

Formative and Impact Evaluation

Formative Evaluation

- An evaluation designed to produce qualitative and quantitative data and insight during the early developmental phase of an intervention, including assessment of:
 - Feasibility of program implementation
 - Appropriateness of content, methods, materials, media, and instruments
 - Immediate or short-term cognitive, psychosocial, psychomotor (skill), and/or behavioral impact of an intervention for a well-define population at risk.

Impact Evaluation

- Can the observed change in the behavioral impact rate, if statistically significant, be attributed to the health promotion program?

External Validity

- The extent to which an observed, significant change in a behavioral or health impact rate, attributed to a feasible, replicable intervention and documented by multiple evaluations with high internal validity, can be generalized to other program settings and to comparable populations at risk

4

Internal Validity

- The extent to which an observed impact, a significant improvement in cognitive, skill, behavioral, health status, or economic indicator rate, can be attributed to a planned intervention
 - Did a planned, replicable intervention (X) produce the significant increase in the impact rate (O)?
 - Is X the only or the most plausible explanation for the changes in O, or can all or part of the observed change be attributed to other factors?

5

Primary Sources of Bias

- Measurement
 - Core sources of bias: poor validity and poor reliability
 - Must be eliminated in order to determine how much change in an impact rate occurred
 - Randomization does not control for this
 - Combines testing and instrumentation threats

6

Primary Sources of Bias

- Selection
 - Sources: participation rate and attrition rate
 - A program must describe eligibility criteria for the defined population at risk
 - Ability to generalize results to a defined population may be severely limited or impossible if not addressed
 - Primary questions:
 - How representative are the results?
 - Can the results of an evaluation be applied to a defined population in this setting?
 - Combines maturation, statistical regression, selection, and participant attrition threats

7

Primary Sources of Bias

- Historical
 - Primary sources:
 - Transient or enduring external historical events
 - Internal programmatic, historical events
 - Specific intervention procedures delivered or not delivered
 - The extent to which planned exposure to the program was documented and the degree of standardization, replicability, and stability of intervention procedures by staff
 - Combines history and maturation threats

8

Evaluation Designs

- Start with the most rigorous design possible and then, if necessary, modify it to fit the setting
- Assess each design to determine the degree to which a program will be able to attribute an observed impact to your program
- Purpose:
 - To create E vs. C group equivalence at baseline and follow-up

9

Evaluation Design

- Categories
 - Nonexperimental
 - Includes one E group but no C group
 - Quasi-experimental
 - Includes an E group and C group created by methods other than randomization and includes observations of both groups prior to and after the intervention
 - Experimental
 - Includes random assignment of an E group and C group, and observations of both groups, prior to and after the intervention

10

One-Group Pretest and Posttest

- Most basis design
- Should not be used to assess program impact
- Can be used for pilot study
- Primary weakness:
 - Extent to which participants selected for the program are similar to typical users

11

Nonequivalent Comparison Group

- Adds an experimental group to the one-group pretest and posttest design
- May improve a program's ability to rule out some alternative explanations of impact

12

Time Series

- Can be used if a program can:
 - Establish the periodicity and pattern of the impact rate being examined for a well-defined population at risk
 - Observe multiple monthly/quarterly data points before and after the new intervention
 - Establish the validity, reliability, and completeness of measurement
 - Collect data unobtrusively as a part of an existing monitoring system
 - Document introduction of an intervention at a specific time and, if planned, withdrawal of the intervention abruptly at a specific time

13

Time Series

- Requires adequate number of observations to document impact rate trends
- To document the significance of a trend in an observed rate, the treatment must be powerful enough to produce shifts in an impact rate considerably beyond the normally observed variations in the rate
- External and internal historical biases increase significantly with extended duration

14

Multiple Time Series

- Impact rates are studied at different times over several years for an experimental group and comparison group
- Can only be applied in situations where retrospective and prospective databases exist and are accessible or where an organization can periodically observe rates from program participants
- Major threat: selection bias

15

Randomized Pretest and Posttest with an E and C Group

- Randomization is used to establish two groups, E and C, not significantly different at baseline for any independent or dependent variables that are predictors of impact
- Random selection
- Random assignment
- Should control the 3 major biases to internal validity, assuming no major implementation problems

16

Quasi-Experimental Designs

- Recommendations to reduce bias:
 - Select internal rather than external comparison subjects
 - Avoid participants' self selection into E or C conditions
 - Consider selecting comparison groups on stable pretest measures
 - Include pretests on the dependent variable and use to adjust the posttest
 - Use sensitivity analyses to explore possible effects of hidden bias and different selection assumptions
 - Significantly reduce or prevent attrition by pilot testing

17

Establishing a Comparison (C) Group

- Matching with external units
- Matching with internal units—participants
 - Apply the one-group pretest and posttest in the same site with standardized methods and procedures
- Participant- or peer-generated C groups
 - Participants match themselves with a nonparticipant friend

18

Establishing a Control Group

- Random assignment
- Delayed treatment
 - Randomly assign participants to either immediate (E) or delayed (C) treatment, with each having an equal opportunity to participate
- Multiple-component program
 - Determine the differential impact of multiple components of a program by exposing subsamples or participants to different components
 - Randomized, factorial study
- New Program
 - Comparison of new program to existing or standard program

19
