of doing the job. Second, Extension releases information by mass media and many times there is no group to become the control.

Therefore, the before and after and after only designs seem to be the most practical ways to evaluate the results of Extension programs.

REFERENCES

- Ary, Donald, Jacobs, L. C., and Razovieh, Asghar, *Introduction to Research in Education*, Holt, Reinhart and Winston, 1985.
- Rossi, Peter H., and Freeman, Howard E., *Evaluation: a Systematic Approach*, Sage Publications, London, 1989.