AM Last Page: Common Evaluation Designs in Medical Education I

Dario M. Torre, MD, MPH, PhD, associate professor of medicine and vice chair, Education, Drexel University College of Medicine, and Barbara Daley, PhD, interim associate dean, College of Nursing, and professor of Adult and Continuing Education Program, University of Wisconsin—Milwaukee

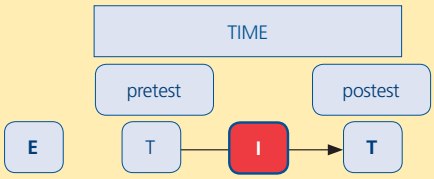
Evaluation design is important to establish that an instructional program produced the measured effects and learning outcomes. Evaluation design should help isolate extraneous factors so that differences or outcomes can be safely attributed to the instructional program. The goal of this Last Page is to describe three common program evaluation designs, along with benefits, drawbacks, and examples.

Internal validity of an evaluation measures the certainty with which educators can ascertain whether the program actually caused the effects they find. Educators should always consider threats to validity in designing program evaluation. Some of these internal validity threats (history, maturation, testing) are included in the description of some of the evaluation designs below.

Legend

E = Experimental group
C = Control group
I = Instructional program
T = Test, measurement, or observation
Rand = Randomization

One-Group Pre-Post Test Design



Example:

One group of students rotating through a medicine clerkship is given an 8-station objective structured clinical examination before and after the administration of a month-long, Web-based instructional program about the clinical presentation and diagnosis of ten common medical diseases

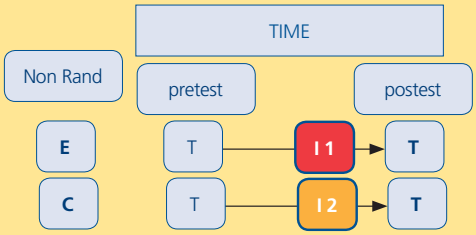
Pros:

- Easy to implement
- May be helpful for formative evaluation (e.g., to gather information that will guide program improvement), particularly if the interval of time is short

Cons:

- Internal validity threats, including both history (within the time that passes before and after the intervention, events may occur that influence the outcome) and maturation (learners may naturally grow during the time of the experiment), should be considered
- Absence of comparison group makes it difficult to assess whether extraneous variables affected the outcome

Non-Equivalent Control Group Pre-Post Test Design



Example:

Students from hospital X (group E) and students from hospital Y (group C) are selected (not randomly) and given a multiple-choice-question (MCQ) knowledge-based pretest on asthma and chronic obstructive lung disease (COPD). Group E experiences a self-directed reading program while Group C experiences a seminar-based program, involving discussion of asthma and COPD. Both groups are tested again one month later through a MCQ knowledge-based test that has the same content, but different questions, as the pretest.

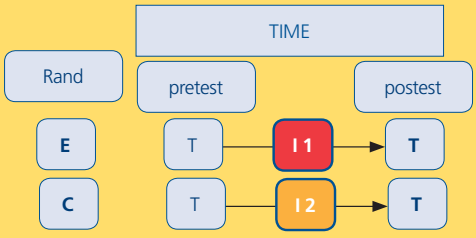
Pros:

- Feasible when randomization is not possible
- Allows for the comparison of two educational interventions
- The use of a pretest allows researchers to assess the comparability of the groups (e.g., are pretest scores the same or different between groups?) at the beginning of the program, since it is important to ensure that the 2 groups are similar at the beginning of the intervention

Cons:

- The use of a pretest can lead to testing effect; that is, students may identify certain content topics that will be on the posttest, based on test items in the pretest (e.g., the use of beta blockers in the treatment of congestive heart failure), regardless of question items being different between pre-and posttest
- Selection bias and dissimilar initial groups may be misleading and influence the outcome

True Control Pre-Post Test Design



Example:

A group of students are randomly assigned to either the E group or the C group at the beginning of the academic year (or at the beginning of a rotation). Group E experiences a small-group discussion instructional program and Group C experiences a video-based program with facilitators. Both groups learn about the diagnosis and treatment of osteoarthritis. Both groups are tested through a MCQ, knowledge-based test before and after administration of the programs.

Pros:

- Randomization assures group equivalence and eliminates selection bias
- Eliminates many of the internal threats to validity, thus yielding stronger conclusions about the outcome

Cons:

- Randomization may be challenging in medical education settings, particularly when classes and rotations are predetermined
- If the pretest is reactive (i.e., the content of the pretest may cause students to focus their study on specific program material), it may influence the outcome of the evaluation

References:
Cresswell JW. Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research. Upper Saddle River, N.J.: Pearson/Merrill Prentice Hall; 2008.
Fitz-Gibbons CT, Morris LL. How to Design a Program Evaluation. Newbury Park, CA: Sage; 1987.
Popham WJ. Educational Evaluation. 3rd ed. Boston: Allyn and Bacon; 1993.
Rossi PH, Freeman HE, Lipsey MW. Evaluation: A Systematic Approach. 6th ed. Thousand Oaks, CA: Sage; 1999.
Author contact: Dario.Torre@DrexelMed.edu