LAURA COLOSI
RACHEL DUNIFON

What's the Difference?
"Post then Pre" & "Pre then Post"

The act of evaluating a program inevitably uncovers conflicting goals: the need for as much rigor in methods as possible while at the same time designing an evaluation that interrupts program delivery as little as possible. It is important to rely on rigorous evaluation design and methods to not only report the outcomes of a program most accurately, but also to lend as much credibility to the results as possible. On the other hand, it is important to take as little time as necessary in order to maintain program participants' trust and comfort in the program setting (Hill and Betz 2005).

Evaluation Designs

The most widely used evaluation design is a traditional pre then post test, where participants are asked a series of questions at both the beginning of a program (pre test) and then again at the program's completion (post test). This design is believed to measure changes in participant knowledge, attitudes, or behaviors regarding whatever the program content is (e.g., disciplining children). In general, evaluators prefer measurement at two time intervals to accurately detect any participant changes from the program participation. There are of course, many variations and permutations of the pre/post design that further strengthen the validity of such an approach, that are not discussed in this brief.[1]

Validity: what is it?

Validity is the strength of our conclusions, inferences or propositions. More formally, Cook and Campbell (1979) define it as the "best available approximation to the truth or falsity of a given inference, proposition or conclusion." In short, were we right? For example, if we are studying the effect of a parent education program designed to teach appropriate and effective discipline for children, and we saw that parents did change their discipline practices at home, we want to make sure our conclusion that there is a relationship between our treatment (discipline program) and our observed outcome (parents behavior at home) is valid, or true.

Threats to validity: Pre/post design

Although early evaluators understood some inherent threats to the validity of the traditional pre and post test approach to evaluation, it was the identification of "response shift

---

[1] For example, see Shadish, Cook and Campbell (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference or *Center for Social Research Methods* at http://www.socialresearchmethods.net

bias" by George Howard in 1979 that captured the biggest weakness of this widely accepted approach. Response shift bias is described as a "change in the participant's metric for answering questions from the pre test to the post test due to a new understanding of a concept being taught (Klatt and Taylor-Powell 2005)."

For example, if a program is designed to teach parents effective discipline for children and some participants don't know that spanking a child is considered inappropriate, they may report at the beginning of the program (pre test) that they are engaging in age appropriate discipline practices at home. However, if over the course of the program they begin to understand the relationship between age of a child and the "appropriate" discipline techniques, they may also report that they understand age appropriate discipline on the post test – and thus, their evaluation data would reflect no program impact at all. If, on the other hand, parents believed they knew appropriate discipline at the time of the pre test but then realized over the course of the program that they had more to learn, they may report less confidence on a post test in their parenting practices, and thus, actually generate evaluation data that appears to show a negative impact of the program. Overall, criticisms of the traditional pre/post test design are that one cannot account for response shift bias, and as a result, this method is more likely to either underestimate the program's effect on participants.

These criticisms of the traditional pre/post design led to the use of a "retrospective pre test" – sometimes refereed to as a **post then pre** design. This type of design is popular because it is implemented at only one point in time – often on a single instrument that queries participants about a given topic "then" (pre-test) and "now" (post test). The theory behind this design is that by testing what participants believe about program content after program completion, their standard of assessing the changes in knowledge, skills or attitudes is consistent, and thus, not subject to a response shift bias (Rockwell & Kohn 1989; Davis 2003). Others note that the post then pre design reduces incomplete data sets (Raidl 2004) and is convenient to administer given the time constraints many program providers face, as well as easier for program participants to complete (Lamb 2005).

It is important to note that many of the threats to validity for either a pre/post or retrospective evaluation instrument can be minimized by paying attention to the wording of each question asked on a survey. For example, if you want to determine whether or not parents exhibit appropriate discipline, approaches, asking a specific question such as "how often do you spank your children?" generates a more accurate measure of *actions* parents take at both program beginning and end, for either a pre/post or post then pre design. This is in contrast to asking more general statements such as "it is appropriate to spank a child over the age of two." This type of question item is subject to response shift, as program participation in and of itself will influence parent's responses, or more likely, parents may not know it is inappropriate until the program's end, and will skew the results as discussed above. More specific questions regarding actions are not as subject to problems of response shift.

Threats to validity: Post then Pre design

Just as the traditional pre/post evaluation method has some limitations, the post then pre method has limitations as well (Lamb 2005; Hill and Betz 2005; Shadish, Cook and Campbell 2002). Hill and Betz (2005) identify several threats to validity using the post then pre method:

- *Recall*: the inability to accurately recall attitudes and behaviors held in the past;
- *Social desirability bias*: the need for people to report change or improvement to fit program expectations or to inflate perceived improvement on those items that are most important to them personally;
- *Effort justification bias*: occurs when participants report improvement (many times subconsciously) to justify the time and energy they have invested in program attendance; and
- *Cognitive dissonance*: occurs when participants report improvement even if it did not occur, to meet their own expectation that they *should have* changed. In other words, if parents expected to change as a result of participation, but did not, they will report a change to resolve an internal conflict and put their mind at ease.

Another important issue to consider is that use of a post then pre design means that evaluation data is only collected from those participants who complete the program, and not from the entire group of participants. This is important because the lack of information gathered from parents who do not complete the program does not allow for any examination of attrition information -- who drops out of a program, why they may have withdrawn, or what content area they may have missed. As a result, evaluations that rely solely on a post then pre design may over-inflate program success by the mere fact that those sampled are only those that successfully complete the program. In terms of program improvement, the data used to make changes is incomplete and relies only on feedback on those who found the program worthwhile or for some other reason were able to complete the program.

Why do we care about validity?

Validity is an important concept to address in any evaluation endeavor. First and foremost, evaluators should take all steps possible to ensure that they are measuring what they intend to measure. Secondly, the more credible the results are, the more useful the evaluation data is to both measure outcomes and improve program delivery or content, and ultimately will better serve those who attend programs for help with a given issue.

Control Groups

It is important to note that perhaps the best way to address the shortcomings of the pre/post is to utilize a control group. In this design, program participants complete a pre-test at the beginning of a program, and a post-test at the end. Additionally, a control group of people not participating in the program also complete a pre-test and a post-test, taken at the same time interval as that of the program participants. In some cases, the control group could be drawn from the waiting list for a program, for example, and should consist of individuals who are as similar as possible to those in the program. For each group, a measure of change is computed by comparing responses from the pre-test to that of the post-test. One then examines if the treatment group experienced greater change than the control group. This is the most rigorous design because it addresses the major shortcoming of the pre/post design, specifically response-shift bias. If such bias (or any other type of bias) exists, it will influence the results of the program and control group equally. Therefore, any pre/post change in outcomes that is found among the program group, and not the control group, can be attributed to the program itself.

Summary

Review of the literature about evaluation design shows that there are strengths and weaknesses for both the traditional **pre/post** and the **post then pre** methods when collecting data from program participants. In general, both methods address the need to assess participant change from program beginning to completion. Whether or not the preferred instrument collects data at two separate points in time or only once at program end is the point of debate among practitioners, evaluation theorists and researchers.

This brief has highlighted the strengths and weaknesses of both approaches. The choice between methods must be made in the context of the type of program delivered, the sensitivity of the program content, comfort levels of participants, confidence of educators to collect data, time limitations, budgetary constraints, skill with data analysis and reporting needs set forth by program funders. It is hoped that the information presented here will inform such choices about evaluation design.

In general, many believe that the **post then pre** method reduces response shift bias because it accounts for changes in learner's knowledge from program content, allowing them to accurately assess what they did or did not know at program outset (Rockwell and Kohn 1989), thereby improving accuracy because the participants can accurately reflect on what they learned (Davis 2003). Additionally, the need to administer the instrument only once relieves time limitations many program providers experience. Note though that the **post then pre** also includes several threats to validity that should be considered when deciding on an evaluation design for a program.

Well-known evaluation theorists recommend using a traditional **pre/post** design when possible, but offer some suggestions on how to increase the credibility of results if a program must rely on a **post then pre** method.[2] First, the collection of supplemental or complementary data (such as participant demographics) would allow for the analysis of participant attrition, audiences served, and other important process related data that could allow for program improvement. In addition, the addition of follow up data (e.g., one month after program) with participants who completed the post then pre instrument would more accurately capture longer tern impacts of program participation and reinforce results from the previous data collected at program end. Finally, it is important to note that using a control group, with a traditional pre/post design can dramatically improve the validity of the results of an evaluation study.

---

[2] For example, see Shadish, Cook and Campbell (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference; Hill and Betz (2005*). Revisiting the Retrospective Pretest*, American Journal of Evaluation, Vol. 26, No. 4.

Table 1. Strengths and Weakness of Evaluation Designs

	Pre/Post	Post then Pre
Source of bias	Response shift	Recall Social Desirability Effort Justification Cognitive dissonance
Concerns about method	Could underestimate program effect Time constraints educator faces Burden on participants to fill out same form twice Comfort level of participants to be "tested" at program start or sensitivity of information gathered (e.g., child discipline)	Tends to overestimate program effect. Measures pre and post at same time, sometimes using same instruments, leads to biases listed above. Does not allow for any attrition data or process evaluation data. Generally not regarded by social scientists as a rigorous or credible method to capture outcomes.
Benefits of method	Viewed as rigorous, lends credibility to results Measures same person at two intervals in time, reduces many sources of bias listed under "post then pre." Objective measure of outcomes, not perceived change of participants.	Easy to administer. Reduces response shift bias. Allows for participant reflection about program effect on particular topic.
Best to determine (Goal of Evaluation)	Quantifiable changes in outcomes, especially behavioral items.	Participant's perceived change due to program attendance.

The Big Question

Perhaps the most important criteria that should be used when choosing an evaluation method is to ask the question, "**what are we trying to measure**?" If the goal of an evaluation is to capture outcome data (changes in parent behavior) then a **pre/post** design is generally viewed as a more accurate measure of change between two time intervals. If the goal is to capture how participants *perceive* the changes they have made in knowledge, skills, attitudes or behavior, then a **post then pre** method may be adequate to capture information on this type of data. While it is important to acknowledge that all self-reported information could be considered somewhat subjective, using a pre/post design measures actions or behaviors at two points in time through carefully worded questions or statements. Utilizing the same questions or statements on a retrospective instrument provides parents *views* on how their behavior has changed at one point in time (program completion). In general, if a program's funding agency requires quantifiable changes in outcomes, the **pre/post** method is more likely to capture those changes, and is generally viewed as a rigorous, credible measure of that type of data.

References

Davis, G. "Using Retrospective Pre-post Questionnaire to Determine Program Impact."
    Journal of Extension, vol. 41, no.4, August 2003.

Hill, L. and Betz, D. "Revisiting the Retrospective Pretest." American Journal of Evaluation,
    Vol. 26. no.4, December 2005; 501-517.

Klatt, J. and Taylor-Powell, E. Synthesis of Literature relative to Retrospective Pretest
    Design. Presentation to the 2005 Joint CES/AEA Conference, Toronto.

Lamb, T. "The Retrospective Pretest: An Imperfect but Useful Tool." Evaluation Exchange,
    vol. 11, no. 2, Summer 2005.

Raidl, et al. "Use Retrospective Surveys to Obtain Complete Data Sets and Measure Impact
    in Extension Programs." Journal of Extension, vol. 42, no.2, April 2004.

Rockwell, S. and Kohn, H. "Post-Then-Pre Evaluation." Journal of Extension, vol. 27, no.2,
    Summer 1989.

Shadish, W., Cook, T. and Campbell, D. (2002). Experimental and Quasi-Experimental
    Designs for Generalized Causal Inferences. Houghton Mifflin Company; New York.

Laura Colosi is an Extension Associate in the Department of Policy Analysis and Management at Cornell University.

Rachel Dunifon is an Associate Professor in the Department of Policy Analysis and Management at Cornell University.

Cornell University
Cooperative Extension