# UF UNIVERSITY of FLORIDA
## IFAS Extension

# Capturing Change: Comparing Pretest-Posttest and Retrospective Evaluation Methods[1]

Jessica L. Gouldthorpe and Glenn D. Israel[2]

## Introduction

The Extension mission has at its core an intention to change the awareness, knowledge, attitudes, or aspirations of community members who choose to participate in Extension programs. These programs are often designed to influence a change in behavior. However, Extension professionals need to find accurate and reliable ways to capture evidence that change has occurred because of a program's activities. Multiple evaluation models exist to capture change. Many government and nonprofit programs use strategies that measure performance through tracking systems and evaluation designs with no comparison group. Performance data are often obtained from participation records, staff observations and client self-reports. These performance measurements help to address whether the program accomplished what it set out to accomplish (Hatry, 1999; Newcomer, 1997; Pratt, McGuigan, & Katzev, 2000), but they are criticized by some as lacking rigor.

Two models that are commonly used in Extension programming to capture change over a short period of time are the *pretest-posttest model* and the *retrospective pretest* (or post-then-pre) model. When deciding which model to use, Extension professionals should keep in mind that each participant has a knowledge base that includes both factual information and perceptions pertaining to factual information. As you read about the strengths and weaknesses of these two design models, consider how each model fits the evaluation situation to select the one that can best measure change in your program (Israel, Diehl, & Galindo-Gonzalez, 2009).

## Prestest-Posttest Model

The pretest-posttest model is a common technique for capturing change in Extension programming (Allen & Nimon, 2007; Rockwell & Kohn, 1989). In this model, a pretest is given to participants prior to starting the program to measure the variable(s) of interest, the program (or intervention) is implemented, and then a posttest is administered to measure the same variable(s) of interest again (Gall, Gall, & Borg, 2003). With measurements being collected at the beginning and end of the program, program effects are often revealed by calculating the differences between the two measures (Pratt et al., 2000).

Imagine that you are trying to identify the change in participants' factual information caused by your program. You would subtract the number of correct responses on that participant's pretest (e.g., 8 out of 20) from the number of correct responses on a participant's posttest (e.g., 15 out of 20). This calculation indicates a 7-point increase in factual knowledge for that person. This suggests that your program has positive effects on changing knowledge.

But what if you asked participants to rate some of their perceptions about a personal habit on a scale of 1 to 5 (with

5 being the highest)? Suppose a participant rates himself as no lower than a 4 on any pretest item, based on his preprogram knowledge perception. However, during the course of the program, the participant realizes that he rated himself too high based on the information that you have presented. So, on the posttest he rates himself on the same items as either a 2 or 3. If we subtract the posttest (score = 2) from the pretest (score = 4), we end up with a negative score (score = -2). Does this mean that the program had a negative impact? Not necessarily; however, it does make interpretation a bit more complex. Your participant is demonstrating response-shift bias, where the frame of reference your participant is using to measure himself has changed, thus making the pretest-posttest comparison invalid (Howard, 1980). Since measuring perceptions in this way opens the door for this type of bias, it is better to use a pretest-posttest evaluation design when attempting to measure factual knowledge or skill sets at two defined points in time, rather than perceptions of change.

## An example

If an Extension professional chose to use a pretest-posttest for evaluating change among new homeowners enrolled in a financial management program, appropriate questions would revolve around reporting knowledge of factual information (i.e., *what percentage of your income is recommended for spending on housing costs*) or current skill sets (i.e., *using the information provided, balance the following checkbook entries*). These questions would be asked prior to the start of the 2-day workshop in order to inform the facilitator about areas that need the most attention during the sessions, and they would be administered following completion of the workshop. The results from the two data collection points would then be compared to determine whether change occurred as a result of participation. While straightforward, there are some advantages and disadvantages associated with using this model.

## Advantages of the pretest-posttest model

- **Multiple data points:** This model provides more information than a posttest-only design. Since this method provides a measure of participant knowledge or behavior prior to the start of programming efforts, it can be helpful in refocusing the information to be presented while providing a point of comparison from beginning to end (Ary, Jacobs, Razavieh, & Sorensen, 2006).

- **Capture of factual information/skill change:** Assessing factual knowledge or current skills can provide a more accurate measurement of change than simply perceptions of change. Therefore, it is important to clearly identify

what you are trying to capture—factual knowledge change or perceptions—and to select the appropriate evaluation method.

- **Accurate behavior measurement:** Routine behaviors (e.g., food recalls) are more accurately reported in pretests for multisession programs because people remember fewer details as time passes (Sudman, Bradburn, & Schwarz, 1996).

## Limitations of the pretest-posttest model

- **Time constraints:**

  - **Instrument creation:** More time is required to create solid items that assess factual knowledge than is needed to capture perceptions.

  - **Program delivery:** It takes time to administer both a pretest and posttest questionnaire (Pratt et al., 2000); therefore, in short educational activities, it may not be worth the time necessary to conduct both.

- **Attendance concerns:** Meaningful pretest-posttest comparisons require that participants be present at the start and end of the program; however, consistent attendance can be difficult to obtain, especially among high-risk groups (Pratt et al., 2000). Without pairs of responses (a pretest and a posttest), comparisons cannot be made and the available data are reduced.

- **Measurement error through response-shift bias:** Meaningful pretest-posttest comparisons require a participant to use the same frame of reference to measure himself against; when this is missing, it makes the pretest-posttest comparison invalid (Howard, 1980). There is also the potential for the limited information a participant has prior to the program to affect his ability to properly judge baseline functioning (Allen & Nimon, 2007; Howard et al., 1979).

## Retrospective Pretest Model

In contrast to the pretest-posttest model, a retrospective pretest (or post-then-pre) design administers the preprogram assessment concurrently with the posttest by asking individuals to recall their knowledge or behavior prior to the program (Allen & Nimon, 2007). In this situation, one must create an instrument with sufficient sensitivity to detect changes in participants (Lynch, 2002) while also choosing words and phrases that assist the participant with recalling their thoughts prior to exposure (Pratt et al., 2000). Upon completion of the program, a participant is

asked to consider a question from two juxtapositions: 1) knowledge or behaviors as a result of participating in the program and 2) reflections on what the knowledge or behavior was prior to the program (Rockwell & Kohn, 1989). There are some times when it is better (and necessary) to utilize a retrospective pretest evaluation design. These situations include measuring change over a very short period of time (i.e., a 4-hour course), attempting to gauge perceptions of change as a result of program participation, attempting to reduce response-shift bias, or trying to evaluate change without having collected baseline data prior to the start of programming efforts (Klatt & Taylor-Powell, 2005).

## An example

If an Extension professional chose to use a retrospective pretest, appropriate questions would revolve around reporting participant perceptions (i.e., using the scale provided [1 = *not at all knowledgeable* and 5 = *extremely knowledgeable*], and based on the information presented during this workshop, how would you rate your knowledge about proper spending practices?). This question acts as the posttest question. Then, the participant is asked to consider their pre-intervention levels (i.e., using the same scale, how would you rate your knowledge about proper spending practices prior to participating in this program?). This question provides the pretest data point. There are some advantages and disadvantages associated with using the retrospective pretest model.

## Advantages of the retrospective pretest model

- **Control for rival hypotheses:** Events that happen outside the program can affect participants' attitudes and behaviors if there is a significant period of time between the pretest and posttest; the retrospective pretest captures the pretest and posttest responses at the same time, thus limiting the impact of outside events on the results (Ary et al., 2006).

- **Stable Instrumentation:** Creating equivalent—but not identical—pretests and posttests can impact participant results; the retrospective pretest uses the same instrument, thus eliminating the potential for a second, potentially different instrument (Ary et al., 2006).

- **Same frame of reference:** Participants taking part in a program can experience a shift in their frame of reference used to answer pretest perception questions (response-shift bias), or they can over- or underestimate pretest reports based on limited pre-intervention knowledge. A retrospective pretest provides a more accurate assessment of the participants' perception of change because both answers are generated within the same frame of reference, and they are able to properly judge their functional baseline (Allen & Nimon, 2007; Howard et al., 1979; Pratt et al., 2000).

## Limitations of the retrospective pretest model

- **Relies on recall:** Retrospective pretests must attempt to minimize the effect that demand characteristics and memory-related problems may have on the recall process (Pratt et al., 2000).

  - Demand characteristics may be problematic when participants have motivation for making the program look good or providing a socially desirable response.

  - Recall can be impacted when the length and/or specificity of the pertinent time period is too broad or undefined.

- **Additional biases:** Retrospective pretests are based on a self-report and, therefore, remain an estimated report. Participants can also exhibit subject bias since they are actively trying to improve their skills and want to see improvement (Pratt et al., 2000).

- **No data on dropouts:** While retrospective pretests have full information for clients who complete the program (Raidl et al., 2004), no information is available for people who start and drop out of the program.

## Conclusion

All good evaluation requires selecting the appropriate tools for the particular circumstance (Klatt & Taylor-Powell, 2005). As suggested by Israel et al. (2009), the time, effort, and intensity of your programming should be factors when determining the quality and rigor of your evaluation. Regardless of which evaluation strategy you choose, it is important to consider the pros and cons for each circumstance, as well as what information you would most like to capture. Then, using thoughtful and intentional craftsmanship, construct an instrument that allows you to capture the change created by your programming effort. In summary, we suggest using a pretest/posttest format when you have the time and want to measure true knowledge change. However, we promote using a retrospective pretest when you are measuring perceptions of knowledge and when time or other factors limit your ability to use a true pretest/posttest.

## References

Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A practical technique for professional development evaluation. *Journal of Industrial Teacher Education*, *44*(3), 27–42.

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education* (7th ed.). Belmont, CA: Thomson Wadsworth.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). New York, NY: Allyn and Bacon.

Hatry, H. P. (1999). *Performance measurement:* Getting results. Washington, D.C.: The Urban Institute Press.

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4*(1), 93–106. doi: 10.1177/0193841X8000400105.

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D., & Gerber, S. L. (1979). Internal invalidity in pretest-posttest self-report evaluations and the re-evaluation of retrospective pretests. *Applied Psychological Measurement*, *3*(1), 1–23.

Israel, G., Diehl, D., & Galindo-Gonzalez, S. (2009). *Evaluation situations, stakeholders & strategies*. WC090. Gainesville: University of Florida Institute of Food and Agricultural Sciences. Retrieved from http://edis.ifas.ufl.edu/wc090

Klatt, J., & Taylor-Powell. E. (2005). *Using the retrospective post-then-pre design.* Quick tips #27. Madison: University of Wisconsin-Extension. Retrieved from http://www.uwex.edu/ces/pdande/resources/pdf/Tipsheet27.pdf

Lynch, K. B. (2002, November). *When you don't know what you don't know: Evaluating workshops and training sessions using the retrospective pretest methods*. Paper presented at the meeting of the American Evaluation Association Annual Conference, Arlington, VA.

Newcomer, K. (1997). Using performance measurement to improve public and non-profit programs. *New Directions for Evaluation*, *75*, 5–14.

Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation*, *21*(3), 341–349.

Raidl, M., Johnson, S., Gardiner, K., Denham, M., Spain, K., & Lanting, R. (2004). Use retrospective surveys to obtain complete data sets and measure impact in extension programs. *Journal of Extension*, *42*(2). Retrieved from http://www.joe.org/joe/2004april/rb2.shtml

Rockwell, S. K., & Kohn, H. (1989). Post-then-pre evaluation: Measuring behavior change more accurately. *Journal of Extension*, *27*(2). Retrieved from http://www.joe.org/joe/1989summer/a5.php

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Fransisco, CA: Jossey-Bass.

## Acknowledgments