



**SAMPLE MATERIAL: Tools, Tips, and Common Issues in Evaluation Design Choices**

**Purpose:** When you are trying to figure out the most rigorous evaluation design that is also appropriate for your program, these summaries of various evaluation designs may be useful. Table 4-1, “Sample Evaluation Designs,” presents the advantages and disadvantages of a variety of evaluation approaches.

**Source:** Excerpted from Sue Allen and Pat Campbell’s “Chapter 4: Tools, tips and common issues in evaluation experimental design choices” (pp. 31-34). In Friedman, A. (Ed.). (2008). *Framework for evaluating impacts of informal science education projects*. The entire guide is available at <http://caise.insci.org/resources>. Last accessed December 18, 2008.

## CHAPTER 4 TOOLS, TIPS, AND COMMON ISSUES IN EVALUATION EXPERIMENTAL DESIGN CHOICES

---

*Sue Allen*

*This chapter deals with several aspects of evaluation that are common to most projects. First we look at some of the choices you can make in evaluation design to gather the impact data your project needs. Experimental designs are very powerful for some purposes, but often present difficulties in ISE settings. We'll discuss those methods first, and then summarize the broader array of strategies you can consider. In Part II, you will find examples of most of the choices described here, including both experimental designs and naturalistic methods.*

### EXPERIMENTAL DESIGN CHOICES

---

Sue Allen

As noted in Chapter 2, experimental study designs are not necessarily more appropriate than naturalistic ones; rather, you should use the most rigorous study designs that are best suited to the nature of the project and its intended outcomes. When a project can consider conducting experimental designs with representative sampling, the following are study designs worth exploring:

- (a) Randomized controlled trial: One theoretically powerful experimental design is the randomized controlled trial (sometimes called “randomized clinical trial,” “RCT,” or “true experimental design”). It is a pre-post study with comparison group, in which participants are assessed before and after experiencing the project materials, and their learning is compared with that of a control group who were also assessed twice but without experiencing the materials. Ideally, audience members are randomly assigned to these two groups.

This design is often summarized as:

R OXO

R O O

where the “R” indicates random assignment, the “O’s” indicate an assessment (often called a pre-test or a post-test), and the two rows indicate that one group experiences the project materials “X” between their pre and post-test, while the other does not.

If properly implemented, this design rules out many competing possible causes of audience members’ learning (such as practice with the assessment), but it is expensive, and potentially taxing for the audience members – especially those who are told they cannot immediately see

the exhibition / film / other deliverable and are then assessed twice for no obvious reason. This study design is rarely used in practice to assess audience member's learning in informal environments.

- (b) Randomized post-only design: An equally powerful design that is somewhat more feasible in many ISE projects is the randomized post-only design. As in the RCT design above, participants are randomly assigned either to a group that experiences the project deliverables or a group that does not experience them (at least, not until after the study). All participants are then assessed once, and any differences are attributed to the effect of the project materials.

This design is often summarized as:

R XO  
R O

This kind of study is somewhat less expensive and taxing to participants than the RCT design, but requires assessing a relatively large number of participants, and it requires that they be randomly assigned to the control or treatment groups (which may be unrealistic). Random assignment can sometimes be achieved by recruiting audience members who are willing to experience two sets of materials (say, exhibitions) in any order; they can then be randomly assigned to see the target exhibition either first or second.

- (c) Using comparisons where possible: It is often feasible to provide evidence of impact with at least some form of comparison. Some kinds of assessment, such as direct questions, card-sorting tasks, or concept maps, may be used before and after participants experience the project's materials (in a pre-post study design without a control group). Alternatively, participants' responses after their experience may be compared to a group of participants who have not yet had the experience, matched if possible by key demographic descriptors such as age and education level (if random assignment is not realistic).

Failing such direct comparisons, it may be possible to compare the measured indicators of participants' learning to rates reported in other literatures, such as front-end evaluation studies, summative evaluations of similar exhibitions, misconceptions literature in science education, etc. Such benchmarks can provide at least some sense of the degree to which a project has been effective as an aid to learning.

Comparisons also strengthen evidence of learning when process-based measures of learning are used. For example, if a particular exhibit engages museum visitors in asking their own questions, this becomes a stronger form of evidence if the frequency or quality of visitors' questioning is compared to that of visitors at "typical" exhibits or in other kinds of settings.

- (d) Cases where comparisons are unnecessary: Some kinds of evidence do not require a comparison to be compelling, particularly when a plausible case can be made that

participants could not already have had the knowledge at the time they experienced the project's materials. Examples are:

1. visitors figuring out an exhibition's main idea(s);
2. viewers making connections between a TV program and their own lives;
3. professionals remembering their experiences in a workshop and their responses over time;
4. visitors sharing something they know that is inconceivable for them to have known previously (such as pieces of information uniquely displayed in an exhibition);
5. participants self-reporting that they had not previously realized something.

Finally, a few notes of caution when planning study designs:

- When planning experimental studies, there is often a trade-off between rigor of the design and the authenticity of the situation being studied. For example, it may be possible to design a fully randomized controlled trial, but the implementation may require that learners be constrained in ways that significantly undermine the informal, free-choice nature of their experience. Such design choices should be made thoughtfully, in consultation with an evaluator early in the project.
- While we value rigor, in experimental designs, case studies, or naturalistic observations, it is even more important that participants not be traumatized or alienated because of over-zealous assessment practices (which would also lessen validity of the results). Evaluators should therefore pilot-test their methods and be sensitive to participants' emotional responses.
- Irrespective of the rigor of their study designs, evaluators should be careful not to over-interpret their data or over-generalize their claims, lest they lead to misguided or simplistic policy decisions that may adversely affect learners in other settings.

*The next section of this chapter summarizes a broad selection of possible evaluation design choices, including the ones discussed in some detail above.*

---

## AN ARRAY OF EVALUATION DESIGN CHOICES

---

Pat Campbell

There are a number of designs that can be used in the evaluation of your program or project. Table 4-1 provides an overview of many of the designs including some of their advantages and disadvantages.

**Table 4-1: Sample Evaluation Designs**

Study Type	Design	Representation (X= treatment; O=measures/evidence; R=random assignment)	Advantages	Disadvantages
Quantitative Case Study	One-shot Post-test only Design	X O	Takes fewer resources Can present a “snapshot” of a point in time	Doesn’t look at change
Quasi-experimental Study	One-shot Pre-test- Post-test Design	O X O	Looks at change over time	Other things besides treatment could be causing change
Quasi-experimental Study	Post-test Only Intact Group Design	X O O	Compares to another group	Doesn’t control for any initial differences in groups
Quasi-experimental Study	Pre-test- Post-test Intact Group Design	O X O O O	Allows statistical control for possible extraneous variables	Doesn’t control for any effect of testing
Experimental Study	Post-test Only Design With Random Assignment	X O R O	Controls for pre test effects Random assignment reduces the chances of extraneous group differences	Random assignment is often not possible in evaluation Doesn’t control for extraneous variables
Experimental Study	Pre-test- Post-test Design With Random Assignment	O X O R O O	Allows statistical control for possible extraneous variables	Random assignment is often not possible in evaluation. Doesn’t control for any effect of testing
Experimental Study	Solomon Four Group Design	O X O R X O O <sub>a</sub> O <sub>b</sub> O <sub>b</sub>	Strongest quantitative design controls for all possible extraneous variable	Random assignment is often not possible in evaluation Very resource intensive
Quasi-experimental Study	Time Series Design	O O X O O	Looks at longer term change	Doesn’t control for extraneous variables
Ethnography	Participant observer examination of group behaviors and patterns	NA	Explores complex effects over time	Resource intensive Story telling approach may limit audience Potential observer bias
Case Study	Exploration of a case (or multiple cases) over time	NA	Provides an in-depth view Elaborates on quantitative data	Limited generalizability
Content Analysis	Systematic identification of properties of large amounts of textual information	NA	Looks directly at communication Allows for quantitative and qualitative analysis	Tends too often to simply consist of word counts Can disregard the context that produced the text
Mixed Methods Study	Use of more than one of the above designs	NA	Can counteract the disadvantages of any one design	Requires care in interpreting across method types.

Adapted from: Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963). Gary Ingersoll, *Experimental Methods (in Encyclopedia of Educational Research (Fifth Edition))*; Harold Mitzel ed. New York: The Free Press 1982. pp 624-631. Lydia’s Tutorial *Qualitative Research Methods* <http://www.socialresearchmethods.net/tutorial/Mensah/default.htm> Accessed April 15, 2007. Writing@CSU <http://writing.colostate.edu/index.cfm> Accessed April 15, 2007.