Non-Researcher's Guide to Evidence-Based Program Evaluation

July 2012

Table of Contents

Course Overview

About This Course

This course provides a basic overview of program evaluation, including the goals of evaluation, types of evaluation, and common study designs. It provides general guidance on how to successfully plan, conduct, and manage an evaluation. The course also covers how to address research quality and disseminate your findings so that the results of your evaluation will have maximum impact.

Intended Audience

NREPP created this course to help program developers and administrators understand in broad terms the steps and activities entailed in a high-quality evaluation. The course is appropriate for those who have organizational resources to conduct an evaluation internally as well as those who will need to hire an external evaluator.

Course Topics

The following topics are covered in this course:

- Why We Evaluate
- Types of Evaluation
- Evaluation Designs
- Research Quality
- The Three Stages of Evaluation
- Hiring an External Evaluator
- Managing an Evaluation
- Evaluation Reporting

Learning Objectives

This course will assist you, the program administrator, to do the following:

- Describe some of the common goals of conducting evaluations.
- Identify the purpose of the three main types of evaluation (process, outcome, and impact).
- Recognize the defining features of the three most common study designs.
- Identify six criteria for measuring research quality.
- Specify the activities that take place during the planning, data collection, and data analysis stages of evaluation.
- Describe key factors to consider when hiring an external evaluator.
- Name management activities important for a successful evaluation.

- List the key components of a well-written evaluation report.

Why We Evaluate

What Is Program Evaluation?

Program evaluation is the systematic process of studying a program (or practice, intervention, or initiative) to discover how well it is working to achieve intended goals. When we evaluate, we gather and analyze information to serve three purposes:

- Program assessment—verifying, documenting, and quantifying program activities and their effects
- Program improvement—finding out where a program may be failing or need improvement
- Strategic management—providing information that can help an agency or organization make decisions about how resources should be applied in the future to better serve its mission or goals

Done well, evaluation provides meaningful results to help you make informed judgments about your program and if and how resources should continue to be allocated for the program's success. You should be able to demonstrate that you are committing resources to efforts that produce tangible benefits for your target population.

Goals of Evaluation

Any number of goals can be accomplished with a well-designed evaluation study. Potential objectives of an evaluation include:

- **Clarify program objectives:** *What are you trying to accomplish? How will you define success?* A well-planned evaluation requires you to clarify assumptions about the links between your target population, the program activities, and the immediate, intermediate, and long-term outcomes you expect. Evaluation helps you decide where you are going and whether you are taking the best path to get there.

- **Assess your program's appropriateness and effectiveness:** *Is the program working as intended? Is it the right way to address the problem?* Evaluation can help you determine whether the program is an effective way to deal with the problem of interest.

- **Advocate for your efforts:** *How can you energize your supporters and win over potential critics?* In the rush to do good things, program developers sometimes forget to explain their actions. Evaluation helps to educate others about your program as well as the underlying philosophy and the results being achieved. In turn, stakeholders and other interested parties will better understand the program and what to expect from it. Because it relies on a

scientific process, evaluation has a major advantage over mere opinion. It is easy to say you know something is working and provide anecdotes that support that view. However, as we grow more sophisticated in our decisionmaking, it becomes increasingly important to conduct evaluations that can demonstrate results empirically.

- **Solidify support for continued financial investment:** *How can you ensure continued funding for your program?* Positive results from a well-planned evaluation can be used to justify a program's existence, maintain commitments from program administrators or the community, and leverage additional resources from funders, community partners, and other stakeholders.

- **Address program cost:** *Is the program easily implemented with a reasonable budget? Are there opportunities for cost savings? Do the results justify the investment?* Evaluation can give valuable information about ways to save money and whether the program is delivering sufficient "bang for the buck."

- **Make informed program improvements:** *Are there any changes that should be made to the program before it is allowed to continue or is replicated elsewhere?* Rarely does a program go from conception to implementation with perfect success. Most of the time, there are lessons to be learned through experience and room for improvement in the future. The information obtained through evaluation enables you to fine-tune your program or make midcourse corrections when needed.

Additional objectives of an evaluation might include:

- **Monitor program fidelity and integrity:** *Are you really doing what you said you would do?* Tracking the number and type of activities you offer, the number and type of participants involved, and your activity-related expenses can help determine if you actually are implementing the program as intended and promised.

- **Provide a blueprint for peers:** *Will other agencies or organizations be able to implement the program?* A quality evaluation provides a road map for others to follow, with the potential barriers and pitfalls identified.

- **Recognize unanticipated or unintended program effects:** *Are there "side effects" that need to be documented and understood?* Evaluation often uncovers program effects—both positive and negative—that were neither anticipated nor intended. These effects are as important to understand as the intended outcomes of the program. However, they can be harder to detect by definition, and may be lost if no evaluation process is in place.

- **Pilot test innovations:** *How can we identify new ideas for the future? How can we demonstrate the value of a new program or intervention?* Prevention and treatment interventions must continually evolve, and those responsible for implementing these interventions must be willing to take reasonable risks

on new ideas. The challenge is to conduct quality evaluations to determine whether promising ideas might pan out in practice.

- **Contribute to the field:** *Does this evaluation help our understanding of future directions?* The lessons that you learn from your evaluation can be shared with other agencies, organizations, and communities. Sharing evaluation findings regularly will contribute to creating a much-needed body of verifiable information about what works in mental health and substance abuse prevention and treatment.

Final Thoughts on Why We Evaluate

Thinking about what you want to get out of an evaluation is an important first step, but it is just the beginning. From conception to completion, a lot of thinking and planning must go into designing an evaluation that works for you. The next seven topics of this course provide a broad overview of some key concepts in evaluation that you should understand before beginning a study.

Types of Evaluation

Determining Your Focus

We just reviewed the many possible motivations for evaluation and the specific goals it can achieve. Some or all of these factors might influence you to make the decision to evaluate. Once you commit to conducting an evaluation, however, you must decide precisely what questions you would like answered and what you want to measure. Keep in mind that a study should be clearly focused, and you may need to undertake more studies over time to get the whole picture about your program.

What information is most important to you? For example, do you want to know:

- Is my program being implemented as intended?

- Are program participants achieving desired outcomes?

- Are our efforts effecting change in the population in which the problem was initially documented?

This section explains the three types of evaluation that are most appropriate to answer these kinds of questions, respectively: process evaluation, outcome evaluation, and impact evaluation.

Process Evaluation

A process evaluation describes how your program is being implemented. This type of evaluation is important for several reasons. First, it can be used to determine the extent to which a program is implemented as intended—a concept known as intervention fidelity, which is among the six research quality criteria we introduce

later in this course. Second, a process evaluation can provide a program developer with information on the implementation process. This information can in turn be used to refine the delivery of a program and improve its quality.

Process evaluation data also may be important in the interpretation of outcome data. If your program is not achieving expected outcomes, it may be because there are problems with intervention fidelity, meaning the program is not being delivered as intended. A process evaluation can help identify discrepancies between program design and delivery and ultimately enable you to measure outcomes that more accurately reflect the program as designed.

Process Evaluation Questions

A process evaluation might ask:

- What are the critical components and activities of the program?

- What aspects of the implementation process are facilitating success or acting as stumbling blocks?

- Are the program providers receiving the proper amount of training and supervision to ensure fidelity?

- To what extent does what is being implemented match the program as originally planned?

- Do program participants understand the program and its intended objectives, and are they able to participate fully in the required components of the program?

- Are there important dosage effects, or differences in outcomes based on how many program modules or sessions were implemented, the length of sessions, attendance rates, or other variables?

- What strengths can be built upon to improve the program? Conversely, are there gaps or deficiencies in the services or activities being provided that need to be addressed?

In itself, a process evaluation usually does not seek to provide evidence that a program is effective, although it may provide insight to why a program is or isn't working.

Outcome Evaluation

An outcome evaluation investigates whether changes have occurred for the people participating in a program. It quantifies the magnitude (how big) and direction (positive or negative) of those changes and the circumstances associated with them. An outcome evaluation also seeks to tie these changes to specific elements of the program. This is a way of testing whether the logic model or rationale for the program is valid. An outcome evaluation essentially asks, "What is my program accomplishing in the short term?" and "Am I meeting my objectives?"

When an outcome evaluation demonstrates the effectiveness of a program in achieving its intended outcomes, it not only argues for continued investment in the program but can provide justification for replication elsewhere.

Outcome Evaluation Questions

Questions asked in an outcome evaluation might include:

- What effect is the program having on its stakeholders or 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 should continue to support this program?

Impact Evaluation

This type of evaluation answers the questions, "Is my program producing long-term, global changes?" and "Am I meeting my long-term goals?" Data are collected about the long-term or wide-reaching impact of the program.

Impact and outcome evaluation are often confused. The key difference is that an outcome evaluation documents short-term or immediate outcomes, while an impact evaluation is focused on long-term, more global changes. For example, an 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, on the other hand, might focus on changes in reports of past 30-day substance use among all adolescents in a particular school district.

Impact Evaluation Questions

An impact evaluation usually asks questions like these:

- What effect is the program having on our long-term goals (e.g., change in the number of reported incidents, change in rates associated with the problem)?

- 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 the result of implementation failure or some aspect of the program itself?

- What degree of confidence is there that the outcomes can be attributed directly to the program?

Evaluation Designs

What Is Evaluation Design?

Evaluation design refers to the structure of a study. There are many ways to design a study, and some are considered to be more rigorous than others. Evaluation designs are differentiated by at least three factors:

1. The presence or absence of a control group (or groups) in the study,
2. How participants are assigned to study groups (i.e., with or without randomization), and
3. The number of times or frequency with which outcomes are measured.

Control Groups

A control group is a group of individuals who participate in the study but do not receive the main intervention being tested. Depending on what the study is trying to establish, the control group may receive something else instead of the intervention or may receive nothing. Using a control group tells you what might have happened to the intervention group over the study period if they had not received the intervention.

By including in the study a group of people who are equivalent to the intervention group in almost all aspects other than their participation in the intervention, you can have greater confidence that the outcomes you observe in the intervention group are in fact associated with the intervention and not other things.

Randomization

Random assignment is a way of distributing the participants in your study (also known as the study sample) among different study groups using a random process, as opposed to assigning participants based on certain criteria. When you use preexisting groups as your study groups or use other methods of assignment, there may be systematic differences between the groups at the outset of the experiment, which is a source of bias.

Random assignment controls for this source of bias and enables you to have more confidence that the outcomes seen in the intervention group are in fact tied to the intervention.

Frequency of Outcome Measurement

The frequency of outcome measurement is another important aspect of evaluation design. Some studies measure outcomes only once, after the intervention. Some studies measure outcomes before and after the intervention. Others go even further and add one or more follow-up measurements weeks or months after the intervention. Within any of the three evaluation designs, there may be one point of measurement or multiple points of measurement.

Measuring outcomes before and after an intervention is desirable, since it allows you to determine how much outcomes change over the course of the intervention and whether these changes persist over time. However, some outcomes may be more appropriate to measure only after the intervention, either because the outcome might not be measurable before the intervention or because the degree of change in the outcome is not as important as establishing that something has occurred (or not occurred) or that a certain threshold has been met once the intervention has been provided.

Categories of Evaluation Design

The three basic categories of evaluation design, in order of least to most rigorous, are:

- Preexperimental,
- Quasi-experimental, and
- Experimental.

Next, we explore in detail what each of these designs entail.

Preexperimental Design

A preexperimental design is defined by the absence of a control group and by the absence of random assignment. Only the program participants are tracked in these types of studies.

There are several variations within the category of preexperimental design, differing only in the number and the timing of outcome measurement points. These variations include:

- Posttest only,
- Pre- and posttest, and
- Pre- and posttest with follow-up.

Posttest-Only Preexperimental Design

In the posttest-only design, data are collected only once from participants, immediately after they complete the program. The image below illustrates what posttest-only data collection looks like in a preexperimental study.

Group	Pretest	Treatment	Posttest
Intervention Group		**X**	**O**

X = intervention is administered
O = measurement is taken

The major drawback of any posttest-only study is that it does not provide a baseline to which postintervention results can be compared. To assess change, you must be able to compare post intervention data with a baseline measure.

However, a posttest-only design can be appropriate if you only need to know—or if it is only feasible to know—if participants have reached an identified outcome (for example, learned a new skill), rather than measure the degree of change. Also, if you have limited resources to spend on the evaluation or limited access to program participants, this could be the best or only option.

Pretest and Posttest Preexperimental Design

A pre- and posttest design measures outcomes among program participants before and after the intervention. The image below illustrates what pre- and posttest-only data collection looks like in a preexperimental study.

Group	Pretest	Treatment	Posttest
Intervention Group	O	X	O

X = intervention is administered

O = measurement is taken

This design is relatively easy to implement. You simply administer the same measure twice, before and after the intervention. The timing of the posttest measure is important. It should allow enough time for your program to have an effect, but not so much time that program effects are diluted or influenced by confounding factors such as participation in other programs or participant social, psychological, or environmental circumstances. (Controlling for confounding factors is an important part of research quality that we discuss later in this course.)

Pretest and Posttest Preexperimental Design With Follow-Up

You can improve on the pre- and posttest design by adding follow-up measurement points—for example, at 3, 6, or 12 months postintervention, or perhaps earlier, depending on the likelihood of finding effects in the short and long term.

The pre- and posttest with follow-up design enables you to gain a more longitudinal (i.e., over time) perspective on your program. The image on the next page illustrates what data collection looks like in this type of design.

Group	Pretest	Treatment	Posttest	Follow-up
Intervention Group	O	X	O	O

X = intervention is administered

O = measurement is taken

Follow-up measurements can tell you if a program has a sustained effect, yielding changes in behavior or knowledge that persist beyond the immediate treatment period. However, it is sometimes difficult and costly to locate participants for follow-up assessments, and you will need to plan carefully how you will ensure participant retention over the follow-up period.

When the same instrument is administered multiple times, the test repetition is also a potential confound. Participants could become bored or annoyed by having to complete the same measure again, or they may be less interested in participating months or years later. These situations can affect the response rate for follow-ups as well as the accuracy and validity of the responses that are obtained. However, these are issues that you may be able to control through planning and careful administration of your measures.

Quasi-experimental Design

While a preexperimental study follows one group of participants receiving the intervention, a quasi-experimental design uses two or more study groups (see image below). The study groups are preexisting sets of people sharing some common variable. For example, participants already enrolled in a program may be compared with individuals waiting to receive the intervention, or with individuals who have chosen some other treatment or no treatment. The term "control group" is commonly used to describe the comparison group in a quasi-experimental design; however, some researchers do not consider quasi-experimental designs to have a true control condition because the study groups are preexisting or self-selected rather than created through random assignment.

Group	Pretest	Treatment	Posttest
Intervention Group	O	X	O
Control Group	O		O

X = intervention is administered

O = measurement is taken

Types of Control Conditions in a Quasi-experimental Design

The control group in a quasi-experimental study may receive a different intervention, selected components of the intervention being tested, or something that merely mimics the time and attention paid to participants (i.e., a placebo). Another alternative is the wait-list control, which means the control group participants receive nothing during the study period but will eventually receive the intervention some time after the study period.

Benefits and Challenges of Quasi-experimental Design

The quasi-experimental design is frequently the most practical option for conducting outcome evaluations in the social services context. By using preexisting or self-selected groups, such as individuals already enrolled in a program, it avoids the additional steps required with random assignment to study conditions, as well as the potential ethical concerns involved in withholding or delaying treatment or substituting a less effective treatment for one group of study participants. The significant limitation of this design is that without randomization, the study groups may differ in important ways that account for some of the group differences in outcomes after the intervention. In other words, this design provides comparatively weaker evidence of program effects than one that uses randomization.

Experimental Design

The most sophisticated evaluation studies use an experimental design, considered to be the gold standard. The hallmark of this design is randomization. Randomization reduces the likelihood that study groups will differ from one another before the intervention, which can be a potential confound. However, randomization by itself cannot guarantee the absence of group differences. To confirm that the randomization was effective, you must test the groups after randomization for demographic differences or any other preexisting differences that may affect outcomes. This step helps to ensure the groups are comparable.

Data collection in an experimental study resembles that of quasi-experimental studies, but with the addition of random assignment. The image below illustrates what data collection looks like in an experimental design.

Group	Pretest	Treatment	Posttest
Intervention Group (Randomly Assigned)	O	X	O
Control Group (Randomly Assigned)	O		O

X = intervention is administered

O = measurement is taken

Benefits and Challenges of Experimental Design

An experimental design will offer the best evidence that your program is responsible for changes in outcomes. It is used most often by researchers in the physical sciences because they can control their lab environments, repeat experiments, and determine causality. For obvious reasons, it is much more difficult to implement valid experimental designs in human services settings.

Often, ethical concerns render experimental evaluation of human services unfeasible. Some human service administrators may be unwilling to allow participants to be randomly assigned to experimental and control groups, as they believe it denies treatment to individuals who need it (although wait-list control groups can remedy this issue).

For some agencies, the main constraint may be the higher cost, as experimental designs require tracking an extra study group through a period of time in which they may not be receiving the benefit of an intervention.

However, an experimental design is worth considering if your resources and circumstances allow it. As evaluation has become more common in the social services arena, especially in light of the trend towards evidence-based programming, many agency directors are willing to work with an evaluator to implement experimental design because they know it is the best way to determine whether the program to which they are committing resources is achieving its intended outcomes.

How To Choose the Right Evaluation Design

To select the design that is right for your circumstances, consider your needs and resources and pick the most rigorous design possible. Each type of design has its advantages and disadvantages, so be realistic about what kind of design you can afford and implement.

Factor #1: What are your research questions?

Start by considering what research questions you want to pose and what information you want to generate from the study. An evaluation that seeks to establish your program's long-term impact may require a different design than one that will focus on immediate effects on study participants.

Factor #2: Can you collect the data?

Think about whether you will be able to collect the data required for a given design. More complex evaluation designs require you to collect more data over multiple measurement points. This lengthens the data collection phase of the study and may require more sophisticated analyses to be conducted once the data are collected.

While undertaking an ambitious study can yield rich information about your program, you may find that your available funding requires a simpler approach. If your program is relatively new and your resources are limited, it is perfectly

acceptable to start with a smaller study and then plan to conduct more sophisticated studies in the future, building on the results from your first study.

Factor #3: Is the experiment feasible?

As we just discussed, in the social services context, program stakeholders may not be willing to allow participants to be assigned to a control condition in which they receive no treatment.

Designs that require a control condition may not be feasible in these cases. Evaluating social service programs, including those designed to achieve substance abuse and mental health outcomes, often means adapting evaluation techniques to fit circumstances like these.

Design Considerations for NREPP Acceptance

In all situations, you should strive to implement the most rigorous design possible. The more rigorous your evaluation—particularly if you are able to use an experimental or quasi-experimental design—the better it will be received by the scientific community and the more weight your findings will carry.

If your ultimate plan is to have your program included in a registry of evidence-based programs, such as NREPP, consider the reviewing body's applicable research requirements. For example, to meet NREPP's minimum requirements for a review, you must be able to submit at least one quasi-experimental or experimental study of the program that shows statistically significant outcomes (additional requirements may apply; for more information on the current minimum requirements, visit the NREPP Web site[1].

Remember that if you are limited to a less ambitious design for an initial study, you may be able to build on the success of that first study to generate support for a more rigorous evaluation down the road.

Research Quality

Why Research Quality Matters

When you disseminate your study findings to the public, the chances are good that your evaluation methods will be critically assessed by others in the scientific community. There are a number of standards people use to determine the quality, or methodological rigor, of a study. It is important to understand what these standards are so that you can address them from the beginning and avoid the common pitfalls of poorly designed or poorly executed studies. These pitfalls are sometimes called "threats to validity" because they detract from the overall validity

---

[1] NREPP Web site, http://www.nrepp.samhsa.gov/

of your findings. Controlling threats to validity is an integral process that must be built into every step of your evaluation.

A high-quality study is not any more (or less) likely than a less rigorous study to show that your program is effective; however, whatever findings it generates will have more scientific value. When you can clearly document that you have conducted a high-quality study, people who take a critical look at your research are more likely to consider your conclusions to be accurate and valid.

The NREPP Model

Next, we go over the rating criteria that have been developed by SAMHSA's National Registry of Evidence-based Programs and Practices (NREPP) as a starting point for understanding quality. If you are interested in having your program reviewed and included in NREPP, you need to be aware of what these criteria measure, as well as what minimum requirements you will need to meet in order to be considered for an NREPP review. If you understand how to achieve high ratings for each criterion, you can build quality into your study and maximize your potential for a positive review. Even if you do not intend to submit your program to NREPP, taking deliberate steps to meet these quality standards can help validate and support the evidence you generate through your study.

About NREPP

NREPP is a searchable online registry of mental health and substance abuse interventions that have been reviewed and rated by independent reviewers. The purpose of this registry is to assist the public in identifying scientifically based approaches to preventing and treating mental and/or substance use disorders that can be readily disseminated to the field. In an NREPP review, two things are evaluated: Quality of Research (the strength of the evidence supporting outcomes) and Readiness for Dissemination (the quality and availability of training and implementation materials). For the purposes of this course, we will focus on Quality of Research (QOR).

NREPP Rating Criteria

NREPP's QOR rating criteria were developed to provide a focused, objective, and efficient way of rating the quality of research studies in the behavioral health field. Additional standards for measuring research quality exist, but these six criteria were specifically chosen to be representative of the fundamentals of quality.

Criterion 1: Reliability of Measures

Outcome measures should have acceptable reliability (i.e., at a level that is conventionally accepted by experts in the field) to be interpretable.

Criterion 2: Validity of Measures

Outcome measures should have acceptable validity (i.e., at a level that is conventionally accepted by experts in the field) to be interpretable.

Criterion 3: Intervention Fidelity

The "experimental" intervention implemented in a study should have fidelity to the intervention proposed by the applicant. Instruments that have tested acceptable psychometric properties (e.g., inter-rater reliability, validity as shown by positive association with outcomes) provide the highest level of evidence.

Criterion 4: Missing Data and Attrition

Study results can be biased by participant attrition and other forms of missing data. Statistical methods as supported by theory and research can be employed to control for missing data and attrition that would bias results, but studies with no attrition or missing data needing adjustment provide the strongest evidence that results are not biased.

Criterion 5: Potential Confounding Variables

Often variables other than the intervention may account for the reported outcomes. The degree to which confounds are accounted for affects the strength of causal inference.

Some common confounds include history (any event that happens between measurements that affects participants), maturation (normal processes of development that affect may account for differences seen in people over time), test-taking effects (resulting from repeated exposure to the same test/instrument), flaws in instrumentation, selection bias (in the absence of random assignment), and statistical regression (tendency for extreme scores on any measure to move toward the mean).

Criterion 6: Appropriateness of Analysis

Appropriate analysis is necessary to make an inference that an intervention caused reported outcomes.

When NREPP reviewers use these criteria, they apply them not to the intervention or to an individual study, but to individual outcomes. Programs may not be equally effective in achieving all their intended outcomes. In addition, one study may use a variety of measurement techniques and analytical methods for different outcomes, and different studies of the same program may measure outcomes in different ways over time. For all of these reasons, NREPP applies the QOR criteria to individual outcomes, sometimes looking at multiple studies for that outcome, but always rating the outcome rather than the study.

Throughout the rest of this course, tips are periodically provided on how to address these criteria during the various stages of your study. For more information on the

criteria and associated rating anchors, please visit the QOR Review[2] section of the NREPP Web site.

NREPP Submission Requirements

Particularly if you are embarking on the first evaluation of your program, keep in mind that NREPP only reviews programs that meet certain minimum requirements, and it may take time to generate an evidence base that meets these requirements. As of the 2011-2012 open submission period, NREPP's minimum requirements[3] specify that interventions:

- Have been evaluated using an experimental or quasi-experimental study design;

- Have demonstrated one or more positive change outcomes in mental health and/or substance use among individuals, communities or populations;

- Have results that are published in a peer-reviewed publication or documented in a comprehensive evaluation report; and

- Provide documentation, such as manuals, guides, or training materials, to facilitate broader public dissemination of the intervention.

Three Stages of Evaluation

Beyond Exploration

At this point, you should have a good idea of:

- Why you want to conduct an evaluation and what you hope it will accomplish,

- What type of evaluation best suits your needs,

- Which evaluation designs might be feasible for your situation and will answer your targeted questions, and

- The quality of research standards you should strive to achieve.

The rest of this course explains what you need to know if you decide to move forward from the exploration stage and begin the process of conducting a formal evaluation study. In this topic, we will discuss the activities you will undertake in each of the three stages of evaluation: planning, data collection, and data analysis.

Planning

Like any large-scale project, a successful evaluation begins with a thoughtful planning process. Preferably, this process includes the development of a written

---

[2] Quality of Research Review, http://www.nrepp.samhsa.gov/ReviewQOR.aspx
[3] Minimum Requirements, http://www.nrepp.samhsa.gov/ReviewSubmission.aspx

plan to document your approach and the decisions made during planning. An evaluation plan specifies what you will evaluate and how. Some funders may require you to submit an evaluation plan, but it is a good exercise to draft one even if it is not required.

The evaluation planning process usually involves determining:

- What questions the evaluation will attempt to answer,
- Who will serve as your study sample (intervention group and control group, if applicable),
- What outcomes, outcome measures, and data sources will be used, and
- The statistical power needed for your study.

The planning process also might include other preliminary work such as obtaining institutional review board approval. Each of these topics is explored in detail next.

Deciding on Evaluation Questions

An evaluation must proceed from a well-defined question or set of questions. Here are some examples of the kinds of questions typically addressed:

- Is the program being implemented as intended?
- Is the program reaching its target population?
- What are the barriers to implementing the program?
- What approaches to overcoming implementation obstacles have been effective?
- Is the program achieving its objectives?
- What are the actual outcomes?
- Is the program cost-effective?

Whether on your own or working with an external evaluator, you will need to invest some time thinking about your desired objectives and then hone in on a specific set of questions, which you will then document in your evaluation plan.

The evaluation questions will help determine what type of evaluation you should conduct as well as your choice of evaluation design.

Defining the Study Sample and Group Assignment

The target population from which you will draw your study sample must be identified early in the planning process, since recruitment can be a time-consuming process. If you are using a preexperimental design, only one group of study participants will need to be identified and recruited. Quasi-experimental and experimental designs will require identifying a control group in addition and determining what the control group will receive.

To identify a study sample, you will need to:

- Determine where you will obtain the participants (e.g., what program sites, what cities or jurisdictions, etc.),

- Define any applicable eligibility criteria (e.g., participants must be between the ages of 21 and 55, participants must not be receiving other treatments for the condition), and

- Build partnerships, if they do not already exist, with the agencies and personnel who can grant you access to the population.

Study participants generally are not recruited until the necessary approvals are received for the study.

Choosing Outcomes

Outcomes say in concrete terms what you will measure to answer the evaluation questions you have posed. Most studies look at multiple outcomes. Some examples of outcomes are "30-day substance use" or "family functioning." Outcomes can be short term, intermediate, or long term. Short-term outcomes might be assessed immediately following program completion, while intermediate and long-term outcomes might be measured six and 12 months following program completion, respectively. In some cases, short-term outcomes may help predict intermediate and long-term outcomes, becoming mediating variables in explaining how your intervention works to achieve its ultimate goals.

For example, your study may hypothesize that short-term changes in knowledge and skills will eventually result in changed behavior.

Outcomes should be:

- Relevant to your program's goal and objectives,

- Important to achieve if your program is to attain its objectives, and

- Indicative of meaningful changes.

Choosing Outcome Measures

For each outcome, at least one appropriate measure (or instrument) must be identified. The scientific usefulness of a measure is characterized in terms of its validity and its reliability. Validity is the degree to which an instrument measures what it purports to measure. Reliability refers to an instrument's consistency of measurement. Ideally, a measure will give consistent results even when it is administered in different places, at different times, or by different people (this latter aspect is known as inter-rater reliability). In addition, related items should produce similar results, showing consistency of results within the measure itself (this is known as inter-item reliability).

When you can document that your measures have demonstrated validity and reliability, you can be much more certain in the quality of the data they yield.

Validity and reliability of measures are key aspects of research quality, and in fact they account for one third of the total Quality of Research score that NREPP assigns when reviewing a given outcome in a study.

Using Existing Measures Vs. Home-Grown

Standardized outcome measures may already exist for the kind of data you wish to collect. If not, you may need to develop your own measure. One advantage to using a standardized measure is that its validity and reliability should be well established in the literature. On the other hand, standardized measures can be costly and may require special training to administer.

Developing your own outcome measure allows you to tailor questions to your program goals, may be less expensive than using a standardized measure, and avoids any restrictions on the qualifications of staff that can administer or score the instrument. Any measure that is "home-grown" will need to be tested before use in a study, so that you can determine its reliability and validity and document its psychometric properties.

Measuring Fidelity

In addition to whatever outcome measures you pick, consider including measures of implementation fidelity in your evaluation. Since the goal of evaluation is to draw conclusions about the program itself—not just a particular instance of its implementation—it is important to be confident that the program was implemented as intended. Fidelity measures are specifically tied to your program components, so it makes sense for you to develop these instruments yourself or work with an evaluator to develop them. Activity logs, record review, observation, and interviews are among the ways to measure fidelity.

When you measure implementation fidelity, you will be answering these questions:

- Is the program being delivered as designed? (e.g., Are core components being implemented in the proper order?)

- Are program recipients receiving the proper "dose" of the program? (e.g., Are all sessions implemented? Is each session of the length specified?)

- Is the quality of program delivery adequate? (e.g., Are providers trained and skilled in delivery of the program?)

Identifying Data Sources

The evaluation planning stage also requires the identification of appropriate data sources. Depending on your targeted outcomes, data sources may include:

- Individuals participating in your program, or others who can provide secondary reports on those individuals (teachers, parents, spouses),

- Staff involved in program implementation,

- Documentation generated by your program, both administrative and operational (e.g., program records, operations handbooks, policies and procedures, training materials, proposals, past evaluation plans or reports, budgets, organization charts, memoranda), and

- Archival (existing) databases and documents; examples include school records, State records, police reports, national datasets, or records from your own agency.

Quantitative Data

The nature of the data you collect will be either quantitative or qualitative (or you may want to collect both). If you are using an experimental or quasi-experimental design, the emphasis of your study will likely be on quantitative data, since these are the data to which statistical tests can be applied.

Quantitative data can be counted and expressed in numerical terms. Sources of quantitative data include surveys, questionnaires, and archival records or databases. Surveys and questionnaires can be self-report (i.e., people give information about themselves), or they may ask third parties, such as teachers, parents, caregivers, or spouses, to provide information about an individual. Whenever possible, it is a good idea to use standardized quantitative instruments that have been well tested in the field and have demonstrated strong psychometric properties.

Qualitative Data

Depending on what questions your study poses, you may also seek to collect some qualitative data. Though not useful for statistical analysis, qualitative data can provide important insights and context (such as attitudes and perspectives of people involved in your program) that cannot be gained through quantitative data collection.

Qualitative data capture information that is difficult to measure, count, or express in numerical terms. Interviews and focus groups are two of the most common ways of collecting qualitative data, although other sources are possible.

Interviews

In interviews, respondents are asked open-ended questions to allow them to express in their own words their experiences with and attitudes toward the program. Interviews typically ask people to describe or explain particular program practices or issues in relation to their personal experiences with them. Interviews can be a primary data source, a first step in developing other data collection strategies (e.g., survey or focus group questions), or a complement to quantitative data collection. Among the strengths of interviews is that they let you gather information you cannot learn through document review or observation. While open-ended techniques such as interviews can provide a fuller picture of respondents' experiences and opinions than surveys, they do tend to be time consuming.

Focus Groups

A focus group is a small-group discussion, facilitated by a trained leader that explores a topic in depth. Focus groups are a quick way to gather data on respondents' impressions and opinions. As with all data collection strategies, successful implementation depends on thorough planning. To prepare for a focus group, ask yourself "What do I want to learn?" and "Why am I conducting a focus group?"

Calculating Statistical Power

Another important step in planning is confirming that your study will be adequately "powered." This has to do with the statistical tests you plan to use. Power refers to the probability that a given test will find a statistically significant difference (either between study groups or within groups over time) when such a difference actually exists. It is generally accepted that power should be 0.8 or greater. In other words, assuming that a statistically significant difference exists, a test ideally should be able to detect it at least 80% of the time.

Power is affected by several factors, including the specified significance level, sample size, and measurement error.

Significance Level

The significance level specified in your statistical calculations can affect the power of your study. The purpose of calculating statistical significance is to quantify the likelihood that an event or change observed in a study has resulted from chance, rather than from the variable being tested. A study finding can be considered statistically significant when the likelihood of chance occurrence is lower than the specified significance level. Two commonly used significance levels are .05 or .01, representing 5% and 1% likelihood of chance occurrence, respectively. Depending on the strength of evidence you expect to be able to demonstrate, the significance level may be set higher or lower.

Sample Size

The sample size, or number of study participants in each study condition, is another factor affecting study power. Other things being equal, effects are harder to detect in smaller samples, making it more desirable to have a larger sample. Increasing sample size can increase statistical power. However, you will need to weigh the tradeoffs associated with a larger sample, such as the higher costs of collecting and analyzing data from more study participants. Also, the sample size may be hard to manipulate if your participants are being recruited from a preexisting pool of individuals.

Measurement Error

Using outcome measures that have a low level of measurement error can positively affect power. To minimize measurement error and improve the power of your

study, use measures that are appropriate for your outcomes and that have been documented to have high reliability and validity.

One last point to make about statistical power is that the precision with which outcomes are measured also can influence the power of your study. Reducing the measurement error in the data can improve the power of your study. By choosing outcome measures for which validity and reliability have been well documented, you may be able to achieve a low error rate and, in turn, a more strongly powered study.

Obtaining Institutional Review Board Approval

Once all of the planning decisions are made, the next step is usually to apply for institutional review board (IRB) approval. IRBs are committees that review research protocols and other materials to ensure the rights, safety, and welfare of human subjects participating in studies.

For research that is federally funded, an IRB review is mandatory, but IRB reviews should be considered for any study using human subjects. Invasion of participants' privacy (e.g., sharing information with those not involved in the study) and data confidentiality issues (e.g., securing study data) are serious concerns. This is why academic researchers almost always send their survey protocols, consent forms, and data collection methods to an IRB for review. The IRB can certify that the rights of subjects will be protected, that any potential adverse effects on participants will be minimized, and the data will be securely managed and maintained.

IRB clearance may be required by your agency leadership or funding source, so be sure you understand what is required before you begin to recruit subjects and collect data. You do not want to get halfway through your study only to find out that you did not meet whatever formal approval was required.

If you are affiliated with a university, you should have access to a university IRB that can review and approve your study procedures. If not, you might ask someone from the funding agency about how you should go about getting an IRB review. While there are private organizations that are in the business of conducting such reviews, the costs can be high (in some cases, between $8,000 and $12,000 for a full review).

Data Collection

The data collection stage of an evaluation is "where the rubber hits the road" and you begin collecting the information you will ultimately analyze and report, following the guidelines set forth in the evaluation plan. Data collection is not simply about administering surveys and checklists. It also involves activities such as:

- Obtaining permission to collect data from participants,

- Developing strategies to retain participants, and

- Putting data management systems in place and implementing procedures to properly manage and secure the data.

Obtaining Participant Consent

Participants have the right to privacy regarding any information they provide to you. Their participation must be both informed and voluntary. In most cases, you will need to obtain informed consent from your participants before collecting information from them. Informed consent enables participants to learn exactly what their participation in the study will involve, including any potential risks or benefits to them, so that they can make an informed, voluntary decision to participate or not.

Informed Consent

Informed consent is usually obtained using a form that contains:

- A description of the research, including its purposes, expected duration of participants' involvement, and procedures;

- A description of any risks and benefits to the participant;

- Information about how information will be kept confidential;

- A statement explaining that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the participant is otherwise entitled, and the participant may discontinue participation at any time without penalty or loss of benefits to which the participant is otherwise entitled; and

- Contact information for the person responsible for conducting the study, along with a statement that this person can be contacted for answers to pertinent questions about the research and participants' rights, and in the event of a research-related injury to any participant.

Assent

For participants younger than 18 years, you may need to obtain their assent as well as consent from their parent(s). Assent is obtaining agreement to participate in a study from an individual who is unable to provide consent, either because they are minors or because they are physically or mentally incapable of making informed decisions. Assent is not always required, but should be considered regardless of requirements. Respecting the rights of study participants should always be a priority.

Retaining Participants

Participant retention is critical to the success of any evaluation that involves more than a single data collection point. Retention strategies must be comprehensive to ensure that data are collected at each time point from a maximum number of participants.

Following are some tips for maximizing participant retention:

- **Collect ample contact information:** Collect enough information to help locate participants over the duration of the study. Also request contact information for three relatives or friends who are most likely to know where to contact the participants at any given time.

- **Stay in contact:** Emails and postcard reminders can be sent to participants prior to each data collection period. These can also serve as reminders to participants to keep their contact information up to date.

- **Provide incentives:** Incentives are most effective when they are age appropriate (e.g., music CDs and movie passes for students, grocery store gift cards for parents) and proportional to the effort required (i.e., not so large as to be considered coercive).

- **Designate a study representative:** Try to identify or recruit one person at each program site to help retain study participants. Study retention can be positively influenced by the presence of a familiar and trusted study representative. Make sure your representative has a strong presence at the program site and established rapport with participants and other stakeholders (e.g., teachers, parents, and program staff).

- **Consider developing a Web site:** Web sites can be used to create an affiliation and sense of belonging for study members, and serve as an effective means for posting such important study information as progress reports and data collection timelines. They can also be used by participants to update contact information and, in some cases, to complete follow-up surveys. Increasingly, researchers and evaluators are using social media outlets such as Facebook and Twitter to retain study participants or maintain ongoing contact with them. While these tools pose potential security and confidentiality issues, (which can be overcome or managed), they may be particularly relevant for an evaluation where the primary subjects are teenagers, a group for whom social media outlets are highly salient.

Managing and Securing Data

It is not an overstatement to say that safe and proper handling of data is one of the most serious responsibilities evaluators have, if not the most important one. The loss or misuse of data threatens the success of your study, and more importantly, it may represent a violation of study participants' rights or even put participants at risk. The information collected in evaluations of social services programs is often highly personal and sensitive in nature, which makes the stakes particularly high.

The Data Collection Plan

Your approach to managing and securing data should be thoughtfully developed and outlined in a data collection plan. The data collection plan identifies the instruments to be used, who is responsible for administering the instruments, the timing and/or frequency of administration, and the specific information collected. Any staff training that might be required related to data collection also might be

discussed in this plan. In addition, the plan should document how you will enter, track, store, and secure data, even if you are only collecting data from a small sample.

Data Management Activities and Responsibilities

Data management is an ongoing process that is best handled by establishing rigorous procedures early in your study, training staff to adhere to these procedures, and conducting periodic quality control checks to make sure the procedures are being followed. You also will need appropriate software and other tools to properly enter and store the data.

Activities and responsibilities involved in data management include the following:

- **Data identification:** When training staff and developing written protocols, emphasize the importance of labeling and identifying data (using information such as participant names or identification numbers, date, location, and the names of staff collecting the data) so that responses can always be identified and traced back to the source.

- **Data tracking:** Use a chart or other tracking system to document the overall progress of data collection. For example, you could use a chart that lists all of the instruments by name, showing the source of the data (e.g., program participant, program document), method of data collection, and dates of administration.

- **Data entry and management:** Evaluations usually generate enough data to require the use of a database. Entering data into a database allows you to know at a glance how many people have responded, when, and if there are any missing participant data. Your database also can be used to carry out data analysis.

- **Data storage and security:** Store any information collected through your evaluation in a secure place, safe from damage or loss. Do not keep written surveys or other papers with confidential information in unlocked filing cabinets. Determine exactly who should be able to access the data and put controls in place to prevent others from gaining access. Any electronic system can potentially fail or lose data, so it is a good idea to institute redundant systems. For example, if the data are originally in paper format, you may want to keep those paper copies for backup. If the data are in electronic format to begin with, keep backup copies of the database on flash drives or CDs. Make sure backup copies are not available to anyone other than members of the research team.

The Confidentiality Plan

The confidentiality plan, usually included as part of the overall data collection plan, details the steps that will be taken to ensure that the data, once collected, are not shared inappropriately. Social science evaluation often involves matters that may be personally sensitive. Mental health and substance abuse intervention studies, for example, may require the collection of information about participants' history with

mental health or substance use problems, the extent of their problems, as well as information about their families and spouses.

In most research, assuring confidentiality is only a matter of following some routine practices: substituting codes for identifiers, removing face sheets (containing such items as names and addresses) from survey instruments containing data, properly disposing of computer sheets and other papers, limiting access to identified data, impressing on the research staff the importance of confidentiality, and storing research records in locked cabinets. In some studies, more elaborate procedures may be needed—either to give subjects the confidence they need to participate and answer questions honestly, or to enable researchers to offer strong, truthful assurances of confidentiality. Such elaborate procedures may be particularly necessary for studies in which data are collected on sensitive matters such as sexual behavior or criminal activities.

It is important to maintain strict confidentiality at all times. Any information collected from or about participants must not be divulged to others without permission, and data must be safely and securely maintained and stored.

Tips for Maintaining Confidentiality

Here are some tips for maintaining confidentiality:

- Design protocols to minimize the need to collect and maintain identifiable information about participants. When possible, gather information anonymously. If you must be able to identify individual participants (e.g., for record-matching purposes), you can use codes or numbers rather than names on research records.

- Store data in a locked and safe location accessible only to the research team. Data files should be maintained in a manner that maintains participant privacy, as discussed above. The right to privacy also extends to the publication or presentation of information. Information should be analyzed and reported in a manner that maintains the privacy of all research participants.

- During the informed consent process, inform potential participants of the precautions you will take to protect the confidentiality of their information, including who will or may have access to it (e.g., you, your evaluator). This will allow participants to decide if protections are adequate.

- When you must tie study data to identifying information, do your best to maintain confidentiality. For example, use codes rather than participant names on study materials, or require anyone engaged in the collection, management, and/or analysis of data to sign a confidentiality agreement.

Data Analysis

Data analysis refers to the use of descriptive and inferential statistical techniques to transform raw data into information that is easier to manipulate, understand, and

report. Ultimately, the purpose of analysis is to understand what is happening with your program or what is changing as a result of those efforts.

To ensure that the conclusions of your study are sound, the analytical approach you choose must be carefully matched to the evaluation questions, the study design, and the data collected. In addition, the analysis should identify and address any potentially confounding variables that could affect the interpretation of the study findings. These two aspects of research quality—appropriateness of analysis and potentially confounding variables—are both among the six Quality of Research criteria used by NREPP.

The care and skill with which data analysis is conducted has a major impact on the quality of your overall evaluation effort. For this reason, if you (or others in your agency) do not possess strong expertise in data analysis, it is often worth hiring experts who can provide a high level of technical competence for this part of the evaluation.

Next, we consider some of the common analytical techniques used in data analysis.

Analytical Techniques

Always begin your analysis by reviewing the ultimate goal of your evaluation and the specific research questions you are trying to answer. To be useful to stakeholders and decisionmakers, data analysis should lead to straightforward, readily interpretable information about program processes and effectiveness.

Analytical techniques range from basic counts to content analysis to complicated inferential statistical analyses. After being collected, data typically are cleaned and entered into databases or spreadsheets. Next, statistical or qualitative analysis software is used to conduct various analyses. Depending on the type of data, you might conduct quantitative analysis, qualitative analysis, or both.

Quantitative Analysis

Quantitative analysis may include descriptive statistics and/or inferential statistics. Descriptive statistics provide a way to summarize large amounts of information using frequencies (counts), percents, ranks, measures of central tendency (e.g., mean, median, and mode), and measures of variability (e.g., range, standard deviation). Inferential statistics allow you to test for statistically significant differences. Your choice of statistical tests will be determined by the research questions you seek to answer and the nature of the data collected.

Qualitative Analysis

Qualitative analysis, on the other hand, is a process for reducing the amount of data as well as aggregating and organizing individual responses into themes that allow you to identify data trends. Qualitative data include information from interviews, focus groups, written documents, observations, and open-ended survey questions. For outcome and impact evaluation studies, quantitative analysis tends

to be the workhorse because it is the only way to show that your program has generated statistically significant results—in other words, that it is "effective"—and that your program is evidence based.

Confounding Variables

A critical part of data analysis is controlling for confounding variables that may affect the validity of the evaluation findings. Using an experimental design with a control condition and random assignment to the study conditions, as we discussed earlier in this course, is the gold standard method of controlling for confounding variables. Whether or not you are able to use a true experimental design with a control group and random assignment, it is important to consider and address the role of any confounding variables.

Common types of confounding variables are described below.

Selection Bias

If participants are not randomly assigned, they are by definition selected based on some other criteria. Those criteria then become an active variable in your study. This selection process is a source of bias in your study. In other words, it may introduce differences between the groups at pretest that may account for any differences you see at posttest. For example, students who volunteer to attend a tutoring session may be more academically motivated than students who do not volunteer to attend this group. Different levels of motivation may affect these two groups as much as, or more than, the tutoring did.

History

Any event, such as a disaster, school crisis, a major layoff in the community, or other programming that participants may seek out or be provided, that happens between measurement points in your study has the potential to affect outcomes. While most of these events cannot be controlled, they should be investigated and disclosed in your final evaluation report.

Maturation

The normal processes of development that occur over time could affect your outcomes, independent of your intervention. For example, as children grow and mature, they are likely to develop longer attention spans, making it easier for them to fully participate in your program. This development makes it difficult to separate program effects from normal developmental growth. This means that differences observed in a group of children between the time they enter kindergarten and the time they enter 1st grade may reflect maturation rather than the effects of program participation.

Testing Effects

The simple act of testing can itself influence outcomes. For example, if a study includes testing participants for drugs, the participants may be less likely to do

drugs simply because they know they will be subject to testing, even if they know the information will be kept confidential and will not be used against them. In addition, administering a test multiple times can affect results simply because the participants become familiar with it or even become bored or irritated with the repetition.

It can be difficult to determine if the effects that you see are caused by your program or the administration of a prior test. The testing effect tends to increase with the number of tests you administer.

Instrumentation

Any flaws in your instruments (or inconsistencies in how they are administered) can skew results. It is important to use instruments with well-documented psychometric properties. As part of your final evaluation report, it is important to present psychometric properties for each measure. This is another highly important aspect of research quality. Two of NREPP's six Quality of Research criteria, representing one third of the total QOR score for a given outcome, focus on the psychometric properties of the instruments.

Statistical Regression

Regression is the general tendency of very high and very low scores on any measure to move toward a more moderate score. This confound can adversely affect your study if you select participants based on extreme scores (e.g., if you decide to study only very depressed adolescents). When you implement your posttest measure, the extreme scores may have improved in part because of regression to the mean and not solely because of the intervention.

Hiring an External Evaluator

Do I Need an External Evaluator?

In the last section, we discussed the stages of an evaluation study and what kinds of decisions and activities are involved in each stage. At some point before beginning an evaluation, you will need to decide whether you (or your agency) have sufficient resources to carry out all of these activities without outside help. This decision should be made with great care, since it can greatly affect the success of your efforts and the quality of the research that is ultimately conducted.

You may find it feasible to conduct your own evaluation if you have a sufficient number of staff trained in relevant areas of evaluation (e.g., instrument design, survey administration, statistical analysis) and can afford to commit these staff to the project. Otherwise, it is appropriate to hire an external evaluator to conduct part or all of the planning, data collection, and data analysis.

Assess Your Readiness To Evaluate

To decide if it is in your or your agency's interest to hire an evaluator, try answering the following questions:

- Do you (or someone in your agency) know how to conduct process and outcome evaluations?

- Do you (or someone in your agency) know enough about evaluation to conduct one with the support of resources that are available in the field (e.g., manuals, toolkits, textbooks, guides)?

- Do you (or someone in your agency) have enough time to conduct a comprehensive program evaluation?

- If you have some expertise, do you have enough expertise to plan and implement an evaluation that meets your information needs or those of your funder?

If the answer to most of these questions is "no," you should consider hiring an external evaluator to plan and conduct your evaluation.

Advantages to Hiring an External Evaluator

Advantages to hiring an external evaluator include:

- **Specialized knowledge:** An external evaluator comes with specialized knowledge and ability. He or she understands how to assess the needs of a community, collect and analyze data, and document program outcomes.

- **Objective viewpoint:** An external evaluator remains unbiased about the program being studied. A good evaluator will point out both positive and negative aspects of a program, offering suggestions on how best to achieve the program's goals.

- **Higher level of credibility:** When you use a qualified external evaluator, the conclusions and recommendations of your evaluation study may carry more weight with funding institutions, the field, and the public. That said, a study is not inherently "better" simply because the evaluator is independent of the program being studied. Keep in mind that the same objective standards of quality, such as the NREPP Quality of Research criteria, apply to any study, regardless of who conducts it.

- **Reduced burden on your agency:** An external evaluator reduces the amount of time and internal staffing resources you need to commit to the evaluation project.

Challenges in External Evaluations

There are also some potential challenges involved in hiring someone else to do your evaluation. Someone who is not familiar with your program may need to spend some time getting acquainted with the logic model and operational details, which

can add significant cost. To minimize the startup factor, consider hiring an evaluator from your community or field who has an understanding of social services generally, and ideally some familiarity with the type of program being evaluated (e.g., theoretical model, populations, outcomes).

The single most important factor in using an external evaluator is finding a good match. Hiring an unqualified evaluator, or one who does not work well with you or your target population and program staff, can be detrimental, leading to such unintended consequences as alienating staff, intruding on the community (e.g., school, agency) in which your program is implemented, or drawing incorrect conclusions about the program.

Assessing Evaluator Credentials

To ensure a good match, take the time to examine and verify potential evaluators' credentials, following up with any references they provide. Look at their past performance conducting evaluations to make sure they have a strong technical understanding of evaluation research and knowledge of statistical methods. In addition, meet with candidates in person to ask questions that will give you an idea of their working style and how they will respond to challenges that typically arise in an evaluation. Also look for strong skills in communication, team building, group processes, and negotiation, which are vital to a successful relationship between the evaluator and your program staff and stakeholders.

If you do hire an external evaluator, be sure to clearly define the roles and responsibilities you expect your evaluator to handle. This can help you avoid later misunderstandings and communication breakdowns. We will come back to this issue in more detail in the next topic, Managing an Evaluation.

Ongoing Assessment

Even if you are confident that a candidate is a perfect match, remember that some issues may not emerge until the evaluation is under way. Continue to assess your evaluator's skills and performance throughout the evaluation. Ask program staff about their interactions with the evaluator. Find opportunities to observe the evaluator yourself, while on site or in meetings. If you find problems, address them with your evaluator immediately and help develop a plan for remedying them.

Finding a Qualified Evaluator

If you need an external evaluator, how and where should you look for one? Referrals from contacts in your professional network are one good way to find a qualified evaluator. In the absence of a referral from a trusted source, you can check with local colleges and universities to see if they have any faculty members or students who would be willing to work with you. Another option is to prepare a formal request for proposals to solicit bids from private consultants and/or research and evaluation firms who specialize in this type of work.

Potential Sources for Finding an External Evaluator

The table below outlines the pros and cons of three potential sources for a hired external evaluator.

Source	Pros	Cons
College or university	Educational institutions offer the expertise of faculty members or graduate students at low to moderate overhead costs. Many small, not-for-profit organizations have found low-cost data collection assistance through colleges or universities. With proper training, graduate students (the more inexpensive option) can create instruments and collect and analyze data. Contact college or university behavioral and social science departments (e.g., criminal justice, social work, public health, psychology) to learn if they have faculty or graduate students to assist you.	May be challenged to meet the variety of expertise and time commitments to complete an evaluation within a grant cycle.
Independent researcher	Researchers who work alone or with minimal staff can provide high-level expertise at low overhead costs. They commonly have the flexibility to be located on site, enabling frequent interaction with program participants.	May be focused on social service or other professional activities and not have sufficient time available to conduct a meaningful evaluation.
Research and evaluation organization	Organizations with research and evaluation departments offer a range of expertise. They are able to meet the multiple needs of an initiative by offering experts across disciplines and having state-of-the-art technology in house.	May be comparatively expensive because of the higher overhead costs associated with larger organizations.

Managing an Evaluation

Planning and Management

Planning and management can be as important to the success of your study as your research methodology. Whether or not you are using an external evaluator, having formal management processes in place makes it much easier to plan for, monitor, and regulate the many ongoing activities and costs involved.

During the planning stage of your evaluation, you will determine the evaluation requirements and develop the evaluation plan. As part of that planning work, it is also helpful to establish the budget, the timeline or schedule, and a monitoring plan. These tools lay out the expectations and provide benchmarks for keeping tabs on your staff's or your external evaluator's progress and performance. The details of your budget, timeline, and monitoring plan will vary depending on whether you are conducting the evaluation or overseeing an external evaluator.

Setting a Budget: Labor Costs

Your evaluation budget should take into account several categories of expenses. Labor is usually the greatest expense and may be higher when you use an external evaluator. As a rule of thumb, expect to allocate 20% of your overall budget to paying an external evaluator. Reputable evaluators may agree to reduce their rate if you allow them access to the data for their own purposes (e.g., further analyses) or to publish the results or some aspect of them.

Setting a Budget: Non-Labor Costs

In addition to labor, evaluation budgets also can include incentives or rewards for participants, if appropriate, and other indirect costs (e.g., postage, telephone, duplication). If you expect travel expenses for anyone involved in the evaluation, these also must be estimated. Be careful about building travel into your study, because extended travel expenses can add up quickly.

Finally, consider whether the evaluation will require any special equipment. For instance, will any standardized instruments (e.g., existing surveys and questionnaires) need to be purchased? Will tape recorders be used for data collection? Will any new data analysis or management tools need to be purchased?

Soliciting Cost Proposals

When developing your budget, it is a good idea to first solicit cost proposals from several evaluators. Cost proposals require that each bidder describe in detail the tasks to be completed, the number of hours each task will take, and the costs associated with each, including labor. You should receive several bids that are approximately the same. If you receive bids that are very different from the rest, you can assume bidders either under- or overestimated the cost of the work. Also, be sure that the cost proposals realistically address all potential costs, as your evaluation can be quickly derailed when necessary expenditures start to go over budget.

Funding Sources

As you set your budget, be certain you have the financial resources to cover the required costs before taking any steps toward implementation. If you are currently receiving a grant, you may be able (or required) to use up to 20% of your grant funds for evaluation. This amount may be sufficient or may need to be

supplemented. If your organization is within or affiliated with a social service agency, find out if the agency has any funds specifically set aside to cover evaluation activities. If there is a set-aside, be clear about the total amount of money available, the types of activities the set-aside is designed to fund (some are earmarked for specific categories of expenses, such as labor or data collection), and if others intend to use the money for their own evaluations.

Government, private, and some nonprofit agencies and foundations also offer funds specifically for evaluation activities. These grants are offered most often in the context of advancing knowledge in a particular area or to support the implementation and evaluation of evidence-based interventions. You may need to do some investigating to find a grant program focused on funding evaluations in your program area.

Timeline

The requirements of your evaluation most often determine its duration. For evaluations that are designed to meet funding requirements, the funder and the duration of the grant generally determine the evaluation timeline.

For example, if you are awarded a 3-year grant to implement a teen violence prevention program in your local school district, you might need to measure and report outcomes to the funder at the end of the 3 years. In the interim, the funder could require that you document certain components of the program, and/or measure short-term or intermediate outcomes and submit the findings in quarterly or semiannual reports. In this situation, the evaluation as funded is expected to terminate at the end of the grant period and will continue only if you choose to and have resources to support ongoing evaluation efforts.

Factors Affecting the Timeline

For evaluations that are not bound by funding requirements, it should be up to you and your evaluator to determine the timeline. If this is the case for your program, there are several issues to consider:

Outcomes

First, think about what outcomes the program or intervention is designed to achieve and when those outcomes realistically might be realized. For example, if you are implementing a substance abuse awareness program in a local high school, you might expect knowledge about the dangers of drug use to be evident in the participants immediately after they complete the program. Lower substance use rates in the high school population, however, could take longer to achieve; in fact, they might not be evident for years.

Determine which outcomes your evaluation is intended to achieve and when they will occur, and develop your timeline accordingly. If you conduct an impact evaluation (e.g., demonstrating long-term outcomes), extend the evaluation timeline outward to a point at which these outcomes can be realized. On the other

hand, if you conduct an outcome evaluation (e.g., demonstrating immediate, short-term outcomes), your evaluation timeline will be more closely related to the duration of your program or intervention. For example, if you implement a ten-week program or intervention, you should gather evaluation data right before (pretest) and right after (posttest) program implementation.

Participant Capacity

Unfortunately, most social services programs (or program sites) have limited participant capacity. That is, they can accommodate only a certain number of participants at any given time. They also tend to have high attrition rates. To maintain a sample large enough to power your study, you may need to consider implementing your program and collecting data multiple times, across a specified timeframe.

For example, you could run your 10-week program three times across an 8-month period. You could also implement the program in multiple sites simultaneously—for example, running your 10-week program in three separate program sites during the same 10-week period. In any case, always extend your evaluation timeline a few months past the end of your data collection period to allow ample time for data analysis and reporting.

Budgets

Budgets and timelines are best developed simultaneously. The money and other resources that are available for the evaluation will have implications for how long it can continue. You and your evaluator must know the extent and duration of available resources. You do not want to discover midway through your evaluation that funds have been exhausted. Be certain from the outset what funds are available and for how long so you can implement your evaluation completely and successfully.

Monitoring Plan

If you are working with an external evaluator, having a written monitoring plan gives you leverage and documentation to effectively manage the business relationship with your contractor. The contract itself usually serves this role. If your evaluation is internally conducted, the monitoring plan may be more perfunctory, since it is likely you already have working relationships and standards within your organization. Even so, a plan can be helpful in that it puts into writing any expectations that are specific to your evaluation.

Frequency and Purpose of Meetings

Since someone outside your organization may not be available as often as someone within your organization, it helps to stipulate when and how often meetings and other interactions will take place. Different types of meetings might be scheduled to discuss or work on different areas, such as the work plan, problems with implementation, or budgetary matters. The meeting schedule should strike a

balance between the amount of contact you need with the evaluator to feel informed about their work and the number of meetings your budget can support. If your evaluator is not local, you should also consider how often you will schedule in-person meetings.

Progress Reporting

It is a good idea to require some type of progress report from your evaluator (or your internal staff), regardless of how often you meet or have telephone or email contact. In the contract, you can specify the purpose, content, and timeline for these reports.

Data Reporting and Management

Decide when and how you will be involved in reviewing the data being collected. With an external evaluator, you may choose to have the evaluator report to you after each data collection (or submission) period, or you may want to be more personally involved in conducting data checks. Your choice may depend in part on the evaluator's knowledge of the data and any contextual issues affecting data quality.

Issues to consider include:

- Are surveys being completed in full? If not, what parts of the survey are missing?
- Are surveys being collected from all participants? If not, why not?
- What attrition rate can be considered reasonable?

In addition, you will want to decide how often you and/or your evaluator analyze and review process evaluation findings.

Invoicing

This is an aspect of monitoring exclusive to external evaluations. Determine how you will monitor the evaluation budget of your evaluator. Will you require monthly or quarterly invoicing? More frequent invoicing might allow you to monitor your budget more closely, but if your budget is small, less frequent invoicing may be acceptable. Also consider what details you want each invoice to contain. Some contracts require little detail about how the money was spent or expenses were incurred, while others specify that all items must be clearly identified and documented.

Contracts

Remember that in managing an external evaluation, your contract is your leverage. It must clearly state each requirement and how it will be monitored. Be sure your evaluator understands the contract requirements before they sign it. If you are managing a large evaluation with a substantial budget, you may want input on the contract from someone with a strong contracts management background. Contracts

are legally binding documents and can be challenged in a court of law; therefore, make sure your contract meets the standards for contracts of its kind.

Creating a Positive Relationship

You can use administrative procedures such as those suggested above to keep your external evaluator in check. However, nothing beats having a positive working relationship, which facilitates communication between you and your evaluator and, in turn, ensures that critical information is available to you in a timely manner. You can increase the likelihood of a positive relationship with your evaluator by hiring an individual or agency you already know or who has been referred by someone you know and trust. If this is not possible, do not hesitate to ask candidates for past evaluation reports and references.

Tips for Managing External Evaluations

Other than hiring someone who is qualified and implementing the management techniques already described, what else can you do to ensure the success of an externally conducted evaluation?

Here are some starting points.

Tip #1: Conduct frequent meetings

Many evaluations fall apart because the assumptions of the evaluator start to diverge from your assumptions, or because important information is not shared soon enough. As soon as the contract is signed, hold a kickoff meeting to clarify expectations, go over any aspects of the contract that might remain in question, and discuss the schedule for ongoing communication. Then begin holding regular meetings either by phone or in person. Daily communication can be done via email. Critical issues such as project expenditures and evaluation challenges are best addressed in person or by telephone.

Meetings should have specific topics and goals, whether related to planning and implementing the evaluation, discussing the progress of the evaluation, talking about problems and solutions, or going over data or early findings. During most meetings, you also should review project expenditures and plan (and confirm) next steps.

Tip #2: Closely monitor expenditures

Be vigilant about monitoring project expenditures. You should continually compare actual expenditures against the budget to make sure that there are no unplanned activities or monies being spent.

Whenever invoices are received, use them to compare actual expenditures to those that were planned for that timeframe. For example, if you agreed upon a 3-month invoicing schedule, you will need to have some reasonable expectation as to what evaluation activities are planned to take place during each 3-month time period. If

labor is billed for unplanned activities, contact your evaluator immediately and ask for an explanation. To manage unplanned expenses, you might require your evaluator to submit a written explanation of any unplanned expenses and the rationale for them with each invoice. That way, you will have documentation you can use to determine if the expense was necessary and reasonable, and can reference later, should a dispute arise.

Tip #3: Use a budget tracking system

There are many commercially available tools for tracking project costs, although some may be too expensive or unnecessarily complicated for your purposes. Start first by determining if your agency's financial tracking system might be leveraged for your purposes. Most organizations have a financial system that maintains records of expenditures. If they do, someone in the financial division may be able to set up a template for you to track evaluation costs, even if the agency is not financially supporting the evaluation.

You can also consult colleagues or other researchers with whom you are familiar who have hired external evaluators and ask them what methods they use for tracking project expenditures. If they have a system already in place, ask them to share it with you. Remember, this system doesn't have to be complicated or sophisticated. It just needs to allow you to manage your budget and project costs.

Tip #4: Set controls to limit unplanned expenditures

Unplanned expenses occur. The key is to set up controls so there are not so many that they derail your budget. Make it clear to the evaluator that the scope of the evaluation cannot be changed without discussion and written approval from you. For example, if an invoice indicates that the evaluator added a program site or conducted focus groups that were not part of the original evaluation plan, you must address this immediately. Changing the scope of the evaluation midcourse not only quickly expends already limited resources, but can jeopardize the validity of your findings.

Tip #5: Ensure costs are in line with your funding source

If you are using external funding to support your evaluation, you have the responsibility to comply with all applicable grant requirements and will be obligated to demonstrate sound management of all expenditures and actions affecting the grant.

Grant-making organizations and agencies generally have guidelines for documenting expenditures, including the nature and extent to which they must be documented. In some cases, you may be asked to provide detailed information in specific cost categories. In others, you may only need to indicate how much money has been spent, in total, in a specified period. Most grants stipulate how grants funds may and may not be spent (i.e., allowable and nonallowable costs).

It is your responsibility to see that grant funds are used to support only those costs that are reasonable, necessary, and allowable; consistent with grant terms and conditions; and represent effective utilization of resources. Be sure you are well versed in the terms and conditions of your grant funds. If your expenditures are not consistent with these requirements, you could lose future support or even be required to repay the money you have spent.

Evaluation Reporting

Purpose of Evaluation Reporting

The evaluator's final task, once the study is complete, is to communicate the study's findings to the appropriate audiences. If evaluation findings are to have any real impact, they need to be disseminated to stakeholders such as funders, program staff, and agency administrators. The findings also must be made relevant and accessible to the unique perspectives and interests of each of these varied groups.

The Final Evaluation Report

Conventionally, evaluators disseminate their findings in a final evaluation report. The purpose of the report is to synthesize everything learned from the evaluation of your initiative. It should comprehensively discuss each of the major components of the evaluation, from planning to implementation to completion. Of course, it should also clearly present the key findings or conclusions of the evaluation. Whenever possible, it is helpful to draw connections between the data and possible areas for program improvement or refinement. For example, the report might make recommendations for how to implement the program with fidelity based on observations made during the evaluation, or explain how differential findings among different subsets of the study population might provide insight for how to tailor the program.

Tailoring the Report for Different Audiences

In choosing what information to focus on, consider the audiences you intend to reach. For example, a report that is primarily intended for the broader social services program and evaluation communities should focus on implementation and outcomes—two areas most likely to be of interest to these more generalized audiences. If a report is intended for your funder, you may need to include financial information that you would not otherwise include in a report intended for dissemination to a broader audience. If a report will be disseminated to a research audience, it should provide sufficient detail that researchers will understand the technical aspects of your evaluation.

If you plan to draft a manuscript that will be submitted to a journal, keep in mind that you may need to adjust the contents and presentation style of your report to meet the journal's requirements. For example, most journals require an abstract to be developed and may have specific requirements for the contents of the report.

Components of an Evaluation Report

A comprehensive evaluation report typically includes the following components:

- **Executive summary:** An executive summary provides a high-level overview of the report. In one or two pages, it provides a concise description of the program, a brief statement about the evaluation goals and methodology, a summary of the most important findings of the evaluation, and any implications or recommendations. An abstract usually replaces the executive summary in reports intended for journal publication.

- **Introduction:** The introduction acquaints the reader with the context and background of the research being described in the report. It should summarize the program and explain the problem it seeks to remedy or treat. In addition, this section should briefly review the existing literature to explain what previous methods to address the problem have found.

- **Program description:** This section gives a complete, detailed description of the program, including theoretical or conceptual origins, core components, a definition of the target population and intended outcomes, and basic implementation requirements.

- **Evaluation focus:** Here the evaluator explains what aspects of the program are being examined and why. Topics covered here include the rationale for the evaluation, activities and strategies followed in implementing the evaluation, and the questions put forth.

- **Procedures:** The procedures (or methods) section describes the methodology of the evaluation in sufficient detail that other investigators will be able to (a) assess the rigor and appropriateness of your methods and (b) carry out a replication if desired. Topics covered in this section include methods for sample size and recruitment, data collection, and data analysis.

  - Sample size and recruitment methods: This section explains how the study sample was selected and recruited and often provides demographic data on study participants. Information should be provided on the number of individuals who were identified as potential participants and the number who elected to participate, as well as the retention rate (how many participants remained through the end of the study).

  - Data collection: This section describes the measures or instruments used in the evaluation (which should be included as appendixes) and provides psychometrics on the reliability and validity of each measure. It should explain how each measure was administered (by whom, how, and when). Also important to include is a discussion of missing data and attrition and any statistical methods or other approaches used to address these issues, along with a justification for these methods, if applicable.

  - Data analysis: This section explains the analytic techniques used in the study. Covered here are the methods used to explore, identify, and address potential confounding variables; documentation of adequate sample size and statistical power; a discussion of the fit between

statistical techniques used and complexity of the data; and documentation that analyses were sufficiently sophisticated for the nature of the data.

- **Results:** The evaluation findings are presented and interpreted in the results section. Interpretation of the findings should reflect a careful process of synthesis and analysis that supports judgments and subsequent recommendations. It is helpful to clearly relate the findings to specific evaluation questions. If necessary to manage report length, and if appropriate for the intended audience, more extensive details related to the findings can be included in supporting appendixes. Outcomes should be presented with accompanying significant levels (p values) and effect sizes to demonstrate the size and statistical significance of the findings.

- **Conclusions and recommendations:** The conclusion of the report discusses the implications of the evaluation findings for your program as well as any program modifications or improvements that are recommended based on these findings. It is often effective to present your conclusions in terms of both the strengths and limitations of the program. This section might also offer some ideas for future research.

Dissemination of Findings

A written evaluation report may be the conventional method of disseminating findings, but it need not be your sole strategy. Try using use other techniques to reach the broadest possible audience.

Oral presentations can be effective in reaching certain decisionmakers and community groups. Depending on what kind of exposure you are trying to achieve, you can try venues that are large (e.g., national conferences and meetings of professional associations) or small (e.g., meetings of local community groups, boards of directors). An interactive presentation with visual renderings of your evaluation results or contextual information (such as graphics and photos) can go far to spark interest. You can offer to answer questions after the presentation and provide handouts for further reading and follow up.

Other venues include Web sites, press releases, newsletters, email blasts, question-and-answer statements, op-ed articles in newspapers, articles in professional journals, written testimony to local or State legislative bodies, or even brown-bag lunches. Video-based presentations shared via CD-ROM or online (e.g., YouTube) and social media are also increasingly being used as communication tools.

Summary

Brief Summary of Concepts

Having completing this course, you should now have a solid grasp of the basic concepts of evaluation. Hopefully, the information we have presented has also given you a good understanding of why evaluation is important and what a high-

quality evaluation can do to document, promote, and enhance your program's impact.

Let's review a brief summary of the concepts we have covered.

- A carefully planned evaluation can yield a range of useful information, from process data (relating to how your program operates) to outcome and impact data (describing effects on participants or systems). We conduct evaluations for many reasons, sometimes to achieve multiple ends at once. For example, an evaluation can justify past funding while showing where the program might be improved or what specific components account for its success. Evaluations also contribute to the field by adding to the body of literature on evidence-based programs.

- The "ideal" (most rigorous) evaluation has an experimental design, using a control group and randomization to study conditions. Because of unique factors affecting research in the behavioral health field, evaluators sometimes must compromise and use a quasi-experimental or preexperimental design. Try to implement the most rigorous design you can given your constraints.

- Attending to the quality of your research is exceptionally important if you want others to regard your conclusions as "evidence based" or scientifically sound. NREPP's rating criteria can be a helpful reference for understanding how research quality is measured.

- Planning, data collection, and data analysis are all important to the success of your evaluation. If you do not have the technical qualifications to handle all of the work involved, seek help from an external evaluator.

- If you do hire an external evaluator, be sure to use sound management practices (planning, budgeting, and monitoring) to keep the project on course.

- To get the most out of your evaluation, take the time to develop a well-written evaluation report, but also consider using other, creative ways to disseminate your findings to your target audiences.

Conclusion

Conducting a high-quality evaluation may sound like a tall order, but it is an effort that pays big dividends. The key to success is to start with a realistic idea of the planning and work involved, because when it comes to evaluations, to quote the old saying, "You get out what you put in."

Wherever you are in the process of thinking about or preparing for a study, we hope that this course has encouraged you to take the next step toward your ultimate evaluation goals.

Appendix A – Further Reading

General Resources

Chronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass.

Clancy, C. M., & Eisenberg, J. M. (1998). Outcomes research: Measuring the end results of health care. *Science*, *282*, 245-246.

Feuerstein, M. T. (1986). *Partners in evaluation: Evaluating development and community programs with participants*. London: MacMillan, in association with Teaching Aids at Low Cost.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. New York, NY: Pearson Education.

Grembowski, D. (2001). *The practice of health program evaluation*. Newbury Park, CA: Sage.

Guba, E., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Beverly Hills, CA: Sage.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York, NY: Holt, Rinehart, and Winston.

Morbidity and Mortality Weekly Report. (1999). Framework for program evaluation in public health*. MMWR 48*(RR11), 1-40. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/rr4811a1.htm

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Rogers, P. J., Hacsi, T. A., Petrosino, A., & Huebner, T. A. (Eds.). (2000). *Program theory in evaluation: challenges and opportunities: New directions in evaluation*, Number 87. San Francisco, CA: Jossey-Bass.

Rosnow, R. L., & Rosenthal, R. (2008). *Beginning behavioral research. A conceptual primer* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Rossi, P., Lipsey, M., & Freeman, H. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Shadish, W. (1998). Some evaluation questions. *Practical assessment, research & evaluation*, *6*(3). Available at http://PAREonline.net/getvn.asp?v=6&n=3

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation*. Thousand Oaks, CA: Sage.

W. K. Kellogg Foundation. (1998). *Evaluation handbook*. Battle Creek, MI: Author. Available at http://www.wkkf.org/knowledge-center/resources/2010/W-K-Kellogg-Foundation-Evaluation-Handbook.aspx-Kellogg-Foundation-Evaluation-Handbook.aspx

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation*. New York, NY: Longman.

Research Methods and Designs

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Houghton-Mifflin.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Houghton-Mifflin.

Cozby, P. C. (2001). *Methods in behavioral research* (10th ed.). New York, NY: McGraw Hill.

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed). Boston, MA: Pearson Education.

Dunn, D. (2009). *Research methods in social psychology*. Malden, MA: Wiley-Blackwell.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Wadsworth Publishing.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole.

Leedy, P. D., & Ormrod, J. E. (2010). *Practical research: Planning and design* (9th ed.). Boston, MA: Pearson Education.

Monette, D. R., Sullivan, T. J., & DeJong, C. R. (2011). *Applied social research: A tool for the human services* (8th ed.). Belmont, CA: Brooks/Cole.

Patten, M. L. (2012). *Understanding research methods* (8th ed.). Glendale, CA: Pyrczak Publishing.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.

U.S. Department of Education. (2011). *What Works Clearinghouse procedures and standards handbook* (Version 2.1). Available at http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

Missing Data and Attrition

Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance, drop-outs, and how to deal with missing data in clinical research: A review. *Physical Therapy Reviews*, *14*(1), 36-49.

Barry, A. E. (2005). How attrition impacts the internal and external validity of longitudinal research. *Journal of School Health*, *75*(7), 267-270.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549-576.

Harris, P. M. (1998). Attrition revisited. *American Journal of Evaluation*, *19*(3), 293-305.

Mason, M. J. (1999). A review of procedural and statistical methods for handling attrition and missing data in clinical research. *Measurement and Evaluation Counseling and Development*, *32*(2), 111-118.

Reliability and Validity

Bannigan, K., & Watson, R. (2009). Reliability and validity in a nutshell. *Journal of Clinical Nursing*, *18*(3), 3237-3243.

Frank-Stromborg, M., & Olsen, S. J. (2004). *Instruments for clinical health care research* (3rd ed.). Boston, MA: Jones & Barlett.

McDowell, I., & Newell, C. (1996). The quality of a measurement: Validity and reliability. In *Measuring health: A guide to rating scales and questionnaires* (2nd ed.) (pp. 29-43). New York, NY: Oxford University Press.

Rush, A. J., First, M. B., & Blacker, D. (2008). *Handbook of psychiatric measures* (2nd ed.). Washington, DC: American Psychiatric Association.

Williams, S. T. (2008). *Mental health screening and assessment tools for children: Literature review*. Report prepared for the Northern California Training Academy and funded by the California Department of Social Services. Available at http://humanservices.ucdavis.edu/Academy/pdf/FINAL2MentalHealthLitReview.pdf

Evaluation Reporting

Torres, R., Preskill, H. S., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting*. Thousand Oaks, CA: Sage.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.

SAMHSA's National Registry of Evidence-based Programs
and Practices (NREPP)

www.nrepp.samhsa.gov