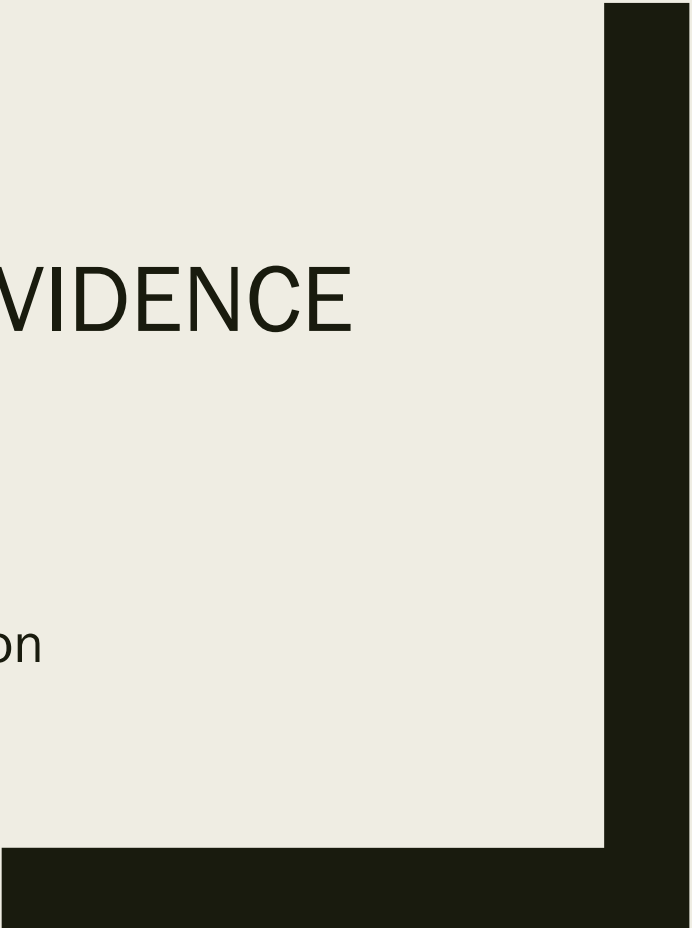


PROGRAM EVALUATION FOR EVIDENCE BASED PRACTICE

Teresa M. Beck, PhD, CTRS, FDRT
Illinois Recreational Therapy Association
November 4, 2019



Learning Outcomes

Upon completion of session participants will:

- *Identify three types of program evaluation*
- *Identify the steps of program evaluation*
- *Identify a minimum of two techniques for program evaluation*
- *Identify one program evaluation model that could be used in participants' respective programs*
- *Develop a minimum of two evaluation questions to take back to participants' respective programs.*

Who currently has a
systematic program
evaluation process in place?

What is Program Evaluation?

- A systematic process of studying a program (or practice, intervention, or initiative) to determine how well it is working to achieve the intended goals
- Gather and analyze information to serve three purposes:
 - *Program Assessment*
 - Documenting program outputs and outcomes
 - *Program Improvement*
 - What works, what does not work, and why things work the way they do
 - *Strategic Management*
 - Make decisions about how resources should be applied in the future to better serve its mission or goals

Goals of Evaluation

These include but are not limited to:

- Clarify program objectives – what are you trying to accomplish? How will you define success?
- Assess your program’s appropriateness and effectiveness – Is the program working as intended? Is it the right way to address the problem?
- Address program costs – Is the program easily implemented with reasonable budget? Do the benefits justify the cost?
- Make informed program improvement – what changes should be made before implementing again or elsewhere
- Monitor program fidelity and integrity – are you really doing what you said you would do?
- Contribute to the field – creating a much-needed body of verifiable information about what works in the disciplines.

Types of Program Evaluation

- Process Evaluation
- Outcome Evaluation
- Impact Evaluation

Process Evaluation

- Describes how your program is being evaluated
 - *Used to determine the extent to which program is being implemented as intended (intervention fidelity)*
 - *Can provide information on the implement process – used to refine the delivery of the program and improves its quality*
 - *Important in the interpretation of outcome data – if program is not achieving expected outcomes, there may be problems with intervention fidelity*

Process Evaluation Questions

- What are the most important components and activities of the program?
- Are there aspects of the implementation process that are enabling or inhibiting success?
- Are the program providers receiving the appropriate amount of training and supervision to ensure intervention fidelity?
- Does the program being implemented match the original design?
- Do program participants understand the program and its intended outcomes?
- Are there differences in outcomes based upon how many times the intervention or treatment was offered, length of sessions, attendance level etc.

Outcome Evaluation

- Investigates whether changes have occurred for the persons participating in the program?
 - *Quantifies how much change and in what direction (positive or negative)*
 - *Seeks to tie these changes to the specific elements of the program*
 - *Essentially asks “What is my program accomplishing in the short term?” and “Am I meeting my objectives?”*
 - *When an outcome evaluation demonstrates effectiveness of a program achieving its intended outcomes (increase in attention span, decrease in aggressive behavior etc.), it argues for continued investment, but also replication elsewhere (EBP)*

Outcome Evaluation Questions

- What effect is the program having on its participants (e.g., changes in knowledge, attitudes or behavior)?
- What unexpected outcomes, if any, have resulted from the program?
- What can be modified to make the program more effective?
- Is there any evidence showing that funders (administration, 3rd party reimbursement) should continue to support the program?

Impact Evaluation

- Answers the questions, “Is my program producing long-term or global changes?”, “Am I meeting my long-term goals?”
- Impact and Outcome evaluation often confused – key difference is that outcome evaluation documents short term or immediate outcomes, and impact evaluation is focused on long-term or global changes.
- Outcome evaluation might examine the extent to which a substance abuse prevention program produced decreases in past 30 day substance use among program participants – an impact evaluation may look at decreases in substance abuse 3 month, 6 month or 1 years post discharge (long-term) or may look at a decrease in past 30 day substance use among all participants in a district school system (global).

Impact Evaluation Questions

- What effect is the program having on our long-term goals (e.g. change in number of incidents reports, decrease in pain medicine post discharge – 3 month, 6 month, 1 year)?
- What effect did the program's activities have on components of the system in which the activities were targeted?
- Were there any negative outcomes? Are they result of implementation failure or some aspect of the program itself?
- What degree of confidence is that that the outcomes can be attributed directly to the program?

What are Evaluation Questions for your Program?

Spend 10 minutes or so with a small group and identify your program and some questions you want answered

Evaluation Designs

- Structure of the Study
- Evaluation Designs are differentiated by at least three factors
 - *Presence or absence of a control group*
 - *How participants are assigned to a study group (with or without randomization)*
 - *The number or times or frequency which outcomes are measured*

Evaluation Designs Continued:

- Control Groups
 - *Group of individuals who participate in the study but do not receive the main intervention being tested – may receive something else instead of the intervention or nothing at all*
- Randomization
 - *Controls for the sources of bias and enables more confidence that the outcomes seen from the intervention group are in fact tied to the intervention*
- Frequency of Outcome Measurement
 - *Only once after the intervention*
 - *Before and after the intervention (pre- and posttest)*
 - *Follow ups months after the intervention*

Categories of Evaluation Design

- Three basic categories of evaluation design, in order of least to most rigorous
 - *Preexperimental*
 - *Quasi-experimental*
 - *Experimental*

Preexperimental Design

- Defined by absence of a control group and absence of random assignment
- Variations in the number and timing of outcome measurements
 - *Posttest only*
 - Data collected only once from participants, immediately after they complete the program
 - Does not provide a baseline to which post intervention data can be compared
 - Posttest only can be appropriate if you only need to know (or if it is only feasible to know) if participants have reached an identified outcome, rather than measure the degree of change
 - Posttest only can also be a good choice if you have limited resources to spend on the evaluation or limited access to program participants

Preexperimental Design Continued

- Pretest and Posttest

- *Measures outcomes among participants before and after the intervention*
- *Relatively easy to implement – administer the same measure twice*
- *Timing of posttest is important – enough time to have intervention have an effect, but not so long that you have other intervening factors that could impact the effect*

- Pretest and Posttest with Follow-Up

- *Can improve pre posttest with follow up (3, 6, or 12 months)*
- *More longitudinal data over time – sustained effect beyond immediate treatment period*
- *Sometime difficult and costly to locate participants for follow-up*
- *Confounding variable – participants bored or annoyed at having to repeat the measurement*

Quasi-experimental Design

- Pre-experimental follows one group of participants, quasi-experimental uses two or more study groups
- Study groups are pre-existing sets of people sharing some common value
 - *Participants already enrolled in a program may be compared to individuals waiting to receive the intervention or with person who have chosen some other intervention or no intervention*
 - *Term control group is used to describe this comparison group*
- Control conditions
 - *Control may receive a different intervention, components of the intervention, or something that mimics time and attention paid to participants*
 - *Wait-list control – control group participants receive nothing during the study, but will eventually received the intervention after the study*

Quasi-experimental Design Continued

- Benefits and Challenges

- *Often the most practical option for implementing outcome evaluations in the social services context*
- *By using pre-existing groups, avoid random assignment as well as ethical concerns of withholding or delaying treatment*
- *Limitation – without true randomization, the study groups may differ in some important way that may account for some of the group differences in outcomes after the intervention*

Experimental Design

- Most sophisticated design – considered the gold standard
- Hallmark of this design is randomization
- Data collection in an experimental studies is similar to that quasi-experimental – difference is the randomization
- Benefits and Challenges
 - *Often used by researchers in physical sciences because they can control the lab environments*
 - *Ethical concerns in dealing with human subjects in withholding interventions*
 - *Has a higher cost*

How to Choose the Right Experimental Design

- What are your research questions?
 - *Immediate outcomes verses long-term impact*
- Can you collect the data?
 - *More complex design requires more data collection over multiple measurement points*
 - *Funding for data collection*
- Is the experiment feasible?
 - *Control groups may not be feasible*

Research Quality – Six Criteria

- Reliability of Measures
- Validity of Measures
- Intervention Fidelity
- Missing Data and Attrition (participants leaving study)
- Potential Confounding Variable
 - *History, maturation, test-taking effects, flaws in instrumentation, selection bias*
- Appropriateness of Analysis

Program Evaluation Models

- Importance Performance
- Satisfaction Based
- Goal and Objective
- Triangulated

Importance-Performance

- Examines the desirability of product attributes
- Participant satisfaction is a function of both patron expectations about attributes of a program they consider important **and** patron judgments about their experience of agency performance on these attributes
- Patrons administered a survey before they experience the program on what they think is important to them about the program
- After they participate, they are given a second survey which asks them to rate how well the agency met their expectations regarding the program
- Program supervisors and administrators are most the most likely audience for this evaluation technique

Satisfaction Based

- purpose is to measure the outcome of engagement (satisfaction)
- worth of programs can be determined by identify the degree to which programs have provided positive engagement for its participants
- Audience - program managers, program supervisors, and administrators
- Data can be used several ways
 - *can document participant-reported program outcomes*
 - *determine if programmatic goals are being met*
 - *can compare data among existing programs/services*

Goal and Objective

- Discrepancy model - examines the discrepancy between what was supposed to happen to what actually happened
- Did the program/services meet its goals/objectives
- Three Steps
 - *Develop hierarchy of goals/objectives ending with participant goals*
 - *Actual operation is compared with its design*
 - *Actual outcomes compared to intended outcomes*

Triangulated

- Data is gathered from multiple perspectives with multiple methods (tri = three)
- Perspectives (participants, parents, staff etc.)
- Methods (Surveys, interviews, observations, testing)
- Purpose is to make disposition decisions about programs (drop, modify or continue as is)
- Audience - program supervisors and administrators

Institutional Review Board

- Institutional Review Board (IRBs) are committees that review research protocols and other materials to ensure the rights, safety, and welfare of the participants
- Federally funded research requires an IRB, but any data collection should get an IRB review as sharing information and data confidentiality are serious issues
- If researcher intends to share evaluation information in any way, should have IRB approval
- If used internally for own program improvement, IRB approval is often not required

Nine Steps of Program Evaluation

1. Purpose: Why Evaluate?
2. Audience: Who is the audience for?
3. Process: How will the evaluation be conducted?
4. Issues: What questions should the evaluation address?
5. Resources: What resources are needed to conduct the evaluation?
6. Evidence: What evidence should be collected?
7. Data-gathering: How is the evidence to be collected?
8. Analysis: How can the evidence be analyzed?
9. Reporting: How can evaluation findings be reported?

Purpose: Why Evaluate?

- A single evaluation often serves a single purpose
- When attempting to serve several purposes, one may never achieve the focus necessary for answering the evaluation question
- Reason for specifying purpose is provide a focus for the evaluation
- Need to determine if the evaluation is being conducted for
 - *Program Assessment*
 - *Program Improvement*
 - *Strategic Management*

Purpose continued:

Examples of purpose statements

- This evaluation is being conducted to determine participant-reported outcomes with program services
- This evaluation is being conducted to help make decisions about the disposition of this program services
- This evaluation is being conducted to evaluate the efficacy of our agency's social justice policies and judge the worth of our agency's distribution of program services across sociometric variable
- This evaluation is being conducted to document the benefits of participating in specific programs

Audience: Who Is the Evaluation for?

- Political considerations – members of the evaluation team
 - *Must have an interest in evaluation findings*
 - *Have the power to use the evaluation findings for making decisions*
 - *Believe the evaluation is worthwhile*
 - *Concerned how the results are used*
 - *Willing to commit time and energy to the evaluation*
- Personal factor
 - *Identifiable person or group who personally care about the evaluation and the findings it generates*
 - *Make a judgement of the worth of the program and subsequent decisions from the evaluation data*

Audience continued:

- Person's position in the organizational structure of the agency helps determine the types of information deemed relevant and therefore desired for evaluation
- Top administrators/boards generally show interest in a summary of broad issue and effects of program services; accountability of resources
- Program directors and middle managers most interested in strategies that effectively meet outcomes
- Direct service personnel mostly concerned the effects of their face-to-face intervention techniques
- Other possible users of evaluation data include:
 - *External funding agencies*
 - *The public*
 - *Clients*
 - *National associations*

Process: How Will the Evaluation be Conducted?

- General techniques
 - *Professional Judgement*
 - *Measurement*
 - *Discrepancy*
 - *Decision Oriented*
 - *Goal free*
 - *Transaction-observation*
- All techniques are either process or preordinate models

Process continued:

- Process Models
 - *Identify a procedure for conducting the evaluation, but do not identify the criteria for making judgements of worth (e.g. goals and objectives)*
 - *Evaluator must provide the criteria for making judgements of worth*

- Preordinate models
 - *Provide the techniques for accomplishing the evaluation and the criteria for making judgements of worth (e.g. use of standards)*

Issues: What Questions Should the Evaluation Address

- Must be concerned with validity and values
 - *Questions must be true indicators of the values interested in getting at the information*
 - *Attendance often used attendance as an indicator of worth – it may be valid indicator of impact, but not of quality, nor can it provide data on outcomes*
- 5 P's of Evaluation
 - *Participants*
 - *Program*
 - *Place*
 - *Policies/Administration*
 - *Personnel*

Resources: What Resources are Available for Evaluation?

- Evaluation costs money
 - *At minimum will consume staff time*
 - *Printing evaluation forms, collecting and analyzing data, and preparing/distributing evaluation report*
- Need to ascertain if there is any data that has already been collected
- Resources available to the agency for data collection
 - *Colleagues*
 - *Universities*
 - *Cooperative extension agencies*
 - *Consultants*

Evidence: What Evidence Should be Collected?

- Evidence is any information that can be used for making judgements of worth about the issues raised
 - *Descriptions of personnel, participants, policies etc.*
 - *Participant judgements about various components of the programs*
 - *Goals and objectives*
 - *Costs*
 - *Program outcomes*
- Need to be concerned is trusting the source and the quality of the evidence
- Need to be concerned with the reliability and validity of the evidence
- Program outcomes – use of a research design

Data Gathering: How is Evidence to Be Collected?

- How is the evidence obtained
 - *How is data actually collected*
 - *When to collect data*
 - *How much data to obtain*
 - *Who to collect data from*
- Questionnaires, interviews, conversations, observation schedules, participant observations, anecdotal data, standardized tests, checklists and rating scales
- What is reliability and validity of data gathering instruments

Analysis: How can the evidence be analyzed?

How can the data be analyzed to determine what they indicate about the program

- Putting data into meaningful patterns
 - *quantitative (numbers – statistics)*
 - Distribution of scores, Mean, median, mode, standard deviation, correlations
 - *qualitative (descriptive)*
 - Compare program results over time; discrepancy comparisons, need reduction, standards, inter-intra agency comparisons

Reporting: How can evaluation findings be reported?

- How can evaluation finding be reported?
 - *formal report (conventional method)*
 - *slides, videotapes, movies, still photos*
 - *audience participation*
 - *Websites, press releases*
 - *Publishing in professional journals*
- The formal summative written report with the familiar spiral binding is the most common

REFERENCES/RESOURCES

- Non-Researcher's Guide to Evidence-Based Program Evaluation. (July 2012). NREPP. http://www.eblcprograms.org/docs/pdfs/NREPP_Non-researchers_guide_to_eval.pdf
- <https://www.samhsa.gov/ebp-resource-center>
- Rossman, J.R. & Schlatter, B.E. (2019). Recreation Programming: Designing, Staging, and Managing the Delivery of Leisure Experiences (Chapter 19-21). Sagamore Publishing.