8

CONDUCTING RESEARCH THAT EVALUATES SERVICES

Carla provided individual therapy to eight children who had experienced posttraumatic stress disorder (PTSD) due to being the victims of trauma-like abuse or witnessing violence between parents. She employed trauma-focused cognitive–behavioral therapy (TF-CBT) because it had been found to be an evidence-based treatment for PTSD, in both children and adults. The model of TF-CBT is founded on the assumptions of cognitive–behavioral therapy (CBT), which emphasizes the examination of thoughts about life events to change the emotional response to the events. Trauma-focused cognitive-behavioral therapy (TF-CBT) is one form of CBT. This option was employed by Carla with her clients because they had all experienced a trauma. Among the components of the treatment used by Carla were psychoeducation, relaxation techniques, cognitive coping, and in vivo gradual exposure. The structure of this treatment was 8-hourly therapy sessions completed one per week for 8 weeks. Carla's credentials for being qualified for this job was the LCSW (Licensed Clinical Social Worker). She has been trained in the use of TF-CBT and has used it for a good number of clients in the past and believes that it is effective. The goal of this treatment for these eight clients was to achieve a life without the consequences of the traumas they had experienced. The objective was to reduce the symptoms of PTSD.

Carla measured client progress with the Child PTSD Symptom Scale. On this scale, the clients examined the symptoms of PTSD and indicated how much they were experiencing each symptom recently. Higher scores indicated a greater problem with PTSD. Scores on this scale can range from a low of 17 to a high of 51, with scores greater than 25 indicating a need for treatment and

scores less than 17 indicating no need for treatment. This scale was administered to each client once before the treatment began and again at the end of the 8-week treatment period. Because Carla knew that these clients had experienced the symptoms of PTSD for many months, she believed that a comparison group research design was not needed. The one-group pretest–posttest design employed in her study was considered adequate because these clients were not expected to overcome the PTSD symptoms without treatment.

The mean pretest score for these eight clients on the Child PTSD Symptom Scale was 29.4, while the mean posttest score was 16.7. The differences between these sets of scores was found to be statistically significant. Carla concluded that her treatment had been effective.

INTRODUCTION

In this chapter, you will review the evaluation of social work interventions. There is a distinction you will see between a human service **program** and a human service intervention. The program is the broader thing to be measured, and the **intervention** is the more specific aspect that is the focus of evaluation. You can conduct a program evaluation of various aspects of the Child Welfare Program in your agency or the Adult Mental Health Program or you can evaluate a single intervention, such as the support group for abused women. Sometime interventions will be referred to as **services**. A broad program will normally have various services (interventions). This chapter reviews the evaluation of interventions, something that social work students and practitioners are more likely to be conducting. But first you will examine a few things about programs and interventions.

This chapter starts with an overview of the human service system, a concept most useful for the evaluation of a broad program. From this overview, you can put into perspective the evaluation of client outcome, the focus of this chapter. Then, you will review various aspects of how to evaluate an intervention, such as tutoring for 14 at-risk middle school students, therapy for eight women who have been abused, or parent training for 14 teenage mothers.

At the completion of this chapter, you will be able to do the following:

1. Describe the parts of the evaluation system with regard to input, process, output, and outcome

2. Identify the five major phases of evaluation research

3. Define and analyze the target behavior that is being treated

4. Review evidence with regard to interventions and target behaviors

5. Describe the intervention with regard to goal, objective, structure, model, and personnel

6. Select a research design that adequately addresses the critical threats to internal validity that should be of concern in a given situation

7. Select a tool for measuring client progress with appropriate attention to reliability and validity

8. Determine if the results of an evaluative study can be generalized from the study sample to a designated study population on a scientific basis and a logical basis

9. Compose an evaluative study hypothesis with attention to critical tips for doing so

10. Determine if a given set of data supported a given evaluative study hypothesis with attention to the two criteria for this endeavor

11. Determine if statistical significance or practical significance has been achieved for a given set of data for a given study

12. Prepare the study conclusions for a given evaluative research study

THE PARTS OF THE EVALUATION SYSTEM

As a reader of this book, you will likely be interested in knowing how successful a given human service has been. The application of the scientific analysis of data regarding client outcome would be expected to be among the questions of special interest. Has participation in your tutoring program had the effect of improving school grades? Have the homeless people who sought the help of your agency been more likely to find a home within 6 months than homeless people nationally? Are your clients less depressed after receiving your therapy than before receiving your therapy?

The issues for intervention research typically focus on one aspect of evaluative research—client outcome. The issue is whether clients are better off as a result of your intervention. There are other issues that are within the purview of program evaluative research. Among the questions that you might pose in program evaluation are the following:

1. Are our services reaching the target population?

2. Are we employing good standards of service?

3. Are we delivering services efficiently?

4. Are our clients satisfied?

5. Have our interventions improved client conditions?

All these are questions related to agency success in one way or another. Funding sources often ask for data on one or more of the above questions, with increasing emphasis on client outcome. If your tutoring program is supposed to be servicing at-risk middle school students, you need to demonstrate this through your descriptive statistics on the clients being served. This reveals the answer to Question 1 mentioned above. You

might be called on to report the credentials of your service personnel to show that your services are being delivered by qualified people. This is one aspect of the standards of service that you will employ. Efficient delivery of service is measured by the amount of money your agency spends for each **unit of service**, such as a cost of $110 per hour of therapy. Client satisfaction is normally measured by simple surveys of clients. Data on this may also be expected by your funding source. And, of course, data on client outcome respond to the last question noted above. This will be the emphasis in this chapter. You may have noticed the two aspects of this question: (1) clients should have improved and (2) their improvement was caused by the service. You will address both these questions in this chapter.

All the above are ways by which success is measured in program evaluations. We are most often going to review success with regard to client outcome. But funding sources also want to know about the quality of services and the efficiency of service delivery. In this section of the chapter, you will review the elements of the service system as well as the depiction of programs and services.

Evaluating Input, Process, Output, and Outcome

One of the models for a service system shows some elements that are labeled **input** because they are energies that come into the system, and some are called **process** (i.e., services) because they refer to what happens to the inputs. **Output** is a term used to refer to the amount of service (output) provided, while **outcome** refers to the benefits that are derived by the processes that were employed. A client with need could be considered an input, while the nature of the service offered could be considered a process. Output would be the amount of service offered, while outcome would refer to the benefit to the client with regard to the need being addressed by the service. For example, a middle school student with poor grades (input) might receive tutoring service (process) in the form of 12 tutoring sessions (output) with the result that her or his grades were improved from an average of 63 to an average of 78 (outcome). Figure 8.1 shows a depiction of the service system.

The assessment of need typically takes the form of a descriptive type of study because the purpose is to describe need. To what extent, for example, is there a waiting list of clients who have asked for your service but have not yet been served? To what extent does your target population have people who qualify for your service because they have the characteristics of youth at-risk for delinquent behavior? These are descriptive inquiries into need.

The review of service characteristics is also an example of a descriptive study because the purpose is to describe the service. What are the credentials of the service providers? What is the average caseload of service staff? Have all clients been given the statement of the nature of the service and their rights as a client? In other words, you are examining the extent to which services have been provided according to plan. There may be a set of

---

FIGURE 8.1 ■ Service System

INPUT [Need] → PROCESS [Quality] → OUTPUT [Efficiency] → OUTCOME [Effectiveness]
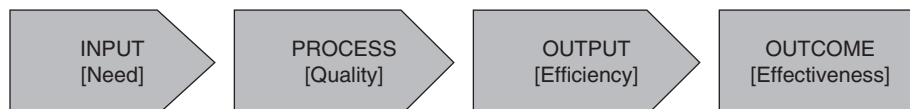
protocols that clarify the plan. To what extent were these protocols employed with the clients in the past 6 months? Sometimes funding sources provide audits of services to determine the extent to which service policies were executed.

The review of output is another example of a descriptive study that can be viewed as helping with the evaluation of service. This requires the quantification of services. The unit of service is the vehicle for this inquiry. There are a number of types of units of service that might be reviewed. The simplest one is a client served. If you served 368 clients in the past year, you have delivered 368 units of service if you have defined the unit of service in this way. If you are providing therapy sessions to clients, a therapy session provided to a single client could be your definition of a unit of service. If you are providing residential treatment, a day of service may be the unit of service.

You can use the units of service as one element of a review of efficiency. The other element is the cost. If you have provided service to 368 clients and your agency spent a total of $368,000 to provide this service, you can easily see that it cost $1,000 per client served. This is a measure of efficiency. If agencies of your type provide service at a cost of only $500 per client, you could be criticized for not being efficient. The measure of efficiency, of course, is not an indicator of outcome, so you might be achieving greater outcome even though you are operating less efficiently.

An evaluation of outcome, of course, requires a measure of benefit to the client. The key for this review is the outcome objective you have established for the service being offered. This could include improved grades in school, reduced anxiety, improved self-esteem, improved marital harmony, reduced health risks, and so forth.

In Practice Exercise A at the end of this chapter, you are asked to reflect on the various parts of the service system. You are called on to select one for a report. You may wish to review those questions now to get a better picture of the service system.

THE PHASES OF OUTCOME EVALUATION RESEARCH

In previous chapters, you have reviewed four major phases of research: (1) research purpose and knowledge base, (2) study methods, (3) data analysis, and (4) study conclusions. Evaluation research will be presented a little differently because of the critical role of the service that is being evaluated. Evaluative research begins with an examination of the nature of the **target behavior** that is being served. This is comparable with the articulation of the purpose and knowledge base. The second part is new. It is the presentation of information about the nature of the service that is being evaluated. If you prefer, you could consider this to be a part of the first phase of research as articulated previously, rather than a separate phase of research. The remaining phases of evaluative research are similar to that which has been presented—study methods, data analysis, and study conclusion.

Here is a very brief and condensed report of how this process may be viewed:

1. *Target behavior:* The target behavior being evaluated is chronic depression among persons who are within 1 year of experiencing divorce. Depression is a condition where individuals lose interest in life, are chronically feeling sad, and have difficulty in fully performing life activities such as dealing with work,

friends, and family. It is viewed as being partly caused by distorted thinking about life events. Divorced persons who become chronically depressed are different from divorced persons who become temporarily depressed. Chronic depression is depression that has existed for at least 6 months.

2. *Treatment:* The **goal** of **treatment** is to improve the client's adjustment to life after divorce, including work, relationships, and family life. The objective is to reduce depression. The model of treatment is CBT. The rationale for this choice is that this model focuses on distorted thinking, believed to be a cause for chronic depression. This treatment was delivered in hourly therapy sessions once weekly for 8 weeks. A credentialed therapist, who was a licensed clinical social worker, provided the therapy.

3. *Study methods:* The study sample consisted of 17 divorced persons recently treated at the Hampton County Family Service Association. The Beck Depression Inventory (BDI) was used to measure each client's depression once before the therapy began and once again at the end of the treatment period. Higher scores on this scale represent higher levels of depression. Scores on this scale that are greater than 30 indicate severe depression, whereas scores from 21 to 30 indicate moderate depression. This research design (one-group pretest–posttest) does not control for normal growth and development over time as the cause of the measured client growth as an alternative to the treatment as the cause. However, this explanation was not considered to be of special concern given the facts that (a) the depression of these clients had existed for more than 6 months without change and (b) the treatment period was only for 8 weeks. Therefore, significant change between the beginning and the end of the treatment period should logically be concluded as being the result of the treatment rather than the result of normal growth and development over time.

4. *Data analysis:* The pretest mean score on the BDI was 32.6, while the mean posttest score (at the end of treatment) was 22.4. This indicates that the average person went from a threshold for depression considered to be severe to one that is considered to be moderate. This indicates, in the opinion of the researchers, that practical (clinical) significance was achieved. These scores were analyzed statistically with the paired t test and found to be statistically significant (t = 2.67; p < .01). Thus, statistical significance was achieved. Because the data went in the hypothesized direction and were statistically significant, the data supported the study hypothesis (posttest scores on the depression scale will be less than the pretest scores).

5. *Study conclusions:* It was concluded that the recently divorced clients of this program of CBT achieved a reduction in depression that was of both statistical significance and practical significance. These findings cannot be scientifically generalized to the population of all divorced persons who are chronically depressed because this study did not employ a random sample of that population. These findings should give encouragement to therapists to employ CBT to chronically depressed persons who have recently been divorced.

In a full report of the research study, each of the aforementioned paragraphs would be elaborated on. This is just a synopsis to show the process of evaluative research from one phase to another.

You will notice that the target behavior section identified the behavior being treated and presented an analysis of its causes. This analysis is related to the choice of the intervention. The condition of depression was defined, so that methods for measuring client progress could be facilitated. The treatment was described sufficiently to be recognized so that the results could be interpreted with regard to what it was that was (or was not) found to be successful. The measure of client progress was identified as well as the research design. These contributions provide information to the reader regarding the credibility of measurement and the adequacy of the research design. Data were given on practical significance—the mean depression score of these clients changed from one threshold of depression to a lower one. Data were also given on statistical significance—the p value ($p < .01$) showed that these data cannot logically be explained by chance. The study conclusions were directly in line with the data results.

Phase 1: Target Behavior Examination

There are three major tasks in this phase of evaluative research: (1) to define the target behavior being treated sufficiently to guide the selection of a tool for measuring client progress, (2) to provide a logical justification for the selection of the service to be provided to the clients, and (3) to examine the evidence about what works in the treatment of this target behavior. This means that you define the behavior, you analyze it, and you seek evidence regarding the treatment of it. The later two are designed to provide guidance on the use of a service to achieve the outcome for the client. If you have evidence that CBT is effective in the treatment of depression, you have a basis for justifying the use of this model for your treatment of persons with depression. If you can link the causes of the behavior, or the special needs of those with this condition, with the nature of your intervention, you have a logical rationale for using your intervention for this target behavior. Treatments in the category of pseudoscience, discussed in Chapter 1, fail to be a source of justification for a treatment because it is not based on systematically collected scientific evidence of the effects of accompanying treatments for given target behaviors.

Target Behavior Definition

If you define stress as a certain type of psychological condition characterized in certain ways, you should find a tool that measures this psychological condition. You should not select a tool that measures life events that lead to stress, because you have not defined stress with regard to life events. Instead, you should measure stress with regard to the psychological condition you have defined.

Here are some examples of **target behavior definition**:

1. *Posttraumatic stress disorder*, according to the *Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (*DSM-5*), evolves after an individual has experienced one or more traumatic events (American Psychiatric Association, 2013). The *DSM-5* defines posttraumatic symptoms within three specific dimensions: (1) reexperiencing, (2) avoidance, and (3) arousal. Reexperiencing

includes behaviors such as recurring memories, nightmares, dissociative episodes, and extreme physiological reactions to reminders of the traumatic event. Avoidance may include avoidance of memories, reminders, and thoughts. It can also include feelings related with the traumatic event. Arousal may be exhibited by behavioral disturbances, angry outbursts, hypervigilance, difficulty concentrating, and sleep problems (American Psychiatric Association, 2013).

2. *Obesity* is defined as a body mass index of 30 or greater and is calculated as weight in kilograms divided by the square of the height in meters.

3. *Adjustment disorder* is defined as a group of troublesome symptoms—such as stress, feeling sad or hopeless, and problematic physical symptoms—that can occur after a person has gone through a stressful life event.

4. *Classroom misbehaviors* refer to behaviors in the classroom that exhibit aggression, defiance of authority, or disruption of the learning of others.

5. *Anxiety* is excessive nervousness, worry, or fear, which may include feelings that things are unreal, and/or a sudden panic state. It often includes apprehension; feeling tense, stressed, or on edge; difficulty concentrating; racing thoughts; and unrealistic fears.

In each of the above definitions, there is guidance on how to select a tool for measuring client progress. When you examine alternatives for the measurement of your target behavior, you should review the definition to see that all aspects of the behavior that are included in your definition are included in your measurement tool.

Target Behavior Analysis

The purpose of **target behavior analysis** is to obtain guidance on the determination of the appropriate intervention for the clients. You can analyze target behavior with regard to causation, special needs, or both. If ignorance is the cause of the target behavior, training would be a logical intervention. If involvement with dysfunctional peers is the cause of drug abuse among youth, an intervention that changes the peer group for the client would make sense. In many situations, the analysis of special needs is better than the analysis of causation. For example, a mental disability may be due to a situation from birth, which cannot be changed. But clients with this condition have special needs. Your service should address these needs.

If your analysis of drug abuse for youth indicates that youthful drug offenders have good knowledge of the consequences of drug use, it would not make sense to provide training on the consequences of drug abuse as your service. That would not be logical. They understand the thing you have decided to train them on. Perhaps your service should address the causes of drug abuse in ways that might make a difference, such as intervening into the social life of the abuser, assuming that you have found out that many youthful drug abusers got into drug abuse because they stared socializing with others who were drug abusers. If you can get them to change their social group, you can have more hope that they will stop their use of drugs.

If the cause of poor parenting for an at-risk group is the lack of good parenting role models in the past, it would make sense for the treatment to include mentoring by parents who can display good parenting. If an outcome of a trauma is a feeling of being the only person who has experienced this trauma, leading to self-blame, it may be that a support group of people with similar experiences would help.

The nature of the problem is sometimes complex with regard to causation or need. Let's examine the theme of stressors, stress, and stress buffers as three concepts to be measured in your study. The theoretical model depicting causation indicated that stressors cause stress to get worse, while social support causes stress to get better. Thus, one intervention could be to reduce stressors, while another could be to increase social support.

Here are four examples of concise statements of target behavior analysis that have appeared in the research reports of social work students:

1. Children who have been exposed to sexual trauma caused by their caregivers are generally unable to reach normal developmental goals and have a high chance of falling behind in their emotional, social, and cognitive advancement. This inability to progress normally is related to dysfunctional thoughts about the trauma that was experienced. The needed treatment, therefore, is a service that combines opportunity to review the traumatic events in light of new ways of thinking about it and provides a supportive environment for the victim.

2. Individuals affected by severe obesity who are seeking bariatric surgery are more likely to suffer from depression or anxiety and to have lower self-esteem and overall quality of life than someone who is normal weight. Bariatric surgery results in highly significant improvement in psychosocial well-being for the majority of patients. However, there remain a few patients with undiagnosed preexisting psychological disorders and still others with overwhelming life stressors who have psychological difficulties. One of the problems for such persons is a feeling of being alone in the world; thus, the improvement of social connections is an important need. A social support group experience is one vehicle for the achievement of this outcome.

3. Drug addiction leads many to feel that there is no hope for recovery, and they must be addicted from now on. This condition is partly due to a lack of knowledge about the success of treatment options. What is needed, therefore, is education on these options and data about their success.

4. Persons who experience adjustment disorder have reacted to life stressors in ways that are more severe than normal, leading to a great deal of feelings of stress. What is needed, therefore, is an intervention that (a) removes stressors from the client's life and (b) improves social support, which reduces stress for those experiencing life stressors.

These concise statements, of course, were supported by a review of the literature and further information that supported the concise statement. They are presented here as examples that illustrate what is meant by target behavior analysis.

Review of Evidence

What works with your target behavior? That is the key question for the examination of **evidence**. As you encounter this task, let's see a definition of this concept. Evidence can be defined broadly as any information that is employed in determining action. However, we are examining evidence regarding outcome evaluation. In this regard, evidence is information collected from sources of a scientific nature that is viewed as credible by key actors in the human service system. The typical example is a research study of the effectiveness of tutoring in the improvement of the grades of at-risk middle school youth. This means that the intervention of tutoring was employed and that grade improvement was measured.

As you saw in the chapter on evidence-based practice, sources of evidence are not equal. Some studies measure outcome once before treatment and once again at the end of treatment, while a more complicated study would compare the gain of clients with the same for a comparison group of people who did not get the intervention. An even more sophisticated study would employ the traditional experimental design whereby a group of people are randomly assigned to be in a treatment group or a control group, with only the treatment group getting the intervention. The difference in measured growth between the two groups indicates the effect of treatment on the target behavior.

As you saw in the chapter on evidence-based practice, the review of a single study on evidence is a weak review. Several levels of evidence were presented in that chapter, including (a) the single article, (b) the traditional literature review, (c) the **meta-analysis**, and (d) the **systematic review**. These levels of evidence were presented in order of their level of scientific sophistication.

You are advised to start your review of evidence by finding systematic reviews. In some cases, the summary of that review is all that you will need. If you fail to find a systematic review, look for the meta-analysis. Failure to find a meta-analysis should lead you to examine traditional literature reviews of evidence for your target behavior and intervention. If there are no traditional literature reviews of evidence, you should find several individual studies. You should not stop your review because you found a study that gave you the results you were seeking. Remember the spirit of scientific inquiry! You should engage in a review of your own that is systematic in its procedures for finding evidence.

There are two key data sources for systematic reviews of human service research. One is the Campbell Collaboration (www.campbellcollaboration.org). It provides systematic reviews of various target behaviors and the treatments that have been studied.

Another source is the Cochrane Collaboration (www.cochrane.org). In one example, information on the treatment of PTSD was reviewed from the Cochrane Collaboration in June 2016. Some of the treatments were TF-CBT, EMDR (eye movement desensitization and reprogramming), and other treatments that were not trauma focused. Here are the author's conclusions:

> The evidence for each of the comparisons in this review was assessed as very low quality. This evidence showed that individual TFCBT and EMDR did better than waitlist/usual care in reducing clinician-assessed PTSD symptoms. There was evidence that individual TFCBT, EMDR and non-TFCBT are equally effective immediately posttreatment in the treatment of PTSD. There was some evidence that TFCBT and EMDR are superior to non-TFCBT between one to four months following treatment, and also that individual TFCBT, EMDR and non-TFCBT

are more effective than other therapies. There was evidence of greater drop-out in active treatment groups. Although a substantial number of studies were included in the review, the conclusions are compromised by methodological issues evident in some. Sample sizes were small, and it is apparent that many of the studies were underpowered. There were limited follow-up data, which compromises conclusions regarding the long-term effects of psychological treatment. (Bisson, Roberts, Andrew, Cooper, & Lewis, 2016)

These conclusions suggest that TF-CBT and EMDR are effective in the treatment of PTSD. Perhaps this is all we might want to know. This review also suggested that treatments specific to PTSD are more effective than no treatment or usual care. This includes TF-CBT and EMDR, both of which have a special focus on the trauma that has been experienced. The quality of the studies was viewed as poor, but we must realize that this source has high expectations regarding the quality of the research that was undertaken. Weak studies are better than no studies at all.

There are many Internet sources for evidence with regard to human service evaluations. Here is a small list:

Cochrane Collaboration: www.cochrane.org

Campbell Collaboration: www.campbellcollaboration.org

National Institute on Drug Abuse: www.nida.nih.gov

Substance Abuse and Mental Health Services Administration: www.samhsa.gov

American Psychological Association: www.apa.org

California Evidence-Based Clearinghouse for Child Welfare: www.cebc4cw.org

There are many more. When you access these websites, go to a place where you can find evidence and enter the key words you need to get what you want. For the APA website, you need to go to Division 12 for information on evidence.

In many situations in the human service, the search for evidence is not simple. For one, the service may not be easily classified to find the evidence. Consider, for example, the following description of the services of an inpatient treatment program: The intervention employed a family focus group, an activity group, a nurse and medication group, and individual therapy. In this situation, you need to know whether any of these individual services is evidence based, and you would report what you find.

Justifying Your Selection of an Intervention

Why did you select a certain service (treatment, intervention) to achieve the improvement of the client's target behavior? That question should be informed by your analysis of the causes of the target behavior, the needs of those who experience it, and the evidence that you found.

We have discussed ways by which you can justify your selection of a certain intervention on the basis of logic. This is done with an analysis of the nature of the target behavior you are treating and the logical connection of that behavior to the service being provided.

If knowledge is needed, training is a logical service. If the client has been negatively influenced by a peer group, some kind of intervention that would change the peer group could be a logical service.

We have also discussed how you can justify this selection on the basis of evidence. This can be a complicated task, but a simple approach is to find credible sources that provide lists of evidence-based practices. If you wish to publish your study, you will have a more complicated challenge than this. For the novice social work researcher, the avoidance of bogus treatments is a key concern. A bogus treatment is one that not only lacks evidence but also makes outlandish claims of success that seem to defy logic. If someone is asserting that a given treatment is successful for the treatment of all mental health disorders, and there is a lack of evidence to support this claim, you can be assured that you have an example of a bogus treatment. Finding credible sources that support a given treatment should deal with that concern.

The author once attended a presentation on equine-assisted therapy in 2017. The presenter was passionate about this treatment. This method of treatment was described along with a dramatic example of success. When you review a passionate presentation with a dramatic example, it is easy to fall into the trap of following a treatment that is not evidence based. When this presenter was asked about the evidence, a vague response indicated that there was evidence of effectiveness but no specifics were presented. After a review of the evidence, I concluded that equine-assisted therapy was not an evidence-based treatment. There was some evidence in its behalf, but the evidence taken as a whole was not supportive. This did not lead me to conclude that this treatment was harmful, or that it would never be found to be an evidence-based practice, but only that the current evidence on it was not positive. Given the number of treatments that are evidence based, it did not seem to me to be wise for a treatment to be used that failed this test. Evidence on treatments are relevant only to the policy issue of what treatments should be supported by agencies, insurance companies, and the government. Such evidence should not lead to the conclusion that this treatment does not ever work for anyone under any circumstances. You may have a treatment that works when you employ it for a given type of client, even if it is not evidence based. Under these circumstances, you should collect your own evidence on effectiveness. When you are seeking new treatments, you should examine those that are evidence based. And you should be sure you are adequately trained on the treatment that you employ.

Phase 2: Description of the Intervention

The description of the intervention (i.e., service or treatment) is not a phase of evaluation research that you have seen in the previous chapters because they dealt with studies that were not evaluating a service. In evaluation research, it is possibly the most important element of the report because the key question is whether a given service had the intended outcome. If you read an evaluation study, you will want to be clear on the intervention that is being evaluated.

The intervention is the composite of the activities that are provided to clients for the purpose of achieving a given outcome. Interventions (services) typically exist within a broad program delivered by the human service agency. We will describe services with regard to (a) broad goal, (b) specific outcome objective, (c) structure, (d) model, and (e) personnel. Each of these elements of the description of the intervention will be

addressed in detail in this section of this chapter. Briefly, the goal is the long-range benefit to the client, while the outcome **objective** is a measured amount of progress toward the achievement of the goal. The structure of the service reveals what it looks like. The model is the theoretical or conceptual framework that connects the target behavior to the structure of the service and serves as a logical justification of it. The personnel of the service refers to the credentials of those who will deliver the service.

You have seen a review of evidence on what works in the previous section on target behavior analysis. You will, of course, be cognizant of this information when you describe your intervention, especially in the discussion of its model. You may or may not have specific evidence with regard to the specific intervention you are employing, but you are expected to connect your intervention to the evidence that is available regarding your target behavior. Perhaps you have evidence that tutoring normally works in the improvement of the school grades for your target population, but you may not have any evidence of the effectiveness of your specific form of tutoring. So you report what your evidence has revealed, and you examine how it connects to your specific form of tutoring. This means that you are using both evidence and logic to justify your selection of your intervention for the treatment of your specific target behavior.

Goal

The **goal** is the major benefit to the client and is more broad or long term than the outcome objective. Both, however, refer to client outcomes, not service processes. The goal of your adolescent parenting program may be to reduce the negative social consequences of adolescent parenting. Some of these consequences may be dropping out of school, poverty, child neglect, and/or poor parenting. The goal of the tutoring service for at-risk middle school students may be the completion of high school. The goal of your marital counseling service may be to achieve marital harmony. All these are statements of broad benefit to the client. The reason for stating the goal of the service is to indicate the connection between the service objective and the broader benefit. You will measure the achievement of your objective but you will not normally measure the achievement of your goal, partly because it will likely take you too much time to do the latter.

While the goal should be broad and long range, it should be specific enough to provide guidance. Take, for example, the phrase, "To improve the quality of life." What does quality of life mean? Couldn't you say that all services are designed to improve the quality of life in some way? This does not provide guidance on the nature of the long-term outcome in your specific treatment situation.

Objective

The **objective** is a measured amount of progress toward the achievement of the goal. If your goal is to achieve high school completion for at-risk middle school students, you may have an objective of improving grades. You will measure grades in your evaluation of this service, with the assumption that the improvement of grades is a measured amount of progress toward the achievement of high school completion.

The objective is the outcome you will measure in your evaluation study. When you state your objective, you should make sure that you have a measure of it. When you examine your measures, you should make sure that each is connected to an objective.

Each objective should be related to the goal. Here is a list of examples of goals and objectives that have appeared in evaluation research papers by graduate social work students:

1. The goal is to prevent PTSD from causing major problems and difficulties in life for children who have been sexually abused.

   • The objective is to reduce trauma symptoms in children who have experienced a potentially traumatic event.

2. The goal of the intervention is to prevent death by opiate overdose.

   • The objective of the intervention is to increase knowledge about Naloxone, an opiate overdose treatment.

3. The goal is to facilitate healthy grief, and prevent chronic depression, for those who have recently lost a close loved one.

   • The objective is to improve feelings of social support.

4. The goal of the intervention will be to help substance users stay drug free.

   • The outcome objective is to increase the amount of hope among substance abusers.

5. The goals of this service is to improve academic performance of participants.

   • The outcome objective is to improve classroom behavior.

In each of the above examples, the objective was measured to determine if the service had been successful. For the first one, symptoms of trauma were measured. For the second one, knowledge of Naloxone was measured, while the third example had a measure of feelings of social support. The fourth and fifth examples had measures of hope and classroom behaviors, respectively.

Structure

The **structure of the intervention** is a concrete statement that shows both the form and the intensity of the service being given. A group therapy service may be described as weekly 1-hour group sessions for a total of 8 weeks. A training service may be described as a single 3-hour training session, or weekly 1-hour training sessions, over a period of 6 weeks. A case management service may be described as a series of actions including assessment, referral, and counseling that are offered for the typical client for a period of 3 months. Each of these descriptions reveals the form of service (group therapy, training, case management) and the intensity of it, such as eight 1-hour sessions, a single 3-hour session, or a series of actions over a 3-month period of time.

Model

The **model of the intervention** is the framework that serves as the logical rationale for offering it in light of what is known about the dynamics of the target behavior being

addressed. If improved knowledge is the need, education would make sense. Thus, you could label this the educational model if you like. In many instances, giving a label to a model is not necessary. You only need to show the connection of the knowledge about the behavior with the nature of the intervention. For example, if modeling of good behavior is the need, a mentor service would make sense. If dealing with a complexity of life challenges is the need, perhaps case management would make sense. If changes in peer groups is the need, then services that connect clients with different peer groups would make sense.

Some services are guided by a major model like the strengths perspective. This might mean that the strengths perspective is viewed as appropriate to all services that are offered by the agency. If that is the case, you could label your model the strengths perspective and briefly describe what this means.

Personnel

The **personnel of the intervention** refers to the credentials of the service providers? Are they persons with a college degree of any kind or persons with a social work degree? Are they licensed in some way (e.g., Licensed Clinical Social Worker)? This does not refer to the names of the staff members who deliver the service nor does it refer to the clients. It refers to the credentials that staff are required to have in order to be qualified to be a service provider.

In summary, the report on the evaluation of a service should describe the service with regard to the goal, the objective, the structure, the model, and the personnel of it. This part of the report will help readers know a lot about how to replicate this service in their own settings. They will also know what it is that was evaluated and shown to be successful or not.

Phase 3: Study Methods

The study methods reveal what data are to be collected from whom with regard to your study purpose. The three key themes are (1) sampling, (2) measurement, and (3) research design. The activities related to these tasks yield critical information for the reader in deciding the uses of your study report. Your method of selecting a study sample reveals the population to whom you can generalize your findings. The tools you employ show the reader whether your measurement was credible. And your research design will provide information on the extent to which one can conclude that the measured client growth was caused by your service rather than something else, such as normal growth over time.

The phase of research known as study methods has been presented in previous chapters as Phase 2 of the research process. In evaluative research, we insert the phase related to the description of the intervention as Phase 2 and shift study methods to the third phase in the process.

Random Sampling and Scientific Generalization

In the previous chapters, you have seen the following information about sampling:

1. The study sample is composed of all those persons from whom data were collected.

2. The study population is composed of the group of persons from whom a sample was selected.

3. If you employ a random sample, you can scientifically generalize your findings from your study sample to the population of persons from whom the sample was selected on a random basis.

4. If you do not employ a random sample, you can generalize your findings to a designated population on a logical basis if you have critical information on the similarities and differences between the members of the study sample and the members of the population.

5. Scientific generalization is superior to logical generalization.

We call generalization from a random sample as "scientific" generalization because with random sampling there is a scientific basis for measuring sampling error. What is **sampling error**? It is the difference in your data between the results you derive from your sample to the same for the study population. The data from your sample will not be identical to the same for the entire population. However, a random sample will yield data that are quite similar between the two groups even though not identical. It is similar enough to conclude that findings from the sample would normally be basically the same as what you would find if you measured your variables for the entire population. This conclusion is more true the larger your sample size. Sampling error is greater for small samples.

A **random sample** is a group of people selected from a study population where each person in the study population has an equal chance of being in the study sample. The students in your research class do not constitute a random sample of all students in your academic program, even though they each had the opportunity to sign up for this course but only some did so. This is not what equal opportunity means in random sampling. What it means is that the researcher drew a sample out of all in the study population on a random basis. The determinant of randomness is in the hands of the researcher, not the sample member.

Here is a challenge question for you: If you have 50 clients in your service program and you can collect outcome data on all 50 persons, should you include all 50 in your study sample or select a random sample of 25 clients from these 50 clients? The answer is that you should include all 50 clients in your study sample. The purpose of the sampling procedure is to be able to generalize your findings to a larger population. If you can include the entire population in your study, please do so. Data on the entire population are better than data from a random sample from that population. Data from the entire population have no sampling error, and they are larger than a sample.

How do you select a random sample? First, you obtain a list of all the names of people in your study population. Second, you select people for your sample on a random basis from the list of the entire population. You can select members randomly by employing a table of random numbers from a research text or just by going to the Internet for help. There are random number generators there.

There is no random sample option when you are evaluating the services for a single client. You have only one person in your sample. Generalization is not normally a major issue being addressed by such a study. Instead, you are evaluating whether a given service worked for this one client. Data from a single client do not lend themselves to the science of sampling.

In this section, we will discuss some specific things about sampling in the evaluative study, starting with the typical type of sample used by social work students—the convenience sample. You are more likely to find random samples in studies that are published.

Convenience Sampling and Logical Generalization

A **convenience sample** is a group of people who were selected for the study out of convenience. Social work students typically select clients with whom they are currently working. This might be a group of clients with the same treatment objective or a single client. Because this is not a random sample, the results of a study from this study sample cannot be scientifically generalized to a larger population of people.

So what can the social work student do? **Logical generalization** is a possibility. This means that you are generalizing on a logical basis rather than on a scientific one. If you have data on the study population, you can compare your study sample with that data to see if the sample seems to be similar to the population. Any major differences will lower the credibility of your logical generalization.

In one example of logical generalization, an inpatient treatment program for chronic alcoholics conducted a follow-up mailed survey for persons discharged from the service a year earlier. They used the addresses in their records but received only a 20% response rate (i.e., only 20% of the study population responded to the survey). The anonymous survey had questions about their use of alcohol and related behaviors. But they wondered if they could generalize their findings to all persons who had been sent a questionnaire. The 20% who responded was not a random sample. Their only option was logical generalization.

So what did they do? The surveys had questions about alcohol use history and family structure. The agency records of all persons sent a questionnaire had similar data. So they compared the sample (those who returned the surveys) with the population (all those who had been discharged 1 year before). The results in this case revealed that the sample was similar to the population on these variables, things that were thought to predict recovery. The staff concluded that the findings from the survey could be generalized to the population on a logical basis.

Using a Tested Instrument for Measuring Client Progress

Measurement refers to the tools used to measure each variable in your study. If you are examining the relationship between gender and income, you must have a way to measure gender and income for each person in your study sample.

In the previous chapters, you have learned the following about measurement:

1. When you conduct a research study, you must have tools to measure each of the variables in your study.

2. You will measure each variable in your study for each person in your study sample.

3. A variable must vary. If all the people in your study sample are men, you do not have data on the variable of gender, because it does not vary.

4. You can measure variables by a tool designed by yourself or one that has been published.

5. There are two general types of measurement: qualitative and quantitative. When you have qualitative measurement, your data are words. When you have quantitative measurement, your data are in the form of either categories (e.g., male or female for the variable of gender) or numbers (e.g., 34 for the variable of age or 21 for the score on the depression scale).

6. The key issue in measurement is credibility, which can be assessed with regard to reliability or validity.

7. Reliability in measurement refers to consistency.

8. Validity in measurement refers to accuracy.

9. Each variable is measured at one level. The levels form a hierarchy from nominal to ordinal, to interval, to ratio.

In this chapter, you will review concepts about measurement that are especially appropriate for evaluative research.

Alternatives for Data Collection

There are several options by which data can be collected. You might review agency records if outcome data are routinely collected and recorded. You might administer a questionnaire with measures of your variables for each individual in your study sample. You might collect data through an interview. In each case, the issue of credibility (reliability and validity) is relevant, but much easier to establish in some situations than others.

If you are lucky, you can simply gain access to agency records where outcome data are routinely recorded. If school grades are the measure of outcome, you will have an easy time collecting data and you might need to pay little attention to ways of testing for reliability and validity. But be sure you have a measure of client outcome, not client characteristics or amount of service delivered, or something like that. Review the situation to see if you have reason to believe that those who record the data for the agency records are doing so accurately.

Many social work researchers will need to administer a questionnaire to individual clients. The questionnaire typically will have scales that are designed to measure the target behaviors. If you have a published scale for measuring anxiety, you will typically have information on credibility (reliability, validity, or both), which you can report. If you have designed the measurement tool yourself, you have the issue of credibility to examine. In these situations, the minimum expectation you encounter is the reporting of **content validity** and **face validity**, each of which are discussed below in the section on reliability and validity.

Your first task is to see if you can find a published scale for measuring your study variable. The published scale is normally superior to a scale that you have developed yourself because it has been designed by an expert on the theme of the scale, and it normally has been tested for reliability and validity.

Most people will assume that grades are appropriate as a measure of school performance. A depression scale, however, is a different matter. There may be readers of your report who wish to know about tests for credibility for your depression scale. They want to know if there is evidence that this tool accurately measures depression. This evidence is a major advantage of using a published scale rather than one that you have developed yourself.

Another form of data collection is the interview. This is more likely to be the vehicle when you are measuring variables qualitatively (i.e., words rather than numbers). It might also be useful when you are measuring a client with a mental challenge who needs things to be explained in careful detail.

Reliability and Validity

The key issue in measurement is credibility. Do readers of your report have confidence that you have measured your study variables in a reasonable way? We have reviewed the issue of credibility with regard to the concepts of reliability (consistency) and validity (accuracy). We have seen that there are several ways to test for reliability. One is the administration of your tool to the same group of people at two points in time to see if the answers are consistent. A test for validity is to examine the correlation of scores on your tool with another tool designed to measure the same thing. You would expect a strong correlation between the scores on the two tests. There are many more tests for reliability and validity.

If you develop your own tool for measurement, you have to face the challenge of considering reliability and validity. At a minimum, you should test for face validity by asking a set of informed people to report whether they believe that your tool accurately measures what you intend for it to measure. If almost all say *Yes*, you can report this as your measure of face validity. You will review other ways of testing for reliability and validity in the chapter on measurement.

Finding a Measurement Tool

In this section, let's examine ways to develop or find a measurement tool. Your first step should be to examine existing tools to see if one is suitable for measuring your study variables. You can search the Internet for such tools. There are various sources from governmental websites that provide information on the nature of many tools for measuring target behaviors. You should first review the purpose of each tool that seems reasonable to see which one is designed to measure just what you wish to measure. You will likely find a copy of the tool through the Internet. You will find books with tools as well. Perhaps your library has many to choose from.

As you review published tools for measuring your variable, you should review several questions. One is the relevance of this scale to the variable you have defined. Another is the credibility of the tool for measuring what you want to measure. Finally, you need to understand the tool adequately to use scores on it to help with practical significance—the amount of change for clients that would be clinically noteworthy.

Here is a checklist of questions for selecting a measurement tool.

- Does this instrument clearly measure your variable as you have defined it?

- Is this tool appropriate for your clients? (Some scales have been developed for children and some for adults.)

- Is your instrument sensitive to the changes you can expect from your clients in the time period they will be measured? (You are more likely to see a change in knowledge from measuring clients before and after three training classes than to find significant change in anxiety with only three therapy sessions.)

- Does your tool have adequate credibility?

- Do you know how to score the instrument?

- Do you know what level of gain would constitute practical significance?

Your second step is to determine how to administer the tool to your clients. This entails making a copy of the tool along with an introduction to the respondent regarding how the information will be used. If you are employing a tool two different times and need to match each person's pretest score with his or her posttest score, you can employ a set of questions for the respondent that provide an anonymous identification number. Privacy, as you will recall, is one of the ethical themes in the use of human subjects in research. The questions might include things such as (a) What is the first letter of your father's middle name? (b) What is the number of letters in your mother's maiden name? (c) What is the last digit of your social security number? You put together the answers to each of these questions to form an identification number that the respondent will know but you do not know. Put the same questions both on the pretest and the posttest questionnaires.

Developing Your Own Tool for Measurement

There are times when the researcher needs to develop his or her special tool for measuring study variables. If you are employing a training program, you will often need to develop your own tool for measuring knowledge of the subject of the training. If your study variables are rather unique, the development of your own measurement tool may be warranted. However, you should always begin your quest for a tool by reviewing the literature to see if a suitable tool has been developed and tested. You do not want to reinvent the wheel. It is likely that any published scale will have been developed by an expert and tested for reliability and validity. Surely this is better than one you develop yourself! The first question, however, is whether that published tool measures what you need to measure. If you are trying to improve the parental skills of parents with an autistic child, a general parental skills tool may not be sufficient. Before you construct your own tool for measuring client progress, review the chapter in this book on measurement.

Research Designs and Causation

The research design specifies the procedures for the measurement of target behavior. You may measure a group of clients before and after the provision of an intervention and compare the two measurements to determine if the clients improved and if that improvement was statistically significant. This illustrates the **one-group pretest–posttest design**. Another design calls on you to measure a group of clients in the same way but compare the gain of the clients with another group of people who did not receive the intervention. In this situation, you are employing the **comparison group design**. There are designs for measuring the progress of a single client. If you are employing the **limited AB single-subject design**, you will measure a single client one time before treatment begins (perhaps the assessment interview) and several times throughout the treatment period (perhaps at the beginning of each treatment session). You cannot, however, measure a single client one time before treatment begins and one time at the end of treatment. That procedure does not constitute any recognized single-subject design because it fails to generate sufficient data to analyze statistically.

There are other designs that will be discussed in a later chapter. The above-mentioned three designs are often used by social work students who are called on to evaluate a service in their research courses. The first and second designs describe group designs, where one

or more groups of people are measured to see if the service was effective. The last one is a single-subject design, where a single client is measured repeatedly over a period of time to determine if the service is working. These will be discussed in a later chapter.

The main reason you need to understand the research design employed in an evaluative study is because of the issue of whether the intervention was the cause of the client growth rather than something else, such as normal growth and development over time. If you use the one-group pretest–posttest design, you will have a measure of growth but will not have evidence that the treatment caused the growth. Maybe the clients were already on the path of recovery and have been making progress weekly. The amount of their growth during the treatment period may not be better than the pace of progress already underway.

On the other hand, suppose you compare the growth of your clients with the growth of a comparison group of people who did not receive the treatment. If you find that your clients did better, you will have evidence of causation. The growth of the comparison group should be a good gauge of the effect of normal growth over time. If your clients did better than this, you can assert that your treatment was the cause of the difference in growth between the two groups because you have controlled for the effects of normal growth over time.

When you make your decision about which research design to employ, you should consider the alternative causations of client growth. One, of course, is the effect of the treatment. But there are other alternatives. One is the effect of normal growth over time that was discussed earlier. Another possible cause is a change in the client's environment. These two causes of client improvement that is outside the effects of treatment are referred to as **threats to internal validity**. So clients may improve because of treatment or they may improve because of something else. The research design employed will help us decide which it is—the treatment or something else. More specifics regarding research designs and causation will be reviewed in a later chapter where you will see which design addresses which of the other causes of client growth.

If you have a good reason to believe that things other than the treatment will be the cause of your measured client gain, you should employ a research design that controls for these other effects. In many situations, you will conclude that it is unlikely that things other than the treatment will affect the target behavior. The issue here is not whether it is possible that something else will cause the measured client growth but whether it is *probable* that they will. If it is not probable, you do not need to worry about these alternatives, but you will report the limitations of your study in the conclusions section of your report, one of which may be the research design employed.

Ethics in the Use of Human Subjects in Research

In several previous chapters, you have seen information on key ethical issues in the use of human subjects in research. In evaluative studies, you will typically be asking clients to complete a questionnaire for your data. In this regard, you should be attentive to the issues of (a) privacy, (b) voluntary participation, (c) justice, and (d) potential harm. For the typical study, like the one illustrated in the research example given in the next section, you will only need to be attentive to the issues of voluntary participation and privacy, because your administration to a questionnaire to clients will not likely require

special attention to the issues of justice and harm. You can attend to the issues of voluntary participation and privacy by having clients complete a questionnaire voluntarily and without their names on the instrument. However, if you need their names for research purposes (e.g., matching pretest and posttest scores), you can address the issue of privacy by keeping the questionnaires confidential.

Phase 4: Data Analysis

The key issue for data analysis in evaluative research is whether the data supported the study hypothesis. From previous instruction, you have learned that you must have data that are in the hypothesized direction and you must also have data that are statistically significant. The second key issue is whether the data are of practical significance. You have learned that you cannot easily make the case that you have practical significance if you do not have statistical significance. If your data can be explained by chance, how can you say they are of practical benefit in the support of the intervention that is being evaluated? What you have discovered is chance!

In this section, you will see a review of tips for constructing the study hypothesis. You will also be given further guidance on statistical analysis, especially for the evaluative study. And you will review some suggestions about practical significance.

Testing the Study Hypothesis

In the previous chapters, you have learned that the study hypothesis is an educated guess about the results you expect to find from your data analysis. You have been given examples like the following:

- Posttest scores for anxiety will be lower than pretest scores.

- Grades at the end of treatment will be higher than grades before the treatment began.

- The proportion of homeless clients who find a home within 6 months of service will be higher than the proportion of homeless people who find a home within 6 months on their own, according to national statistics.

Let's examine a few tips about writing the study hypothesis for the evaluation research study. Here are three such tips:

1. Your hypothesis should identify the client's target behavior that will be measured in your study (e.g., scores on the anxiety scale, grades in school, marital satisfaction scores, and getting a home.)

2. Your hypothesis should report what is expected (e.g., anxiety scores will go down, grades will go up, acts of aggression will go down).

3. Your hypothesis should not contain assertions of causation, for example, "Because of the tutoring service, clients grades will go up."

When you present your statement of hypothesis and your client data, another researcher should be able to test your study hypothesis, provided that he or she knows a statistical test that is appropriate. They will not have to ask you any further questions.

One more tip. Keep the statement of the hypothesis concise, like the ones mentioned above. You do not see unnecessary words that might be called *normal talk* among practitioners. For example, you did not see a statement such as, "Given the importance of anxiety to people who have recently been the victims of family violence, we expect that our clients will see their anxiety scores go down during the treatment they receive because our treatment addresses the causes of anxiety for victims of violence."

Your first step in testing your hypothesis is to submit your data to statistical analysis. With regard to the example about anxiety scores, you will need to determine if the posttest scores are lower than the pretest scores. If not, then you can conclude that your data failed to support your hypothesis. The issue of statistical significance is no longer relevant.

If your data fail to go in the hypothesized direction, you can report that your data failed to support your hypothesis. For example, if posttest scores are not better, you cannot say your data supported your hypothesis. If your data are in the hypothesized direction, you test for statistical significance. If your data go in the hypothesized direction but fail to achieve statistical significance, you must report that you data failed to support your hypothesis. On the other hand, if your data are in the hypothesized direction and are statistically significant, you can report that your data supported your study hypothesis.

You have been instructed that the value of p (converted to a whole number) represents the number of times in 100 that your data would occur by chance. The expression "$p < .05$" means that your data would occur by chance less than 5 times in 100. This is the normal standard in the social sciences for finding statistical support for your hypothesis. Always remember that you have two conditions for determining if you data supported your hypothesis: (1) the direction of the data and (2) statistical significance.

Practical Significance

Practical significance is a judgment about the extent to which the measured results are noteworthy. It is informed by knowledge of the nature of the target behavior that is being measured. You must first decide how much of a change is sufficient to be of practical significance. You might have the aid of information about various thresholds of functioning according to the scale that is being used. You might find, for example, that a score greater than 30 on your depression scale represents a severe level of depression, likely requiring hospitalization. You might learn that a score between 20 and 30 represents major depression and a need for therapy, whereas a score less than 20 represents no need for therapy. If your client moved to a lower threshold, you could claim this as evidence of practical significance.

You should become familiar with what a given score on your scale represents by examining the nature of your scale. If your scale consists of 10 things an abusing partner might do to a spouse and find that each of these items is significant, perhaps you could make the

case that a change of only one point on this scale is of practical significance. On a different scale, you might make the case that a change of 25% constitutes practical significance by explaining what this level of change looks like in real life. What you cannot do, and assert an informed opinion, is to just say that it seems to you that this level of client gain is of practical significance. You have to offer a rationale.

You should decide on your standard for practical significance before you analyze your data. This decision should be independent of the actual data you have for the process to be objective. Remember that research is a process of discovery, not justification. If you know your data before you decide on your standard for practical significance, wouldn't this fact create a bias with regard to the selection of the standard?

Phase 5: Writing the Study Conclusions for the Evaluative Study

In the previous chapters, you have reviewed the following suggestions about the conclusions for your study:

- Your first task is to summarize your data with regard to your research question.

- Do not draw conclusions that outrun your data. Reserve your opinions that are not supported by the data for other venues.

- Discuss the limitations of your study, with a special reference to the kind of sample you employed, the manner in which you measured the variables, and the study design that was used.

- Discuss the implications of your findings for practice. How might practitioners make use of these results?

- In the explanatory or evaluative study, do not conclude that practical significance was found if your data were not statistically significant, given the fact that your data can be explained by chance.

Another thought about the last item mentioned above. If you believe that your data are of practical significance but fail to be statistically significant, you are very likely to have a small sample size. In this situation, you may declare your data not to be amenable to statistical analysis and conclude that another study should be conducted with a larger sample. You should not, however, conclude that your service was found to be effective in your study.

A few more general suggestions are in order. First, don't say you found "nothing" because your data failed to support your hypothesis. Failure to find support for your hypothesis is an important finding. Perhaps we should rethink the rationale for the hypothesis. Second, your conclusions should be tentative. You have not really found the final proof of anything. In fact, saying that you "proved" something is not in good form. Instead, you found, or failed to find, support for your hypothesis. This is more tentative than saying that you have proved something.

This chapter has discussed evaluative research, so your conclusions will be about your study that evaluated a service. This means that you will add the issue of practical

significance to the themes you will address in your conclusions section. You need to provide a rationale for declaring that you did, or did not, find results that were of practical significance.

A consideration in the determination of practical significance is the extent of the resources that were employed by the service that is being evaluated. Suppose you found a 10% improvement in anxiety for a minimal service (e.g., weekly 1-hour therapy sessions for 3 weeks). Could these data support your claim for practical significance? Probably so, because the gain is minimal but so is the expenditure of resources. But now think of the same amount of gain for an extensive service such as mental health treatment given in a hospital for 30 days. Now, you would find it more difficult to convince others that you have practical significance.

RESEARCH EXAMPLE: ARE THE SERVICES OF THE NEW HORIZONS TREATMENT PROGRAM EFFECTIVE IN THE REDUCTION OF DEPRESSION FOR ADULT CLIENTS?

In this research example, you will review the entire process of evaluative research. This means that you will review (a) the development of the research question and knowledge base, (b) the selection of the research methods to be employed, (c) the analysis of the data, and (d) the drawing of conclusions. Each of the first two phases of the evaluative research process will be presented to you, but you will undertake the analysis of the data and the drawing of conclusions yourself.

A service within the New Horizons Treatment Program will serve as the example. Included in the services of this program is individual therapy for adults clients who are seeking services to reduce their depression. A group of clients provided this service constituted the study sample.

Phase 1: The Research Question and Knowledge Base for the New Horizons Study

The research question for this study is as follows: Is CBT offered by the New Horizons Treatment Program effective in the reduction of depression for adult clients? Because this is a research question for an evaluative study, it identified both the target behavior being treated and the treatment that was being evaluated. CBT was the model of therapy that was offered to the clients who were included in this evaluation study. The knowledge base examined for this study entailed the review of literature on the definition of depression, the causes of depression, and the results of studies of various treatments for depression. In particular, there was a focus on the evidence regarding the use of CBT in treatment.

What Is Depression?

You have probably discussed depression with others on many occasions, and you should have an idea of what it means. But your conception may be incomplete as a guide

for an evaluation research study. For this reason, you would likely seek guidance from the literature. Tools that have been developed to measure depression will likely have a definition for this concept. The Internet will have multiple sources that you can use to find your definition. In the study reviewed here, depression was defined as follows:

> Depression is a feeling that can be depicted as sadness, worthlessness, and guilt. These feelings tend to be accompanied by a desire to be alone along with a loss of appetite, sleep, and general activity. Clinical depression is defined as a state of depression that has seriously interfered with normal human functioning for at least 6 months.

Depression as depicted above can be distinguished from normal feelings of sadness, which is short term and likely related to a stressful life event. As depicted above, depression accompanies a feeling of having given up on life in the normal sense. Ambition and hope are low, and the state of depression has been evident for a long period of time, like 6 months or more. A critical role for the definition of the target behavior is to provide guidance on the selection of a tool to measure it. You can examine the tool used to measure depression for the New Horizons study (identified in the section "Study Methods" of this report) to see if you believe that it does an adequate job of measuring depression from the definition given above.

Causes and Consequences of Depression

According to information from the National Institute of Mental Health, major depressive disorder (clinical depression) is a serious mood disorder that affects how a depressed person feels, thinks, and handles daily activities, such as sleeping, eating, and working. To be diagnosed with depression, these symptoms must be present for at least 2 weeks (National Institute of Mental Health, 2018). Among the risk factors (i.e., potential causes) of depression are (a) a family history of depression, (b) major life changes, (c) trauma, (d) stress, and (e) certain physical illnesses or medication (National Institute of Mental Health, 2018).

Among the treatments for depression is psychotherapy. According to the National Institute of Mental Health, there are several types of psychotherapy for the treatment of depression. Examples of evidence-based approaches include CBT, interpersonal therapy, and problem-solving therapy (National Institute of Mental Health, 2018).

What Works in the Treatment of Depression?

The next task is the review of the evidence regarding the treatment of depression. The federal government is one source of information on evidence-based practices for the treatment of various target behaviors. For example, the Substance Abuse and Mental Health Services Administration (samhsa.gov) lists both CBT and interpersonal therapy as being among the evidence-based practices for the treatment of depression (https://www.samhsa.gov/treatment/mental-disorders/depression#evidence-based).

Another source is the Society of Clinical Psychology, which lists the following evidence-based treatments for the treatment of depression: (a) CBT, (b) behavioral therapy, (c) problem-solving treatment, (d) interpersonal psychotherapy, (e) reminiscence therapy, and (f) cognitive bibliotherapy (https://www.div12.org/treatments)

These sources provide a number of treatments that are supported by evidence. Evidence-based practice is the theme of one of the other chapters in this book. The treatment employed in the treatment of depression at the New Horizons Treatment Program was CBT, one of the treatments listed in each of the above-mentioned sources.

Phase 2: Description of the New Horizons Treatment Program

You have reviewed how the intervention (or service) can be described comprehensively with regard to goal, objective, structure, model, and personnel. A major task in the report on the study of the New Horizons Treatment Program is the identification of the intervention with regard to these five concepts.

What should be the goal of treatment for the New Horizons Treatment Program? Should it be (a) to reduce depression, (b) to provide good quality services, (c) to enhance social and psychological functioning so as to improve self-sufficiency, or (d) none of the above? As you answer this question in the practice exercise, keep in mind the distinction between the goal and the objective, and keep in mind that they both focus on outcome rather than on process. What should be the objective? Is it one of the options mentioned above with regard to the goal statement? If not, what should be the statement of the objective of treatment?

The service of the New Horizons Treatment Program is a form of psychotherapy that is delivered in 1-hour sessions, once per week, for a period of 8 weeks. The therapy sessions are delivered at the agency.

The therapy model employed in this treatment is CBT. The Beck Institute is an entity that was developed by the founder of this model. According to the Beck Institute, CBT is a "time-sensitive, structured, present-oriented psychotherapy directed toward solving current problems and teaching clients skills to modify dysfunctional thinking and behavior" (Beck Institute, 2016). It is based on the cognitive model that suggests that the way in which we perceive situations is more closely connected to our behaviors than the situation itself (as viewed objectively). One important part of CBT is helping clients change their unhelpful thinking and behavior that lead to enduring improvement in their mood and functioning. A variety of cognitive and behavioral techniques are employed in CBT. It also employs problem-solving techniques as well as techniques from a variety of treatment models (Beck Institute, 2016). A key assumption of this model is that dysfunctional thinking causes psychological problems such as depression, so the treatment should address thinking patterns.

The staff who delivered this therapy at the New Horizons Treatment Program were licensed clinical social workers, each of whom have the LCSW credential and at least 2 years of clinical experience. These two qualifications are required by this agency for the position of therapist for this program.

Phase 3: Study Methods for the Evaluation of the New Horizons Treatment Program

The individuals from whom data were collected for this study consisted of the first 15 persons who sought treatment for depression from the New Horizons Treatment Program in the month of February 2017. These individuals were from the target population of persons treated by this agency who have reported to have been depressed

at a level that interfered with their satisfaction with life for a minimum of 6 months. The New Horizons Treatment Program has provided the same treatment as revealed in this study to the same target population for the past 2 years to approximately 250 individuals.

The measurement tool was the revised version of the Beck Depression Inventory—II (BDI-II). In its current version, this scale is designed for individuals who are 13 years old and above. It is designed to measure depression and is composed of items relating to symptoms of depression (e.g., sadness, hopelessness, irritability), cognitions indicative of depression (e.g., feelings of guilt), as well as physical symptoms such as fatigue, weight loss, and lack of interest in normal activities of life such as sex. It has 21 questions about how the subject has been feeling in the past 2 weeks. Each question has a set of four possible responses, ranging in intensity. For example, the respondent is asked, on one item, to select one of the following four options with regard to the theme of sadness: 0 = *I do not feel sad*, 1 = *I feel sad*, 2 = *I am sad all the time and I can't snap out of it*, 3 = *I am so sad or unhappy that I can't stand it*. The individual gets 0 points for selecting the first option, 1 point for the second, 2 for the third, and 3 for the last option. This scale has been subjected to numerous studies of reliability and validity and has been shown to have passed the test in good form (see, e.g., Visser, Leentiens, Marinus, Stiggelbout, & van Hilten, 2006). There are several thresholds of functioning exhibited by scores on the BDI. Scores of 0 to 13 are considered minimal, while scores from 14 to 19 are considered mild. Treatment is not normally indicated for people at either of these levels. At the next level, scores from 20 to 28 are considered to be moderate depression, and scores of 29 and above are considered to be severe depression.

The research design employed in this study was the one-group pretest–posttest design. This means that all the 15 clients in this study were given the BDI before the treatment began and were given this same scale at the end of the treatment period. The difference in these two sets of scores were compared to test the study hypothesis.

Phase 4: Analysis of Data for the Evaluation of the New Horizons Treatment Program

The hypothesis for the study of the New Horizons Treatment Program was as follows: *Posttest scores on depression will be lower than pretest scores*. The pretest scores, posttest scores, and gain scores for each client are given in Exhibit 8.1. The statistic for testing this hypothesis will be the paired *t* test, because we have a set of matched scores for one group of people. You should remember that a statistical test will generate the value of *p*, which is considered statistically significant if its value is less than .05 (a *p* value of .05 means that these data would occur just by chance only 5 times in 100).

When you examine the data, you have two major questions to answer: (1) Are the data statistically significant? (2) Are the data of practical significance? The value of *p* is the key to statistical significance (*p* < .05). Practical significance is a matter of informed opinion. Is the client gain enough? Did it meet our expectations? On what basis do we answer these questions? It is not sufficient to say "I believe that practical significance was achieved?" Instead you could say "Practical significance was achieved because . . ." It will be the words after *because* that are critical. This is where you justify why you believe practical significance was achieved.

EXHIBIT 8.1
DEPRESSION SCORES FOR CLIENTS OF THE NEW HORIZONS TREATMENT PROGRAM

Client Number	Pretest Score	Posttest Score	Gain Score
1	31	28	3
2	34	30	4
3	30	20	10
4	28	28	0
5	41	30	11
6	26	25	1
7	31	25	6
8	38	28	10
9	35	22	13
10	45	33	12
11	41	30	11
12	30	30	0
13	31	31	0
14	38	31	7
15	44	28	16

What would, in your opinion, be a gain on the BDI that would be of practical significance in this study, given the extent of the treatment offered? Would it be a movement from one threshold to another in functioning (e.g., from severe depression to moderate depression)? Would it be a certain percentage of gain? What does one point of change mean, and how can this knowledge guide your determination of practical significance?

Phase 5: Study Conclusions for the Evaluation of the New Horizons Treatment Program

Study conclusions are based on the analysis of the data from the study. You have not seen that analysis because this task will be the first thing you do in the first practice exercise for this chapter. In the previous chapters, you have been given a set of tips about how to report your study conclusions. You have been instructed to summarize your findings on your research question and to be sure your conclusions do not outrun your data.

For example, you should not draw conclusions that are not supported by your data just because you believe these statements are true or have been supported by the literature. Your conclusions for an evaluative research study should be supported by your data.

You have been instructed to speak about the limitations of your study and the implications for practice. And you have reviewed the theme of practical significance, a topic that belongs in the conclusions section of the research report.

What follows are the two practice exercises for this chapter. In Practice Exercise 1, you will examine the data from the evaluation of the New Horizons Treatment Program, and you will provide a report on this study. In Practice Exercise 2, you will report what a familiar human service agency is doing with regard to the evaluation of input, process, and outcome.

Chapter Practice Exercises

Practice Exercise 1: Report on the New Horizons Treatment Program Evaluation

In this practice exercise, you will examine the data from the New Horizons Treatment Program and provide a report on key aspects of this evaluation study. You have been given information on this evaluation study with regard to (a) the articulation of the research question and knowledge base, (b) the research methods employed in the study, and (c) the intervention being evaluated. You have also been given the data from that evaluation study. Your tasks are to analyze the data that have been given, to draw conclusions for this study, and to report on all phases of this evaluation research study.

Competencies Demonstrated by Practice Exercise 1

1. The ability to use an Internet website to analyze data in an evaluative research study that includes matched scores from a single group of clients

2. The ability to report whether a set of analyzed data supported the study hypothesis with regard to both the direction of the data and statistical significance

3. The ability to draw appropriate conclusions from an analyzed set of data for an evaluative research study

4. The ability to report on all phases of evaluative research regarding (a) the research question and knowledge base, (b) the intervention being evaluated, (c) the study methods employed, (d) the analysis of data, and (e) the study conclusions

5. The ability to report on how ethical issues were addressed in an evaluative study

Activities for Practice Exercise 1

6. Review if the information on the evaluative research example (New Horizons Treatment Program) is sufficient to report (a) the research question and knowledge base, (b) the description of the intervention being evaluated, and (c) the study methods employed. In addition, review the data given in Exhibit 8.1.

7. Analyze the data in Exhibit 8.1 to determine if the study hypothesis was supported using the instructions from the chapter appendix on how to use an Internet website for this purpose.

8. Decide on the conclusions that should be drawn from the study of the New Horizons Treatment Program.

9. Provide a report on the evaluation of the New Horizons Treatment Program using the questions delineated in the instructions for your practice exercise report.

What You Will Report From Practice Exercise 1

Your task in your practice exercise report is to prepare answers to the questions given below.

A: Target Behavior

- What is the target behavior? What is your definition of it?

- What are some of the treatments that have been found to be effective in the treatment of this target behavior (include literature reference citations)? Is the treatment used by the New Horizons Treatment Program among the treatments found to be effective?

B: Intervention

- How would you describe the New Horizons Treatment Program with regard to goal, objective, structure, and personnel?

C: Study Methods

- What is the study population for the New Horizons Treatment Program evaluation? What is the study sample? Can you generalize your findings to this population on a scientific basis? Explain.

- What tool was used to measure the target behavior identified in the treatment objective? Name the tool and briefly describe it. Does this tool do an adequate job of measuring the target behavior (e.g., "Is it congruent with the target behavior definition?" "Has it passed any tests of reliability and validity?")?

- Do you believe that these 15 clients of the New Horizons Treatment Program would have achieved a noteworthy gain in depression scores if they had not had the intervention during the period of time of the treatment (8 weeks)? Explain.

- Does the research design employed in this study do an adequate job of addressing possible causes of client gain on the target behavior other than the intervention, such as normal growth over time? Explain. (*Note:* If there are no such causes that should be of special concern in this situation, you should say so in your answer.)

D: Data Analysis

- What was the hypothesis that was tested in the study of the New Horizons Treatment Program?

- What statistical test was employed to test the study hypothesis?

- Did the data support the hypothesis? Be sure to answer this question with specific data on the mean scores and the value of *p*.

E: Conclusions

Provide the first paragraph of your statement of conclusions. This includes only a summary of the study results and the basic conclusion that was warranted. You are not asked to discuss the implications of this study for practice or the strengths and limitations of the study methods employed.

Practice Exercise 2: Evaluating Various Aspects of Your Agency's Service System

In this exercise, you are called on to examine one aspect of a familiar service system, such as input, process, output, or outcome. The task is to report what is known from agency records regarding one of these aspects of system evaluation. You

should select one of the following five questions to answer and provide a report.

1. What information does the agency have regarding the evaluation of input (need), and what does the agency do with this information?

   a. Is there a need assessment report that provides data on the extent to which the agency services are addressing the needs of the target population (e.g., abused children, alcoholics, homeless people, persons with eating disorders, mental health of military families, etc.)? These data should show a breakdown of client characteristics that show the extent to which the services are being delivered to this population.

   b. Does the agency have information on unmet needs, such as a waiting list of applicants for service or a report from the U.S. Census Bureau that shows how many people are in the target population for this agency as compared with the number that have been served.

2. What information does the agency have on the evaluation of service process, and what does the agency do with this information?

   a. Are there service protocols that show the agency's service policies with regard to the nature of the service being provided, such as clients being informed of their rights and clients receiving each of the various aspects of the service program (e.g., case management, counseling, and tutoring)? Are data reported on the extent to which these policies have been implemented?

   b. Does the agency monitor services to see to it that the approach of the agency to service is being delivered? For example,

if the agency reports that it provides strengths-based services, how does it demonstrate that this is happening with data?

   c. Does the agency report about standards of service regarding things such as the credentials of the service providers?

3. What information does the agency have on service output, and what does the agency do with this information?

   a. How does the agency define the unit of service for a given program? Is it a client served, a time unit (an hour of therapy), an episode of service (e.g., a child abuse investigation completed), or something else?

   b. Does the agency report the number of units of service that it provides?

   c. Does the agency report the cost per unit of service? If not, can you compute it by dividing the total cost of the program (e.g., $153,000) by the number of units of service delivered?

4. What information does the agency have on client outcome, and what does it do with this information?

   a. How does it define the outcome objective (e.g., to improve school grades, to reduce anxiety, to improve compliance with the medical treatment plan, etc.)?

   b. What data are collected on this outcome?

   c. What are the results?

5. Does the agency measure client satisfaction? If so, how does it do so (e.g., via a mailed survey conducted when the client is discharged, a suggestion box in the agency lobby, etc.)? What are the themes that are included in this information? What have been the results?

CHAPTER REVIEW

Chapter Key Learnings

1. The systems model can help illustrate the nature of evaluation research in a comprehensive fashion, including the assessment of need (input), the examination of service standards (process), the examination of service efficiency and quantity (output), and the evaluation of client outcome.

2. The normal phases of research include the knowledge base, the study methods, the analysis of data, and the drawing of conclusions. In evaluative research, another phase is included: the description of the intervention (service). Another special aspect of evaluation research is the essentiality of reporting on the research design in reference to the issue of causation.

3. The knowledge base for the evaluation study includes special attention to the definition and analysis of the target behavior being measured as well as the evidence on what works in its treatment. The definition provides special guidance for the measurement of the target behavior, while the analysis offers guidance on the justification of the chosen treatment.

4. There are good sources of evidence for what works in the treatment of various target behaviors, the systematic review of evidence being at the top of the list of scientific approaches to making decisions.

5. The selected intervention can be justified on the basis of evidence or logic or both. Evidence as a justification refers to scientific outcome studies that show that a treatment is evidence based for a given target behavior. Logical justification refers to the logical analysis of information. For example, if scientific studies reveal that persons who are depressed are more likely to exhibit dysfunctional thinking about life events, a treatment that addresses dysfunctional thinking would be logically justified.

6. The intervention should be described with regard to the goal, the objective, the structure, the model, and the personnel. These themes provide a comprehensive report on the nature of the intervention that is being evaluated.

7. The random sample provides the basis for scientific generalization of study findings. A convenience sample might be generalized on a logical basis if there is sufficient information that shows that the relevant characteristics of the study sample are similar to the characteristics of the study population.

8. There are times when the researcher is fortunate enough to have a measure of client outcome that is found in agency records such as school grades or disciplinary actions. But target behaviors that deal with social and psychological functioning require some attention to the science behind the scale to be employed. When you conduct an evaluative research that addresses such a variable, you should first seek a published scale for measuring client progress because it has been designed by experts and has typically been tested for reliability and validity. If an appropriate published scale cannot be located, you can develop your own tool for measurement of target behavior provided that you adhere to certain principles of survey measurement.

9. Among other issues in the selection of a means of measurement for psychosocial variables such as depression, marital harmony, and so on are the issues of reliability and validity. Reliability refers to the consistency of a measurement tool, which can be tested by giving the same group of people the tool at two points in time and checking the data for consistency. Validity refers to the accuracy of a tool and can be tested by computing the correlation between the designated tool and a published tool that measures the same thing. A positive correlation would be evidence of the validity of the designated tool.

10. Among the various research designs available for the evaluative researcher are the one-group pretest–posttest design, the comparison group design, and the limited AB single-subject design. A critical issue addressed by the selection of the research design is causation. Did the treatment cause the client's measured growth or was it caused by something else, like normal growth over time? The one-group pretest–posttest design does not address normal growth over time as the cause of the measured client growth because you only have a measure of whether the client improved, but this improvement might be due to normal growth over time. The comparison group design, however, does address this potential cause because the growth of the comparison group is a measure of the effect of normal growth over time. The

limited AB single-subject research design does not control for normal growth over time because you only have a measure of client growth but not a measure of the trends in the target behavior by the client over time.

11. The study hypothesis for an evaluative study identifies the target behavior being measured and the expected direction of the outcome. It should avoid assertions of causation or nonessential phrases. The hypothesis should be a simple statement such as, "Posttest scores for anxiety will be lower than pretest scores."

12. For your data to support your study hypothesis in evaluative research, you must find that the data showed client growth and the amount of growth was found to be statistically significant.

13. Practical significance in evaluative research is an informed opinion on the extent to which the study results were noteworthy. Is a 20% gain in client functioning good news or is it a disappointment? A study can have statistical significance without having practical significance; these are two different issues. Given the nature of a failure to find statistical significance, it would be difficult to argue that data that are not of statistical significance are, in fact, of practical significance.

14. Study conclusions in evaluative research should summarize the findings of the study and the implications of these results for social work practice or policy.

Chapter Discussion Questions

1. What is one of the major advantages of a careful definition of the client's target behavior in evaluative research? In other words, what other aspect of the research process is aided by a careful definition rather than a nebulous one?

2. What is one of the major advantages of a good target behavior analysis, where you examine the causes of the target behavior? In other words, what other aspect of the research process is aided by a good analysis?

3. Suppose that you searched for evidence regarding a social work practice and failed to find any research on this practice. How would this affect you?

4. Think of a study that you might do with a group of clients where you have a convenience sample? What makes this a convenience sample rather than a random one? To whom could you generalize your findings on a logical basis? Explain.

5. Think of a tool that you might design to measure client progress with this group of clients. This is not a tool that you selected from published tools but one that you will design yourself.

   a. What is your definition of the target behavior in this example?

   b. Present one item that would be on the scale you designed.

6. Can you have practical significance with your data if you do not have statistical significance? Explain.

7. When you are writing the conclusions section of your research report, what should you do with your own opinions? Should you present them in the same light as the results of your study, avoid presenting them, or something else?

Chapter Test

1. Which of the following are included in the evaluation of the service system?
   a. The efficiency of service delivered
   b. The needs of those in the target population
   c. The outcomes of services
   d. All of the above

2. Service standards refer to which of the following:
   a. Service input
   b. Service process
   c. Service output
   d. Service outcome

3. The best place to obtain guidance for the selection of a tool for measuring client progress would be
   a. The sampling procedure employed
   b. The analysis of the causes of the target behavior

   c. The definition of the target behavior to be treated.
   d. The analysis of client need

4. The sampling procedures employed in a study are helpful for which issue?
   a. The question of what caused the clients to improve
   b. The generalization of study results
   c. Whether chance is a good explanation for study results
   d. How to define the target behavior

5. The systematic review is a vehicle for
   a. Reviewing various methods of selecting a study sample
   b. Reviewing research designs
   c. Reviewing evidence
   d. Reviewing optional statistical tests

6. Which of the following would constitute appropriate statements of the structure of an intervention?
   a. Interpersonal therapy will be delivered in hourly sessions once per week for 6 weeks
   b. Interpersonal therapy is designed to reduce the anxiety of those who are the victims of violence
   c. Both of the above
   d. None of the above

7. Which of these statements would be the best statement for the outcome objective of the support group service for victims of rape?
   a. To reduce the social and psychological consequences of the experience of rape
   b. To improve feelings of social support
   c. To deliver support group services once per week for 5 weeks.
   d. To provide support group services using qualified personnel

8. You can examine test–retest reliability by
   a. Giving your measurement tool to the same group of people at two different points in time and examining the results to see if there is a good correlation between the scores for these two points in time
   b. Examining the average correlation among items on a measurement tool

c. Both of the above
d. None of the above

9. If you select a random sample of members of the Nebraska Chapter of the National Association of Social Workers, you can scientifically generalize your findings to what population?
   a. All social workers in the state of Nebraska
   b. All social workers in the nation
   c. All members of the Nebraska Chapter of the National Association of Social Workers
   d. None of the above

10. Does the one-group pretest–posttest research design control for normal growth and development as an alternative explanation of the study results?
    a. Yes, because it provides for two different measurements of the target behavior
    b. Yes, because it provides for a measurement of gain on the target behavior
    c. No, because it is not likely to generate results that are statistically significant
    d. No, because it does not measure trends in target behavior before treatment or have another mechanism for comparing client growth with normal growth over time

ANSWERS: 1 = d; 2 = b; 3 = c; 4 = b; 5 = c; 6 = a; 7 = b; 8 = a; 9 = c; 10 = d

Chapter Glossary

Comparison group design. A research design whereby the gain of a group of treated persons is compared with the gain of a group of non-treated persons.

Content validity. The extent to which the scale employed does a good job of including all aspects of the concept begin measured, with

special reference to the definition of the concept in the study report.

Convenience sample. A sample of people who were selected for the study out of convenience.

Evidence. Information collected from sources of a scientific nature that is viewed as credible by key actors in the human service system

Face validity. The extent to which key informants agree that the measurement tool does a good job of measuring the concept intended.

Goal. The major, long-term, benefit to the client of the intervention (service, treatment, etc.).

Input. The resources that are employed in the provision of the intervention.

Intervention. The composite of the activities that are provided to clients for the purpose of achieving a given outcome. This term is used in the same way as service and treatment in this book.

Limited AB single-subject design. A single-subject research design in which a single client is measured one time on the target behavior before treatment begins and several more times during the treatment period.

Logical generalization. Generalizing study findings from the sample to the population on a logical basis, based on information showing the similarities (and the differences) between people in the sample and people in the population.

Meta-analysis. An examination of the data from a group of published studies to determine a summary of the extent of the effect of a given treatment on a given target behavior.

Model of the intervention. The conceptual framework that serves as the logical rationale for offering the intervention in light of what is known about the dynamics of the target behavior being addressed.

Objective. A measured amount of progress toward the achievement of the goal. It is the thing that will be measured with regard to client outcome.

One-group pretest–posttest design. A research design in which one group of clients is measured before and after the intervention with regard to the objective of treatment.

Outcome. The benefit to the client.

Output. The amount of service that is provided, such as 254 hours of therapy.

Personnel of the intervention. The credentials of the service providers.

Practical significance. A judgment about the extent to which the measured results are noteworthy.

Process. That which happens to the clients in order to achieve the intended outcomes (i.e., the service).

Program. An organized set of services designed to achieve a social goal, such as reduction of child abuse, improvement of mental health, prevention of suicide, and/or improvement of family relations.

Random sample. A group of people selected from a study population on a random basis.

Sampling error. The difference in your data between the results from your sample and what would be the results if you had data from the entire population.

Services. A set of professional activities designed to achieve a specific objective for a given set of clients. This term is used in the same way as intervention and treatment in this book.

Structure of the intervention. A concrete statement that shows both the form and the intensity of the services being given.

Systematic review. A review of existing evidence about a problem or an intervention that includes all meta-analyses that have been found and usually includes studies that have not been published.

Target behavior. The behavior of the client that is being treated.

Target behavior analysis. The analysis of the client's behavior that is being treated.

Target behavior definition. The definition of the client's behavior that is being treated.

Threats to internal validity. Potential causes of the client's measured growth that are independent of the intervention, such as normal growth over time.

Treatment. What is done to the client in order to achieve the objectives. This term is used in same way as service and intervention in this book.

Unit of service. The concrete measure of the amount of a service, such as the number of hours of counseling provided or the number of child abuse investigations completed.

CHAPTER APPENDIX: STATISTICAL ANALYSIS OF THE DATA FOR THE NEW HORIZONS TREATMENT PROGRAM

This part of Practice Exercise 1 will provide you with guidance on the statistical analysis of the data given in Exhibit 8.1. You will use GraphPad for this purpose. Each step in this process is give below.

Step 1: Access GraphPad QuickCalcs by clicking on the following link:

https://www.graphpad.com/quickcalcs/

Step 2: Select *continuous data*.

Then click *continue*.

Step 3: Select *t test* to compare the two means.

Then click *continue*.

Step 4: You will be at the screen that says **Choose data entry format.**

1. Select *enter up to 50 rows*.

2. Enter your data in the two rows with the pretest score for the first client in the first column first row and the posttest score for the first client in the second column first row. Then, do the same for the second client and so forth. In other words, enter the data as you see it in Exhibit 8.1.

3. Select *paired t test*

4. Click *Calculate now*

Step 5: Review the results.

- One of the options is ***p* value and statistical significance**. This will give you the value of *p*.

- Another option is **intermediate values used in calculations**. Under this option, you will see the value of *t*.

- The means for Group 1 (pretest score) and Group 2 (posttest score) will be given on this page.

You should report the mean pretest score, the mean posttest score, the value of *t*, and the value of *p*. The values of *t* and *p* should be put in parenthesis following the reporting of the mean pretest and posttest scores.