NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Access Health NYC Training

Measuring, Quantifying & Communicating Our Impact

CAROLYN BERRY, PHD

New York University Langone Health

May 12, 2021

Stakeholders and Purpose

Who cares?

- Who is asking for evaluation?
  - Current funders
  - Supervisors/organizational leadership
  - Board of directors
  - Program staff

- Who are you trying to impress?
  - Same as above
  - Future/potential funders
  - Target population
  - Policy makers
  - Publications

What do they want to know?

- What outcomes interest various stakeholders?
  - Outputs (services provided, number of touches)
  - Proximal or short-term outcomes
  - BIG IMPACT
- What type of information speaks to them?
  - Stories or numbers?
  - Pie charts/bar graphs or statistical analyses?
- What degree of rigor is necessary to persuade them?
- How concerned will they be about potential bias?

Evaluation Approach

Program evaluation vs monitoring

- Process evaluation and program monitoring may overlap to some degree and the two may inform one another
  - Measurement
  - Process/outputs and outcomes

- Differ in important respects:
  - Time frame: time-limited study vs. ongoing activity
  - Purpose: largely external vs largely internal
  - Approach: qualitative/mixed vs. largely quantitative

Types of evaluation activities

- Broad categories include:
  - Needs assessment
  - Program theory explication and assessment
  - Process/implementation evaluation
  - Outcome/impact evaluation
  - Efficiency analysis

- Appropriate evaluation activities largely determined by status of program, policy or intervention

Logistical considerations

- **Resources**
  - Money, funding
  - Staff time
  - Volunteers, interns
  - In-house expertise
  - Existing data
  - Internal support
- **Time frame**
  - Program start date
  - Evaluation due date
  - Time to outcome

Internal evaluation

- **Advantages**
  - Timely
  - Buy-In
  - “Insider” perspective
  - Access
- **Disadvantages**
  - Time consuming for staff
  - Bias
  - *Perception* of bias
  - Lack of expertise
  - Too close for comfort

External evaluation

- **Advantages**
  - Perspective
  - Credibility
  - Expertise
- **Disadvantages**
  - Cost
  - Time
  - Lack of substantive expertise

NEEDS ASSESSMENT

What is “needs assessment”?

- The term “Needs Assessment” (NA) is used to mean many types of assessment.

- Consider the term a broad rubric that encompasses many evaluation activities:
  - Needs identification, needs assessment
  - Assessment of demand, preference, acceptability
  - Feasibility study

- **All** so-called NA activities occur *before full implementation of* an intervention

Phases of needs assessment

Phase 1: Pre-assessment

- What do we already know?
- What do we need to find out?
- Should we launch a formal assessment?

Phase 2: Formal assessment of needs

- Identify and prioritize needs
- Delineate causal factors
- Needs identification/assessment

Phase 3: Post-assessment

- Move to development of potential solutions
- Demand, preference, acceptability, feasibility assessment

Typical data collection methods

- **Primary data**
  - Surveys of target populations and/or key informants
  - Focus groups, key informant interviews

- **Secondary data**
  - External data (e.g., census, government surveys)
  - Internal data (e.g., program monitoring, tracking)
  - Literature
  - Data from other agencies

PROGRAM THEORY EXPLICATION

What is program theory?

- *"a plausible and sensible model of how a program is supposed to work" Bickman, 1987*
- Program theory identifies *"program resources, program activities, and intended program outcomes, and specifies a chain of causal assumptions linking program resources, activities, intermediate outcomes, and ultimate goals" Wholey, 1987*

- Theory of change, logic models, log frames, theory based evaluation

Definition of a logic model

"A systematic and visual way to present and share your understanding of the relationships among the resources you have to operate your program, the activities you plan to do, and the changes or results you hope to achieve."

W.K. Kellogg Foundation Handbook (1998)

Two key components of program theory?

- **Process theory**
  - refers to the “how”
  - focused on implementation and operations
  - includes resources/inputs, activities, outputs

- **Impact theory**
  - refers to the “why”
  - focused on effects of the intervention
  - includes outcomes and impacts

Logic Model Framework

Resources/ Inputs	Activities	Outputs	Outcomes	Impacts
Resources dedicated to or consumed by the program	The process, tools, events, technology, and actions that are an intentional part of the program implementation	The direct products of program activities	Benefits or changes for participants during or after program activities	The fundamental intended change occurring in organizations, communities or systems as a result of program activities
What you need to carry out the activities you have planned	What you do with the resources you have	What the activities produce (e.g., services, products)	What you expect to happen as a result of the outputs – short and long term benefits or changes for participants	The fundamental long-term changes you are seeking

Outputs vs. Outcomes

Outputs	Outcomes
Direct and measurable products of a program's activities and services; they are often expressed in terms of volume and units delivered	The results or impact of the activities and services. Outcomes often represent the results of multiple outputs, each outcome usually corresponds to more than one output
# of nutritious meals served to seniors	Better health status
Attendance at fitness sessions	Weight loss; greater flexibility; enhanced well-being
# of organizations represented at coalition meetings # coalition meetings	Increased cross-agency referrals

PROCESS EVALUATION

Main differences between process and outcome evaluation

	Process Evaluation	Outcome Evaluation
Timing	Any time after program implementation begins	Preferably when program is “ready”
Methods	Mixed	Mixed with definite lean towards quantitative
Logic Model	Resources, Activities, Outputs	Outcomes & Impacts
Purpose	Assess implementation; assess fidelity	Assess effectiveness
Design	Descriptive; no counterfactual	Research design; includes counterfactual
Statistics	Descriptive stats only; percentages, means	Inferential statistics

OUTCOME EVALUATION

Outcome/impact Evaluation?

Outcome/impact evaluation investigates whether the intervention caused demonstrable effects on specifically-defined target outcomes

- Did it work? How and for whom?
- Inform future implementation/expansion
- Convince interested parties findings are valid

Approaches to design

- **Informal** designs
  - "Storytelling"
  - Self-evaluation
  - Expert judgment
  - Descriptive (includes after-only single group designs)
- **Formal** designs based on social science
  - "Experimental"
  - Based on the scientific method
  - Requires a counterfactual (pre-test or comparison/control group)

Three "pillars" of outcome evaluation

- **Research Design**
  - Overall structure
  - Use of a counterfactual (pre-test and/or comparison/control group)
  - Internal validity
- **Sampling**
  - Who ends up in the study
  - External validity
- **Measures**

What's wrong with this picture?

A new program offering social dance classes at senior centers. The goals are to increase the physical and social activity of seniors.

- After six months of the program, a survey of participants showed that 70% rated themselves as "physically active" and 65% rated themselves as "socially active"

The conclusion is that the social dance classes were successful at achieving the stated program goals.

How can we attribute findings to the program?

- Establish a sound **counterfactual**
  - Pre-test
  - Comparison/control group

- Three basic "experimental" designs
  - Pre-test/post-test (reflexive, single sample, program group only)
  - Quasi-experimental design with non-equivalent comparison group
  - "True" experiment (randomized control trial)

Does this help?

A new program offering social dance classes at senior centers. The goal of the program is that more than half of the participants are physically and socially active.

- After six months of the program, a survey of participants showed that 70% rated themselves as “physically active” and 65% rated themselves as “socially active”

The conclusion is that the social dance classes were successful at achieving the stated program goals.

Does this help? Better

A new program offering social dance classes at senior centers. The goal of the program is that more than half of the participants are physically and socially active.

- Upon signing up for the program, 50% of participants rated themselves as “physically active” and 45% rated themselves as “socially active”.
- After six months of the program, 70% of participants rated themselves as “physically active” and 65% rated themselves as “socially active”.

The conclusion is that the social dance classes were successful at achieving the stated program goals.

Does this help? Better

A new program offering social dance classes at senior centers. The goal of the program is that more than half of the participants are physically and socially active.

- A survey of participants showed that 70% rated themselves as “physically active” and 65% rated themselves as “socially active”.
- A survey of seniors at centers without the classes found that 45% rated themselves as “physically active” and 55% rated themselves as “socially active”.

The conclusion is that the social dance classes were successful at achieving the stated program goals.

Does this help? BEST!

A new program offering social dance classes at senior centers. The goal of the program is that more than half of the participants are physically and socially active.

	Pre-test	Post-test
Social dance classes	50%	70%
No social dance classes	45%	45%

Single sample designs

- Weakest design by far, but:
  - Only design possible for full coverage programs
  - Logistical/resource constraints may require
  - Comparable comparison group impossible to identify

- **Must** have a pre-test
  - Multiple pre-tests or time series strengthens considerably

Strengthening single sample designs

- Short time frame between pre- and post assessments
  - What else could have happened in an hour??

- Logic, theory, previous literature, common sense
  - Tie closely to logic model with chain of events

- Compare to regional, state, national trends

Strengthening single sample designs

- Rule out other potential causes by talking to experts, program administrators, participants

- Multiple data collection points

- If a pre-test is not possible, then at least ask participants for attribution
  - How much did your physical activity increase as a result of the social dance classes?

QUESTIONS?