Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

PART V



The Outcome Stage

Chapter 9, the only chapter in this section, introduces several types of evaluations used during the outcome stage of a program or practice intervention. The overall purpose of outcome evaluations is to determine whether the intervention resulted in positive outcomes for the clients receiving it. Client outcomes are also discussed in the context of whether there is enough evidence of a causal relationship existing between an intervention and the client outcomes.

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

CHAPTER 9

Is the Intervention Effective?

A central question to ask at this stage is, how do we know that our intervention actually helps clients?

This chapter addresses this outcome question. To many people, the evaluation of programs and practice interventions mostly means evaluating whether the intervention resulted in positive outcomes for clients. Outcome questions have a familiar ring to them: Did the clients reach their goals? Did they make significant progress? Is there evidence that the program made a difference in their functioning?

Although outcome studies are presented as only one of three general types of evaluations in this text (i.e., planning studies, implementation studies, and outcome studies), it must be said that they are ultimately the most important type. If progress on clients' goals is not evident after the completion of a program or practice intervention, a provider will have real difficulty justifying its continuation. This is the case no matter how well the program was planned, how qualified the staff members are who provide services, and even if the program's quality is superior. This is also true for practice interventions. If practice interventions do not lead to substantial gains for clients in the intended areas, then the intervention needs major modifications or should be discontinued and replaced with another approach.

This chapter describes two critical ingredients in outcome evaluations: a clearly defined intervention and measurable client outcomes. Several types of evaluation designs are used in outcome evaluations and are presented in the following discussion. The advantages and limitations of each design and what each design can claim are described, with examples of outcome evaluations given to help illustrate the relevant concepts. (See Figure 9.1.)

Based on the logic model, outcomes for clients are directly linked back to several activities completed in the planning stage. During that stage, clients' problems, needs, and the causes of their problems were identified and assessed. Then goals and measurable objectives were crafted to measure what the clients' situations would look like when these problems were resolved. Developing an intervention came next, created specifically to help clients reach these goals. Now in the outcome stage after the intervention is implemented, we focus on measuring the clients' goals to determine if they have been achieved.

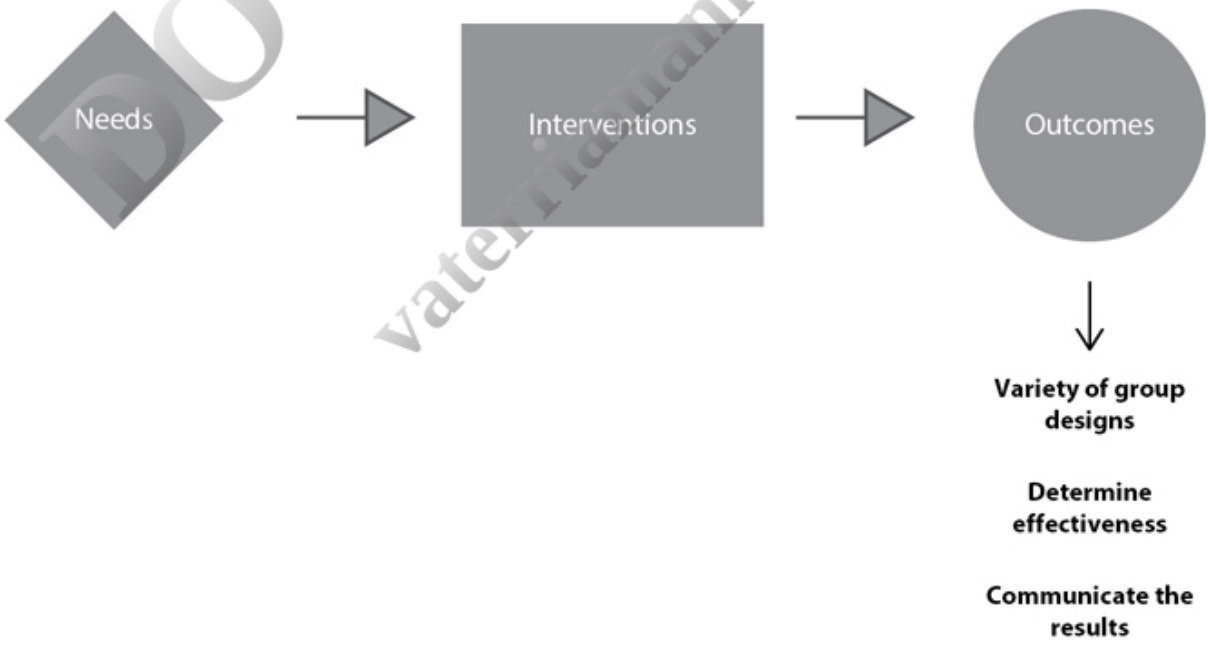


Figure 9.1 Evaluation of the intervention.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

THE NATURE OF OUTCOMES

Program and practice outcomes can focus on many things. Outcome measures may be identified for staff members' performance, material resources, recruiting of more clients, or other concerns. Ultimately, however, the outcomes for clients are the focus of human service outcome evaluations. Outcomes are the entities that help us determine whether our interventions were effective in helping our clients. Chapter 7 explained how to identify, describe, and measure goals as outcomes. This chapter continues that discussion with planning how to implement outcome evaluations.

Outcomes Are Not Outputs

The outcomes of clients, which are the ultimate interest of a program or practice evaluation, have several distinct characteristics. Unfortunately, many agencies focus almost exclusively on outputs rather than outcomes in their performance reports made available to the public. For example, an agency report may focus on the number of staff hours of counseling provided, the number of clients who use the services, or the costs of a specific program.

Excerpts From an Annual Report of a County Government

The following statistics are from the annual report of the Cabarrus County (2007) government:

- 322,000 people visited county-operated parks.
- 14,706 people participated in athletics, competitions, and other recreational events held in county parks.
- 7,801 calls and personal contacts came into the Veterans' Services office.
- 976,939 miles were driven by county vans.
- 81,847 rides were provided to disabled and/or elderly residents by the county transportation service.
- 35,335 individuals visited the Department of Social Services for services or assistance.
- 5,085 families in a crisis received assistance through the Department of Social Services.
- 3,885 children were protected from abuse, neglect, or dependency through provision of child protective services.

Note that all these indicators are outputs, even though some may be misrepresented as outcomes. The last output in the example, "3,885 children were protected from abuse, neglect, or dependency," is incorrectly stated and should be rephrased "3,885 children were provided with child protective services." Based on a follow-up inquiry, the agency reported that it did not have any evidence to conclude that the children were protected. If that information had been available, it would have been an outcome.

Outcomes Ultimately Focus on the Client Not Staff Interventions

A frequent problem for many who are new at crafting outcomes for clients is confusing client outcomes with outcomes for the staff that are helping them. For example, consider the statement "The client will be referred to X agency for assistance whenever the program cannot help her." Note that this statement refers more to what a staff member will do (referral activities) than to what the client does. Many stakeholders may think that the goals of the staff members are outcomes when it is more accurate to say that staff activities are the interventions that assist clients in achieving their goals or outcomes.

Articulating and documenting outcomes for staff activities can be helpful, especially because client outcomes usually depend on the success of staff in reaching their goals. Outcomes for staff members can be viewed as an appropriate evaluation focus at the implementation stage, such as when we ask whether staff members use evidenced-based practice interventions or provide quality services. Yet, these outcomes are not client outcomes.

Quiz: Are These Outcomes for Clients or Staff Members?

Try to pick out which of the following goals are for clients and which are for the staff members.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

1. The clients will be transported to the grocery store in the agency van every Monday.
2. The clients will purchase groceries for the evening meal.
3. The client will have access to a planned meal written out on a piece of paper.
4. The client will pay for the groceries using the money available for shopping.
5. The clients will have enough money to buy the groceries.

The correct answers to the outcomes in the preceding quiz are that #2 and #4 focus on clients, and the others focus on staff members. The fifth outcome could be intended either for the clients or for staff, depending on who has control over the funds for shopping; unfortunately, it is presented in an ambiguous way. Overall, the five examples show how dependent client outcomes are on staff outcomes.

Outcomes Are Logically Linked to the Interventions

As has been mentioned several times before, it is important that the outcome measures be directly linked to the intervention based on the logic model. This should be the case both conceptually and practically. Often, having an easily and well understood intervention helps in crafting realistic outcome measures. An example of this important issue is illustrated with the community organization, First Things First, mentioned in Chapter 7. This community organization wanted to reduce the number of divorces in their community; this was their initial outcome measure. Their intervention was to offer a premarital weekend retreat format at several different houses of worship to prepare recently married couples with help in staying together. The retreat curriculum included five major components: communication between the couple, having realistic expectations of a partner, finances, health issues, and sexuality and intimacy. The sponsoring organization recruited speakers to lead the retreat who had expertise in each topical area. The evaluator devised a pretest–posttest measure of outcomes for the retreat based on what could be realistically expected from participation in each retreat component. Specific questions were devised based on the specific objectives of each speaker. For example, the topic of having realistic expectations of your partner included questions such as, “Do you know what your spouse expects of you?” “Do you and your partner know how to resolve conflicts when they arise?” and “Do you know how to meet your partner’s needs?” The community organization hoped that the program would have an impact, however small, on the divorce rate. However, they needed help in considering a less global outcome, such as that the program participants, not the larger community, would be staying together, particularly during the challenging years of early marriage. The organization came to realize that the community’s divorce rate had multiple causes, many of which were well beyond its influence and ability to change.

Outcomes Can Be Determined at Different Points in Time

The point in time to measure a client outcome is an important factor to consider. For example, outcomes can be measured when a client completes a program or at some future time. Outcomes can be short-term (e.g., when a client completes the program), intermediate (e.g., three months later), or long-term (e.g., twelve months later). The time selected and the specific outcome expected depend on the amount of time realistically needed for the program to have such an impact.

Example of Outcomes at Different Points in Time

A crisis center for transient youths decided to use short-, intermediate-, and long-term outcomes for the youths they sheltered. Their outcome measures were as follows:

Short term: reduction in runaway behavior, physically aggressive behavior, and serious physically aggressive threats while residing at the crisis center.

Intermediate term: maintaining residence in a safe, least restrictive environment for two months after discharge and increased family access to community services.

Long term: maintain residence in a safe, least restrictive environment for six months after discharge, increased school attendance, and reduction in involvement with the juvenile justice system.

Unfortunately, most evaluations are set up to measure client outcomes only when clients complete the program or otherwise terminate from the agency. This is an important time to measure outcomes, as it provides a measure of how much change occurred while the client was in the program. Yet this single short-term outcome measure usually reveals little, if anything, about the long-term effects of a program. It could even be considered a client output, not a client outcome. Inpatient programs for drug-addicted clients, for example, are often criticized for not providing a

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

client outcome. Inpatient programs for drug addicted clients, for example, are often criticized for not providing a long-lasting impact on recovery. Clients who leave a program may be drug free, but the true test of their success is likely determined over time, once they return to their community and reconnect with their social supports, both previous and new. Unfortunately, many substance abuse programs do not conduct a follow-up evaluation to determine whether their services led to long-term recovery or a return to addiction and a lifestyle that supports addiction.

VARIED WAYS TO MEASURE OUTCOMES

Martin and Kettner (1996) identify what they perceive as four kinds of outcome or performance measures: numeric counts, standardized measures, scales of level of client functioning, and client satisfaction indicators. *Numeric counts* are generally used to measure client conditions, status, and behaviors. Examples include attendance at school, frequency of drinking alcohol, and numbers of violent and other antisocial behaviors. *Standardized measures* usually measure such things as client feelings, attitudes, and perceptions. Examples of such scales include scales on self-esteem, clinical anxiety, hope, and happiness. *Level-of-functioning scales* measure client and family functioning. Examples include scales of adaptive behavior, employment behavior, wellness, and social integration. Martin and Kettner (1996) also refer to client satisfaction as an outcome; however, they view client perceptions of the services received as only an intermediate outcome, not a final one.

Example: Child Welfare Outcomes Documented by Children's Bureau

The U.S. Department of Health and Human Services' Children's Bureau (2017) includes child welfare outcomes as part of its annual report to Congress. The report also provides data on the states' performance in meeting the needs of children and families who are in contact with the child welfare system. Federal data-reporting systems are used to gather data on seven outcomes: reduced recurrence of child abuse and neglect, reduced child abuse and neglect in foster care, increased permanency for children in foster care, reduced time in foster care before reunification, reduced time in foster care before adoption, increased placement stability, and reduced placement of young children in group homes and institutions. The executive report and additional information can be found on https://www.acf.hhs.gov/cb/resource/cwo-10-14

Practical Measures of Outcomes

Ideally, virtually every agency has identified client outcomes that they are helping clients accomplish. Because the outcomes are the ultimate reason for the agency's existence, it makes sense to always keep them in the forefront of the practitioners' work, their conversations with clients, their referrals, and their education of the public. One agency, for example, does just that. Community Choices Inc. (2006) helps pregnant women addicted to alcohol and/or drugs by offering group treatment. The outcomes of this group treatment intervention are evident in much of its material, including the brochure handed out to prospective clients and referral agencies. Outcomes that they highlight for clients are becoming drug free; maintaining or regaining custody of children; having a healthy, drug-free baby; becoming self-sufficient; finding a job; finding permanent housing; improving parenting skills; and feeling better about themselves. The open and direct communication about the outcomes has many benefits, including helping the women identify what they want or need, helping them select this agency because of its desired outcomes, and communicating the agency's purpose and intentions to the public.

Outcome Indicators Used for Clients in a Domestic Violence Agency

The following client outcome indicators are from White (2005):

- Lower homicide rate (chosen by the county government)
- Perceived degree of safety
- Knowledge of a safety plan that could be used in an emergency
- Acknowledgment of a broad definition of domestic violence
- Favorable responses to client satisfaction questions
- Referral contact rate (when client follows through and contacts another agency he or she is referred to)

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- Recidivism rate (returning to agency a second time or more)

Litzelfelner (2001) gives another example of a need for practical outcomes when the overall goal of a program is not measurable. She points out that the goal of child protective services to protect children from abuse is not usually measurable. Abuse and re-abuse of children are often undiscovered or unreported. She offers what she views as practical alternative measures in a set of four general areas related to child abuse: child safety, child permanency, preserving families, and child health and development. For each area, she identifies several measurable objectives and activities to reach the objectives. For example, for child safety, she suggests two objectives as examples: (a) 90 percent of children served will not re-enter the system within two years following their initial discharge (based on reports of the agency's home visit and case management services), and (b) each quarter, 85 percent of caregivers participating in the parent education program will demonstrate three nonphysical behavior management techniques during the agency's home visits. As an exercise, please identify the weaknesses in this approach to measuring child abuse. Then develop an alternative approach that could be another way to measure child abuse.

Outcomes can also be measured by qualitative methods. One student who was working with a group of emotionally disturbed children in a public school decided to use journaling to measure progress on their personal goals related to self-esteem (McLean, 2007). She encouraged her group members to keep an ongoing journal of their impressions discussed in the group. They were given ten minutes at the end of each session to write about their impressions. During most sessions, members were asked to journal on the topics that had been discussed such as, What are the benefits of self-esteem? or how can mistakes be opportunities for growth? Journaling during the first and last sessions was used to evaluate their progress in working on their personal goals. During the first session, members were asked to journal on their goals and expectations for the group. The topic for them to focus on was, what do I want to get out of this group? The final session was used to journal on the topic: How has the group changed my self-esteem? Quantitative ratings were also used to evaluate their progress on goals during the last session.

CRITERIA FOR CHOOSING OUTCOME MEASURES

What determines whether an outcome measure is adequate or acceptable? Martin and Kettner (1996) offer seven criteria: the outcome measure's usefulness, validity, reliability, precision, feasibility, cost, and unit costs.

Usefulness

First, the measure needs to be useful and relevant to stakeholders. This is a common-sense criterion that raises an important question: How will the outcome measure be useful? Will it offer helpful information to stakeholders to support the program or assist them in improving, expanding, or even reducing it?

Validity

Is the outcome measure valid? Validity is one of the two psychometric standards for determining whether a measure is satisfactory. Validity refers to whether the outcome measure reflects what it is supposed to measure and whether it does so accurately. Accuracy is the key issue. The more valid that a measure is, the better it is. For example, an outcome measure for father involvement should include several different functions or activities, not just one. Some will involve direct contact with their child (e.g., how many direct contact hours in an average week, quality of the specific activities), and other functions may involve no contact but direct responsibilities (e.g., attending a parent–teacher conference, arranging a dental appointment for the child). From this example, it is clear how complex outcomes like father involvement are and the enormous challenges they pose for creating accurate measures. Father involvement requires at the very least a certain amount of contact. The quality of that contact, whether direct contact or indirect responsibilities, is also important to consider.

Reliability

The third criterion for choosing an outcome measure is reliability. Reliability is a second key psychometric standard in determining whether a measure is satisfactory. Reliability refers to the internal consistency of the measure. It asks two questions: whether the measure of the outcome consistently measures it from one time to another and whether the measure is consistently applied among different people (e.g., do different clients consistently understand the measure in the same way?). Reliability is weak if the questions asked in a measurement instrument are ambiguous, with words likely

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

to be understood and interpreted differently at different times or by different respondents. For example, satisfaction with a program can be interpreted in many ways.

Precision

Precision is the fourth criterion. Precision refers to how well the outcome measure captures the incremental changes in clients' condition, either quantitatively or qualitatively. Quantitative measures are at the interval, ordinal, or nominal levels. Qualitative measures are at the prenominal level. For example, incremental quantitative measures are the number of times truant children attend school (interval level) and their average letter grade (ordinal). An example of an incremental qualitative measure is the subtle variations in how a parent uses nonviolent strategies in disciplining their children.

Feasibility

A fifth criterion in deciding on outcome measures is their feasibility. Is the measure viewed as practical and acceptable to all stakeholders? Some may be opposed to a measure because it is at odds with their cultural values or ethics. For example, a measure of assertiveness needs to consider the ways that various cultural groups view confrontation with authority. Feasibility may also be a problem if it is likely to demand a lot of the attention and time of staff members in explaining or administering it.

Cost

The cost of an outcome measure is another criterion. Does the measure require some major start-up or maintenance costs? Another possible cost factor is the purchase of any standardized scales owned by other people or organizations.

Unit Cost Reporting

Unit cost reporting is the final criterion. Can the measure be described as a cost per unit of outcome? For example, the cost of a client graduating from a program could be used as a cost per unit of outcome. Numeric counts lend themselves to this type of cost more than other measures, such as standardized scales.

OUTCOMES AND PROGRAM COSTS

Sometimes outcome studies are conducted that focus directly on costs. Some refer to these evaluations as cost–benefit or cost-effectiveness evaluations. Cost-related evaluations are increasingly important to consider, given the diminishing funds for human service programs and the increasing call for financial accountability. One ethical dilemma with cost–benefit studies is that it is too difficult to directly translate the benefits of a program, such as reduced child abuse or treating and managing severe mental illness, into financial savings. Although immediate financial savings are evident in such things as fewer health care visits and avoiding legal costs in a court case, the benefits of protecting a child from harm go well beyond material factors and may have implications for the life of the child and for the child's offspring. In brief, although fiscal matters have obvious value, their value is limited and should be subordinate to the value of basic human needs.

Cost-effectiveness evaluations, like cost–benefit studies, may have more utility in human service programs than cost-benefit studies. In this case, the client outcomes resulting from a program can be analyzed according to cost. The cost, for example, of obtaining adoptions, on average, for special needs children could be considered in terms of costs. The costs could then be compared to the costs of such an outcome in the previous year, or two different programs could be compared based on the costs incurred in completing a successful adoption.

It is complicated but possible to calculate costs in a human services program. For example, one could categorize costs into direct (e.g., staff salary) and indirect (e.g., secretarial salary, telephone bills). Costs could also be divided into fixed (e.g., building maintenance) and variable (e.g., salaries), and recurring (e.g., rent) and nonrecurring (e.g., equipment). Determining the costs of providing services to an average client must take all these costs into account, which sometimes becomes a real challenge to figure out. For example, in the 1970s, with the initiation of deinstitutionalization of people with mental retardation, some legislators argued in favor of continuing institutional services rather than establishing a community service system because they viewed institutional services as having lower per capita costs. This debate led to the discovery that it was easier to calculate the per capita costs of community-based services because they were discrete

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

(e.g., residential, vocational, recreational). In contrast, the cost of institutional services, encumbered by extensive capital costs such as maintenance of buildings and grounds, was more difficult to calculate accurately.

EVIDENCE-BASED INTERVENTIONS

Once the outcome measures have been carefully selected for an evaluation, another challenging task is to demonstrate that the measures are logically and empirically linked to the intervention that will be used. Interventions that are empirically linked to important outcome measures are considered evidence-based interventions. This raises a critical point about evidence-based interventions—they are interventions that have, to some degree, a demonstrated a causal relationship to specific client outcomes.

Evidence-based interventions were defined in Chapter 2 as integrating individual practice expertise, the best-available external evidence from practice research, and the values and expectations of clients (Gambrill, 1999, p. 346). This definition suggests that evidence-based should not be either defined simply by one criterion or based solely on one source. Instead, the interventions should have multiple factors in common. One such factor, a necessary one, is that evidence-based interventions have the best evidence available that they are effective in helping clients, and strong evidence is available to support the claim in the measures of client outcomes. Thus, we can say that evidence-based interventions should be grounded one way or another in outcome evaluations.

O'Hare (2015) describes four important tools for determining evidence-based practices: they use critical-thinking skills, obtain scientifically credible evidence, seek multiple sources of evidence whenever possible, and seek diversity-sensitive evidence.

Critical Thinking Skills

Critical-thinking skills are used to determine whether an intervention is evidence-based. Critical thinkers are natural skeptics about how well an evaluation is conducted, whether someone else's evaluation or one's own. Critical thinkers are also skeptical about whether their interventions actually work. As mentioned in Chapter 1, Gibbs and Gambrill (1996) identify several types of problems that evaluators experience when they fail to be critical thinkers, including the selection of weak or inappropriate interventions.

Scientifically Credible Evidence

Evidence-based interventions must have documentation that is scientifically credible. This means, in part, that there is scientific evidence of a causal relationship between the intervention and clients' outcomes. Such evidence for programs is most likely to come from group designs, especially experimental and quasi-experimental designs, which are known to be superior to other evaluation strategies (Cook & Campbell, 1979). Evidence for practice interventions is most likely to come from a version of a single-system design, especially a design that controls for as many extraneous influences as possible.

Scientifically credible evidence and critical thinking share many common values. Both approaches involve skepticism. Both carefully distinguish between questions of fact and questions of value. Both use caution when inferring what may have caused improvement in the client outcomes and when to generalize to other client groups, and both emphasize the need for measurement of clients' needs, interventions, and outcomes. Finally, both perspectives are likely to continually ask, does this intervention work?

Seeking Multiple Sources of Evidence

Triangulation, a process of using multiple methods to measure something, is also important to evidence-based interventions. Evaluators who obtain two or more measures of an outcome for comparison purposes triangulate the findings. The process also adds confidence that the two measures are valid because of their similarity. The more sources available that document that the intervention has a causal relationship to client outcomes, the better. An example is to find out the frequency of a nonresidential father's contact with his children in an average week by seeking a report from two or more sources (e.g., the father, the children's mother, the father's current partner).

Criterion-related validity is another way to triangulate the findings of an evaluation. Criterion-related validity is a means of determining whether the scores of an outcome measure are valid by correlating the scores with those of another measure of the same construct (Babbie, 2016). The other measures of the same construct are external criteria. For example, an external criterion could be a behavioral indicator that correlates with a verbal response of a client to an interview question. If a short-term outcome measure of an evaluation is regular attendance at a drug treatment program,

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

interview question. If a short-term outcome measure of an evaluation is regular attendance at a drug treatment program, the evaluator could compare the client's report of his or her attendance with a report of attendance obtained from the drug treatment agency. In another example, the outcome measure could be a depression scale developed by an agency, and the scores could be correlated with the scores of the same clients using a well-established measure like the Beck Depression Scale (Beck, Steer, & Brown, 1996).

Evidence Is Diversity Sensitive

If there is evidence that a program or practice intervention is effective with one group of people, it is extremely important not to automatically conclude that it will be effective with other groups. This is a fallacy that has resulted from the misuse of previous findings, especially evidence that if a specific approach helps white, middle-class men, then it can be generalized to being effective in helping people of other races and social classes as well as white women. Evidence must be obtained skeptically for each group before arrival at such a conclusion. The issue of generalizing the results of a group design to other people is referred to as the extent that the group design has external validity.

Thyer and Meyers (2003) recommend that the most realistic way to generalize the results of a group design to other client groups is to use the process of replication. Replication involves repeating a group design with different client groups and at various times. If the results of a group design reveal that an intervention works for one client group, the intervention could be tried with a different client group that has similar characteristics. If the same result occurs, a claim that the intervention is effective would be somewhat strengthened. This could be followed by further replication, such as with similar clients in another agency or by other evaluators with similar clients in other geographic areas. Also, if the results of the group design indicate that an intervention is effective with a group of men, the intervention could be tried with a group of women. Or if an intervention works for young-adult clients, it could be tested with middle-aged adults. Repeated replication of a group design can be extremely valuable in exploring the extent to which an intervention can be generalized to others. Such replication can also be used to determine the limitations of an intervention and the characteristics of clients with whom it may not work and with client circumstances that are contraindicated.

Example of Setting Standards for Evidence-Based Practice

Harding and Higginson (2003), in a systematic review of articles from relevant professional journals, found twenty-two interventions for helping cancer and palliative care patients and their caregivers. They analyzed the twenty-two interventions based on the type of patient and caregiver populations, the evaluation design used, and findings. They graded the evaluations as follows:

Grade I (strong evidence): randomized control trials
Grade II (fairly strong evidence): group designs with a comparison group
Grade III (weaker evidence): retrospective or observational studies
Grade IV (weak evidence): cross-sectional studies, Delphi exercises, consensus of experts

They concluded that there was a lack of outcome evaluation designs used to evaluate the effectiveness of the interventions. Instead, most articles contributed more to understanding feasibility and acceptability issues than to effectiveness. The authors encouraged programs to conduct more experimental and quasi-experimental designs to determine whether interventions effectively helped cancer patients and their caregivers.

Decisions on the sources of evidence for social work evaluations are important. Evidence can come from the reports and experiences of administrators, staff members, and faculty; clients' records; observations that document evidence; interviews and questionnaires that document perceptions of evidence; case examples of clients; other types of recordings; and, most important, scientific studies. Sources will vary in relevance and importance depending on their purposes. Evidence-based interventions that are described next are concerned with interventions in which there is scientifically credible evidence of a causal relationship between the intervention and clients' outcomes.

DETERMINING A CAUSAL RELATIONSHIP

How do we know whether the program was the cause of the progress in client outcomes? The answer to this question is complicated. Three conditions must be met before a causal relationship can be claimed. *The first condition* is that the program (or practice) intervention must precede clients' improved outcomes. This condition makes logical sense: a provider could not ethically claim responsibility for clients' progress if the progress occurred prior to the introduction of the program.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The second condition needed to claim a causal relationship is that an association is found between the introduction of the intervention and the client's improvement in his or her outcome measures. An association means that the client's improvement becomes evident after the intervention is implemented. To be able to determine whether improvement occurs in the outcome measure during the period of implementation, we need to obtain a measure of the outcome both before (i.e., pretest measure, or baseline measure) and after (i.e., posttest measure) the intervention is introduced.

An association between the intervention and clients' outcomes can be determined through one of two methods: a test of statistical significance or an established standard of clinical significance. *Statistical significance* can be calculated by applying a statistical test, which determines whether the probability of an association is high enough to claim significance. Usually, a probability of 95 percent is expected, with the likelihood of error at 5 percent.

Clinical significance is defined by clinical criteria established by a professional group or agency. In this case, the amount of improvement in client outcomes is satisfactory when it meets the established clinical criteria. Clinical significance may be preferred over statistical significance in many circumstances, as it is based on clinical judgment, not mathematical probability. However, statistical significance is usually expected in summative evaluations when evidence of a program's impact is expected to be based on scientific principles.

To illustrate these points, consider the example of a recovery group of clients with alcohol abuse problems. The group has as its desired outcome reduced alcohol consumption. Assume that the recovery group is a weekly treatment group that emphasizes cognitive behavioral therapy and a modified version of the twelve-step program of Alcoholics Anonymous. The blood-alcohol level of group members is the measure used to determine whether there is improvement. If statistical significance is the standard to be used, then there must be a statistically significant difference between pretest and posttest scores of the average blood-alcohol level of group members. A statistical test, such as a *t*-test, can tell us whether the two measures were significantly different according to probability theory. The various types of *t*-tests and other statistical tests that can be used to determine statistical significance are discussed in the next chapter.

Let's determine whether group members improved significantly according to clinical criteria. Let's assume that the agency sponsoring the group intervention has a clinical standard based on the definition of satisfactory improvement reported in the professional literature. The agency has found studies that have reported that an intervention is not fully successful until a person totally withdraws from alcohol. Therefore, the agency decided that clinical significance would not be fully achieved until the group members totally withdrew from drinking; in other words, the members would not show enough improvement if they fell short of this high standard.

The third condition for claiming the intervention is the cause of the client's improvements is the most difficult to achieve. This condition is the claim that the improvement occurring in the clients' outcome measure is due, at least in part, to the program intervention. We need to consider that many other factors, such as a change in family circumstances or joining an Alcoholics Anonymous support group, can have an influence on the outcome as well. Usually, an intervention is not expected to be totally responsible for clients' improvement. If we are interested in determining how much of the improvement is due to the intervention, we would have to use an experimental design. A less dependable method is a multiple regression test, which is discussed in the next chapter.

Using the example of the recovery group mentioned earlier in this section, an experimental design could be used that includes an intervention group (i.e., clients randomly selected and assigned to the recovery group) and a control group (i.e., another group randomly selected and assigned to receive traditional individual services). If the outcome scores of the intervention group members (their average blood-alcohol level) positively changed significantly from pretest to posttest and the outcome measure of the control group members changed less or not at all, then it can be claimed that the intervention provided to clients had a causal influence. The amount of influence that it had on the recovery group members is determined in two steps. First, the change in the recovery group members' outcome score from the pretest to posttest is calculated as the amount of change that occurred in that group. This change would then need to be reduced by the amount of change, if any, that occurred in the outcome measures for the control group.

Quasi-experimental and pre-experimental designs can also be used to explore whether a social work intervention has a causal influence on changes in client outcomes. In such cases, only tentative conclusions can be made about the intervention's influence or impact. However, in most cases, an experimental design cannot be realistically used for ethical reasons, as field-based evaluations are not amenable to the controls that an experimental study requires. Quasi-experimental, pre-experimental, and experimental group designs are all reviewed next for program interventions.



GROUP DESIGNS FOR PROGRAMS

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Several group research designs are available for conducting outcome evaluations of programs. These designs are known as group designs because they measure the impact of an intervention on a group of clients, not on one client. Two overall categories of designs are possible in an evaluation study: cross-sectional and longitudinal. A cross-sectional design collects data from participants at one time only, and a longitudinal design collects data from research participants at two or more times. Longitudinal designs provide the most powerful evidence of a causal relationship between a program intervention and client outcomes.

Internal and external validity are important standards to use when evaluating the strength of a group design. Internal validity addresses whether the intervention and no other factors are responsible for improvement in the client outcome variable. Several extraneous variables could be responsible for improved outcome in addition to or instead of the intervention. For internal validity to be strong, for example, factors independent of the intervention present in clients' lives (e.g., receiving services from another professional agency, joining a self-help group) could be responsible for improvement in the outcome measure in addition to or instead of the intervention.

External validity addresses the issue of generalizing the results of a group design to other people. In this case, if the participants in the group design were randomly selected from a large population (e.g., a client group of one hundred or more), generalizations to the larger group are possible. However, if participants selected for the group design were drawn from a smaller population or were selected using nonprobability sampling, generalizations are not advisable.

Experimental designs, quasi-experimental designs, and some pre-experimental designs are longitudinal designs that can be used to explore a causal relationship between a program and its impact on client outcomes. Several group designs are described in Campbell and Stanley (1973). Six designs are highlighted here because they are among the most practical ones for outcome evaluations of programs in human service agencies:

Pre-Experimental Designs

1. One-group posttest-only design
2. One-group pretest–posttest design

Quasi-Experimental Designs

3. One-group pretest–posttest design with a comparison group
4. Time-series design

Experimental Designs

5. One-group pretest–posttest design with a control group
6. Posttest-only control group design

All these group designs can be used to explore the impact of a program on a group of clients, whether the intervention focuses on individual, family, small group, or community systems. All these designs require that there be a clearly defined program intervention that can be replicated and at least one client outcome measure. These are the central elements of any group design. Other elements are included in some of these designs and not others.

Concepts Relevant to Group Designs

a. *Clients' outcomes*: outcome measures of what clients are to achieve
b. *Program intervention*: an intervention defined clearly enough to be replicated
c. *Pretest or baseline measure*: a measure of an outcome for clients prior to the intervention
d. *Posttest measure*: a measure of an outcome for clients after implementation of the intervention
e. *Intervention group*: group of clients receiving the program intervention.
f. *Comparison group*: a group of clients in a group design similar but not identical to the group receiving the program intervention; this group does not receive the program intervention
g. *Control group*: a group of clients in a group design considered identical to the group receiving the program intervention, based on either random selection and assignment or matching; this group does not receive the program intervention

Let's look at each of the six group designs and the concepts incorporated into each one. Diagrams of each design are also included in Table 9.1.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Table 9.1. Pre-Experimental, Quasi-Experimental, and Experimental Designs

DIAGRAM	DESCRIPTION
$X\ O_1$	1. *One-group posttest-only design*: intervention provided to intervention group followed by a posttest measure of client outcome
$O_1\ X\ O_2$	2. *One-group pretest/posttest design*: intervention provided to intervention group preceded by a pretest measure and followed by a posttest measure of client outcome
$O_1\ X\ O_2$ $O_3\quad O_4$	3. *Pretest/posttest design with a comparison group*: intervention provided to the intervention group preceded by a pretest measure and followed by a posttest measure of client outcome; also, a comparison group that does not receive the intervention preceded by a pretest measure and followed by a posttest measure of client outcome
$O_1\ O_2\ O_3\ X$ $O_4\ O_5\ O_6$	4. *Time-series design*: multiple measures of client outcome before and after the intervention provided to the intervention group
$R–\ O_1\ X\ O_2$ $R–\ O_3\quad O_4$	5. *Pretest/posttest design with a control group*: intervention provided to the intervention group preceded by a pretest measure and followed by a posttest measure of client outcome; also, a control group that does not receive the intervention preceded by a pretest measure and followed by a posttest measure of client outcome. The R indicates that the two groups were randomly selected from a larger group and randomly assigned to either the intervention or control group. Therefore, they can be viewed as identical groups of people.
$R–\ X\ O_1$ $R–\quad O_2$	6. *Posttest-only control group design*: intervention provided to the intervention group followed by a posttest measure of client outcome; also, a control group that does not receive the intervention followed by a posttest measure of client outcome. The R indicates that the two groups were randomly selected from a larger group and randomly assigned to either the intervention or control group. Thus they are viewed as identical groups of people.

Notes: X symbolizes the program intervention, O represents a measure of the client outcome, and R indicates that the two groups were randomly selected from a larger group and randomly assigned to either the intervention or control group.

One-Group Posttest-Only Design

Pre-experimental designs are often important to consider initially because they are easy to implement and take less time than stronger designs. This design involves collecting data at only one point in time after implementation of the intervention to the intervention group. Thus, this is a cross-sectional rather than a longitudinal design. During time 1 the intervention is introduced, and during time 2 the posttest measure of client outcome occurs.

This design is a good one for a beginning exploration of program impact, and it can simply answer the question of whether it is possible that the program had an impact. A one-group posttest-only design has several weaknesses. Most important, there is no pretest measure to determine whether any improvement occurred in the outcome variable during the time of the intervention. Furthermore, none of the extraneous influences (e.g., environmental changes) are controlled for. An important conclusion derived from this design if the client outcome measure is not positive or favorable is that the intervention does not work.

One-Group Pretest–Posttest Design

The pretest–posttest design builds on the previous design by adding a pretest measure before the initiation of the intervention. This design, like the first one, is pre-experimental. However, it is longitudinal because data are collected from the clients at two different times—before and after the introduction of an intervention. The design is helpful in revealing how much improvement occurred between the pretest and the posttest. During time 1, the pretest score of the outcome variable is measured, during time 2 the intervention is introduced, and during time 3 the posttest score of the outcome variable is measured.

An example of a pretest–posttest design was described earlier in this chapter. The community organization that sponsored the premarital weekend retreat to reduce the number of divorces in the community was essentially interested in determining whether participants made progress between the time that the retreat began and ended. The pretest–posttest design served that purpose and helped them further refine the curriculum they offered.

With this design, we can only determine whether improvement occurred during the time that the program was introduced, not whether the program was necessarily responsible for any change in the client outcome measure. The most important weakness of this design is that it does not have a comparison or control group to control for extraneous influences.

Example of an Evaluation with a One-Group Pretest–Posttest Design

The purpose of an evaluation conducted by Lynch (2007) was to explore the impact of an independent-living skills-

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

The purpose of an evaluation conducted by Lynch (2007) was to explore the impact of an independent-living skills-training program on the resilience, social support, and life skills of sixteen ethnically diverse foster-care teenagers.

The teenagers' scores on standardized measures of resilience, social support, and life skills were compared before and after they participated in the training program. The improvements in scores from pretest to posttest were found to be statistically significant for social support but not for resilience and life skills.

Pretest–Posttest Design with a Comparison Group

The pretest–posttest design with a comparison group builds on the two previous designs. It is longitudinal in nature, as data are collected of the client outcome measure both as a pretest and as a posttest. However, another group similar to the first group is organized as a comparison group that does not receive the intervention. This design compares the progress of both groups during the intervention period. As Table 9.1 indicates, the design is diagrammed with two different groups: one that receives the program intervention and the other that does not. At time 1 a pretest score of the outcome variable is measured for both groups, and during time 2 the intervention is introduced for the group assigned to receive it and nothing happens for the comparison group. The comparison group could receive services that are usually available as an alternative to nothing. At time 3, a posttest score of the outcome variable is measured for both groups.

Example of Pretest–Posttest Comparison Group Design

Engagement services during the intake period were introduced to female clients in a substance abuse treatment program as a motivational strategy to draw them into treatment (Comfort, Loverro, & Kaltenbach, 2000). An evaluation was conducted to determine whether the engagement services resulted in an increased use of treatment services. The evaluation used a pretest–posttest design with a comparison group. The engagement services involved an engagement specialist offering van transportation to the agency and child care during intake interviews and making telephone calls to remind clients of intake appointments. Although transportation and child care were always offered to the group, members infrequently used child care. Nevertheless, the results revealed that clients receiving engagement services were more likely than those in the comparison group to use treatment services.

A comparison group can be drawn from a larger pool of clients, such as those on a waiting list or clients who are receiving traditional services during the time of the study. This type of design can reveal with more confidence than the previous ones whether the program had a causal influence on client outcomes. The most important weakness of the design, however, is that the clients selected to be in the program and those selected for the comparison groups are likely to have many differences that the design does not control for. Such differences could have a considerable influence over the measures of client outcomes. This design also creates a potential ethical problem that needs to be addressed, as one could convincingly argue that the clients in the comparison group should receive the program intervention as well. They may, for example, have as great a need for the intervention as does the group receiving the intervention.

A comparison group should be as similar as possible to the intervention group, particularly with variables or characteristics important to a study. For example, both groups should have similar demographic characteristics (e.g., age, gender, race, ethnicity), and both groups should be similar in terms of the important variables of the evaluation, such as their mental illness diagnosis or the problems that the program addresses. Comparison group studies should report on these characteristics and indicate when the two groups are significantly different according to specific characteristics. This reporting helps the reader consider whether any of the differences in characteristics could have influenced their respective client outcome scores.

Time-Series Design

A time-series design is another type of quasi-experimental design. This design has many of the features of the previous designs and is also different from the other designs in that it has several pretest and posttest outcome measures, not one of each. The design involves obtaining several client outcome measures before the introduction of an intervention and several additional measures after the intervention has been implemented. Table 9.1 diagrams a time-series design.

A major advantage of a time-series design is that it overcomes some of the ethical problems of the designs that use comparison and control groups, as it has neither. An important feature of the multiple pretest and posttest measures of this design are the data trends that can help determine the extent to which the intervention, as opposed to other factors external to it, is the causal agent. Stated another way, its multiple measures of client outcomes provide more opportunities to determine whether extraneous influences exist.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

determine whether extraneous influences exist.

Pretest–Posttest Design with a Control Group

The fifth design, a pretest–posttest design with a control group, is the classical experimental design. It is the most powerful design for determining causality. It is similar to the pretest–posttest design with a comparison group with one exception. Clients are first randomly selected from a larger pool of prospective clients and then randomly assigned to either the intervention or the control group. Because of the two steps of random selection and assignment, the two groups of clients can be treated as statistically identical because the random selection and assignment tend to balance out the extraneous differences between the two groups. With this design, pretest and posttest measures are obtained from both groups before and after the intervention group receives the intervention.

This design is diagrammed in Table 9.1. Prior to time 1, clients are randomly selected and assigned to either receive the services of the program or to a control group. During time 1, a pretest score of the outcome variable is measured for both groups. During time 2, the intervention is introduced for the group assigned to receive it, and either nothing happens for the control group or it could continue to receive existing agency services. During time 3, a posttest measure of the client outcome variable is obtained for both groups.

Example of a Pretest–Posttest Control Group Design

Ciffone (2007) evaluated whether a curriculum-based suicide prevention program was instrumental in changing unwanted attitudes and in reducing students' reluctance to seek mental health treatment among groups of students in two demographically diverse high schools. He used a pretest–posttest control group design. Students were randomly assigned to treatment and control groups by the classes that they were assigned. Those in the treatment group were provided a three-day presentation by the school social workers who spoke in their health classes. Those in the control classes had no outside speakers in their health classes during the same period. The pretest and posttest instrument consisted of eight questions. The attitudes toward suicide were measured partially by the item, "Most teens who killed themselves were probably suffering from mental illness." Students' responses to the item in the intervention group changed from "no" to "yes" when their attitudes improved. Two other items measured help-seeking attitudes or reluctance to seek mental health treatment: "If suicidal thoughts crossed my mind regularly, I would seek out and talk to a friend about these thoughts" and "If I was very upset and suicidal thoughts crossed my mind, I would be willing to talk with a professional counselor about these thoughts." Higher percentages of students in the intervention group were favorable to these statements in the posttest than those in the control group.

This design is quite powerful in determining whether the program intervention influences the outcome measure for the client group in the program because it controls for all extraneous influences. Nevertheless, the design poses even more serious ethical challenges than any of the other designs. It can be viewed as unethical to randomly select and assign clients to receive or not receive an intervention, particularly without their consent. Most likely, if agency resources can provide a program intervention for all clients, the decision to select some clients should be made based on professional criteria, not mathematical probability. This dilemma could be worked out by randomly selecting clients from a waiting list and then randomly assigning one group to receive the intervention and the other group to continue to wait for services offered later. However, it would be important to explain this arrangement to clients in an informed consent protocol. Also, other unanticipated consequences must be considered, such as exposing the vulnerabilities of clients to life-threatening crises in the control group during a waiting period.

Posttest-Only Control-Group Design

The sixth design, a posttest-only design with a control group, is the same as the experimental design except that it does not have a pretest for either group. The design depends completely on the initial random selection and assignment steps to ensure that participants in the two groups are identical, which is particularly important because there are no pretest measures to determine whether the two groups begin essentially at the same point in their outcome measures.

This design is preferred over the pretest–posttest design with a control group if there is concern that administering a pretest measure could influence the outcome measure. For example, a test about a knowledge area used as a pretest measure may also be used as a posttest measure to determine how much the clients learned during the intervention phase. An intervention could be safe-sex training, and the outcome measure could be a test that measures how well clients understand safe sexual practices. The problem here is that the test could be discussed informally among clients during the intervention phase to determine the correct answers and the clients could then report these answers in the posttest phase.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

intervention phase to determine the correct answers and the clients could then report these answers in the posttest phase. In this case, testing factors rather than actual learning from the intervention could be responsible for any improvement in the test from pretest to posttest.

In other instances, this design could also be preferred over the pretest–posttest design with a control group if the pretest seemed unnecessary or undesirable. It may be undesirable or not logistically possible to implement, or an evaluator may find it too costly to administer a pretest measure because it would involve time and effort of the staff members to interview clients or administer a questionnaire as a pretest.

Table 9.1 diagrams the posttest-only control-group design. Prior to time 1, clients are randomly selected and are either assigned to receive the program's services or assigned to a control group. During time 1, a pretest score of the outcome variable is not measured for either group. During time 2, the intervention is introduced for the group assigned to receive it, and either nothing happens for the control group or it continues to receive existing agency services. During time 3, a posttest measure of the client outcome variable is obtained for both groups.

This design, like the classic experimental design, is quite powerful in determining whether the program intervention influences the outcome measure for the client group in the program because it controls for all extraneous influences. However, this design also poses serious ethical challenges for the same reasons that it does for the classic experimental design. The clients could be randomly selected from a large waiting list and then randomly assigned either to receive the intervention or to wait for services to be offered later. However, it would be important to explain this arrangement to clients in an informed consent protocol, and a plan to address unanticipated consequences must be considered, such as the control-group clients' vulnerability to a life-threatening crisis during the waiting period.

It is important to note that all group designs have another major limitation if they only measure client outcomes immediately after clients complete a program. Such outcome measures, as mentioned earlier, are usually only short-term outcomes. Therefore, programs are encouraged to consider one or more additional outcome measures at later points in time (e.g., three and six months later) to determine whether clients' progress from participating in a program is sustained over time.



OUTCOME EVALUATIONS FOR PRACTICE

The previously described outcome evaluations are generally used in program evaluations. Practice evaluations tend to use different designs. Many of these designs are described next.

Client Satisfaction and Effectiveness

It is important that practitioners hold themselves accountable for the effectiveness of their interventions even when others may not do so. This is an ethical obligation mandated by the NASW's (2017) Code of Ethics (see Chapter 3). The code points out that practitioners are to "monitor and evaluate policies, the implementation of programs, and (our) practice interventions" (section 5.02[a]).

There are some basic questions that practitioners can periodically ask clients, such as the following:

- Are my interventions helping you? In what ways?
- Are my interventions helping you in the areas in which you are seeking help (e.g., related to the presenting problems, identified needs, solutions desired, expressed hopes)?
- Are my interventions helping you meet your goals identified in our initial work together?
- How would you describe your progress so far in achieving your goals? How have my interventions played a part in helping you or in hindering you?

The specifics of how clients communicate that the practitioner's interventions have helped (or have not helped) can be especially important. For example, did they simply respond to the practitioner's questions or did they also spontaneously offer additional comments? Spontaneous comments may be more authentic and more likely to be valid indicators of their views.

These types of questions are considered client satisfaction questions, but they are also quite relevant to the question of effectiveness; they focus on client outcomes and their connection to the social worker's intervention. They ask did the intervention have a role in bringing about any positive changes to clients' goals or outcomes? In some instances, clients' responses to these questions may be the only available information. In this regard, they have limitations, especially because they are subjective and not necessarily reflective of actual outcomes. For example, clients could report

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

accomplishments that did not actually happen (e.g., discontinued drug use) or overestimate what happened (e.g., reporting completion of homework every day, when it may be sporadic). That is why objective measures of outcomes are also extremely important to obtain, such as checking whether drug use is evident in a client's urine or obtaining an independent report of the client's drug use from a spouse or another reliable source.

As discussed in Chapter 8, however, it can be argued that clients' reports on their outcomes and how the interventions have affected them are a necessary but not sufficient indication of intervention effectiveness. Conversely, obtaining an objective measure of a client's outcomes without some concurrence from the client can be viewed as overlooking the importance of the clients' confirmation of their change.

Single-System Designs

Practice evaluations of outcomes can be conducted in several ways, and many of them revolve around variations of a single-system design. A single-system design is a variation of a time-series design on a practice level in that it involves several measures of the outcome variable both before and while an intervention is being introduced. A single-system design is a practice evaluation tool that measures whether there is a causal relationship between the practitioner's intervention and a client's outcome measure. This design is typically used with one client system at a time whether it is an individual, family, group, or community. The single-system design uses a graph as a tool to visualize the clients' progress, and it is intended to be used by both the client and the social worker in determining whether any progress has been made. The graph helps simplify the explanation for how much progress has been made and it does not require a statistical test for explanation, although one can be used.

An single-system design graph consists of two or more phases. A baseline phase (phase A) is almost always needed as well as one or more intervention phases (e.g., phases B, C, and D). The client outcome measure is plotted along the vertical axis, and a timeline for recording measures of the outcome measure is identified along the horizontal axis (e.g., daily, weekly). Figure 9.2 illustrates an ABC design with three phases: a baseline phase, a phase when the first intervention is introduced, and a phase when the second intervention is introduced.

A single-system design can be conducted using a variety of designs, including AB, ABC, ABAB, and BAB. The simplest design is an AB design, which consists of a baseline phase and one intervention phase. In this case, the client outcome measure is obtained and plotted along a horizontal, standardized timeline, such as weekly or daily intervals during both phase A (baseline) and phase B (intervention phase). For example, a woman with clinical depression wants to develop new friendships but is resistant to getting out and meeting people. The social worker attempts to help her by providing supportive services, as support has been missing in the client's life and is thought to be what she needs to begin attending social occasions in her church and local community center. The frequency of attending these social occasions each week was the client outcome indicator, measured every week while the client was in contact with the social worker's agency. The social worker's intervention, introduced during phase B, is the support service. This AB design requires measurement of the frequency of the client's attendance weekly for a few weeks to obtain a baseline measure that is stable or does not fluctuate highly. This outcome indicator continues to be measured weekly during phase B, when the social worker's intervention is implemented. The intent of the AB design is to discover whether the client attends more social occasions after the intervention is introduced than she did before.

Another single-system design is the ABC design, illustrated in Figure 9.2. This design begins with a baseline phase (phase A), followed by the first intervention phase (phase B), and then followed by another intervention phase (phase C). The ABC design is particularly useful when the first intervention does not have sufficient impact by itself. A second intervention can then be introduced in phase C to either replace or supplement the first intervention. Often the combination of the two interventions can have a significant impact when neither intervention has much impact by itself.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

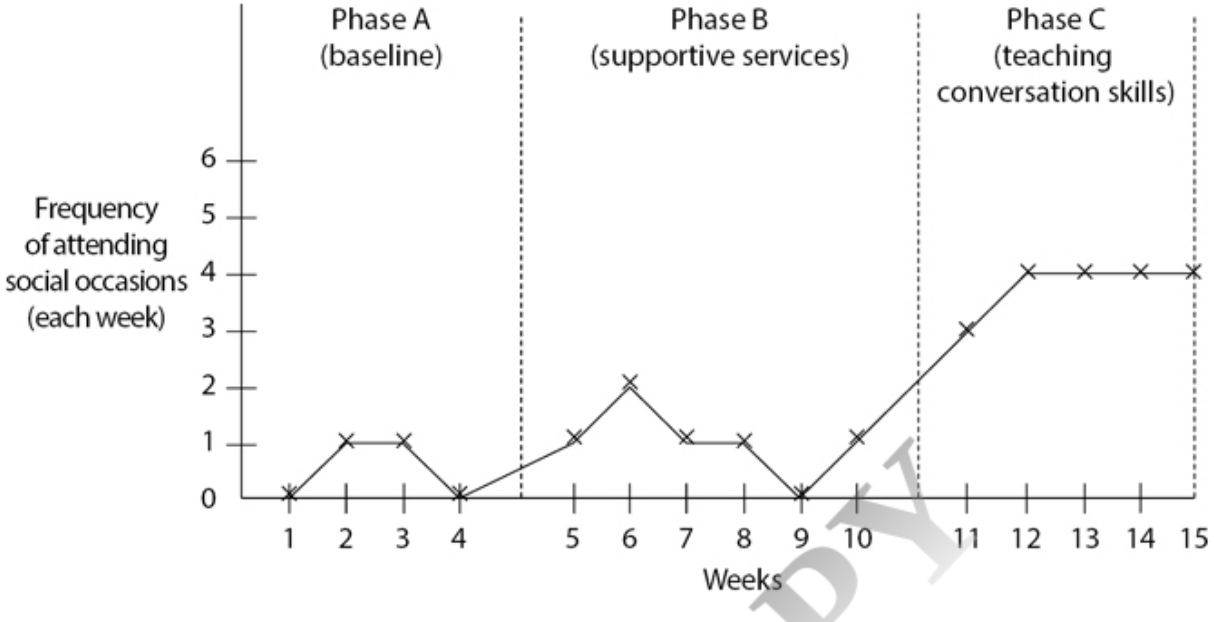


Figure 9.2 An ABC design.

Using the same example of the woman with severe depression, providing support services weekly could be the first intervention to encourage more social contacts (during phase B). If this intervention does not succeed, teaching the client conversation skills could be added as a second intervention (during phase C). The ABC design can plot the client's attendance at social occasions during phases A, B, and C to determine whether attendance increases during implementation of the first intervention and later adding the second intervention. In Figure 9.2, combining the two interventions, support services and teaching conversation skills, visibly documents greater social contact for this client.

Another variation is the ABAB design, which consists of a baseline phase followed by an intervention phase, then another baseline phase when the intervention is removed, and finally the reintroduction of the intervention phase (see Figure 9.3). This design has a special additional feature of withdrawing the intervention after phase B to find out whether other factors outside the agency (e.g., a self-help group, a new friend, joining a new church or synagogue) may be responsible for any continuation of client progress evident in phase B. If progress on the client goal declines during the reintroduction of phase A, this provides further evidence that the intervention and improvement are linked. Finally, the intervention can be reintroduced in the second phase B to find out whether progress reappears.

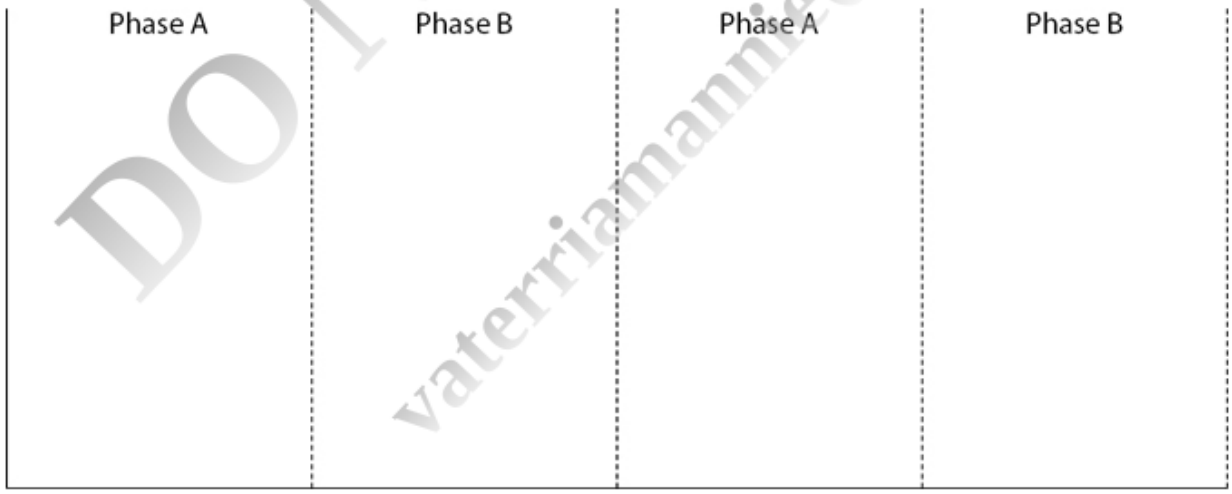


Figure 9.3 An ABAB design.

Although the ABAB design can, in some ways, be an alternative to having a comparison or control group, it can also pose ethical problems if the intervention is withdrawn while the client still needs it. Similarly, it would also be unethical to withdraw the intervention if a client wishes to continue it or if withdrawal could cause harm. However, temporary

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

withdrawal of the intervention may be less of an ethical concern if the social worker or client became unavailable because of a vacation or lengthy sickness. Withdrawal may also be less of a concern if the intervention is mandated (involuntary) and if the client wants a temporary break from services.

An ABAB design has another limitation in that it works only when the outcome measure can be reversed. If a client is taught a new set of behaviors, once he or she has learned the behaviors, they may not be reversed in many cases. For example, if a parent is taught new assertiveness skills in phase B, withdrawal of the intervention is not likely to lead to an immediate reversal of the newly developed skills.

Sometimes a BAB design is relevant if the social worker cannot begin with a baseline phase. This may be the case particularly when it is urgent to introduce the intervention during the first contact with the client. With this design, the baseline phase can be introduced after the intervention has been implemented for a while, and the intervention can be reintroduced in a third phase. However, like the ABAB design, withdrawal of the intervention could pose ethical problems and may not work with irreversible outcome measures.

Goal Attainment Scale

A goal attainment scale (GAS) is a versatile evaluation tool that can be used to measure client outcomes. It is quite useful in evaluating the extent to which the worker's interventions affect clients' goals or outcomes in a favorable way. The GAS works well as an outcome measure with any of the single-system design designs described previously. As Table 9.2 indicates, a GAS is a five-point scale ranging from the most unfavorable to the most favorable outcome thought likely at a future follow-up time. The middle point on the scale is considered the expected level of success. The client and social worker use the right-hand column to fill in qualitative descriptors of each of the five points on the scale, with the expected level of success usually determined first.

Table 9.2. Goal Attainment Scale

LEVELS OF PREDICTED ATTAINMENT	SCALE
Most unfavorable outcome thought likely (–2)	
Less success than expected (–1)	
Expected level of success (0)	
More success than expected (+1)	
Most favorable outcome (unlikely but still plausible) (+2)	

A GAS has many advantages as an outcome tool. It is an incremental scale that can be used to measure degrees of progress. It can be individualized to a client's specific or unique circumstances. And it is used most effectively when the client and worker collaborate in the development and use of the scale. When the tool is used in this way, it can be a great motivator in getting clients to work on their goals. The scale can also become a useful discussion topic when discrepancies become evident between what the worker and client perceive as the client's level of success. It can also become the focal point of an informal or formal contract between the worker and client.

Table 9.3 illustrates the use of a GAS in a parent-training program in which several parenting skills were introduced and practiced in small groups for parents who had been abusing their children. This is a building-block approach that begins with reading; learning comes next, followed by using or applying a parenting skill. Note that the scale reflects a decision to set an expected level of success for these parents after three months of simply learning one parenting skill. Although this may not seem like a far-reaching level of success, this program documented evidence that learning one skill would realistically take this long to internalize.

Table 9.3 Goal Attainment Scale With Goal of Improving Parenting Skills

LEVELS OF PREDICTED ATTAINMENT	SCALE
Most unfavorable outcome thought likely (–2)	Ignore parenting skills altogether
Less success than expected (–1)	Read a book on parenting skills
Expected level of success (0)	Learn one new parenting skill
More success than expected (+1)	Use one new parenting skill with children
Most favorable outcome (unlikely but still plausible) (+2)	Use two new parenting skills with children

Notes: Date and rating at construction: January 5; Date and rating at follow-up or evaluation time: April 15.

Cox and Amsters (2001) effectively used the GAS as a multidisciplinary measure of client outcomes for rural and remote health services. They found several principles to be important in using a GAS. The goals need to remain realistic and relevant to both the clients' capacity or potential and the agencies' resources. In addition, the levels must be mutually exclusive and measurable. They found that the GAS had limitations as well; most important, it was not effective if staff were not fully trained in its use.

A GAS can be developed as either a qualitative or quantitative measure or a combination of the two. The scale in Table

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

A GAS can be developed as either a qualitative or quantitative measure or a combination of the two. The scale in Table 9.3 uses both types of measures; for example, it measures the number of parenting skills learned and takes into account four different behaviors that overlap: ignoring, reading, learning, and using. The scale could also include more than two variables at each level even though it did not in Table 9.3.

A GAS can be implemented with the following principles and sequential steps kept in mind (Dudley, 2011):

1. Consider first whether the GAS is the most appropriate design to use with a specific client. Among other considerations, a GAS depends on clients meeting with a worker for more than one session and having measurable goals.
2. Make sure that you cover and that the client understands all the pertinent issues of informed consent, including the purpose of the selected tool, how it works, how the client can benefit, and the other aspects of informed consent.
3. Help the client identify the problem, need, and its causes.
4. Identify a goal that can resolve this problem and need that is agreeable to the worker and client.
5. Set a realistic date for the goal to be achieved, referred to as an evaluation time.
6. Then work together to construct specific predictions for a series of five levels of reaching the goal (most unfavorable, less than expected, expected level of success, more than expected, and best anticipated success) by recording them on the scale. Usually the easiest way to do this is to first identify the expected level of success, followed by the outer limits of the most unfavorable and most favorable outcome. Finally, identify the less-than-expected and more-than-expected levels.
7. Now the GAS is ready to be used by measuring the client's actual level of attainment in reaching the goal during the first session as a baseline measure.
8. Finally, the client's actual level of attainment in reaching the goal can be measured during the last session. At this time, discuss the progress that has been made and how the GAS was helpful in measuring this progress.

Target Problem Scale

The GAS has limitations. For example, some problems may not easily translate into goals, or clients may not be either willing or able to identify goals to work on. Also, a GAS does not work well if a client is seen for only one session. An alternative outcome scale that can be used is a target problem scale (TPS). In brief, a TPS helps clients identify their problem(s), apply an intervention to overcome or cope with them, and then periodically rate the problem to measure any deteriorations or improvements. A TPS is helpful, for example, if the client remains preoccupied or overwhelmed with multiple problems and is not prepared to immediately pursue goals to overcome the problems.

Like a GAS, the TPS is an incremental scale. It measures the degree of severity of the problem, and the scale is individualized to a client's specific circumstances. Figure 9.4 describes a TPS (Marlow, 2011). In this scale, target problems are identified in the far-left column. This form provides space for up to five problems. Each problem is then assessed in terms of its degree of severity for each of six sessions. The five-point severity scale measures problems from "no problem" to "extremely severe." The next column to the right is the improvement scale used to calculate the degree to which a problem has changed for better or worse. Finally, an overall improvement scale can be used to determine the extent to which improvement has occurred.

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Target Problems	Degree of Severity Session # 1	2	3	4	5	6	Degree of Change (1 to 5 scale)	Overall Improvement*
1.								
2.								
3.								
4.								
5.								

Severity Scale	Improvement Scale
NP = no problem	1 = Worse
NVS = not very severe	2 = No change
S = severe	3 = A little better
VS = very severe	4 = Somewhat better
ES = extremely severe	5 = A lot better

*__The Overall Improvement Rating__ is the total of the degree of change, divided by number of target problems, which results in a number that reflects the client's overall improvement on all problems.

Figure 9.4 Target Problem Scale.

The overall improvement rating is obtained by totaling the degree of change scores and dividing by the number of target problems. This results in a number that reflects clients' overall improvement on all problems.

An example of a TPS used with parents who were lacking strong parenting skills is described in Figure 9.5. As the figure indicates, the parents had multiple problems, including overlooking their parental neglect, getting to appointments late, spending too little time with their child, getting the child to bed late, and yelling at each other. Figure 9.5 summarizes the clients' progress or lack of it after six sessions. In brief, the TPS indicates that their difficulty in facing parental neglect improved from "extremely severe" in the first two sessions to somewhat better or "not very severe" by the sixth session. Overall, most of their progress on other problems over six sessions was in getting to counseling sessions on time and in spending more time with their child. A little improvement was evident in not yelling at each other so often.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Target Problems	Degree of Severity Session # 1	2	3	4	5	6	Degree of Change (1 to 5 scale)	Overall Improvement*
(Examples)								
1. Overlook parental neglect as a problem	ES	ES	S	S	NVS	NVS	3	
2. Come late to appointments	S	S	S	NP	NP	NP	2	
3. Ignore their child's need to spend time with parents	ES	NVS	NVS	NVS	NVS	NVS	3	
4. Get child to bed late	ES	ES	ES	VS	ES	ES	0	
5. Parents yell at each other	S	S	S	S	S	NVS	1	
								9/5 = 1.8

Severity Scale	Improvement Scale
NP = no problem	1 = Worse
NVS = not very severe	2 = No change
S = severe	3 = A little better
VS = very severe	4 = Somewhat better
ES = extremely severe	5 = A lot better

*__The Overall Improvement Rating__ is the total of the degree of change, divided by number of target problems, which results in a number that reflects the client's overall improvement on all problems.

Figure 9.5 Target Problem Scale for a family.

Sequencing Blocks

Other means of evaluating a client's outcomes, positive or negative, are also available, even though they may not be readily known or used. For example, sequencing blocks are a qualitative exercise that can help clients explore the causal relationships among prior behaviors that resulted in a positive or negative outcome. In this case, the outcome behavior is known and the behavior that caused it is to be discovered. Sequencing blocks can be used to help clients gain insight into prior behaviors and interactions that could trigger a current problem behavior that is being addressed (Laura Kroll, graduate student placed at Alexander Youth Network, Charlotte, NC, personal communication, April 15, 2006). This is an exercise that can use narrative from the discussions between a worker and client about a problem behavior, such as cursing out a teacher or getting angry and walking off a job. The discussion typically begins with the current problem behavior and its larger context. Then the discussion can work its way in reverse, going back to what happened just before the problem behavior occurred, and then what happened before that. The conversation eventually can get back to the initial set of behaviors or incidents in the overall episode. Figure 9.6 portrays typical sequencing blocks. As the discussion between the social worker and client proceed, individual behaviors and interactions are written into the blocks in the order that they occurred.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

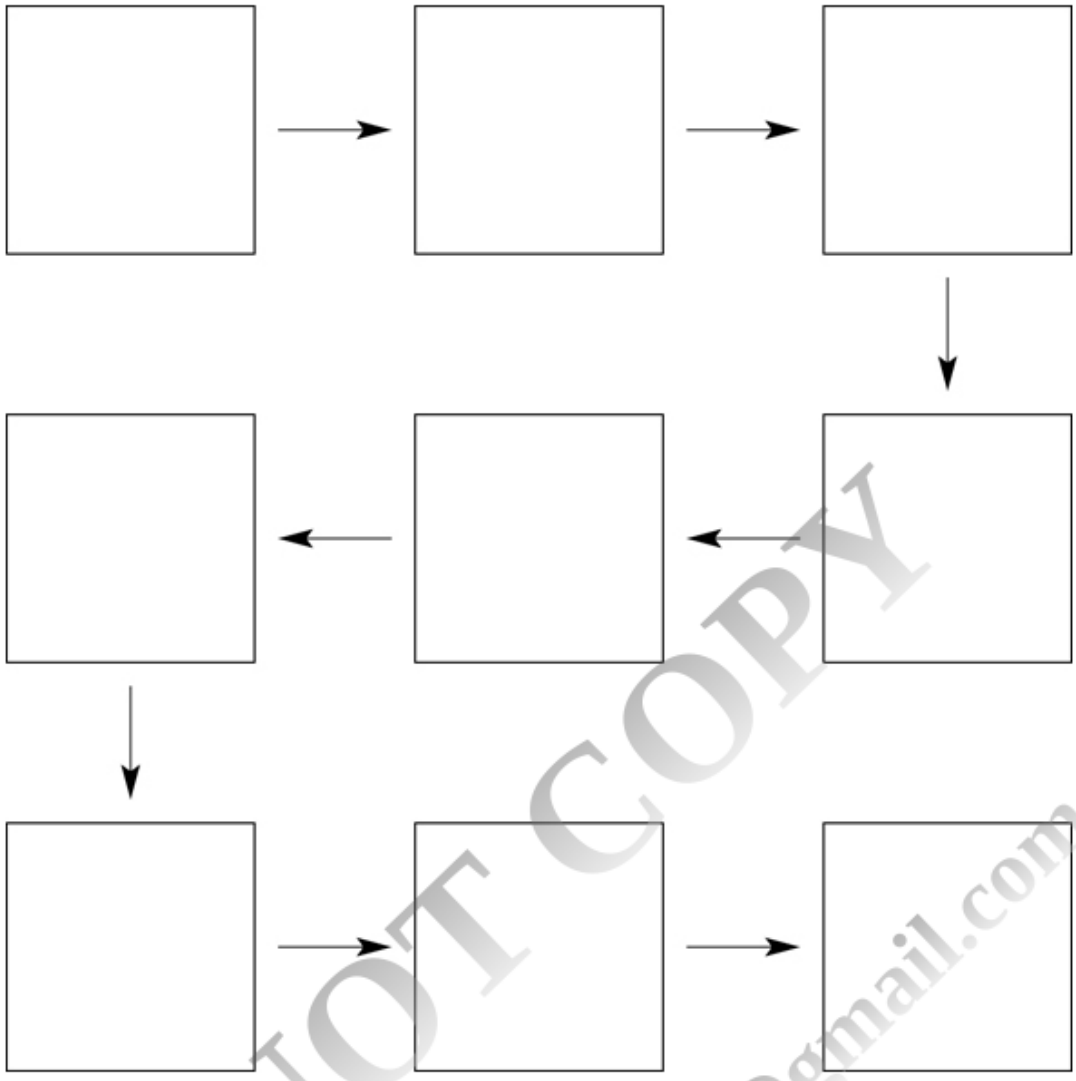


Figure 9.6 Sequencing blocks.

Figure 9.7 describes an example of a teenager who was expelled from school for cursing out his teacher. In this case, the teenager, John, was meeting with a school social worker about the incident upon returning to school from a suspension. The sequencing blocks were used to help him gain insight into the earlier triggers and the precipitating factors contributing to the problem. Their discussion began with the problem behavior and worked its way back one behavior and interaction at a time. It eventually got back to a trigger in which the teacher perceived, incorrectly, that John had initiated a conflict with a peer around opening a container of applesauce at lunch on a field trip. The sequencing block exercise helped John understand not only that his teacher had misperceived the cause of the conflict but also how he could have avoided the interchange with the teacher until after calming down and how he may have been able to avoid the incident in the first place, by avoiding sitting with the peer who got him into trouble.

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

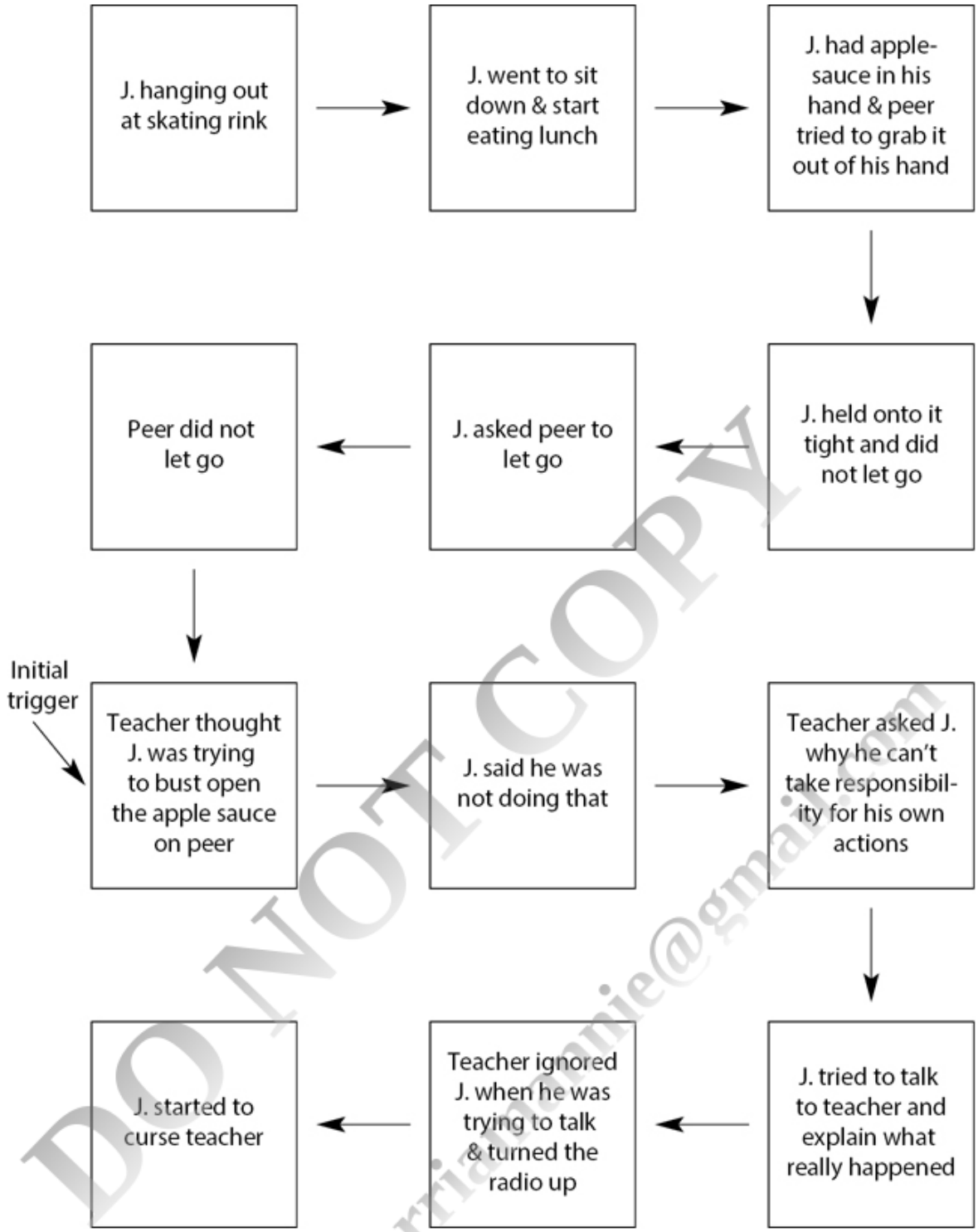


Figure 9.7 Sequencing blocks for a teenager.

Sequencing blocks can also be used to identify prior behaviors or decisions that result in positive outcomes for a client. For example, consider a client with a disability who has difficulty meeting women but who is interested in a friendship and had an unusually positive experience one night in meeting a new acquaintance. The client initially explained this discovery as a "miracle." The worker and client used this positive outcome to look back at prior behaviors and decisions made by the client that may have been positive triggers that could be tried again.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

SUMMARY

The chapter is devoted to introducing several types of evaluations that can be used during the outcome stage of a program or practice intervention. The overall purpose of outcome evaluations is to determine whether the intervention resulted in a positive outcome for the clients and whether there is enough evidence that the improvements in the clients' outcomes are significant. Significance is explained as being determined in either of two ways. One way is to calculate significance using a mathematical test and probability theory. The other way is to use a prior clinical standard and then determine whether the clients' changes met or exceeded that clinical standard. Client outcomes are also discussed in the context of whether there is enough evidence of a causal relationship existing between an intervention and the client outcomes. Three conditions are described and required before a causal relationship can be claimed. Ultimately, during the outcome stage, the purpose of an evaluation is to determine whether the intervention is effective in helping the clients it serves.

The chapter discusses what client outcomes are, how they can be measured, and the criteria for determining whether they are adequate or acceptable. Several criteria are described including the outcomes' usefulness, validity, reliability, precision, feasibility, cost, and unit costs.

Six group designs are introduced that can be used to determine the extent to which *program* interventions have a causal effect on client outcomes. They include two pre-experimental designs, two quasi-experimental designs, and two experimental designs. Experimental designs are the strongest designs that can be used for claiming a causal relationship, but ethical concerns are also likely to exist with these designs. Pre-experimental designs are the weakest designs, but they are useful in beginning to explore causality when the intervention is in an earlier stage of development. Several types of evaluations are also described in the chapter for evaluating client outcomes for *practice* interventions. These evaluations include single system designs, GASs, TPSs, and sequencing blocks. These designs can be used in a variety of situations to determine how effective the social workers' practice is in helping each client system that they serve.

KEY TERMS

Baseline measure
Clinical significance
Comparison group
Control group
Experimental designs
Goal attainment scale
Intervention group
Longitudinal study
Measurable client goal
Posttest
Pre-experimental designs
Pretest
Practice intervention
Program intervention
Quasi-experimental designs
Sequencing blocks
Single-system design
Statistical significance
Target problem scale

DISCUSSION QUESTIONS AND ASSIGNMENTS

1. Review the six group designs summarized in this chapter and then answer the following questions (Appendix C has answers):
   - Which design seems ideal? Why do you think so?
   - Which design is the easiest and least expensive? What can be concluded from this design?
   - Why are multiple measures of the client outcome variable obtained when conducting a time-series design? What are the benefits of the multiple measures?
   - Which design may be the most practical for determining whether the program intervention was responsible for positive changes in the outcome variables? Why?
   - What is useful about a pretest–posttest design?

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

- Under what circumstances would an evaluator consider a posttest-only control-group design?

2. This is an assignment that can be completed in class or as an assignment to video and bring to class.

   A. Role-play a worker–client relationship using a goal attainment scale (GAS). Conduct a first session when the GAS is introduced. Identify a goal that you and your client want to work on. Make sure that you mutually agree on the goal. Also identify a goal that both you and your client agree is measurable. Explain the GAS to the client so that it can be readily understood. Consider using a simpler version that fits your specific client: maybe a three-point scale, simpler words, modified wording (e.g., such as vision of the future or dreams rather than goals).

   B. Later, conduct a session in which the worker and client assess how well the client did in reaching the defined goal. Have your client rate her or his progress along with your rating, and then discuss why your ratings are different if they are. Finally, explore what the specific benefits of the evaluation tool are to the client.

3. Design a program or practice intervention and a set of outcome measures for a group of clients by following three steps:

   A. Select a group to focus on (e.g., teen fathers, older widows, LGBT teens).

   B. Identify two problems/needs and two outcome measures for the group.

   C. Design a program or practice intervention to help the group reach these two outcomes.

4. Critique White's (2005) choices of outcome indicators for clients experiencing domestic violence in a domestic violence agency (in a box earlier in the chapter). Which ones do you think are most useful and why?

5. Assume that you are working in an agency serving people with a dual diagnosis of depression and alcohol abuse. Make up what you understand to be their specific problems/needs related to depression and alcohol abuse (e.g., for depression, the client has no social contact with others). Then identify two outcome indicators for clients that could be used to measure the impact of a program on their depression and identify two outcome indicators determining whether their alcohol abuse problem/need is brought under control.

6. This is a policy-related assignment. Set up two teams of students: one assigned the task of identifying and arguing for the advantages of keeping welfare recipients on the welfare rolls and the other on removing them from the welfare rolls. Each team should begin by identifying two or three desired outcomes of either keeping or reducing, respectively, the number of welfare recipients. For example, a possible outcome for reducing the rolls could be savings to the taxpayer, while an outcome for maintaining recipients on welfare could be to insure that they have health insurance.

7. This is a major class assignment. Carry out the following steps in completing this assignment.

   A. Identify a need for a practice evaluation in your field agency and a client system that can benefit from an evaluation. Inform the client system about the evaluation and encourage them to participate.

   B. Select the type of evaluation and specific tool that most readily fits your client's situation. Choose a tool from the following list: a single-system design, a goal attainment scale, a target problem scale, or sequencing blocks. Feasibility is one important consideration in your selection. For example, consider the number of sessions that you will likely see your client, the overall amount of time that your agency will serve the client, and any difficulties you may have in identifying a measurable goal(s) for the client. Discuss the assignment thoroughly with your field instructor to make sure that she or he understands the assignment and the evaluation tool that you will use.

   C. Obtain informed consent either in written or oral form. Make sure that all the pertinent issues of informed consent are covered and understood by the client (e.g., purpose of the selected tool; how it works; how the client, worker, and agency can benefit; expectations that you have for the client's participation; informing the client that this is a class assignment to be turned in to your instructor; ensuring confidentiality; reminding the client of the option to say no or to withdraw after the evaluation begins). Whenever your field agency has a protocol for obtaining informed consent, use it to complete this assignment.

   D. Devise and implement your practice evaluation.

   E. Finally, describe what you have learned about evaluating your practice from doing this assignment and what advantages and limitations you see in using this tool.

   F. Turn in a paper or give a presentation that includes a description of how you implemented each of the previous steps. Also, include the instruments that you used.

REFERENCES

Babbie, E. (2016). *The practice of social research* (14th ed.). Boston, MA: Cengage Learning.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Babbie, E. (2016). *The practice of social research* (14th ed.). Boston, MA: Cengage Learning.
Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory* (2nd ed.). San Antonio, TX: Psychological Corporation.
Cabarrus County. (2007). *Performance report 2006*. Concord, NC: Author.
Campbell, D., & Stanley, J. (1973). *Experimental and quasi-experimental designs for research*. Chicago. IL: Rand McNally.

Ciffone, J. (2007). Suicide prevention: An analysis and replication of a curriculum-based high school program. *Social Work, 52*(1), 41–49.
Comfort, M., Loverro, J., & Kaltenbach, K. (2000). A search for strategies to engage women in substance abuse treatment. *Social Work in Health Care*, *31*(4), 59–70.
Community Choices Inc. (2006). *Helping mothers have healthy babies and families*. Charlotte, NC: Author.
Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Skokie, IL: Rand McNally.
Cox, R., & Amsters, D. (2001). Goal attainment scaling: An effective outcome measure for rural and remote health services. *Australian Journal of Rural Health, 10,* 256–261.
Dudley, J. R. (2011). *Research methods for social work: Being producers and consumers of research* (upd. 2nd ed.). Boston, MA: Pearson.
Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice. *Families in Society, 80,* 341–350.
Gibbs, L., & Gambrill, E. (1996). *Critical thinking for social workers: A workbook*. Thousand Oaks, CA: Pine Forge Press.
Harding, R., & Higginson, I. J. (2003). What is the best way to help caregivers in cancer and palliative care? A systematic literature review of interventions and their effectiveness. *Palliative Medicine, 17,* 63–74.
Litzelfelner, P. (2001). Demystifying client outcomes. *Professional Development, 4*(2), 25–31.
Lynch, C. J. (2007). *Exploring the implementation of a transitional services program for adolescents in the Texas Foster Care System*. Unpublished doctoral dissertation, University of Texas, Austin, TX.
Marlow, C. (2011). *Research methods for generalist social work* (5th ed.). Stamford, CT: Thomson Learning.
Martin, L. L., & Kettner, P. M. (2009). *Measuring performance of human service programs* (2nd ed.) Thousand Oaks, CA: SAGE.
McLean, T. (2007). *Evaluating progress on personal goals of a children's group using journaling: Proposal for the formation of a new group*. Unpublished manuscript. University of North Carolina, Charlotte, NC.
National Association of Social Workers. (2017). *Code of ethics* Washington, DC: Author.
O'Hare, T. (2015). *Evidence-based practice for social workers: An interdisciplinary approach*. (2nd ed.) Chicago, IL: Lyceum Books.
Thyer, B., & Meyers, L. (2003). Linking assessment to outcome evaluation using single system and group research design. In C. Jordan & C. Franklin (Eds.), *Clinical assessment for social workers: Quantitative and qualitative methods* (2nd ed., pp. 385–405). Chicago, IL: Lyceum Books.
U.S. Department of Health and Human Services. (2017). Child welfare outcomes 2010–2014: Report to Congress. Executive summary. Retrieved from https://www.acf.hhs.gov/cb/resource/cwo-10-14
White, M. (2005, March 16). A demonstration of evaluation as a central aspect of social work practice at the Women's Commission. Presentation to a social work research class, University of North Carolina, Charlotte, NC.

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

PART VI



Final Steps in Completing an Evaluation

Evaluations are not complete until two additional major tasks are addressed. Chapter 10 describes the process of analyzing the data obtained from an evaluation. Data analysis is discussed for qualitative and quantitative findings. This chapter also addresses data analysis when both qualitative and quantitative findings are collected. The final task of an evaluation is covered in Chapter 11 and involves preparing and disseminating reports of findings.

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

CHAPTER 10



Analyzing Evaluation Data

James R. Dudley and Jeffrey Shears

How do we transform evaluation data into useful information?

The analysis of evaluation data comes next in the evaluation process. Analyzing the data begins in step 5 (implement the evaluation) and continues into step 6 (prepare a written or oral report) of the Seven Evaluation Steps. In step 5, once quantitative data are collected, they are coded and entered into a computer program such as SPSS (Statistical Package for the Social Sciences) and prepared for analysis. Qualitative data are usually prepared for analysis in narrative form. In step 6, the data are analyzed and transformed into understandable findings and recommendations for stakeholders.

Although the basic principles of data analysis are virtually the same for evaluations and other kinds of research, differences are evident in emphasis and use. Such differences are the primary focus of this chapter. Data analysis is also closely linked to the topic of the next chapter, writing a report and disseminating the findings. The choice of which data analysis tools to use depends on the kind of report and dissemination strategy that are most helpful to the stakeholders and others interested in the results.

FORMATIVE OR SUMMATIVE EVALUATIONS AND DATA ANALYSIS

Evaluations depend on quantitative and qualitative research methods and sometimes involve both. As discussed frequently in previous chapters, the methods used depend on the evaluation's purpose and the stage of program or practice development in which they occur. One way to begin to decide whether to mostly use quantitative or qualitative data analysis is to ask whether the evaluation is formative or summative.

Formative Evaluations

As mentioned in Chapter 2, formative evaluations mostly focus on planning for a new program or practice intervention or improving its implementation. Outcomes evaluations are also possible but only in a tentative way. Formative evaluations tend to be exploratory in nature and may use either qualitative or quantitative methods, or both, depending on their purposes. Formative evaluations are typically conducted by the agency administering the intervention, which allows for considerable flexibility in choosing the methods of data collection and analysis. Formative evaluations do not require one specific type of procedure, as they can be conducted to identify small glitches in programs, breakdowns in delivering services, or an unanticipated drifting away from the original program design. Also, the links can be explored between the interventions and client outcomes to determine how well the intervention seems to be succeeding in helping clients reach their goals.

Examples of formative evaluations include a needs assessment questionnaire with forced-response and open-ended questions, or observations of interactions between clients and practitioners, management strategies, practice philosophies, or costs associated with a program. A more detailed example of a formative evaluation may help. An agency wanted to understand why a program for eating disorders for teenage girls has not been successful in having participants complete the twelve-week program; only one fourth of these teenagers participated long enough to complete the treatment. A formative evaluation can provide some meaningful information to finetune this program. Here exploring questions that are critical of the program could be beneficial in identifying glitches in the delivery of services. Conducting a more in-depth evaluation of the clients and their needs, and obtaining a profile of those who are successful and those who fail to complete the program are important examples of a formative evaluation.

Summative Evaluations

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Summative evaluations focus on the outcomes of programs and practice interventions and attempt to answer the ultimate question: Did the intervention reach its goals? These evaluations have a finality to them because they attempt to measure whether an intervention was effective. Summative evaluations are also typically conducted by an external agent, such as an independent evaluator or a governmental agency, to ensure that they are objective. Further, they primarily depend on choosing from a more limited set of research design options, such as a quasi-experimental or experimental design. These designs work best when they use quantitative measurement instruments with strong psychometric properties.

If an evaluator's question is to determine whether an intervention is effective, then this is considered a summative evaluation and thus draws from quantitative research methods. Summative evaluations seek to explain how the independent variable, a program or practice intervention, influences the dependent variables, the desired client outcomes. For example, does participation in a parental support group influence parental attitudes? Does group therapy improve marital stability? Do parental education classes significantly improve parents' ability to support early infant development?

STAGES OF INTERVENTIONS AND DATA ANALYSIS

The stage of development of a program or practice intervention, whether planning, implementation, or outcome, also provides some clarity about what data collection and analysis methods are used and whether quantitative and/or qualitative data analyses are preferred.

The Planning or Input Stage

One period in which data analysis is needed is during the input or needs assessment stage of a proposed program or practice intervention. Needs assessments, as discussed in Chapter 6, are usually proposed when there is a concern that a new intervention may be important to consider for a specific target population. During this stage, various evaluation strategies, such as handing out questionnaires or analyzing secondary data, might be used to examine the specific needs of a community. Evaluators could also conduct focus groups or individual interviews with community members or consult other local experts to clarify where services are needed and how to better allocate resources. For example, a needs assessment of a new target population to be served by a county's mental health system might involve interviewing prospective clients, giving out questionnaires to staff members who work with mental health clients, or conducting focus groups with other stakeholders who have a vested interest in an issue. These strategies can be qualitative or quantitative in nature. In general, it is wise to combine different strategies. Royse, Thyer, and Padgett (2009) discuss the importance of triangulation, or seeking multiple sources and comparing them. For example, interviews with people who are homeless, those who work with the homeless, and other key informants could all be conducted and compared to discover their similarities and differences.

Because of the varied nature of needs assessments, there are many types of data analyses that can be used to answer questions about the perceived needs of a community. Using focus groups with stakeholders, for example, could involve video or audio recording their responses to a few open-ended questions and analyzing the responses through a content analysis of the words that are used. Another possibility is to organize their responses into themes at a more abstract level to discover any possible associations among these themes. Lune & Berg (2017) and others present other impressionistic approaches that can be used to evaluate qualitative data collected from evaluations.

In other instances, strategies for assessing the needs in a community can be quantitative in nature. The data resulting from a questionnaire with forced-response questions or existing secondary data collected from a previous evaluation are examples. Let's assume that there is a concern that more public housing is needed for homeless people in a local area and that existing shelters keep detailed records on these people. An evaluator might be able to complete a needs assessment by developing a descriptive account or profile of the homeless population. The descriptive analysis could give some insight into the severity of the issue. The descriptive statistics could include simply calculating mean (average) scores along with standard deviations and ranges; or frequency distributions of such variables as age, gender, income, family structure, and length of time being homeless. This information would be informative to stakeholders, as it provides concrete data for making informed decisions about the needs of homeless people in the community.

Another possibility for data analysis is to compare the quantitative responses of two different groups of interviewed people. In this case, bivariate analysis would be used to help determine whether the responses of the two groups are significantly different and the amount of difference that is evident. Given the previous example of a needs assessment of prospective mental health clients, you may find that groups of prospective clients and mental health workers agree that specific types of services are needed in a satellite office somewhere in the county, but they may disagree about where the location should be. Using this information, the evaluator can report on how much consensus and difference exists in the responses of the two stakeholder groups.

Implementation Stage

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Another point in time in which data analysis is needed is during the implementation stage of an intervention. This is the stage in which the most varied types of evaluations can occur, as described in Chapter 8. A range of evaluation activities can be considered that are similar in some ways to those mentioned for the input stage. Evaluators could conduct focus groups, interviews, administer questionnaires, conduct observations, or undertake a combination of these activities. However, the intent of these evaluations is very different from those conducted during the input stage. Evaluations during the implementation stage are intended to find out whether an intervention works as planned; how well it works; and whether some components, such as a staffing issue, an intake glitch, or ambiguity in how a specific service is delivered, need to be changed or at least fine-tuned.

For example, an intake unit responsible for assessing and diagnosing clients' problems or needs in a mental health outpatient clinic may, after their assessment, refer clients to various staff members and recommend various kinds of services; yet, the staff members may treat the recommendations differently than they are proposed. Some may ignore them, others may interpret them one way, and still others may interpret them another way. Clarity and consistency are needed in how intake referrals are prepared by intake workers, understood by the treatment team, and used by them.

Because of the varied nature of evaluations that occur during the implementation stage, there are several types of data analyses that can be used to answer questions about the intervention's implementation. In the example of intake services at a mental health clinic, interviews with intake workers and treatment staff could include a set of open-ended questions to determine whether there are differences in their perspectives or understandings of procedures.

As an illustration, an open-ended question could be, "What do you see as the relationship between the recommendations provided by the intake workers and the decisions made by the treatment staff on how services are to be implemented?" Because these would be open-ended questions with qualitative responses, a content analysis could be conducted to categorize comments by keywords. The categories could then be cross-tabulated by whether they were made by intake or treatment workers. If an examination of specific words failed to reveal important distinctions in comments, then themes could be teased out of the comments at a more abstract level of analysis and cross-tabulated.

Another possibility for data analysis is to compare the quantitative responses of two different groups of interviewees. In the example of intake issues at a mental health clinic, a set of forced-response statements can be used. For example, "The intake workers' recommendations are optional to implement," and "The recommendations of intake workers are usually not feasible." Response categories for such statements could include strongly agree, agree, disagree, and strongly disagree. In this case, bivariate analysis would help determine whether the responses of the two groups are different and, if so, by how much. Using this information, the evaluator can report on how much consensus or difference exists in the responses of the two stakeholder groups on the intake workers' recommendations.

Outcome Stage

Outcome evaluations emphasize different kinds of data analysis than those used during the planning and implementation stages. Outcome studies focus on the relationships between the intervention and the desired client outcomes. This focus is used not only to determine whether there is an association between the intervention and outcome but also to go a step further and determine whether the intervention caused the outcome measure to change favorably.

For example, in a substance abuse agency, the intervention might be group treatment. The evaluator wants to conduct an outcome evaluation to find out whether the intervention is responsible for any progress in the clients' short- and long-term decreased substance use. An association will need to be found between the intervention and clients' outcomes, as well as evidence of a causal relationship. Several types of controls are needed to determine whether there is a causal relationship between group treatment and client progress in discontinuing substance use. This could include being able to rule out other environmental events in the clients' lives and other changes beyond external events. As was discussed in Chapter 9, a quasi-experimental or experimental group design would need to be used to identify any extraneous influences.

Data analysis strategies used in the outcome stage are the most complex, and several are reviewed later in the chapter to highlight when to use each strategy and what each strategy can accomplish. Each strategy also has limitations that are also reviewed. For example, some statistical strategies can be used to find out whether there is an association between an intervention and client outcomes. Other strategies can be used to determine whether a causal relationship exists and to what extent a causal relationship is evident.

SUMMARY OF PERTINENT TOOLS FOR QUALITATIVE DATA ANALYSIS

As reported earlier in this chapter, many evaluations are qualitative in nature or at least have qualitative components. They may collect data using open-ended questions, in-depth interviews, focus groups, unstructured observations, or existing data sources. Three qualitative data analysis strategies are especially relevant for such evaluations: case studies, reducing responses to fewer general categories, and theme analysis.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Case Studies

Case studies involve one or a few research participants, but they are intended to provide stakeholders with insights into an entire group of people that the case example tends to represent. Case studies are detailed portrayals of an individual that can illuminate how a larger group of people that the individual represents is functioning. Case studies are also possible with other social units such as a cultural group, a family, or neighborhood (Dudley, 2011). Case studies can be prepared using many different research strategies, such as in-depth interviews, participant observation, and existing documents. Case studies typically do not go beyond a descriptive state. A few examples of case studies follow.

A Mexican boy recently immigrated to the United States with his mother, who was an undocumented worker. The boy was having extreme difficulty adapting to his peers and teachers in the public school in his new community. He was finally suspended. He ended up in the juvenile court system after torturing and killing stray animals. A case study describing his difficulties in adapting to a U.S. school could provide valuable insights into the struggles that other immigrant children like him may face as they attempt to adapt to a new environment in the United States.

A teenager living secretly with HIV was experiencing severe depression because she did not know what to do. She did not want to seek treatment from a health clinic because she was uninformed about the treatments that could be made available to her, and she did not want to disclose her condition to her family for fear of being severely punished or disowned. To make things even worse, she had no one she could trust to talk to about what she was going through. She began to have suicidal thoughts and contemplated running away. A case study described the challenges that this teenager faced and the obstacles associated with attempting to get help; it provided valuable insights into how teenagers with HIV or AIDS could be helped in the early stages of this disease.

Case studies can be an important approach to consider because they offer insights about what might be typical, challenging, or possibly misunderstood about specific types of clients who participate in a specific program. A case study could also illustrate how a client was especially successful in participating in and using the specific services of a program. Quality assurance evaluations use the case study approach to explore how a few randomly selected clients function in a program and interact with providers. Case studies also can be used in a needs assessment to gain a fuller understanding of the social circumstances of prospective clients. Case studies used to conduct a needs assessment are described in Chapter 6.

The data analysis involved in preparing a case study begins with careful planning. A case or cases need to be selected and specific topics for exploration chosen. Such decisions are determined by carefully considering what the evaluator and other stakeholders want to accomplish and what is practically possible. Other issues are important once data have been collected. A case description should be prepared and organized to tell the story about an individual (or a family, a group, a neighborhood) in the most effective way. Constructing an outline of the story can help identify the most important topics and further questions to ask. Another task is to determine when enough data has been collected on each chosen topic or when additional data is needed to fill out the story. As a case study is prepared and all the data on topics are assembled, another consideration is to discover and describe how the topics may be related so that the story can reflect a more holistic account and perspective. It should be kept in mind that preparing a case study is an effort in creativity and art as well as science. The important thing to remember is to tell the story that needs to be told in the most creative, accurate, compelling, and comprehensive way.

Reducing Responses to Fewer General Categories

Another strategy for analyzing qualitative data is to create a set of general categories (usually no more than ten) and assign research participants' responses to these categories. This option is mostly used with data collection strategies that elicit brief qualitative answers. Qualitative methods that lend themselves to brief answers include open-ended questions in interviews and questionnaires, some questions asked in focus groups, and some semi-structured observations. Condensing many short responses into fewer general categories is useful for succinctly summarizing and organizing responses together that have similar meaning. For example, a client satisfaction study could have the question, "What do you like about living here?" asked of residents of a group home for people with developmental disabilities. The responses of a large sample would be too numerous to meaningfully grasp in their original form, so a smaller number of general categories of what they liked can be used, such as "types of food or beverages," "social activities," and "hygiene activities."

In some ways, this strategy is like a frequency distribution that summarizes the quantitative responses to a forced-response question. The major difference is that the quantitative responses are already in categorical form, whereas the qualitative responses are assigned to newly created categories based on their similarities (e.g., taking baths and brushing teeth can fit into the "hygiene activities" category). The new categories created from this qualitative strategy are usually at the prenominal level of measurement and could be used as a set of response categories for a forced-response question in a later study.

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

In review, the steps involved in reducing qualitative responses of an open-ended question into fewer newly created categories are as follows (Dudley, 2011):

1. Prepare the responses for analysis by listing all of them together under the same question.
2. Carefully review all the responses to a question to make sure that each one is understandable.
3. Code responses using the same code for responses that are similar. Coding can be done using different letters in the left-hand margin or assorted colors.
4. Group similar responses together based on your coding and give each group an overall name, which should generally describe each of the responses in a group (e.g. "hygiene activities" is the group name selected for responses such as "taking a bath," "washing in warm water," "brushing one's teeth").
5. Count the number of responses that fall into each group.
6. Report the results in quantitative form (e.g., hygiene activities, 27; social activities, 19; field trips, 7; etc.).
7. Although the groups or newly created categories provide a useful summary of the responses, it may be helpful to add in parentheses a few examples of the responses.

It is important to involve two evaluators who work independently of each other in creating the new categories and assigning responses to the categories. Two reviewers are used to ensure that there is inter-rater reliability.

Theme Analysis

A third qualitative data analysis strategy is to identify common themes or patterns that are prevalent in narrative data. Theme analysis is useful in analyzing lengthy narrative material from data collection strategies such as participant-observation, in-depth interviews, focus groups, and unstructured observations. This strategy can also be used to analyze existing data from sources such case records and group treatment summaries that are in narrative form.

Example of One Theme Deriving from the Responses of a Focus Group

A social work student (Leary, 2003) conducted a focus group study of volunteers of several churches that assisted homeless people in overnight stays at their churches during the winter months. When the volunteers began discussing the needs of the homeless people they witnessed, one theme they reported was that the homeless people wanted to have conversations with church volunteers who would listen to them and help them. They referred to this theme as a perceived need of homeless people to have meaningful discussions with volunteers. An abbreviated version of four of the comments found in the narrative that reflected this theme were as follows:

- I think one of the big needs is just to listen to them, be compassionate, and be there for them. I remember one evaluation meeting . . . when you have the homeless at the meeting with you, and I remember one of the men saying, "Just don't feed us and leave us."
- I would have loved to have one of the clergy members, deacons, priests, whatever, come and say an opening prayer with our people as it can be a show of force that this whole congregation is behind this program. . . . I think that is a really important part, to have a show of force, that clergy and staff support the program 100 percent. I think the guests actually get a sense of being special when we are all present.
- During Lenten season, when we would have Lenten meals on the same night as the shelter program, one of the things that was very interesting was that . . . they were much more likely to sit down with [one of the pastors]. Some of the issues that were shared . . . [included] offering the possibility of some counseling services and providing that there. I see that as caring, a very crucial kind of caring, to their situation, and, it's definitely a help.
- A lot of them have a lack of self-confidence; they don't believe in themselves anymore. Mainly they need someone to talk to, to reassure them they are still a human being and worth something.

Conducting a theme analysis usually takes more time, covers more data, and can be more intellectually demanding than the previous two strategies. Like the strategy of reducing responses to fewer general categories, it is important to involve two independent evaluators in coding themes to ensure inter-rater reliability. If the two people disagree on a specific theme or how it is reflected in the narrative, a third person can be involved to help reach a consensus.

The major stages involved in conducting a theme analysis include the following (Dudley, 2011):

DO NOT COPY
vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

1. Prepare narrative data for analysis by typing it into electronic files.
2. Carefully review and become familiar with all narrative data to make sure they are understandable.
3. Assign a code (e.g., using symbols or magic markers) to each point in the narrative that reflects a more general theme.
4. Group the coded narratives reflecting each theme on the same page (e.g., copy and paste to move data around the file).
5. After carefully reviewing the narrative for each theme, give the theme an accurate label in a few words, a phrase, or a sentence.

Theme analysis can involve additional steps, such as examining how the coded narrative of each theme might vary. For example, some of the narratives illustrating a theme may describe it in a positive way, while others describe it in negative terms. Identifying variations in how a theme is depicted can help further delineate meaning evident in the theme.

Example of Identifying Variations in a Theme

In one student's evaluation (Grantham, 2001) of the needs of patients in a hospice setting, several people identified the theme of having legal issues. On further examination, she found that variations in the comments about legal issues seemed to almost always refer to one of three things: financial problems, a legal will, or advanced directives.

Another step that can be taken in a theme analysis involves capturing how different themes may be related to one another according to what the narrative suggests for each respondent. In these instances, relationships among themes can be explored in follow-up evaluations as hypotheses.

SUMMARY OF PERTINENT TOOLS FOR QUANTITATIVE DATA ANALYSIS

Descriptive Analyses

It is important for evaluators to ask descriptive questions to understand more about the clients participating in a program. For example, what are the ages of the clients? What is the educational level of those who complete the program? What is the demographic profile of those who drop out of the program? Other descriptive questions might be, what is the level of stress of the clients? How many children attended the program? What is the educational level of participants who do not complete the program? And, how many sessions did both the fathers and mothers participate in together?

Descriptive analysis, or univariate analyses, is used to answer descriptive questions. The analyses are considered univariate because the descriptive question focuses on only one variable at a time. Univariate questions provide a means of describing something about the sample, such as its variability or range. The level of measurement of the variables in descriptive research will inform you about which univariate analyses to use. If the level of measurement of the variable is ordinal or nominal, for example, then percentages and frequency distributions are appropriate. If variables are at the interval/ratio level, then you can report the mean, range, and standard deviation. (See Table 10.1.)

Table 10.1. Review of Levels of Measurement

LEVEL	EXAMPLES	EXAMPLES OF MEASUREMENT PROCEDURES	MATHEMATICAL OPERATIONS PERMITTED
Nominal	Gender, ethnicity, political affiliation	Categories	Counting the number in each category and comparing sizes of different categories
Ordinal	Likert scales, social class, military rank	Categories and ranking with respect to one another	All the above plus statements of greater than and less than
Interval/ratio	Age, years of education, household income	All the above plus a description of numerical distances between scores	All the above plus all other mathematical operations

Levels of Measurement

Nominal-level data are the simplest form in which numerical labels are used. Nominal data are categorical, and categories are distinct, mutually exclusive, and exhaustive of all possible circumstances (e.g., category of gender is either female or male). The categories or scores in nominal level data cannot be ranked, ordered, or manipulated mathematically (e.g., gender cannot be ranked with one being higher than the other).

A higher, more sophisticated level of measurement is ordinal-level data. Ordinal data are ranked or ordered from lowest to highest. A Likert scale that ranges from strongly disagree to strongly agree is an ordinal-level measures. The limitation

DO NOT COPY vaterriamannie@gmail.com

Printed by: vaterriamannie@gmail.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

to highest. A Likert scale that ranges from strongly disagree to strongly agree is an ordinal level measures. The limitation

with ordinal-level data is that there is no way to determine the distance between the various levels. For example, it is not possible to calculate a numeric distance between strongly disagree and disagree.

Although there are four levels of measurement, we do not usually differentiate between interval and ratio data. As a result, interval and ratio data are combined and presented together. Unlike nominal and ordinal measures, interval/ratio data permit classification, ranking, and signifying the exact distance between scores to be defined. Measures such as age in years, number of days on a job, or annual income are interval/ratio data.

Frequency Distributions

A frequency distribution is a statistic that describes the number of times that each value of a variable is observed in a sample. For example, in a question using a Likert scale, a frequency distribution describes how many participants selected "very satisfied," "satisfied," "dissatisfied," and so on. Frequency distributions can be presented as actual frequencies or as percentages. Usually percentages are easier to grasp with the large samples. Often, however, frequency distributions list both frequencies and percentages. Frequency distributions are an appropriate statistic for use with categorical variables, including both nominal and ordinal variables. A frequency distribution could also be used with an interval/ratio variable if the interval/ratio responses are reconstituted and reduced to categories.

Measures of Central Tendency

Three types of measures of central tendency are possible to use in evaluations: *mean, median, and mode*. The mode is the easiest measure to use, as it is the most frequently occurring response or value of a variable. In addition, the mode can be used regardless of the level of data measurement. The median is the middle point, or the fiftieth percentile of the data, and can be used with ordinal and interval/ratio data.

The most commonly reported measure of central tendency in research is the mean. Keep in mind that the mean score is an average and that its interpretation is limited, as it can be extremely sensitive to outliers. As a result, the mean score might not be very close to most scores in a sample. This can be problematic when the mean is reported by itself. For example, we often see the mean score reported for a standardized test, income, or housing prices. As a result, we should be mindful that stakeholders may need an explanation of a mean score and its limitations. When analyzing interval/ratio data, we should report not only the mean but also the range or standard deviation. This added measure will give stakeholders an idea of how the data are distributed and will make the mean score more meaningful.

Measures of Variability

Measures of variability can be either a range or a standard deviation. The *range* of scores is the lowest and highest score. The range is important to report when presenting interval/ratio data, as the mean score can be sensitive to extremes in scores. Knowing the range of scores can indicate some information about the distribution of scores and help determine whether the mean approximates most people in the sample. For example, if the mean age of mothers in a support group is thirty years, but the range is from nineteen to sixty-five years, the mean score may not be close to some of the women's age in the group, as a few sixty-five-year-old women can substantially increase the mean age of the sample.

When you add the *standard deviation* to the presentation, it gives a clearer picture of the sample distribution. The standard deviation is a measure of variability that informs how all the scores relate to one another and to the mean. The standard deviation is useful in interpreting the percentage of scores that are close to the mean score. In normally distributed data, 68 percent of scores fall within plus or minus one standard deviation of the mean. Further, 95 percent of the scores fall within two standard deviations of the mean, and 99 percent fall within three standard deviations. Returning to the age of mothers, if the data are normally distributed and the mean age is thirty and the standard deviation is three, then you would expect that 68 percent of mothers are between the ages of twenty-seven and thirty-three (30 ± 3, where 3 equals one standard deviation). Similarly, 95 percent of mothers in the study are between twenty-four and thirty-six years old (30 ± 6, where 6 equals two standard deviations). Further, you could conclude that 99 percent of mothers are between

DO NOT COPY vaterriamannie@gmail.com