Overview of Evaluation Design

Design: The evaluation design provides an outline of how the evaluation will be conducted, laying out the sequence of program implementation and data collection. In its most basic form, the evaluation design tells when information is collected and from whom. It also answers the question (about evaluation findings): "compared to what?" Choice of design is keyed to the evaluation question, and should also take into consideration lifecycle stage and feasibility. In choosing an evaluation design, keep in mind that there is a tradeoff between the strength of evidence and the feasibility of collecting and analyzing data. Try to document the greatest causal effect possible without going beyond what's appropriate to your program's lifecycle stage. In reporting results, always keep claims in line with the evaluation design used to generate results. For example, when claiming effectiveness, make sure the design supports such a claim by including measures of change, not just post measures, and a strategy that can rule out other possible causes, such as a comparison group.

One way of communicating your chosen design is the "tic-tac-toe method" where **X**'s stand for programs or activities and **O**'s stand for observations (data collection points). When multiple observations or measures are used on one occasion (e.g., at the end of the program) you can use subscripts to distinguish among measures.

For example:

Type of Design	**"Tic-tac-toe" representation**
Post-only	X O
Pre-Post	O X O
Longitudinal	O X O O O O
Mixed Methods	O_1 O_2 X O_1 O_2

As long as your Design description is concise and understandable, it can be represented in whatever form you are most comfortable. The next page provides a grid showing common evaluation designs (in the "tic-tac-toe" format) and their implications.

Implications of common evaluation designs[§]			
Design	**Description**	**Advantages**	**Disadvantages**
A. Post-only X O	Collect data after a program occurs	Provides a snapshot or cross-sectional view at a particular point in time	Doesn't show change over time.
B. Post-only with comparison group X O O	Collect data after a program occurs from both participants and a comparison group	Allows for comparison with another group	Doesn't distinguish program effect from possible effects of initial differences in groups.
C. Post-only with control group and random assignment R X O R O	Randomly assign eligible individuals to either program or comparison group; measure after program participation for both groups.	Random assignment is most effective way to eliminate effects of any possible systematic bias in selection into program. Post-only design eliminates any bias which may be introduced by pre-testing.	Random assignment often not possible in program evaluation. Doesn't distinguish program effect from effects of outside factors or from effects of testing.
D. Retrospective post-then-pre X OO	Collect data after program participation; ask participants about both post and pre states.	Eliminates "shift bias" which occurs when participants shift in their understanding and interpretation of the same questions.	Risks introducing recall bias.
E. Simple pre-post* O X O	Collect data both before and after program participation	Shows change over time	Doesn't distinguish program effect from effects of other factors that could be causing change. (However, other potential competing theories may be ruled out in other ways)
F. Pre-post with comparison group O X O O O	Collect data before and after program participation from both participants and comparison group	Combines strengths of B. (pre-post) and C. (comparison group) to show change and help distinguish program effect from effect of some other possible influences on outcomes.	Doesn't eliminate possibility of bias in selection into program and/or comparison group.
G. Pre-post with control group and random assignment R O X O R O O	Randomly assign eligible individuals to either program or comparison group; collect data before and after program participation for both groups.	Combines strength of E. (random assignment) with strength of F. (pre-post with comparison) for best possible isolation of program impact from other factors that might influence outcomes.	Random assignment often not possible in program evaluation. Doesn't distinguish program effect from effects of testing.

[§] *Excerpted and adapted from Sue Allen and Pat Campell'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. A report from a National Science Foundation Workshop.*

*Pre-post designs may be: (1) matched at the individual level (2) matched at the group level or (3) unmatched.