

# THE PRACTICE OF RESEARCH IN SOCIAL WORK

Rafael J. Engel  
Russell K. Schutt

4e

Do not copy, post, or distribute

Copyright ©2017 by SAGE Publications, Inc.

This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.



# Single-Subject Design

## Foundations of Single-Subject Design

*Repeated Measurement*

*Baseline Phase*

Patterns

Internal Validity

*Treatment Phase*

*Graphing*

## Measuring Targets of Intervention

*What to Measure*

*How to Measure*

## Analyzing Single-Subject Designs

*Visual Analysis*

Level

Trend

Variability

*Interpreting Visual Patterns*

*Problems of Interpretation*

## Types of Single-Subject Designs

*Basic Design: A-B*

*Withdrawal Designs*

A-B-A Design

A-B-A-B Design

*Multiple Baseline Designs*

*Multiple Treatment Designs*

*Monitoring Client Progress*

*Single-Subject Designs*

*With Group Interventions*

**Research in the News: Exercise Your Brain?**

Implications for Evidence-Based Practice

Single-Subject Design in a Diverse Society

Ethical Issues in Single-Subject Design

Conclusion

Key Terms

Highlights

Discussion Questions

Practice Exercises

Web Exercises

Developing a Research Proposal

A Question of Ethics

*Jody was a 40-year-old married White female who had been ill since 22 years of age. She had multiple hospitalizations including 1 year in the state hospital. In the past year, she had six hospitalizations. She suffered from hallucinations, delusions, psychomotor retardation, apathy, flat affect, and avoidance of people and social situations. She was suicidal, inactive, and unable to do minimal self-care. She lived with her husband, who took care of her. She was a high school graduate and had attended 1 year of training as a dental assistant before her illness forced her to drop out of school. She took neuroleptic medications interspersed with brief periods of noncompliance due to paranoid ideas about the medicine. (Bradshaw, 1997, p. 438)*

It is not unusual for social work practitioners to have clients such as Jody who have a mental health condition such as schizophrenia. As practitioners, we often think we know when a client is improving. Yet, when we use our own subjective conclusions, we are prone to human error. In this chapter, you learn how single-subject designs can be used to systematically test the effectiveness of a particular intervention as William Bradshaw did with Jody and subsequently with other participants (Bradshaw 1997, 2003; Bradshaw & Roseborough, 2004), as well as how it can be used to monitor client progress.

Single-subject (sometimes referred to as single-case or single-system) designs offer an alternative to the group designs described in Chapter 7. The focus is on a single participant whether that participant is an individual, a group, or a community, and yet, as we will illustrate, single-subject designs can be used with groups. The structure of these designs, which are easily adapted to social work practice, makes them useful for research on interventions in direct and community practice. The process of assessment, establishing intervention goals and specific outcomes, providing the intervention, and evaluating progress prior to termination, parallels the process of using single-subject designs, which depend on identifying the focus of the intervention, taking preintervention measures, providing the intervention, taking additional measures, and making decisions about the efficacy of the intervention. Because of these parallels, single-subject designs can be used not just for research but also adapted to evaluating practice and monitoring client progress.

Contrast this design with group designs. In Chapter 7, we noted that group designs do not naturally conform to practice, particularly when the practice involves interventions with individuals. The analysis of these designs typically refers to the “group’s average change score” or “the number of subjects altering their status.” By describing the group, we miss each individual’s experience with the intervention. Once a group design is implemented, it is difficult to change the nature of the treatment, yet individual participants within the group may not respond to the particular type of treatment offered.

In this chapter, we begin by taking you through the basic features of single-subject designs. Next, we describe ways to measure changes in the case that is the target of the intervention. We then describe different designs and connect them to their uses for social work research, practice evaluation, and client monitoring. Finally, we end the chapter with a discussion about the implications of single-subject designs for evidence-based practice and the ethical issues associated with single-subject designs.

## 2 Foundations of Single-Subject Design

The underlying principle of a single-subject design as a social work research method is that if an intervention with a client, a group of clients, an agency, or a community is effective, it should be possible to see a change in status from the period prior to intervention to the period during and after the intervention. At a minimum, single-subject research designs include

- repeated measurements to identify a client’s status,
- the baseline phase or the time period prior to the start of the intervention, and
- the treatment phase or the time period during the intervention.

Furthermore, the baseline and treatment phase measurements are usually displayed using graphs.

## Repeated Measurement

Single-subject designs require the repeated measurement of a dependent variable, that is, the target or the focus of the intervention such as a status, condition, or problem. The target is measured at regular time intervals such as hours, days, weeks, or months, prior to the intervention and during the intervention. The preferred method is to take measures of the target with the client prior to implementing the intervention, for example, during the assessment process, and then continue during the course of the intervention. Gathering information may mean withholding the intervention until the repeated measures can be taken. Alternatively, repeated measures of the dependent variable can begin when the client is receiving an intervention for other concerns. For example, a child may be seen for behavioral problems, but eventually communication issues will be a concern. The repeated measurement of the communication issues could begin prior to that specific intervention focus.

Sometimes it is not possible to delay the intervention either because there is a crisis or because to delay would be unethical. Yet you may still be able to construct a set of preintervention measures using data already collected or by asking about past experiences. Client records may have information from which a baseline can be constructed, although you are limited to the information that is available on the record. Some client records, such as report cards, may have complete information, but other client records, such as case files, may or may not. Another option is to ask the client or, if permission is granted, ask significant members of the client's network about past behaviors. Trying to construct measures by asking clients or family members assumes that the information is both remembered and reported accurately. Generally, behaviors and events are easier to recall than moods or feelings, yet the recall of even behaviors or events becomes more difficult with the passage of time and probably should be limited to the preceding month. Although recognizing the limits of these retrospective data collection methods is important, the limitations should not preclude using the information if that is all that is available, particularly for evaluating practice.

There are other times when using retrospective data is feasible. Agencies often collect quite a bit of data about their operations, and these data can be used to obtain repeated measurements. For example, if an agency director was trying to find an outreach method that would increase the number of referrals, previous monthly referral information could be used and the intervention begun immediately. Or if an organizer was interested in the impact of an empowerment zone on levels of employment in a community, the preintervention employment data are likely to exist.

## Baseline Phase

The **baseline phase** (abbreviated by the letter A) represents the period in which the intervention to be evaluated is not offered to the respondent. During the baseline phase, repeated measurements of the dependent variable are taken or reconstructed. These measures reflect the status of the client (or agency or community) on the dependent variable prior to the intervention. The baseline phase measurements provide two aspects of control analogous to a control group in a group design. First, in a group design, we expect the treatment group to have different scores than the control group after the intervention. In a single-subject design, the subject serves as the control as the repeated baseline measurements establish the pattern of scores that we expect the intervention to change. Without the intervention, researchers assume that the baseline pattern of scores would continue its course. Second, in a control group design, random assignment controls for threats to internal validity. In a single-subject design, the repeated baseline measurements allow the researcher to discount most threats to the internal validity of the design.

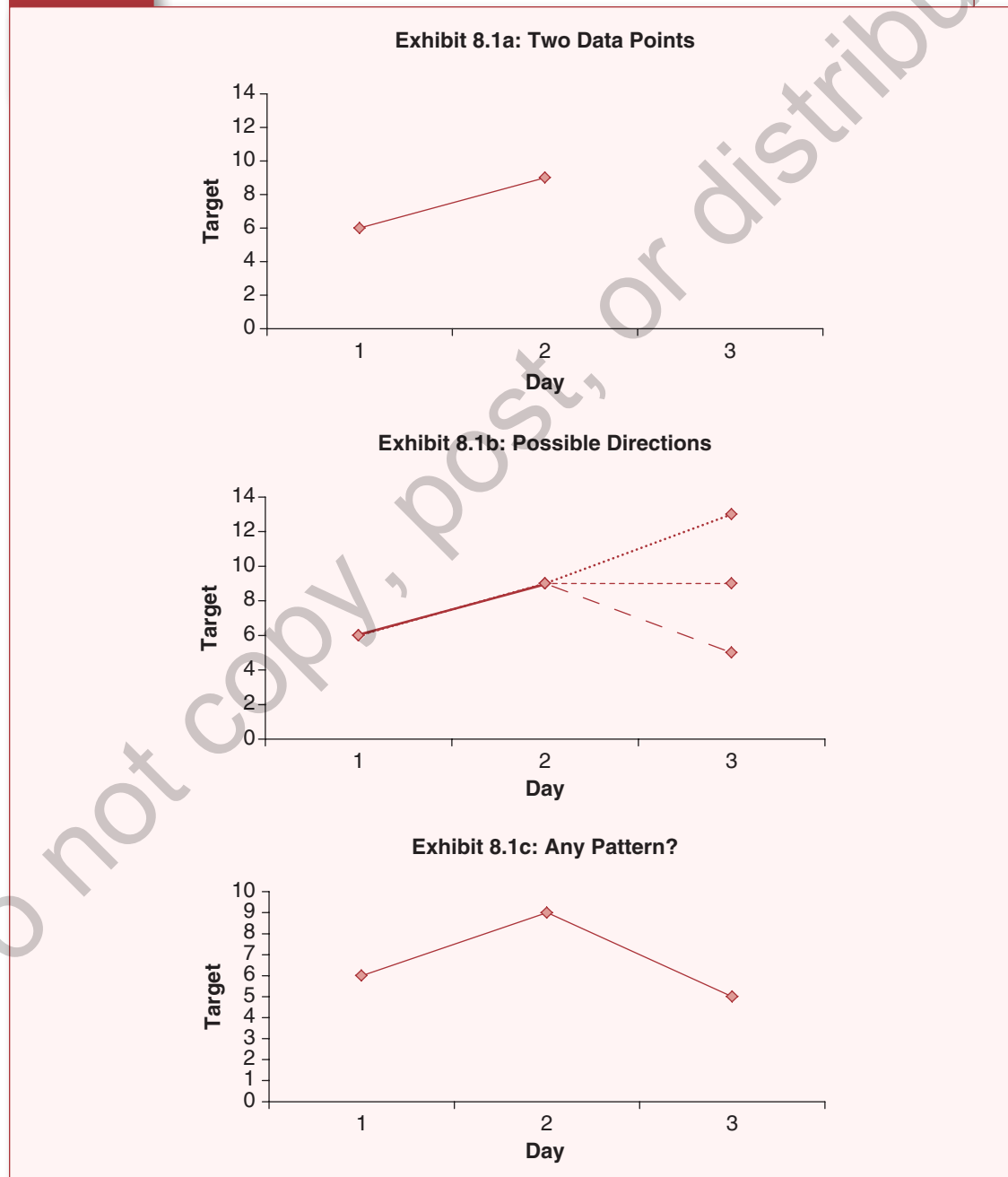
**Baseline phase** The initial phase of a single-subject design, typically abbreviated by the letter A; it represents the period in which the intervention to be evaluated is not offered to the subject. During the baseline phase, repeated measurements of the dependent variable are taken or constructed.

## Patterns

In the baseline phase, measurements are taken until a *pattern* emerges. You have found a pattern when you can predict with some certainty what might be the next score. Predicting the next score requires a minimum of three observations

in the baseline stage. When there are only two measures, as shown in Exhibit 8.1a, can you predict the next score with any certainty? The next data point could be higher, lower, or the same as the previous data points (see Exhibit 8.1b). With three measures, your certainty increases about the nature of the pattern. But even three measures might not be enough depending on the pattern that is emerging. In the graph shown in Exhibit 8.1c, is the pattern predictable? You probably should take at least two more baseline measures, but three or four additional measures may be necessary to see

### Exhibit 8.1 Predicting a Pattern



a pattern emerge. As a general rule, the more data points, the more certain you will be about the pattern; it takes at least three consecutive measures that fall in some pattern for you to have confidence in the shape of the baseline pattern.

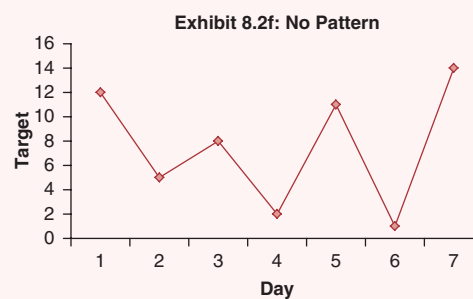
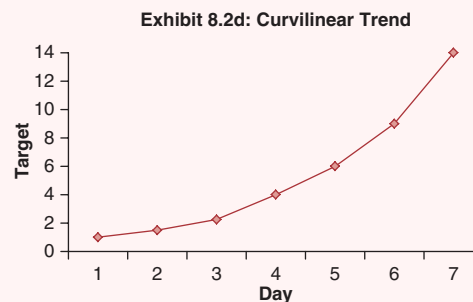
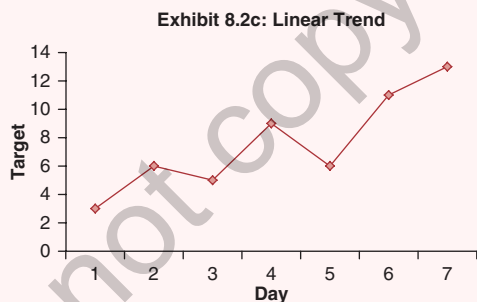
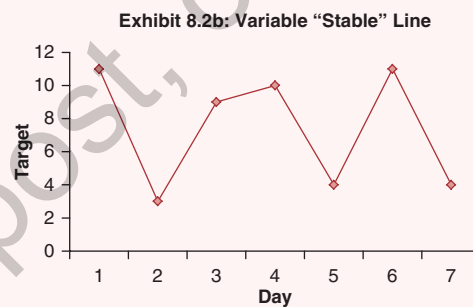
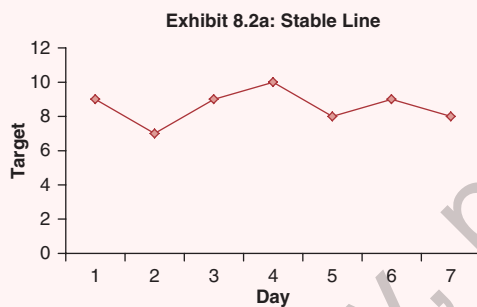
The three common types of patterns are a stable line, a trend line, and a cycle. A **stable line** (see Exhibit 8.2a) is a relatively flat line, with little variability in the scores so that the scores fall in a narrow band. This kind of line is desirable because changes can easily be detected, and it is likely that there are few problems of testing, instrumentation, statistical regression, and maturation in the data. A wider band or range of scores (see Exhibit 8.2b) is more difficult to interpret than a stable line with little variation.

A **trend** occurs when the scores may be either increasing or decreasing during the baseline period. When there is a linear trend (see Exhibit 8.2c), the scores tend to increase (or decrease) at a more or less constant rate over time. A trend may also be curvilinear (see Exhibit 8.2d) so that the rate of change is accelerating over time rather than increasing or decreasing at a constant rate.

**Stable line** A line that is relatively flat, with little variability in the scores so that the scores fall in a narrow band.

**Trend** An ascending or descending line.

### Exhibit 8.2 Different Baseline Patterns



**Cycle** A pattern reflecting ups and downs depending on time of measurement.

A **cycle** (see Exhibit 8.2e) is a pattern in which there are increases and decreases in scores depending on the time of year (or week or month) when the measures are taken. For example, homeless shelter occupancy varies by the time of year with increased use in winter months and decreased use in summer months.

There are times when no pattern emerges (see Exhibit 8.2f). With such baseline patterns, it is important to consider the reasons for the variability in scores. Is it due to the lack of reliability of the measurement process? If so, then an alternative measure might be sought. The client may be using a good measure, but not reporting information consistently, such as completing a depression scale at different times of day. Or the variability in scores may be due to some changing circumstance in the life of the client.

### Internal Validity

Findings of causality depend on the internal validity of the research design. Identifying patterns by taking repeated baseline measurements can control for several threats to the internal validity of single-subject designs. Specifically, problems of maturation, instrumentation, statistical regression, and testing may be evaluated because patterns illustrative of these threats to internal validity should appear in the baseline. When the measurement in the baseline phase is reconstructed from existing data or memory, these threats to internal validity are problematic.

When baseline measures are stable lines, these threats may be ruled out, but it is more difficult to rule out some threats if the pattern is a trend, particularly if the trend is in the desired direction. For example, if maturation is a problem, you would expect to find a trend line and not a horizontal line. Perhaps you have a client who has suffered a loss and you are measuring sadness. At the first baseline measurement, you find a high degree of sadness. At the second baseline measurement a week later, you find a lower level of sadness, and a third baseline measurement finds sadness has declined below the level of the second week. This pattern suggests that there may be a maturation effect as improvement is in the desired direction and occurs before the intervention. This does not mean that an intervention would not be effective, but it may be more difficult to demonstrate its effectiveness.

If statistical regression and testing effects occur, the impact is likely to appear initially in the baseline measures. A high score obtained from a measurement may be lower in a second measurement because of statistical regression or because of the respondent's acclimation to the measurement process. If there were only one or two baseline measures, the line might reflect these effects. But with multiple measures, the effect of statistical regression should occur in the beginning of measurement, and continued measurement should produce a baseline pattern. The testing effect should be observable early in the baseline measurement process, as the subject adjusts to the testing requirements, and not with subsequent measurement.

The most significant threat to internal validity is history. Repeated measurement in a baseline will not control for an extraneous event (history) that occurs between the last baseline measurement and the first intervention measurement. The longer the time period between the two measurement points, the greater the possibility that an event might influence the participant's scores. At the end of the study, the researcher should check with participants to determine whether some other event may have influenced the results.

### Treatment Phase

The **treatment phase** (signified by a B) represents the time period during which the intervention is implemented. During the treatment phase, repeated measurements of the same dependent variable using the same measures are

**Treatment phase** The intervention phase of a single-subject design.

obtained. Ultimately, the patterns and magnitude of the data points are compared to the data points in the baseline phase to determine whether a change has occurred. The length of the treatment phase should be as long as the baseline phase (Barlow, Nock, & Hersen, 2009; Tripodi, 1994).

## Graphing

The phases of a single-subject design are usually summarized on a graph. Graphing the data facilitates monitoring and evaluating the impact of the intervention. The  $y$ -axis represents the scores of the dependent variable, whereas the  $x$ -axis represents a unit of time, such as an hour, day, week, or month. Although you may make your graph by hand, both statistical software and spreadsheet software have the capacity to present data on graphs. For example, Deochand and colleagues (2015) provide simple instructions to using Microsoft Excel 2013 to create graphs for single-subject designs.

Care in graphing is important as the act of graphing can create visual distortions that can lead to inaccurate conclusions. When making a graph, it is important to make the axes as proportionate as possible in order to minimize distortions. You want to ensure that the graph will visually represent small but meaningful changes.

## 2 Measuring Targets of Intervention

Three questions to answer about measurement are (1) what target to measure, (2) how to measure the target, and (3) who will do the measuring. With each decision, there are important issues to consider. There should be some certainty based on theoretical literature, empirical support, or practice experience to suggest that the chosen intervention is an appropriate method to address the target.

### What to Measure

The dependent variable in a single-subject design is the target, that is, the concern or issue that is the focus of the intervention. The target and intervention are usually established as part of the research project and specified prior to any client interaction. In contrast, social work practitioners using single-subject design methods to evaluate practice or monitor their work typically arrive at the target problem through their interaction with clients or client systems. Clients may start with some general problem or need that, through the processes of assessment and discussion, becomes narrowed to a specific set of treatment goals. Similarly, a community organizer may identify the general needs of a community, and through discussion and meetings, specific outcomes are identified.

The target may focus on one specific concern or different aspects of that concern. For example, with an adolescent who is having behavioral problems in school, you may decide to measure the frequency of the behavioral problems. Or you may hypothesize that the adolescent's behavioral problems are caused by poor family communication and low self-esteem, so you may choose to measure all three concerns. The target problems can be measured simultaneously or sequentially.

Single-subject design is applicable to other systems as well, such as agencies and communities. An agency director may decide to evaluate the efficacy of different methods to improve agency functioning or examine the extent to which a community-based program produces changes in the community. The choice of the target becomes a question of determining the information that is important to the agency or community.

### How to Measure

Once the target or outcome of the intervention has been identified, you must determine how you will measure it. Generally, in a research study, operationalization occurs prior to the beginning of the study; when practitioners evaluate practice or monitor clients, operationalization takes place through client and practitioner discussion. For example, if you are evaluating the impact of positive parenting techniques on altering a child's behavior, you would identify jointly with the parents a behavior such as tantrums. You would then guide the parents to be able to distinguish a tantrum



from other behaviors or verbal expressions. This engagement is particularly important because there may be gender and ethnic differences in how a general problem may manifest itself (Nelson, 1994).

Measures of, status, or functioning are often characterized in four ways: frequency, duration, interval, and magnitude:

- **Frequency** refers to counting the number of times a behavior occurs or the number of times people experience different feelings within a particular time period. Based on the prior example, you could ask the parents to count the number of tantrums their child had each week. Frequency counts are useful for measuring targets that happen regularly, but counting can be burdensome if the behavior occurs too often. However, if the behavior happens only periodically, the counts will not be meaningful.
- **Duration** refers to the length of time an event or some symptom lasts and usually is measured for each occurrence of the event or symptom. Rather than counting the number of tantrums in a week, the parents could be asked to time the length of each tantrum. The parents would need a clear definition as to what constitutes the beginning and end of a tantrum. A measure of duration requires fewer episodes than do frequency counts of the target.
- The **interval**, or the length of time between events, may be measured. Using a measure of interval, the parents in our example would calculate the length of time between tantrums. Just as a clear operational definition was necessary for the duration measure, the parents would need a clear definition when measuring the interval between tantrums. This kind of measure may not be appropriate for events or symptoms that happen frequently unless the intent of the intervention is to delay their onset.
- The **magnitude** or intensity of a particular behavior or psychological state can be measured. A scale might be used by the parents to rate or score the intensity of the tantrum—how loud the screaming is, whether there is rolling around on the floor or hitting, and the like. Often magnitude or intensity measures are applied to psychological symptoms or attitudes such as measures of depressive symptoms, quality of peer interactions, or self-esteem.

**Frequency** In a single-subject design, counting the number of times a behavior occurs or the number of times people experience different feelings within a particular period.

**Duration** The length of time an event of some symptom lasts; usually it is measured for each occurrence of the event or symptom.

**Interval** Used in single-subject design, a measure of the length of time between events, behaviors, or symptoms.

**Magnitude** Measuring the intensity of a particular behavior or psychological state.

Social work researchers and practitioners have a variety of alternative methods available to measure the target problem. Standardized scales cover a wide range of psychological dimensions, family functioning, individual functioning, and the like. Another option is to collect data based on clinical observations. Observations are particularly useful when the target problem involves a behavior. A third option is to develop measures within the agency, such as a goal attainment scale. Regardless of how the data are collected, the measures should be reliable and valid. In particular, the reliability and validity of the instruments should have been tested on subjects of the same age, gender, and ethnicity as the client who is the focus of the single-subject design (Nelson, 1994).

## Who Should Measure

It is important to consider who will gather the data and to understand the potential consequence of each choice. Participants or clients can be asked to keep logs and to record information in the logs. Participants can complete instruments at specified time points, either through self-administration or an interview; or the social work researcher may choose to observe the participant's behavior.

A particular problem in gathering the data is the issue of reactivity. You do not want the process of collecting data to influence a subject's behavior. If you ask a subject to keep a log and record each time a behavior occurred, the act of keeping the log may reduce the behavior. Observing a father interacting with his children might change the way he behaves with them. Staff, knowing that supervisors are looking for certain activities, may increase the number of those activities. Changes due to reactivity may be short in duration and observable in the baseline, so repeated measurements

in the baseline might mitigate this problem (Tripodi, 1994). Nonetheless, it is important to recognize that there might be reactivity and to choose methods that limit reactivity.

Yet reactivity is not always a problem for either research or practice. If you were testing an intervention to improve a father's interaction skills with his children and you decided to observe the interactions, reactivity is likely to occur. The father, knowing that he is under observation, is likely to perform at his best. But in this case, reactivity is useful for the researcher who wants to see what the father thinks is the best way to interact with his children. It could be that the "best" is not very good, and the intervention would work on improving those skills.

Reactivity may have clinical utility for practice interventions. If keeping a log enhances the impact of the intervention, then this finding could be integrated into the actual intervention. But we would still have to test whether different methods of gathering data produce different outcomes.

Another measurement concern is the feasibility of the measurement process. Repeatedly taking measures can be cumbersome, inconvenient, and difficult. Is it going to be possible to use the method time and time again? Is the method too time-consuming for the subject and/or the researcher or practitioner? Will continuous measurements reduce the incentive of the subject to participate in the research or treatment?

The choice of measurement must be sensitive enough to detect changes. If the measuring scale is too global, it may be impossible to detect incremental or small changes, particularly in such target problems as psychological status, feelings, emotions, and attitudes. In addition, whatever is measured must occur frequently enough or on a regular basis so that repeated measurements can be taken. Unless the research is designed to last a long time, it will be impractical to take repeated measures of infrequent events or behaviors.

To assess the influence of cognitive-behavioral therapy with Jody and other participants, Bradshaw (1997) measured three outcomes simultaneously: symptomatology, psychosocial functioning, and hospitalizations. The symptoms measure was based on the severity of the symptoms (magnitude), the psychosocial functioning measure had behavioral anchors and was characterized by magnitude, and hospitalizations were the actual number that occurred in each month (frequency). The rating scales had been evaluated previously for their reliability and validity. To assure reliability and validity, the ratings were done by a psychotherapist and two case managers trained to use the measures. Each client was rated by two observers (usually the psychotherapist and one case manager), providing a measure of interrater reliability. Information about the hospitalizations was obtained from the clients. Consistent with a focus on practice, Bradshaw (1997) measured outcomes specific to the client; in Jody's case, he measured social supports, employment, and her comfort with her sexuality.

## 2 Analyzing Single-Subject Designs

How might we analyze data from a single-subject design? One option is to use a statistical technique such as the two-standard deviation-band, chi-square analysis, or time series to analyze the data (see Barlow et al., 2009; Bloom, Fischer, & Orme, 2009; Borckardt et al., 2008). Charles Auerbach and Wendy Zeitlin Schudrich (2013) have developed open-source software, SSD for R, incorporating visual graphic display capacity and a range of statistical tests for single-subject designs. In this chapter, we concentrate on visual analysis.

### Visual Analysis

Visual analysis is the process of looking at a graph of the data points to determine whether the intervention has altered the subject's preintervention pattern of scores. Three concepts that help guide visual inspection are level, trend, and variability.

#### Level

**Level** relates to the amount or magnitude of the target variable. Has the amount of the target variable changed from the baseline to the intervention period? Changes in level are used when the observed scores fall along relatively stable lines. The

**Level** Flat lines reflecting the amount or magnitude of the target variable; used in a single-subject design.

simple method is to inspect the actual data points to see if the intervention data points differ from the baseline points. In Exhibit 8.3a, it appears that the actual amount of the target variable—*anxiety*—has decreased.

Alternatively, the level of the phase scores may be summarized by drawing a line at the typical score for each phase separately. For example, the level may be summarized into a single observation using the mean (the average of the observations in the phase) or the median (the value at which 50% of the scores in the phase are higher and 50% are lower). The median is typically used in place of the mean when there are outliers or one or two extreme scores that greatly alter the mean. The mean of the baseline scores is calculated, and a horizontal line is drawn across the baseline phase at the mean. Then the mean of the intervention scores is calculated, and a horizontal line is drawn at the mean score across the intervention phase (see Exhibit 8.3b). The summary line for the baseline phase is compared to the summary line for the intervention phase.

Changes in level are typically observations that fall along relatively stable flat lines and are not used when the baseline pattern is either ascending or descending. Imagine the case, displayed in Exhibit 8.3c, where there is an ascending trend in the baseline phase and a descending trend in the intervention phase. As you can see, the direction has changed, but the mean for each phase may not have changed or may have changed only insignificantly.

### Trend

When the baseline data points reflect trend lines, there are two different questions depending on the nature of the lines in the treatment phase. The first question is, Has the intervention altered the direction of the trend? If the baseline trend line is ascending, is the trend line descending in the treatment phase? When the direction does not change, you may be interested in whether the rate of increase or decrease in the trend has changed. You might ask the second question: Does the intervention alter the slope, that is, the rate of increase or decrease of the line?

Visually inspecting the lines might provide an answer but trends can also be represented by summary lines. One approach is to use ordinary least squares (OLS) regression to calculate a regression line that summarizes the scores in the baseline and another regression line to summarize the scores in the intervention phase. The baseline OLS regression line is extended into the intervention phase, and the two lines are visually examined to determine whether the trend has changed (see Exhibit 8.4a). The increasing level of anxiety reflected in the baseline has stopped and the level of anxiety has dropped. Since the calculation is quite complicated, statistical software can be used to produce OLS regression lines.

William Nugent (2000) suggests a simpler approach to represent the trend in a phase. When the trend is linear (as opposed to curvilinear), draw a straight line connecting the first and last data points in the baseline phase with an arrow at the end to summarize the direction. A similar line would then be drawn for the points in the intervention phase. These two lines could then be compared. In the case of an outlier, the line should be drawn either from the second point to the last point if the first point is the outlier or from the first point to the second to last point if the last point is the outlier. The same methods can be used to summarize nonlinear trends except that two lines are drawn, one representing the segment of the first point to the lowest (or highest) point and the second line from the lowest (or highest) point to the last data point.

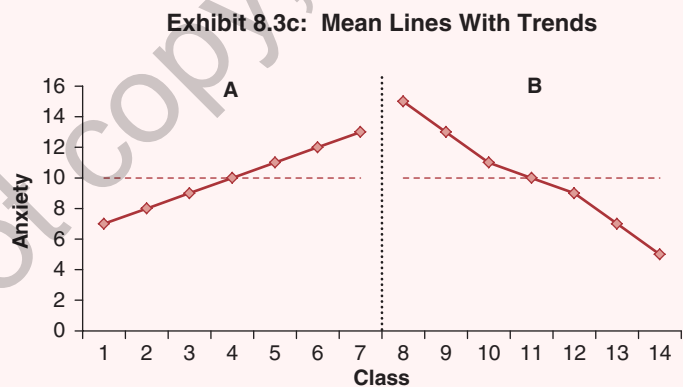
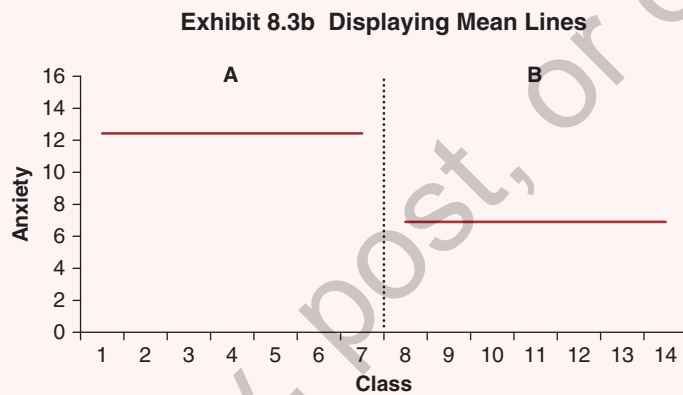
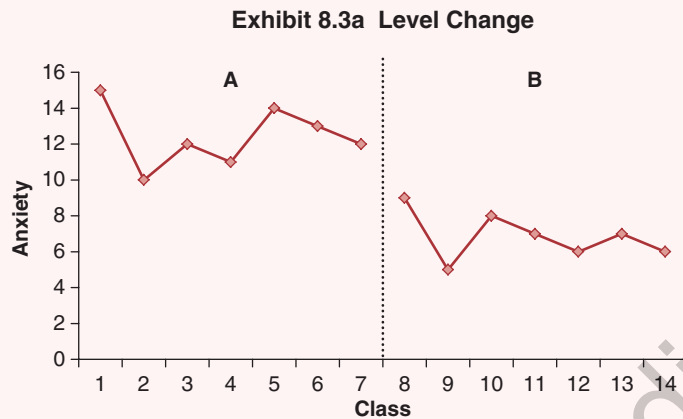
Exhibit 8.4b illustrates the use of Nugent's method. A line was drawn through the first and last time points in the baseline; this line was extended into the intervention phase. A similar line was drawn through the first and last time points in the intervention phase. A comparison of the lines suggests that the level of anxiety was no longer increasing, but had stabilized at a much lower score.

### Variability

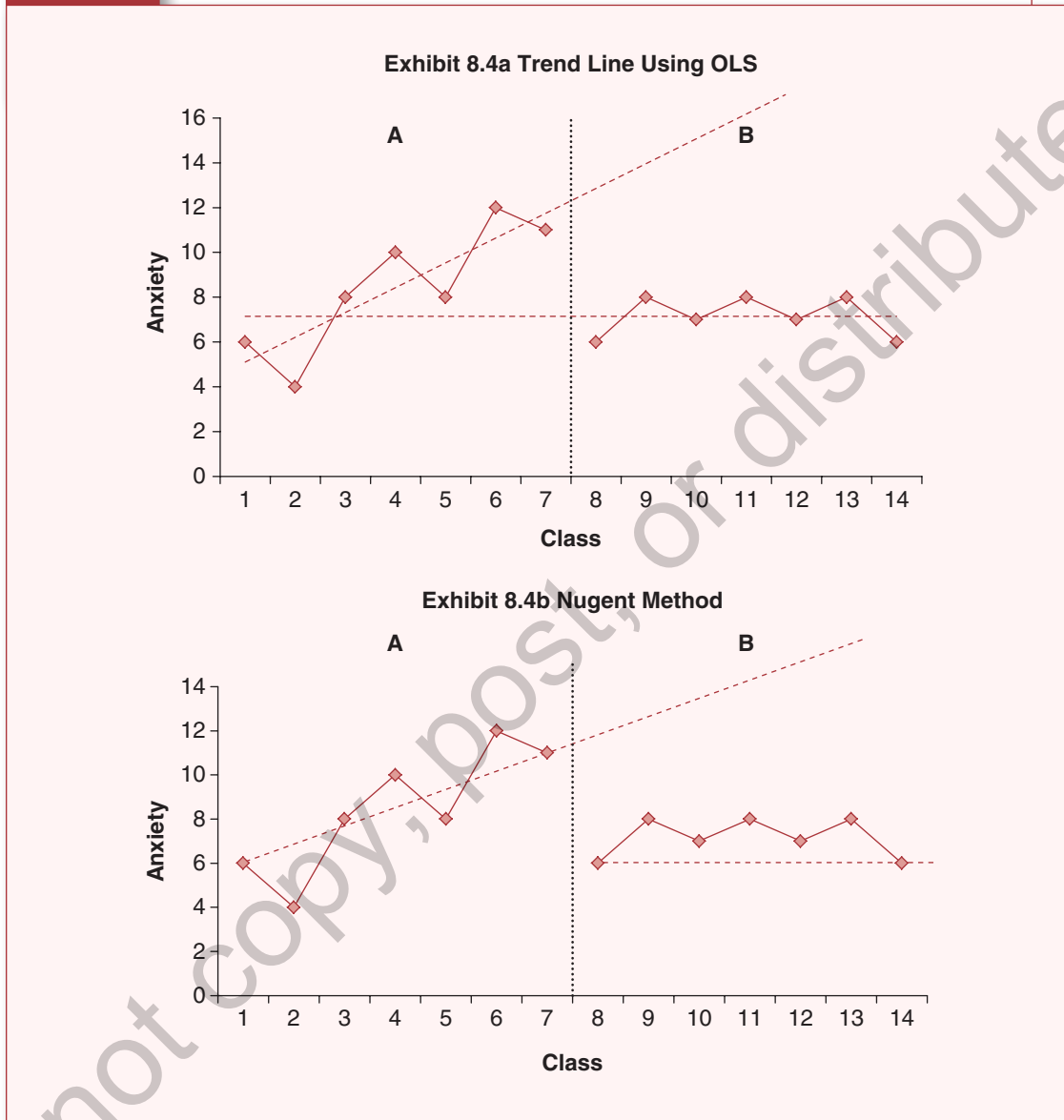
**Variability** The extent to which cases are spread out through the distribution or clustered in just one location.

The **variability** of scores, or how different the scores are within the baseline and intervention phases, may be evaluated. Widely divergent scores in the baseline make the assessment of the intervention more difficult, as do widely different scores in the intervention phase. There are some conditions for which the lack of stability is the

## Exhibit 8.3 Level



problem, so creating stability may represent a positive change. One way to summarize variability with a visual analysis is to draw range lines (see Exhibit 8.5). Whether the intervention had an effect depends on what goal was established with the client. As you can see in this graph, the only change has been a reduction in the spread of the points. But this does not mean that the intervention has not been effective because it depends on the goal of the intervention.

**Exhibit 8.4** Displaying Trend Lines**Interpreting Visual Patterns**

What conclusions might be made from level and trend patterns? Exhibit 8.6a displays a situation in which there is a stable line (or a close approximation of a stable line), so the *level* of the target is of interest. The target in this exhibit is the amount of anxiety, with lower scores being desired. For Outcome A, the intervention has only made the problem worse, for Outcome B the intervention has had no effect, and Outcome C suggests that there has been an improvement, although the effects of history may explain the change. Exhibit 8.6b illustrates outcomes of a stable baseline and a trend line in the treatment phase. Outcome D represents a deteriorating trend, while Outcome E reflects an improving trend.

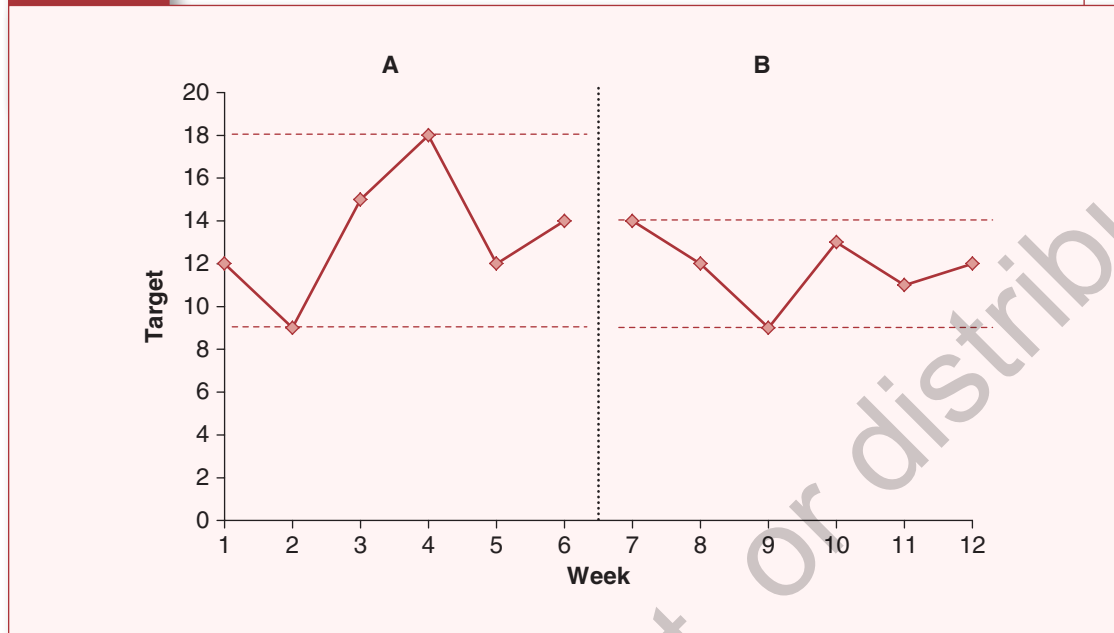
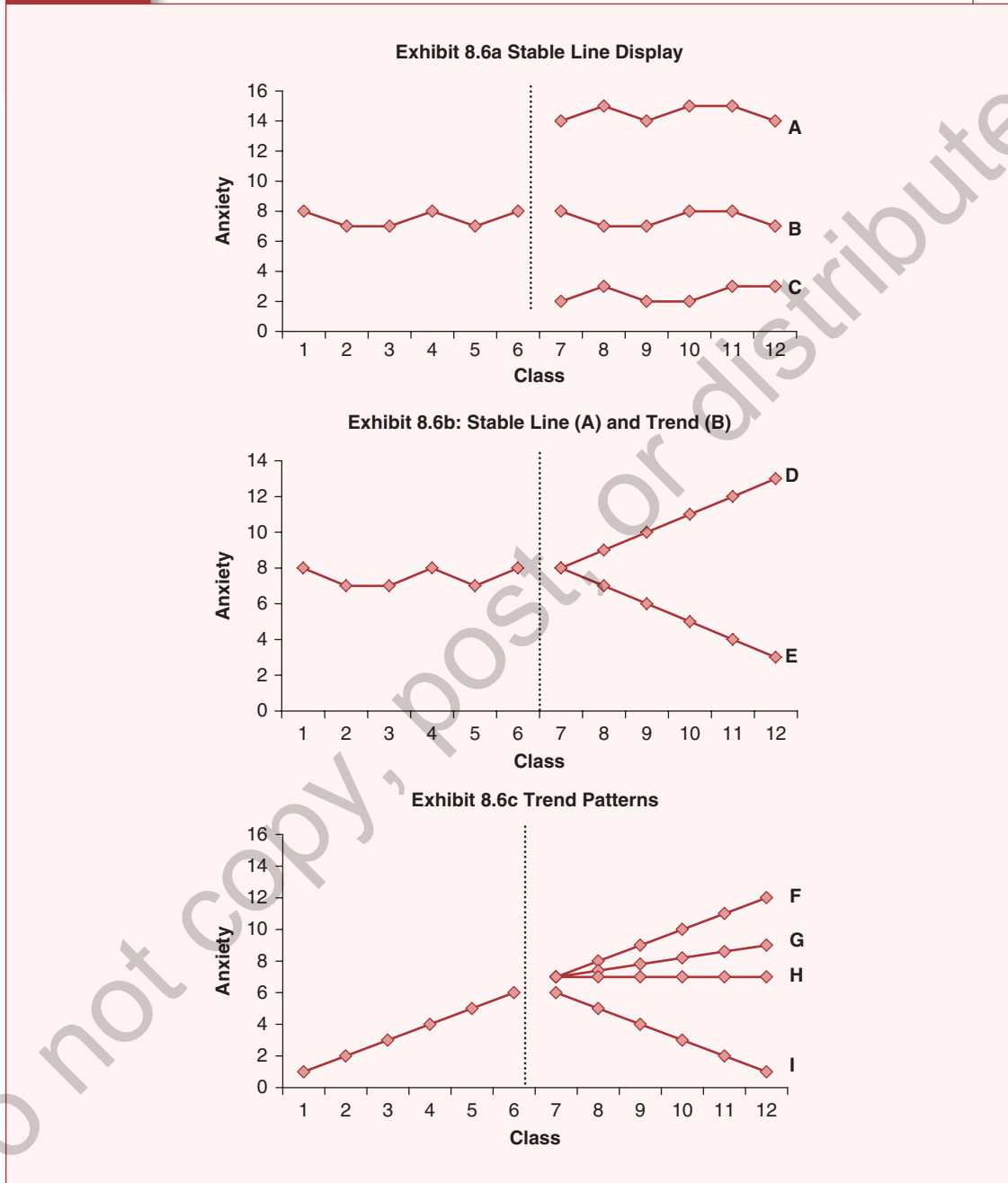
**Exhibit 8.5** Variability and Range Bars

Exhibit 8.6c displays common patterns when there is a trend in the baseline; the baseline phase is marked by an increase in anxiety from week to week. In the case of Outcome F, the intervention had no effect on the level of anxiety. For Outcome G, there was no change in the direction of the trend, but the rate of increase of anxiety has slowed, suggesting that the intervention has had some effect, but has not alleviated it. Outcome H represents the situation in which the intervention has improved the situation only to the extent that it is not getting worse. Finally, for Outcome I, the intervention has resulted in an improvement in the subject's status.

Regardless of whether you use visual inspection or one of these statistical approaches, the overriding issue is the clinical significance (or practical significance) of the findings. Has the intervention made a meaningful difference in the well-being of the subject? Although clinical significance at times is subjective, there are several principles you might apply to reduce the uncertainty:

- *Setting criteria.* One simple method is to establish with the client or community the criteria for success. If the intervention reaches that point, then the change is meaningful.
- *Cut-off scores.* A second method, particularly useful for psychological symptoms, is whether the intervention has reduced the problem to a level below a clinical cut-off score. For example, if you are using the Center for Epidemiologic Studies Depression scale (described in Chapter 4), you would determine whether the depressive symptom scores fall below the cut-off score for depression for that particular scale. Visual inspection or a statistical test may lead you to conclude that the intervention did reduce the number of reported symptoms of depression, but the number did not fall below a cut-off score for depression. Is it a clinically meaningful change if the client is still depressed?
- *Costs and benefits.* A third way to view practical significance is to weigh the costs and benefits to produce the change. Do efforts to increase employment in a community result in sufficient change to be worth the cost and effort to produce the improvement in employment?

**Exhibit 8.6** Typical Baseline-Intervention Patterns



**Problems of Interpretation**

The examples presented up to now have been quite neat, but when you are engaged in real practice research or evaluation, you are less likely to obtain such clear patterns. Because you are relying on visual judgment, there is the real possibility of coming to the wrong conclusion even if you are using systematic approaches like those suggested by Nugent

(Borckardt, Murphy, Nash, & Shaw, 2004). There are several possible concerns that make conclusions from visual analysis less certain including (1) discrepant scores, (2) delayed change, and (3) improvement in the baseline.

When scores in the baseline widely differ (discrepant scores; see Exhibit 8.2f), it is harder to determine whether there is any pattern. Measures used to create a summary line may poorly represent the data points. Therefore, judging whether the intervention has made a difference becomes more difficult.

A second issue is how to interpret changes in the intervention phase that are not immediately apparent (*delayed change*). For example, the changes in anxiety displayed in Exhibit 8.7 take place several weeks into the intervention. Is the change due to the intervention or some extraneous event or factor unrelated to the intervention? There is no easy answer to this question. It may depend on the nature of the intervention and when it is hypothesized that change will occur. The alternative interpretation that “something extraneous happened” (i.e., history) to produce the effect is equally plausible.

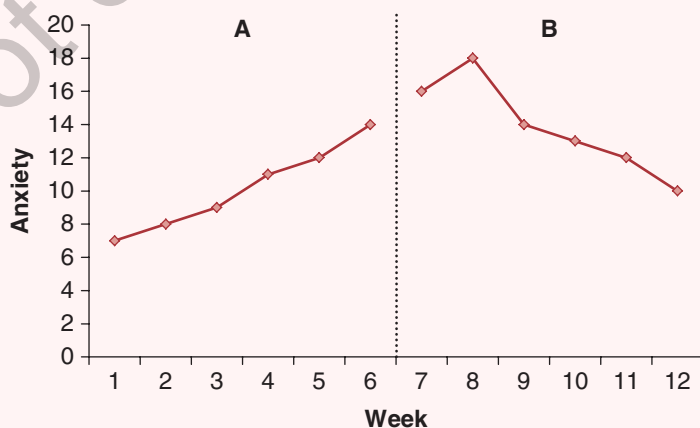
Another difficult judgment occurs when there is improvement in the target problem scores during the baseline phase even prior to the onset of the intervention. This improvement may occur for a variety of reasons, including the impact of an event or the passage of time (i.e., maturation). The effectiveness of the intervention may then depend on whether there is a shift in level or in the rate of the improvement. In Exhibit 8.8a, you see a pattern in which the intervention had no impact, as the improvement continues unchanged after the intervention has begun. Based on the pattern of scores in Exhibits 8.8b and 8.8c, there may have been an intervention effect on the target problem. In Exhibit 8.8b, there was a shift in level, whereas in Exhibit 8.8c, the rate of improvement has accelerated. Of course, these changes may still be due to an event occurring between the last baseline measure and the first intervention measure.

## 2 Types of Single-Subject Designs

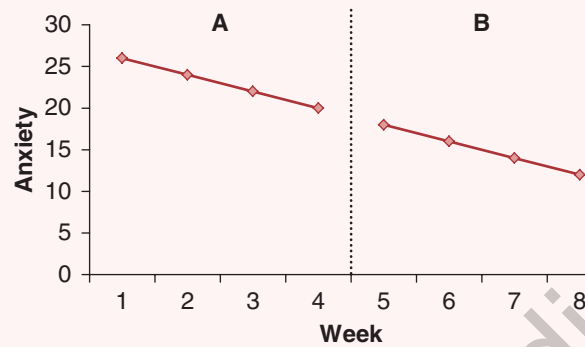
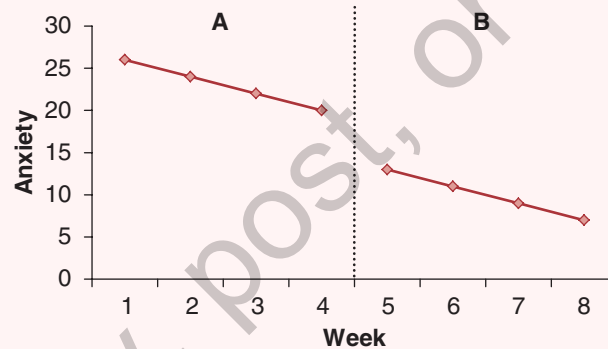
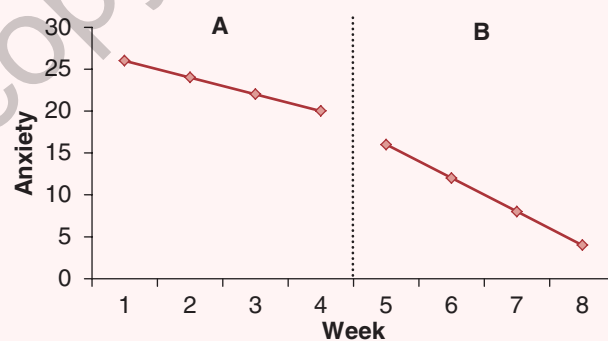
Single-subject designs may be used for research, as a method to assess practice outcomes, or as a tool to monitor client progress. There are more constraints when using a single-subject design for research purposes than when using it for practice evaluation; monitoring client progress has even fewer constraints.

The goal of a research experiment is to test the efficacy of an intervention on a particular target and, therefore, to enhance social work knowledge about what works. The intervention has already been specified, as has the target(s) that

**Exhibit 8.7** Delayed Change





**Exhibit 8.8** Improvement in the Baseline**Exhibit 8.8a: No Change in Pattern****Exhibit 8.8b: Change in Level****Exhibit 8.8c: Change in Rate**

will be evaluated. The measures should be reliable and valid indicators of the target(s). The baseline should include at least three data points, and measurement should continue until a pattern emerges. The baseline measures should also be collected during the course of the experiment. To establish causality, the design should control for all internal validity threats, including history.

The focus of practice evaluation is to describe the effectiveness of the program or particular intervention approach. Increasing knowledge about a particular treatment approach may be a goal, but that is secondary to the overall purpose of evaluation. Practice or program evaluation is conducted to provide feedback about the program to agency staff and funders, making demonstrating a causal relationship less important. The specific target and the appropriate intervention emerge from the interaction of the social worker with the client rather than being established before the interaction. As in a research study, the measures should be reliable and valid indicators of the target. Ideally, the baseline should include at least three measures and be characterized by a stable pattern, but this may not be possible; only one or two measures may be available. Unlike the case in a research design, the baseline measures may be produced through the recollection of the client, significant others, or client records. Finally, controlling for causality is less important.

The purpose of monitoring is to systematically track a client's progress. Monitoring provides ongoing feedback that may be more objective than just relying on the practitioner's impressions. The information can be used by the social worker to determine whether the intervention should continue without change or whether the intervention should be modified. As with practice evaluation, the target problem and intervention are not specified in advance; rather, they emerge through the client–social worker interaction. Ideally, the measures are reliable and valid indicators. There may not be any baseline, or the baseline may be limited to a single assessment. When the techniques are used to monitor a client's progress, threats to internal validity are not a concern.

Keep these distinctions in mind as you read about the various designs. Some designs can be used for both research and practice evaluation, while other designs are more suited for monitoring.

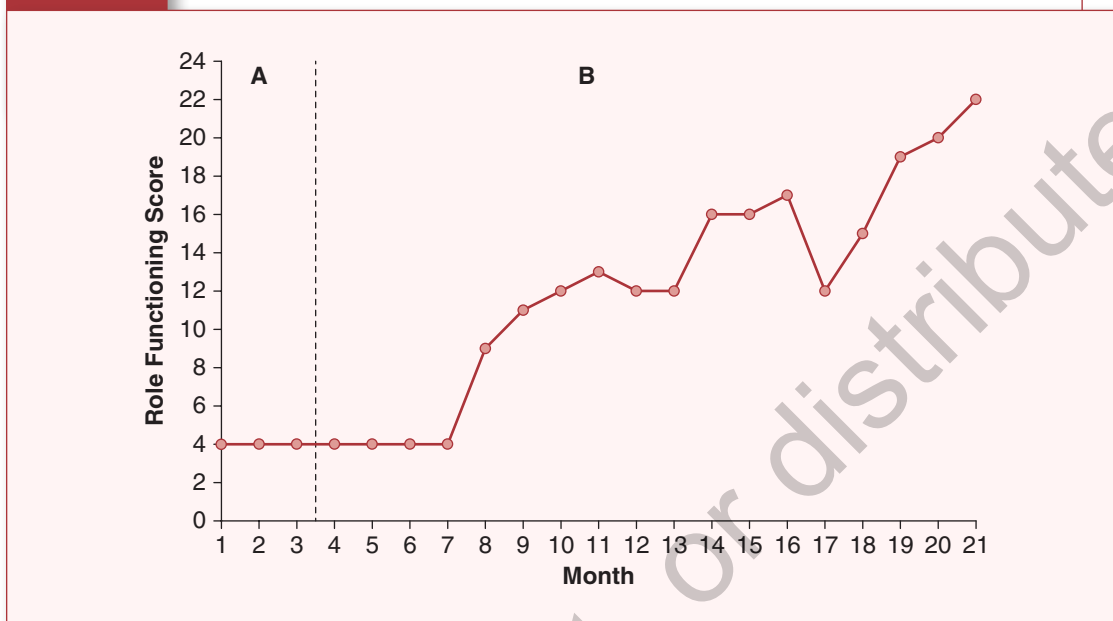
## Basic Design: A-B

The *A-B design* is the basic single-subject design and is used for all three purposes: research, evaluation, and client monitoring. It includes a baseline phase with repeated measurements and an intervention phase continuing the same measures. Bradshaw (1997), in his work with Jody and three other clients who also had schizophrenia, used an A-B design to test the effectiveness of cognitive-behavioral therapy (CBT). Bradshaw collected 3 months of baseline data on symptomatology, role functioning, and psychosocial functioning using standardized scales. This was followed by 18 months of intervention, with data collected monthly. As a research study, Bradshaw (a) set the length of the study and the particular outcomes prior to contact with the participants, (b) used standardized measures for which there was evidence of reliability and validity, and (c) had a minimum of three baseline measures.

The results for Jody's role functioning scores are presented in Exhibit 8.9. The three baseline scores did result in a stable line; subsequently there is a delayed change but after an additional four months, role functioning shows improvement. Based on the evidence of just this graph (remember Bradshaw had three other subjects), you could argue that the intervention had a positive effect on role functioning, but you could also plausibly suggest that some event (history) might have occurred after month seven that resulted in a positive change.

This example points to the limits of the A-B design. The design cannot rule out history, so it is impossible to conclude that the treatment *caused* the change. It is important that the participant and the researcher debrief to review whether any events transpired that might offer an alternative explanation for the results. The A-B design provides evidence of an association between the intervention and the change, and given that some threats to internal validity are controlled, it is analogous to a quasi-experimental design.

The A-B design can also be used to monitor a client's progress. Consider a case in which a social worker is meeting with the parents who are having problems with their 17-year-old daughter. They note that for several months their daughter has been squabbling constantly with her brother and being rude and sarcastic with her parents. The social worker suggests an intervention based on negative rewards may lead to a behavioral change. The parents agree to use a point system, with points being accrued for poor behavior. Once a certain number of points are attained, the child will begin to lose certain privileges. The parents are instructed to count and record the total number of rude and sarcastic comments, as well as sibling arguments begun by the daughter, every 3 days over a 15-day period. The intervention begins on the 16th day, with the parents explaining how she might get negative points and face the consequences of accumulating points.

**Exhibit 8.9** A-B Graph

Source: Bradshaw, W. (1997). Evaluating cognitive-behavioral treatment of Schizophrenia: Four single-case studies. *Research on Social Work Practice*, 7, 419–445.

## Withdrawal Designs

There are two withdrawal designs: the *A-B-A design* and the *A-B-A-B design*. By withdrawal, we mean that the intervention is concluded (*A-B-A design*) and there is a planned, systematic follow-up or the intervention is stopped for some period of time before it is restarted (*A-B-A-B design*). The premise is that if the intervention is effective, the target problem should improve only during the course of intervention, and the target scores should worsen when the intervention is removed. If this assumption is correct, then the impact of an extraneous event (history) between the baseline and intervention phase would not explain the change.

However, this premise is problematic for social work research. Ideally, the point of intervention is to reduce or eliminate the target problem without the need for ongoing intervention. We would like the impact of the intervention to be felt long after the client has stopped the intervention. Practice theories, such as behavioral or cognitive-behavioral treatment, are based on the idea that the therapeutic effects will persist. This concern, referred to as the **carryover effect**, may inhibit the use of these designs. To be used for research, the implementation of each of the withdrawal designs may necessitate limiting the length of the intervention and ending it prematurely. Or the design can be used to test the assumption that the impact of the treatments persists after it is discontinued. If the designs are being used for evaluation, it is unnecessary to prematurely withdraw the intervention; rather, the second baseline provides important follow-up information.

**Carryover effect** The impact of an intervention persists after the end of the treatment process.

## A-B-A Design

The *A-B-A design* builds on the *A-B design* by integrating a posttreatment follow-up that would typically include repeated measures. This design answers the question left unanswered by the *A-B design*: Does the effect of the

intervention persist beyond the period in which treatment is provided? Depending on the length of the follow-up period, it may also be possible to learn how long the effect of the intervention persists.

The follow-up period should include multiple measures until a follow-up pattern emerges. This arrangement is built into a research study. For practice evaluation, the practicality of this depends on whether the relationship with the client extends beyond the period of the actual intervention and on the ease of collecting information. For example, the effect of an intervention designed to reduce problem behaviors in school might be amenable to repeated measurement after the end of the intervention given that the client is likely to still be in school. Some involuntary clients are monitored after the end of the intervention period. The effects of community practice interventions or organizational changes are more amenable to follow-up repeated measurements.

However, a voluntary client who has come to a family service agency for treatment of depression might be more difficult to locate or might be unwilling to go through repeated follow-up measurements. Nevertheless, do not be dissuaded from trying to obtain follow-up measures. Some clients may not find the continued monitoring cumbersome, particularly if they understand that they may benefit. The methods of collecting data may be simplified and adapted to further reduce the burden on ex-clients, such as using phone interviews rather than face-to-face interviews.

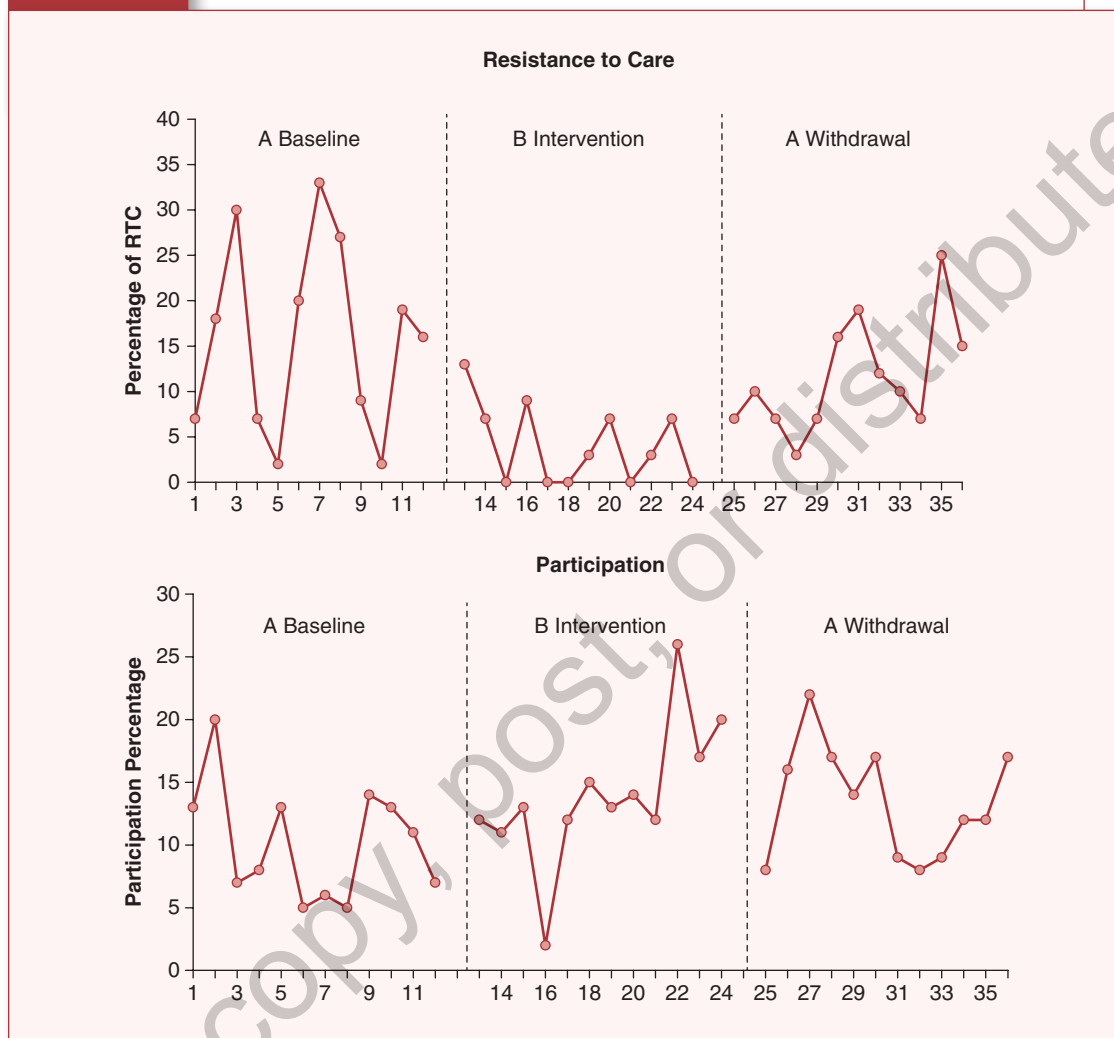
Claire O'Connor and colleagues (O'Connor, Smith, Nott, Lorang, & Mathews, 2011) explored the effectiveness of a video-simulated family member presence on reducing resistance to care (RTC) and increased participation in a variety of tasks with a nursing resident with dementia. The two A phases involved treatment as usual with staff operating as they might typically do with a resident. The intervention used during the B phase involved playing one of several very brief videos with a family member encouraging the relative to comply with staff and participate in taking medicine, eating, and the like; the video was played right before the staff came to interact with the participant. The baseline period included 12 data points over 10 days, the intervention was conducted for 14 days to get 12 data points, and the withdrawal period covered 15 days to obtain 12 data points. The results are graphed on Exhibit 8.10; RTC was reduced during the intervention period and increased when the video was withdrawn while participation increased during the intervention phase and slightly worsened with withdrawal of the intervention.

### A-B-A-B Design

The A-B-A-B design builds in a second intervention phase. The intervention in this phase is identical to the intervention used in the first B phase. The replication of the intervention in the second intervention phase makes this design useful for social work practice research. For example, if, during the follow-up phase, the effects of the intervention began to reverse (see Exhibit 8.11a), then the effects of the intervention can be established by doing it again. If there is a second improvement, the replication reduces the possibility that an event or history explains the change.

Just as with the A-B-A design, there is no guarantee that the effects will be reversed by withdrawing the intervention. If the practice theory holds, then it is unlikely that the effects will actually be reversed. So it may be that this first intervention period has to be short and ended just as evidence of improvement appears. Even if the effect is not reversed during the follow-up, reintroducing the intervention may demonstrate a second period of additional improvement, as displayed in Exhibit 8.11b. This pattern suggests that the changes between the no-treatment and treatment phases are due to the intervention and not the result of history.

Kam-fong Monit Cheung (1999) used an A-B-A-B design to evaluate the effectiveness of a combination of massage therapy and social work treatment on six residents in three nursing homes. Measurements included an assessment of activities of daily living and the amount of assistance received. Each phase took 7 weeks, with the massage therapy applied in Weeks 8 through 14 and Weeks 22 through 28. In the first 7 weeks (the A phase), residents received their usual social work services; in the second 7 weeks (the B phase), residents received massage therapy and social work services. In the third 7-week period (the second A phase), residents received just social work services; in the fourth 7-week period (the second B phase), massage therapy resumed. The measurements at the baseline were retrospectively constructed from client, nursing aide, and social work assessments. Subsequent measurements were taken from logs and reported behavior by the clients.

**Exhibit 8.10 A-B-A Design**

Source: O'Connor, C. M., Smith, R., Nott, M. T., Lorang, C., & Mathews, R. M. (2011). Using video simulated presence to reduce resistance to care and increase participation of adults with dementia. *American Journal of Alzheimer's Disease & Other Dementias*, 26, 312–325.

## Multiple Baseline Designs

In the previous designs, the individual baseline scores serve as the control for the impact of the intervention. Yet the withdrawal designs suffer from the problem that often the target behavior cannot be reversed, and it may not be ethical to withdraw treatment early. A solution to these problems is to add additional subjects, target problems, or settings to the study. This method provides social work researchers with a feasible method of controlling for the effects of history.

The basic format is a *concurrent multiple baseline design*, in which a series of A-B designs (A-B-A or A-B-A-B designs) is implemented at the same time for at least three cases (clients, target problems, or settings). While the data are collected at the same time, the unique feature of this design is that the length of the baseline phase is staggered (see Exhibit 8.12) to control for history across the three cases. The baseline phase for the second case extends until the intervention data points for the first case become more or less stable. Similarly, the intervention for the third case does not

## Exhibit 8.11 A-B-A-B Designs

Exhibit 8.11a: A-B-A-B Design With Reversal

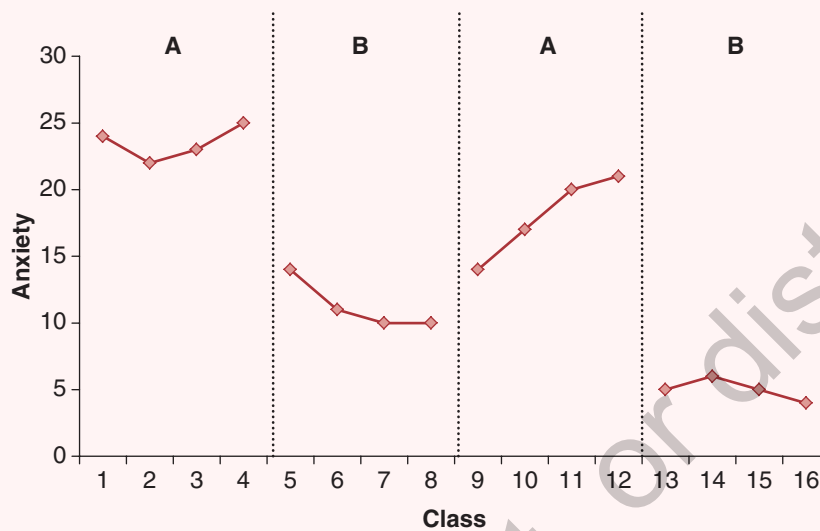
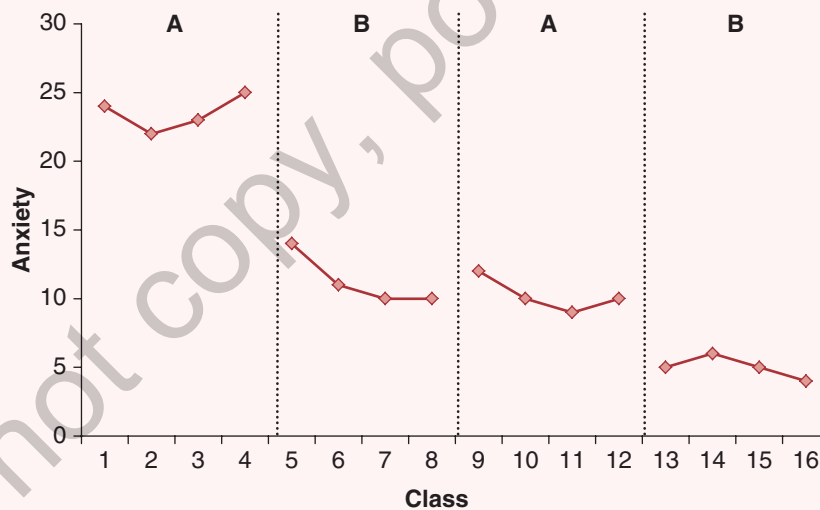


Exhibit 8.11b: A-B-A-B No Reversal



begin until the data points in the intervention phase for the second case become stable. The second and third cases act as a control for history in the first case, and the third case acts as a control for the second case.

One problem with a design requiring that all subjects start at the same time is having enough available subjects. An alternative that has been used is a *nonconcurrent multiple baseline design*. In this case, the researcher decides on different lengths of time for the baseline period. Then as clients or subjects meeting the selection criteria become available, they are randomly assigned into one of the baseline phases. For example, Matthew Jarrett and Thomas Ollendick (2012) used the nonconcurrent multiple baseline design to evaluate the impact of a treatment integrating parent management

**Exhibit 8.12 Multiple Baseline Design**



training and cognitive behavioral therapy on anxiety with children with attention-deficit/hyperactivity disorder. They randomly assigned clients to a baseline phase of 2, 3, or 4 weeks.

Multiple baseline designs have features for establishing causality. The design introduces two replications so that if consistent results are found, the likelihood that history is causing the change is reduced. If history does impact on all three cases, the effect of the event may be picked up by the control cases. The pattern of change in Exhibit 8.13 suggests that something occurred that affected not only Client A but also simultaneously Clients B and C, as they reported changes and improvement even before they received the intervention.

**Exhibit 8.13** Multiple Baseline Design With History?





**Across subjects.** When a multiple baseline is used across subjects, each subject receives the same intervention sequentially to address the same target problem. After Bradshaw's 1997 study, he tested the effectiveness of CBT with schizophrenia with seven participants using a concurrent multiple baseline A-B design (Bradshaw, 2003). All participants started the program at the same time. The baseline phase lasted 6 months for two participants, 9 months for two participants, and 12 months for the remaining three participants. During the baseline phase, all participants received treatment as usual.

**Across target problems.** In this case there is one client, and the same intervention is applied to different but related problems or behaviors. The application of the intervention as it relates to the target problems or behaviors is staggered. For example, Christina Johnson and Jeannie Golden (1997) used a multiple baseline design to examine whether an intervention using both prompting and reinforcement would have a positive impact on different aspects of peer interactions for a child with language delays. The three behaviors measured were social response (verbal or nonverbal efforts to join in play with another child), approach behavior (approaching another child using vocal expressions or gestures), and play organizer (the child organizing play by specifying an activity, its rules, or inviting another child to play). The baseline period for social response lasted 3 sessions, the baseline for approach behavior overlapped these 3 sessions and continued for 7 more sessions, and the baseline for play organizer overlapped the above two baselines and continued for 4 more sessions, lasting 14 sessions. Measuring these different behaviors for different periods allowed Johnson and Golden to determine which behaviors were influenced by the intervention while controlling for external events.

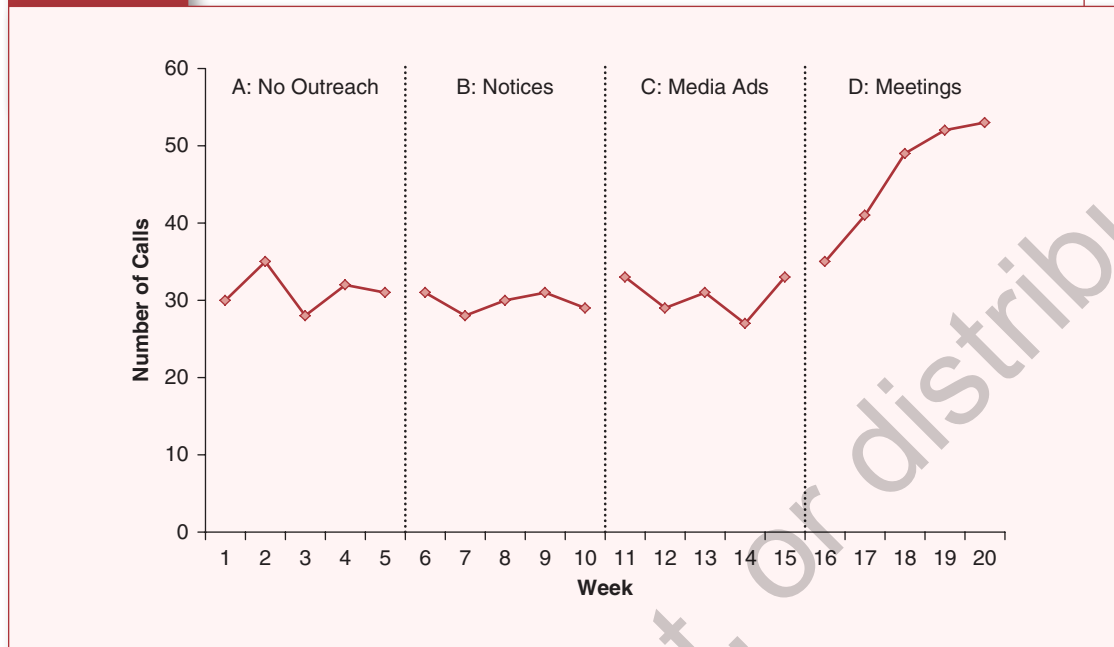
**Across different settings.** Multiple baseline designs can be applied to test the effect of an intervention as it is applied to one client, dealing with one behavior but sequentially applied as the client moves to different settings. You might imagine a client with behavioral problems in school, at home, and at play with friends. A behavioral intervention might be used, with the application of rewards introduced sequentially across the three settings, starting with home, then school, and then play.

## Multiple Treatment Designs

In a multiple treatment design, the nature of the intervention changes over time, and each change represents a new phase of the design. One type of change that might occur is the intensity of the intervention. For example, you might be working with a family that is having communication problems. The actual amount of contact you have with the family may change over time, starting with counseling sessions twice a week, followed by a period of weekly sessions, and concluding with monthly interactions. In this case, the amount of contact declines over time. *Changing intensity designs* are characterized by A-B<sub>1</sub>-B<sub>2</sub>-B<sub>3</sub>.

Another type of changing intensity design is when, during the course of the intervention, you add additional tasks to be accomplished. For example, older adults who lose their vision in later life need to relearn how to do different independent activities of daily living taking into account their vision loss. The intervention is learning independent self-care. The B<sub>1</sub> may involve walking safely within the house, the B<sub>2</sub> may add methods for using a checkbook, the B<sub>3</sub> may add a component on cooking, and the like.

In an *alternative treatment design* (A-B-C-D), the actual intervention changes over time so that each phase represents a different intervention. For example, a student designed a study to determine the effectiveness of outreach efforts on the number of phone calls received by a help line (information and referral). The baseline period represented a time in which there was no outreach; rather, knowledge about the help line seemed to spread by word of mouth. The B phase represented the number of calls after the agency had sent notices about its availability to agencies serving older adults and families. During the C phase, the agency ran advertisements using radio, TV, and print media. Finally, during the D phase, agency staff went to a variety of different gatherings, such as community meetings or programs run by different agencies, and described the help line.

**Exhibit 8.14** Multiple Treatment Design

The student found (see Exhibit 8.14) that the number of phone calls did not increase appreciably after notices were sent to other professionals or after media efforts, but calls did increase dramatically in the final phase of the study. This graph demonstrates how tricky the interpretation of single-subject data can be since only adjacent phases can be compared. One plausible explanation for the findings is that sending notices to professionals and media efforts at outreach were a waste of resources in that the notices produced no increase in the number of calls relative to doing nothing, and advertising produced no increase relative to the notices. Only the meetings with community groups and agency-based presentations were effective, at least relative to the advertising. An alternative interpretation of the findings is that the *order* of the activities was essential. There might have been a carryover effect from the first two efforts that added legitimacy to the third effort. In other words, the final phase was effective only because it had been preceded by the first two efforts. If the order had been reversed, the impact of the outreach efforts would have been negligible. A third alternative is that history or some other event occurred that might have increased the number of phone calls.

Multiple treatment designs might also include interactions where two treatments are combined. An interaction design often parallels experiences with clients or agency activities in which interventions are combined or done simultaneously. In the previous example, the agency outreach effort might have included its baseline (A), notices to agencies (B), media efforts (C), and then a combination of the two (B-C phase).

## Monitoring Client Progress

When monitoring a client's progress, the A-B design is recommended for the baseline information it provides. But there are times when establishing a baseline is not possible, other than to have a single point based on an initial assessment. Nonetheless, to ascertain whether a client is making progress, a form of monitoring should be done. Therefore, a social worker might use a B or a B-A design.

The *B design* (see Exhibit 8.15a) has only an intervention phase; during the course of the intervention, the social worker takes repeated measurements. This design can be used to determine whether the client is making progress in

the desired direction. If the client is not making progress, the social worker may decide to change the type of intervention or the intensity of the intervention. For example, if you were working with a client who had symptoms of depression, but after 4 weeks there was no reduction in these symptoms, you would change the intensity or type of intervention. Or it might be that the symptoms reduced somewhat, but then leveled off at a level still above a cut-off score. As a result, you might again alter the nature of the intervention.

With a B design, the actual improvement cannot be attributed to the intervention. There is no baseline, and therefore, changes might be due to different threats to internal validity, reactivity to the measurement process, or reactivity to the situation.

If a period of follow-up measurements can be introduced, then a *B-A design* is a better alternative (see Exhibit 8.15b). The intervention period is followed by a period of no intervention for the specific target. Although it is harder to get repeated measurements of a client after the intervention has concluded, if treatment about other concerns continues, then follow-up measures are possible. Having reduced depressive symptoms to an acceptable level, the social worker may address social support network building with the client. Measurement of the depressive symptoms might still continue.

### Single-Subject Designs With Group Interventions

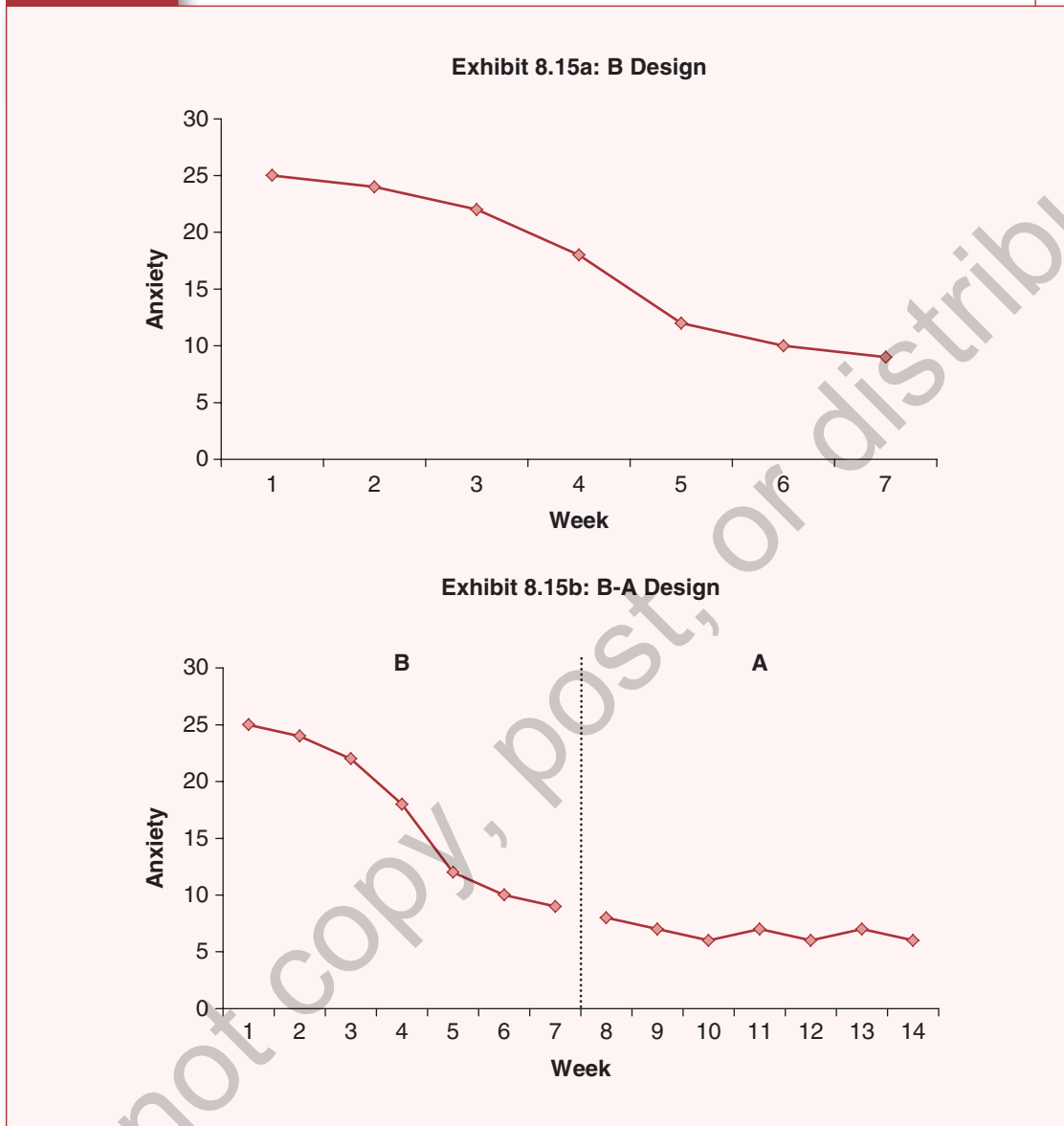
Though it appears that single subject designs are meant for treatments and interventions targeted to individuals, all of these designs have been applied to group interventions. Single subject designs have been used to assess both group processes such as attendance, cohesion, verbal participation, and group interactions and treatment outcomes (Macgowan & Wong, 2014). For example, James Hall and colleagues (Hall, Dineen, Schlesinger, & Stanton, 2000) evaluated the efficacy of a social skills training group on improving different social skills with six developmentally disabled adults. To assess the influence of the intervention, they used a multiple baseline design across skills gathering data for 8 weeks plus a 1 year follow-up. The intervention was provided in a group format, four times a week for 1 hour per session over 6 weeks. Baseline information was collected three times in week one for all six skills with the onset of intervention for a particular skill staggered throughout subsequent weeks. Subsequent data were collected once a week. So the baseline for social conversation consisted of three data points collected in Week 1; the baseline for social invitations included these three data points plus the fourth data point. The intervention was provided during Weeks 2 through 8. The results were provided in a series of graphs, one for each participant.

## 2 Implications for Evidence-Based Practice

Single-subject designs offer a range of evidence to assess the impact of different interventions. The most rigorous designs control for threats to internal validity, while monitoring designs demonstrate client outcomes but without the ability to suggest it was the intervention that mattered. Therefore, understanding the differences in these designs is crucial to weighing the evidence derived from such studies.

One benefit of single-subject design is the focus on the individual as opposed to a group as an aggregate. The evidence derived from single-subject designs differs from that of group designs in that the question of interest is different (Johnston, Sherer, & Whyte, 2006). In a single-subject design, the question is, Does an intervention work for an individual? In contrast, the questions in a group design are, Does the group average change? Does the treatment group average differ in comparison to a second group? In a group design, the impact on any one individual is obscured by the impact on the group.

This different focus is particularly important for social workers because much of their practice involves interventions with individuals or with groups of individuals. Given the focus on the individual, cultural and other contextual variables are considered in evaluating outcomes (Arbin & Cormier, 2005). Single-subject designs are likely to give greater consideration to client characteristics such as gender, age, ethnicity, sexual orientation, or class. Therefore, the evidence may be quite compelling because it reflects more accurately findings from actual practice.

**Exhibit 8.15** Two B Designs

However, the strength of single-subject designs with their focus on an individual is also suggested to be their weakness. How are we to judge findings about a single individual? How is evidence about that single individual relevant to other clients? We can think about this criticism as a statistical problem and/or as a problem about building the generalizability of the findings. The statistical problem is being addressed by statisticians who are developing meta-analytic methods to assess single-subject design research; these methods are designed to take the findings of many single-subject design studies and aggregate them (Jenson, Clark, Kircher, & Kristjansson, 2007).

The problem of generalizability of single-subject design research is not unlike that of group design research—it is an issue of external validity. Ideally, we want to take what has been tested in one research context and apply the findings to different settings, clients, or communities; to other providers; and even to other



## Research In the News

For  
Further  
Thought ?

### EXERCISE YOUR BRAIN?

Katie Hafner (2008) reported on the many exercises that baby boomers are taking up to keep their brains active, to reduce forgetfulness, and, it is hoped, to lessen the risk of Alzheimer's disease. These include video games, on-line cognitive behavioral training, and taking on new brain stimulation challenges like learning to play the piano.

1. Design a single-subject study to test the impact of one of these brain exercise games on a friend. What indicators could you use to track the impact of the game?

Source: Hafner (2008).

problems related to the target concern of the research. To do so when the sample consists of a single subject engaged in a particular intervention provided by a particular individual is challenging. To demonstrate the external validity of single-subject design requires replication of the research conditions and to extend the assessment to other targets and settings.

Barlow et al. (2009) suggest that three sequential replication strategies be used to enhance the external validity of single-subject design: direct replication, systematic replication, and clinical replication.

**Direct replication.** **Direct replication** involves repeating the same procedures, by the same researchers, including the same providers of the treatment, in the same setting, and in the same situation, with different clients who have similar characteristics (Barlow et al., 2009). The strength of the findings is enhanced by having successful outcomes with these other clients. When the results are inconsistent, differences in the clients can be examined to identify characteristics that may be related to success or failure.

**Direct replication** Used to enhance the generalizability of a single-subject design; the single-subject design is repeated using the same procedures by the same researchers and the same providers, in the same setting, and in the same situation with different clients.

**Systematic replication** Repeating a single-subject design in different settings, using different providers and other related behaviors to increase generalizability.

**Clinical replication** Used to enhance generalizability of single-subject designs; clinical replication involves combining different interventions into a clinical package to treat multiple problems.

**Systematic replication.** The next step is **systematic replication**, which involves repeating the experiment in different settings, using different providers and other related behaviors (Barlow et al., 2009). Systematic replication also increases the number and type of clients exposed to the intervention. Through systematic replication, the applicability of the intervention to different conditions is evaluated. Like direct replication, systematic replication helps to clarify conditions in which the intervention may be successful and conditions in which the intervention may not be successful.

**Clinical replication.** The last stage is **clinical replication**, which is defined as combining different interventions into a clinical package to treat multiple problems (Barlow et al., 2009). The actual replication takes place in the same setting and with clients who have the same types of problems. In many ways, findings from practice evaluation can enhance clinical replications.

For any replication effort to be successful, the treatment procedures must be clearly articulated, identified, and followed. Failing to adhere to the treatment procedures changes the intervention, and therefore, there is not a true replication of the experiment.

Bradshaw's efforts at demonstrating the effectiveness of CBT to treat schizophrenia represent the contribution such systematic research makes to advancing practice-related knowledge. His 1997 study was completed with four participants who had been referred for outpatient therapy after discharge from psychiatric hospitalizations. His 2003 study included 10 participants (3 did not complete the study) randomly selected from a county mental health system. In this study, the intervention period was twice as long as in the first study, but followed the same model of CBT. Finally, in collaboration with David Roseborough, Bradshaw reported findings from a third study of 30 clients (8 left the study) who received CBT for 18 months. The ongoing replication and findings of positive results has provided support for this intervention with clients with schizophrenia.

Social work practitioners can be active in building this evidence as part of their ongoing practice. Integrating systematically single-subject designs can provide additional clinical evidence for practice. You can become your own researcher!

## 2 Single-Subject Design in a Diverse Society

Throughout this chapter, we have noted instances when special attention must be paid to issues of diversity. These issues are not unique to research, but are relevant to practice. That is no surprise because single-subject design is so closely aligned to a practice model (Staudt, 1997). Researchers and practitioners must understand that how problems are identified and defined may depend on client characteristics such as gender, ethnicity, sexual orientation, and class. Measures must be acceptable and applicable (reliable and valid) to different population subgroups. Similarly, issues regarding informed consent are relevant for all population subgroups (J. Martin & Knox, 2000; Nelson, 1994).

Single-subject designs may be useful methods for engaging diverse groups that have been underrepresented in research and in particular experimental group designs or clinical research trials. Because they are often practice based, they may more easily mitigate distrust of the researcher. Because they focus on the individual, as opposed to the group, single-subject designs can more easily incorporate cultural factors and test for cultural variation (Arbin & Cormier, 2005).

## 2 Ethical Issues in Single-Subject Design

Like any form of research, single-subject designs require the informed consent of the participant. The structure of single-subject designs for research involves particularly unique conditions that must be discussed with potential participants. All aspects of the research, such as the purpose, measurement, confidentiality, and data collection, are a part of the information needed for informed consent. In particular, the need for repeated baseline measurements and the possibility of premature withdrawal of treatment are particularly unique to single-subject design research.

Participants must understand that the onset of the intervention is likely to be delayed until either a baseline pattern emerges or some assigned period elapses. Until this condition is met, a needed intervention may be withheld. Furthermore, the length of the baseline also depends on the type of design. In a multiple baseline design, the delay in the intervention may be substantial. The implications of this delay must be discussed as part of obtaining informed consent.

When a withdrawal or reversal design is used, there are additional considerations. The structure of such designs means that the intervention may be withdrawn just as the research subject is beginning to improve. The risks associated with prematurely ending treatment may be hard to predict. If there is a carryover effect, the subject's condition may not worsen, but it is possible that the subject's condition or status may indeed worsen. Given this possibility, the use of an A-B-A-B design as opposed to the A-B-A design is preferable for the purpose of research.

Obtaining informed consent may not be limited to the use of single-subject design for research purposes. As we noted in Chapter 3, the National Association of Social Workers (NASW; 2008) *Code of Ethics* does not distinguish between the need for informed consent in research and the need for informed consent for practice evaluation:

5.02(e) Social workers engaged in evaluation or research should obtain voluntary and written informed consent from participants, when appropriate, without any implied or actual deprivation or penalty for refusal to participate; without undue inducement to participate; and with due regard for participants' well-being, privacy, and dignity. Informed consent should include information about the nature, extent, and duration of the participation requested and disclosure of the risks and benefits of participation in research.

Others suggest that informed consent may not be necessary. For example, Royle, Thyer, Padgett, and Logan (2006) suggest that written informed consent is not necessarily required for practice evaluation because the intent is not to provide generalized knowledge or publish the results.

Even if written informed consent is not required when using these tools for practice evaluation and monitoring, social workers using these tools should be guided by practice ethics. According to the NASW (2008) *Code of Ethics*, social work practitioners should, as a part of their everyday practice with clients,

provide services to clients only in the context of a professional relationship based, when appropriate, on valid informed consent. Social workers should use clear and understandable language to inform clients of the purpose of the services, risks related to the services, limits to services because of the requirements of a third-party payer, relevant costs, reasonable alternatives, clients' right to refuse or withdraw consent, and the time frame covered by the consent. (Section 1.03[a])

Therefore, if such techniques are going to be used as part of the overall intervention, clients should be aware of the procedures.

## 2 Conclusion

Single-subject designs are useful for doing research, evaluating practice, and monitoring client progress. Single-subject designs have been underutilized as a research tool by social work researchers. Yet researchers using these designs can make a unique contribution to social work practice knowledge because so much of practice is with individuals. Done systematically, the success or failure of different interventions can be evaluated with distinct clients and under differing conditions. Furthermore, single-subject designs may be useful for understanding the process of change and how change occurs with particular clients.

Applying these techniques to your own practice can be of benefit to your clients. As Aaron Rosen (2003) warns, "uncertainty regarding the effectiveness of any intervention for attaining any outcome pervades all practice situations, regardless of the extent and quality of empirical support" (p. 203). Therefore, if you monitor what you do, you will add to your own clinical experience, which enhances your future work with clients.

## Key Terms

Baseline phase 189  
Carryover effect 204  
Clinical replication 214  
Cycle 192  
Direct replication 214

Duration 194  
Frequency 194  
Interval 194  
Level 195  
Magnitude 194

Stable line 191  
Systematic replication 214  
Treatment phase 192  
Trend 191  
Variability 196

## Highlights

- Single-subject designs are tools for researchers and practitioners to evaluate the impact of an intervention on a single system such as an individual, community, or organization.
- Single-subject designs have three essential components: the taking of repeated measurements, a baseline phase (A), and a treatment phase (B).
- Repeated measurements control for many of the potential threats to internal validity. The period between the last baseline measure and the first treatment measure is susceptible to the effect of history.
- The baseline phase typically continues, if practical, until there is a predictable pattern. To establish a pattern requires at least three measurements. The pattern may include a stable line, an increasing or decreasing trend line, or a cycle of ups and downs dependent on time of measurement.
- Researchers often measure behaviors, status, or level of functioning. These measures are typically characterized by frequency (counts), duration (length of time), interval (time between events), or magnitude (intensity).
- Reactivity to the process of measurement may impact the outcomes, and efforts to limit reactivity are important.
- Data analysis typically involves visually inspecting graphs of the measurements. A researcher may look for changes in level (magnitude), rate or directional changes in the trend line, or reductions in variability. The most important criterion is whether the treatment has made a practical (or clinical) difference in the subject's well-being.
- Generalizability from single-subject designs requires direct replication, systematic replication, and clinical replication.

## Discussion Questions

1. Visual analysis is used to communicate the impact of an intervention in visual form. What are the three primary ways that the pattern of scores established during a baseline or intervention stage may be viewed? When is each of them best used? What information is conveyed, and what information may be omitted by choosing each one of them over the others?
2. Single-subject designs lack the inclusion of additional subjects serving as controls to demonstrate internal validity. How do the measurements during the baseline phase provide another form of control?
3. Social work research seeks to confirm an intervention's effectiveness by observing scores when clients no longer receive the intervention. Yet the carryover effect may necessitate using a withdrawal design—ending a treatment prematurely—to do this successfully. Debate the merits of the withdrawal design in social work research. What are the advantages and disadvantages? Do the benefits outweigh the risks or vice versa?

## Practice Exercises

1. Go to the book's study site ([www.sagepub.com/engelpsw4e](http://www.sagepub.com/engelpsw4e)), and choose two research articles that include some attention to causality (as indicated by a check in that column of the article matrix). For each article describe the following:
  - a. What type of design was used? How does the author describe the design? Was it suited to the research question posed and the specific hypotheses tested, if any? Why do you suppose the author chose the particular design?
  - b. Did the design eliminate threats to internal validity? How did the design do this? Are you satisfied with the internal validity conclusions stated by the author? Why or why not?
  - c. What is the setting for the study? Does the setting limit the generalizability of the results to other similar settings or to the broader population? Is reactivity a problem? Are there other threats to external validity?



2. Stress is a common occurrence in many students' lives. Measure the frequency, duration, interval, and magnitude of school-related stress in your life in a 1-week period. Take care to provide a clear operational definition of stress and construct a meaningful scale to rate magnitude. Did you notice any issues of reactivity? Which of the measurement processes did you find most feasible? Finally, do you believe that your operational definition was sufficient to capture your target and detect changes?
3. Patterns detected in the baseline phase of single-subject designs also emerge in the larger population. Obtain a copy of a newspaper, and locate stories describing contemporary issues that can be described as having the pattern of a stable line, a trend, and a cycle. Is information provided about the number of observations made? If so, does this number seem sufficient to warrant the conclusion about what type of pattern it is?

## Web Exercises

1. Visit the Northwest Regional Education Laboratory's website (<http://educationnorthwest.org>). Search for "School Improvement Research Series." Choose School Improvement Program Research Series Materials, then Series V, and finally, click on Close-Up #9 School Wide and Classroom Discipline. Select three of the techniques that educators use to minimize disruption in educational settings, and then suggest a single-subject design that could be used to evaluate the effectiveness of each technique. Bear in mind the nature of the misbehavior and the treatment. Which of the designs seems most appropriate? How would you go about conducting your research? Think about things such as operationalizing the target behavior, determining how it will be measured (frequency, duration, magnitude, etc.), deciding on the length of the baseline and treatment periods, and accounting for threats to internal validity.
2. Search Social Work Abstracts for articles describing single-subject designs. Try to identify the type of design used. Read over the article. How well did this design satisfy the need for internal validity?

## Developing a Research Proposal

If you are planning to use a single-subject design,

1. What specific design will you use? How long will the study last? How will the data be collected? How often?
2. Discuss the extent to which each source of internal validity is a problem in the study. Will you debrief with participants to assess history?
3. Discuss the extent to which reactivity is a problem. How will you minimize the effects of reactivity?
4. How generalizable would you expect the study's findings to be? What can be done to improve generalizability?
5. Develop appropriate procedures for the protection of human subjects in your study. Include a consent form.

## A Question of Ethics

1. Use of single-subject methodology requires frequent measurement of symptoms or other outcomes. Practitioners should discuss with clients before treatment begins the plan to use de-identified data in reports to the research community. Clients who do not consent still receive treatment—and data may still be recorded on their symptoms in order to evaluate treatment effects. Should the prospect of recording and publishing de-identified data on single subjects become a routine part of clinical practice? What would be the advantages and disadvantages of such a routine?
2. The A-B-A design is a much more powerful single-subject design than the A-B design because it reduces the likelihood that the researcher will conclude that an improvement is due to the treatment when it was simply due to a gradual endogenous recovery process. Yet the A-B-A design requires stopping the very treatment that may be having a beneficial effect. Under what conditions do you think it is safe to use an A-B-A design? Why do some clinicians argue that an A-B-A-B design lessens the potential for ethical problems? Are there circumstances when you would feel it is unethical to use an A-B-A-B design?