5.1. INTRODUCTION

This chapter examines the prototype of single-case experimental designs—the A-B-A design—and its many variants. The objective of this chapter is to familiarize the reader with the advantages and limitations of each design strategy, while providing illustrative examples from the many different applications of these designs reported in the literature. The development of the A-B-A design will be traced, beginning with its roots in the clinical case study and in the application of "quasi-experimental designs" (Campbell & Stanley, 1966). Procedural issues discussed at length in chapter 3 will also be revisited here as they apply to each of the specific design options reviewed. Both ideal and problematic examples will be used for illustrative purposes.

Since the publication of the earlier editions of this book (Hersen & Barlow, 1976; Barlow & Hersen, 1984) the literature has been replete with examples of A-B-A designs. However, there actually has been little change with respect to basic procedural issues. Therefore, we have retained most of the original design considerations and illustrations discussed in these earlier editions, but have added some more recent examples to illustrate how these designs continue to be a valuable tool for clinical researchers.

Limitations of the case study approach

Descriptions of uncontrolled case studies have appeared in the psychoanalytic, psychotherapeutic, and psychiatric literatures for many decades (see chapter 1). Despite the development of applied behavioral methodology (presumably based on sound theoretical underpinnings) in the late 1950s and early to mid-1960s, the case study approach was still the primary method for reporting the efficacy of innovative treatment techniques (cf. Ashen, 1963; Barlow, 1980; Barlow et al., 1983; Lazarus, 1963; Ullmann & Krasner, 1965; Wolpe, 1958, 1976).

There is no doubt that the case history method has long held a valuable place in the clinical research armamentarium (e.g., Freud, 1933; Watson &
the target behavior is clearly specified, and repeated measurement is taken throughout the baseline (A) and treatment (B) phases of experimentation. As in all single-case experimental research, the A phase involves a series of baseline observations of the natural frequency of the target behavior(s) under study. In the B phase the treatment variable is introduced, and changes in the dependent measure are noted. Thus, *with some major reservations*, changes in the dependent variable are attributed to the effects of treatment (Barlow & Hersen, 1973; Campbell, 1969; Campbell & Stanley, 1966; Cook & Campbell, 1979; Hersen, 1982; Kazdin, 1982; Kratochwill, 1978).

Let us now examine some of the important reservations. In their evaluation of the A-B strategy, Wolf and Risley (1971) argued that “The analysis provided no information about what the natural course of the behavior would have been had we not intervened with our treatment condition” (pp. 314–315). That is to say, it is very possible that changes in the B phase might have occurred regardless of the introduction of treatment or that changes in B might have resulted as a function of correlation with some fortuitous (but uncontrolled) event. For instance, at the same point in time when the design changed from A to B, the person under study may have experienced some significant change in their living situation, begun another treatment, started a new relationship, etc. When considered in this light, the A-B strategy does not permit a full experimental analysis of the controlling effects of the treatment inasmuch as its correlative properties are quite apparent. Indeed, Campbell and Stanley (1966) referred to this strategy as a “quasi-experimental design.”

Risley and Wolf (1972) presented an interesting discussion of the limitations of the A-B design with respect to predicting, or “forecasting,” the B phase on the basis of data obtained in A. Two hypothetical examples of the A-B design were depicted, with both showing a mean increase in the amount of behavior in B over A. However, in the first example, a steady and stable trend in baseline is followed by an abrupt increase in B, which is then maintained. In the second case, the upward trend in A is continued in B. Therefore, despite the equivalence of means and variances in the two cases, the importance of the trend in evaluating the data is underscored. Some tentative conclusions can be reached on the basis of the first example, but in the second example the continued linear trend in A permits no conclusions as to the controlling effects of the B treatment variable.

In further analyzing the difficulties inherent in the A-B strategy, Risley and Wolf (1972) contended that:

The weakness in this design is that the data in the experimental condition is compared with a forecast from the prior baseline data. The accuracy of an assessment of the role of the experimental procedure in producing the change rests upon the accuracy of that forecast. A strong statement of causality therefore requires that the forecast be supported. This support is accomplished by elaborating the A-B design. (p. 5)
Such elaboration is found in the A-B-A design discussed and illustrated in section 5.3 of this chapter.

Despite these aforementioned limitations, it is shown how in some settings, such as where control-group analysis or repeated introduction and withdrawals of treatment variables are not feasible, the A-B design can be of some utility (Campbell & Stanley, 1966; Cook & Campbell, 1979). For example, the use of the A-B strategy in the private-practice setting has previously been recommended in section 3.2 of chapter 3 (see also Barlow, Hayes, & Nelson, 1984; Hayes, Barlow, & Nelson-Gray, 1999).

The use of the A-B strategy in field experiments where more traditional forms of experimentation are not at all possible (e.g., the effects of modifying traffic laws on the documented frequency of accidents) has been a valuable tool for many years, as discussed in detail by Campbell (1969). Indeed, the A-B design can be used to study groups of individuals and the use of repeated assessment of the dependent variable offers a valuable advantage over standard pre- and post-intervention assessment because the experimenter can show that behavior was stable over the course of baseline, and changed precisely when the intervention was initiated in the B phase (e.g., Cox, Cox, & Cox, 2000; Porritt, Burt, & Poling, 2006). However one uses the quasi-experimental design, Campbell cautioned the investigator as to the numerous threats to internal validity (history, maturation, instability, testing, instrumentation, regression artifacts, selection, experimental mortality, and selection-maturation interaction) and external validity (interaction effects of testing, interaction of selection and experimental treatment, reactive effects of experimental arrangements, multiple-treatment interference, irrelevant responsiveness of measures, and irrelevant replicability of treatments) that may be encountered. The interested reader is encouraged to read Campbell's (1969) classic article for a full discussion of the issues involved in large-scale retrospective or prospective field studies.

In summary, it should be apparent that the use of a quasi-experimental design such as the A-B strategy while more rigorous than the case study, ultimately yields rather weak conclusions. This design is subject to the influence of a host of confounding variables and is best applied as a last-resort measure when circumstances do not allow for more extensive experimentation. Examples of such cases will now be illustrated.

**A-B with follow-up**

An improvement on the A-B design is the inclusion of follow-up assessment. One early example of an A-B design with follow-up was provided by Epstein and Hersen (1974), who assessed the effects of reinforcement on the frequency of gagging in a 26-year-old psychiatric inpatient. The patient's symptoms had persisted for approximately 2 years despite repeated attempts at medical intervention. During the baseline (A) phase, the patient was instructed to record the time and frequency of each gagging episode on an index card, collected by the
experimenter the following morning at ward rounds. Treatment (B) involved presenting the patient with $2.00 in canteen books (exchangeable at the hospital store for goods) for any decrease in gagging from the previous daily frequency. In addition, zero rates of gagging were similarly reinforced. In order to facilitate maintenance of gains after treatment, no instructions were given as to how the patient might control his gagging. Thus, emphasis was placed on self-management of the behavior. At the conclusion of his hospital stay, the patient was requested to continue recording data at home for a period of 12 weeks. In this case, treatment conditions were not withdrawn during the patient’s hospitalization because of clinical considerations.

The results of this study are plotted in Figure 5.1. Baseline frequency of gagging fluctuated between 8 and 17 episodes per day but stabilized to some extent in the last 4 days. Initiation of the reinforcement procedures resulted in a decline to zero episodes of gagging within 6 days. However, on Day 15, frequency of gagging rose again to seven daily episodes. At this point, the criterion for obtaining reinforcement was reset to that originally planned for Day 13. Renewed improvement was then noted between Days 15–18, and treatment was continued through Day 24. Thus, the B phase was twice as long as the A phase but it was extended for obvious clinical reasons.

The 12-week follow-up period revealed a zero level of gagging, with the exception of Week 9, when three gagging episodes were recorded. Follow-up data were corroborated by the patient’s wife, thus precluding the possibility that

![Graph showing frequency of gagging during baseline, treatment, and follow-up.](image-url)

**FIGURE 5.1** Frequency of gagging during baseline, treatment, and follow-up. (Figure 1, p. 103, from: Epstein, L. H., & Hersen, M. (1974). Behavioral control of hysterical gagging. *Journal of Clinical Psychology, 30*, 102–104. Copyright 1974 by American Psychological Association. Reproduced by permission.)
treatment only affected the patient's verbal report rather than diminution of actual problem behavior.

Although treatment appeared to be the effective ingredient of change in this study, particularly in light of the longevity of the patient's condition, it is conceivable that some unidentified variable coincided with the application of reinforcement procedures and actually accounted for observed changes. The A-B design does not permit a definitive answer to this question. It is notable that this design (baseline, treatment, and follow-up) could be easily carried out in an outpatient setting with minimal difficulty and with no deleterious effects to the patient.

A-B with multiple target measures and follow-up

The clinical researcher often is interested in using multiple measures of a single target behavior, or in changing multiple target behaviors. The A-B design is well-suited for this purpose. An example of an A-B design with multiple target measures and follow-up assessment is provided by de Kinkelder and Boelens (1998), who studied the effects of habit reversal on stuttering behavior in two boys. Habit reversal training consists of teaching a person to become aware of the undesirable behavior to be modified and training in the performance of a behavior that is incompatible with the undesired behavior. In this study, the boys were recorded having conversations at school, at home, and at a speech clinic and the primary dependent variables included the percentage of syllables stuttered from among the first 300 syllables recorded at each assessment point, as well as the overall rate of speech and naturalness of speech as rated by independent coders. Each case began with a two- or three-week baseline phase (A) during which the boys' speech was recorded multiple times in each setting. After baseline behavior was observed to be stable, the intervention phase (B) was initiated, which consisted of four procedures: awareness training, training in regulated breathing, training in positive attitudes, and social support from parents. The treatment phase lasted for 21 sessions for the first boy and 31 sessions for the second.

Examination of the data in Figure 5.2 indicates that there were substantial decreases in stuttering for both boys during the treatment phase. Moreover, data collected for both boys also showed that improvements occurred across all three settings, that there were also substantial improvements in speech rates and naturalness across settings, and that these improvements appeared to be maintained at post-treatment and follow-up assessments. This, then, from a design standpoint, fits in nicely with Kazdin's notion of repeated assessment with marked changes and stability information improving the quality of case study. But, in spite of this, the A-B design does not allow for a clear demonstration of the controlling effects of the treatment. For that we require an A-B-A or A-B-A-B strategy.

Another example of the use of an A-B design with multiple target measures and follow-up assessment is provided by O'Donohue, Plaud, and Hecker (1992). These experimenters tested the effects of a contingency management program
FIGURE 5.2 Stuttering frequencies during baseline, treatment, posttreatment, and follow-up for each child in each setting (clinic, home and school). The frequencies are shown as a function of time since the beginning of the study. The arrows indicate the end of awareness training. BL = Baseline. Figure 1, p. 264, from: de Kinkelder, M., & Boelens, H. (1998). Habit-reversal treatment for children's stuttering: Assessment in three settings, *Journal of Behavior Therapy and Experimental Psychiatry, 29* (3), 261–269.

on engagement in out-of-home behaviors in a 42-year-old woman with agoraphobia. The subject in this case had not left her home for the previous 7.5 years. During a 30-day baseline phase (A), the experimenters collected data on total time spent out of the home. The experimenters also collected data on the subject's level of depression as measured by the Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) as well as level of fear and avoidance on the agoraphobia subscale of the Fear Questionnaire (Marks & Mathews, 1979), both administered at several points over the course of the study. Data also were collected during the baseline phase on the subject's engagement in pleasurable activities while in the home, and this information was used to develop the intervention.

After the 30-day baseline phase, an 18-day intervention phase (B) was initiated during which the subject agreed to only engage in several identified pleasurable/rewarding activities while outside of her home (e.g., watching television, but only at a friend's house). The experimenters collected data over a 60-day follow-up period, which was itself followed by 18-months of more traditional cognitive behavior therapy at the request of the subject. Thus, this design might be considered an A-B-A (given the extensive follow-up phase) or an A-B-A-C (given the addition of a second intervention phase). Given measurement occurred much less frequently over the second intervention phase, we focus more specifically on the A-B phases and follow-up period described in more detail in this report.

The results of the initial intervention are presented in Figure 5.3. As shown, time spent outside the home remained at a zero level across the baseline phase (A), but increased substantially during the intervention phase. This increase was maintained across the two-month follow-up period. Results also revealed that BDI scores had a decreasing-increase trend over the baseline phase, but that this trend was reversed during the intervention phase, and continued on a decreasing trend over the course of follow-up. Interestingly, fear and avoidance decreased during baseline, and increased over the course of the intervention. However, they decreased once again during follow-up and continued to do so during the last phase. The increases observed on the Fear Questionnaire during the treatment phase may have been due to the nature of the intervention, which required the subject to engage in feared activities.

**A-B with follow-up and booster treatment**

In our next illustration of an A-B design, clinical considerations necessitated a short baseline period and also contraindicated the withdrawal of treatment procedures (Harbert, Barlow, Hersen, & Austin, 1974). However, during the course
of extended follow-up assessment, the patient's condition deteriorated and required the reinstatement of treatment in booster sessions. Renewed improvement immediately followed, thus lending additional support for the treatment's efficacy. When examined from a design standpoint, the conditions of the more complete A-B-A-B strategy are approximated in this experimental case study.

Harbert et al. (1974) examined the effects of covert sensitization therapy on self-report (card sort technique) and physiological (mean penile circumference changes) indices in a 52-year-old male inpatient who complained of a long history of incestuous episodes with his adolescent daughter. The card sort technique consisted of 10 scenes (typed on cards) depicting the patient and his daughter. Five of these scenes were concerned with normal father-daughter relations; the remaining five involved descriptions of incestuous activity between father and daughter. The patient was asked to rate the 10 scenes, presented in random sequence, on a 0-4 basis, with 0 representing no desire and 4 representing much desire. Thus, measures of both deviant and non-deviant aspects of the relationship were obtained throughout all phases of study. In addition, penile circumference changes scored as a percentage of full erection were obtained in response to audio-taped descriptions of incestuous activity and in reaction to slides of the daughter. The inclusion of objective data that do not rely on self-report is an important aspect of the single-case experimental design. Three days of self-report data and 4 days of physiological measurements were taken during baseline (A phase). It is notable that while most single-case studies have historically used observations of motor behavior, the use of physiological or other behavioral measures (e.g., Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Nock & Banaji, 2007) can be valuable additions to such designs (see chapter 4).

Covert sensitization treatment (B phase) consisted of approximately 3 weeks of daily sessions in which descriptions of incestuous activity were paired with the nauseous scene as used by Barlow, Leitenberg, and Agras (1969). However, as nausea proved to be a weak aversive stimulus for this patient, a "guilt" scene—in which the patient is discovered engaging in sexual activity with the daughter by his current wife and a respected priest—was substituted during the second week of treatment. The flexibility of the single-case approach is exemplified here inasmuch as a "therapeutic shift of gears" follows from a close monitoring of the data. Follow-up assessment sessions were conducted after termination of the patient's hospitalization at 2-week, and 1-, 2-, 3-, and 6-month intervals. After each follow-up session, brief booster covert sensitization was administered.

The results of this study are presented in Figures 5.4 and 5.5. Inspection of Figure 5.5 indicates that mean penile circumference changes to audiotapes in baseline ranged from 18% to 35% (mean = 22.8%). Penile circumference changes to slides ranged from 18% to 75% (mean = 43.5%). Examination of Figure 5.5 shows that non-deviant scores remained at a maximum of 20 for all three baseline probes; deviant scores achieved a level of 17 throughout.

The introduction of standard covert sensitization, followed by the use of guilt imagery resulted in decreased penile responding to audio-tapes and slides
FIGURE 5.4 Mean penile circumference change to audiotapes and slides during baseline, covert sensitization, and follow-up. (Figure 1, p. 83, from: Harbert, T. L., Barlow D. H., Hersen, M., & Austin, J. B. (1974). Measurement and modification of incestuous behavior: A case study, *Psychological Reports*, 34, 79-86. Copyright 1974 by Psychologica Reports. Reproduced by permission.)

(see Figure 5.4) and a substantial decrease in the patient's self-report of deviant interests in his daughter (see Figure 5.5). Non-deviant interests, however, remained at a high level.

Follow-up data in Figure 5.4 reveal that penile circumference changes remained at zero during the first three probes but increased slightly at the 3-month assessment. Similarly, Figure 5.5 data show a considerable increase in deviant interests at the 3-month follow-up. This coincides with the patient's reports of marital disharmony. In addition, non-deviant interests diminished during follow-up (at that point the patient was angry at his daughter for rejecting his positive efforts at being a father).

As there appeared to be some deterioration at the 3-month follow-up, an additional course of outpatient covert sensitization therapy was carried out in three weekly sessions. The final assessment period at 6 months appeared to reflect the effects of additional treatment in that (1) penile responding was negligible, and (2) deviant interests had returned to a zero level.
FIGURE 5.5 Card sort scores on probe days during baseline, covert sensitization, and follow-up. (Figure 2, p. 84, from: Harbert, T. L., Barlow, D. H., Hersen, M., & Austin, J. B. (1974). Measurement and modification of incestuous behavior: A case study. Psychological Reports, 34, 79–86. Copyright 1974 by Psychological Reports. Reproduced by permission.)

5.3. A-B-A DESIGN

The A-B-A design is the simplest of the experimental analysis strategies in which the treatment variable is introduced and then withdrawn. For this reason, this strategy as well as those that follow, are most often referred to as withdrawal designs. Whereas the A-B design permits only tentative conclusions as to a treatment's influence, the A-B-A design allows for an analysis of the controlling effects of its introduction and subsequent removal. If after baseline measurement (A) the application of a treatment (B) leads to improvement and there is deterioration after it is withdrawn (A), one can conclude with a high degree of certainty that the treatment variable is the agent responsible for observed changes in the target behavior. Unless the natural history of the behavior under study were to follow identical fluctuations in trends (see chapter 3, section 3.6 on cyclic variation), it is most improbable that observed changes would be due to any influence (e.g., some correlated or uncontrolled variable) other than the treatment variable that is systematically changed. Also, replication of the A-B-A design in different subjects strengthens conclusions as to power and controlling forces of the treatment (see chapter 10).

Although the A-B-A strategy is acceptable from an experimental standpoint, it has one major undesirable feature when considered from the clinical context. Unfortunately for the patient or subject, this paradigm ends on the A or
Figure 5.6 Serum sodium concentration documented through baseline (B1), intervention (A1) and post treatment (B2) phases. The time interval from the last baseline (B1) assessment to the first intervention (A1) point was one week and the time interval from the last intervention (A1) point to the first post-treatment (B2) point was 2 weeks. (Figure 1, p. 245, from: Thomas, J. L., Howe, J., Gaudet, A., & Brantley, P. J. (2001). Behavioral treatment of chronic psychogenic polydipsia with hyponatremia; A unique case of polydipsia in a primary care patient with intractable hiccups. *Journal of Behavior Therapy and Experimental Psychiatry*, 32, 241–250.)

example, Thomas and colleagues (Thomas, Howe, Gaudet, & Brantley, 2001) used in A-B-A design in examining the treatment of polydipsia (compulsive fluid consumption) and chronic hiccups (for 20 years) using an intervention composed of education about hiccups, reinforcement for decreased fluid consumption, and relaxation training. Results, presented in Figure 5.6, revealed a significant improvement during the intervention (B) phase, but no return to baseline functioning during the reintroduction of the A phase. This is good for clinical reasons as it suggests maintenance of treatment effects; however, such findings preclude the experimenters from drawing firm conclusions about the controlling effects of the intervention.

**A-B-A from child literature**

A classic example of an A-B-A design is provided by Walker and Buckley (1968) in their functional analysis of the effects of an individualized educational
program for a 91/2-year-old boy whose extreme distractibility in the classroom interfered with task-oriented performance (see Figure 5.7). During baseline assessment (A), percentage of attending behavior was recorded in 10-minute observation sessions while the boy was working on programmed learning materials. Following baseline measurement, a reinforcement contingency (B) was instituted whereby the boy earned points (exchangeable for a model of his choice) for maintaining his attention (operationally defined for him) to the learning task. A progressively increasing time criterion for attending behavior over sessions was required (30 to 600 seconds of attending per point). The extinction phase (A) involved a return to original baseline conditions.

Examination of baseline data shows a slightly decreasing trend followed by a slightly increasing trend, but within stable limits (mean = 33%). Institution of the reinforcement procedures led to an immediate improvement, which increased to its asymptote in accordance with the progressively more difficult criterion. Removal of the reinforcement contingency during extinction resulted in a decreased percentage of attending behaviors to approximately baseline levels.
were examined when applied to a 10-year-old boy with mental retardation whose “talking-out” behaviors during special education classes was disruptive, especially because other children were copying his talking out behaviors. Baseline observations of talk-outs were recorded by the teacher (reliability checks indicated 84% to 100% agreement) during five daily 15-minute sessions. During these first five sessions (A), the teacher responded naturally to talk-outs by attending to them. However, in the next five sessions (B) the teacher was instructed to ignore talk-outs but to provide increased attention to the child’s productive behaviors. The third series of five sessions involved a return to baseline conditions (A), and the last series of five sessions consisted of reinstatement of contingent attention (B).

The results of this study are plotted in Figure 5.8. The use of equal phases facilitates the analysis of results. Baseline data are stable and range from three to five talk-outs, with three of the five points at a level of four talk-outs per session. Institution of contingent attention resulted in a marked decrease that achieved a zero level by Sessions 9 and 10. Removal of contingent attention led to a linear increase of talk-outs to a high of five. However, reinstatement of contingent attention once again brought talk-outs under experimental control. Thus application and withdrawal of contingent attention clearly demonstrates its controlling effects on talk-out behaviors. This is twice-documented, as seen in the decreasing and increasing data trends in the second set of A and B phases.

**FIGURE 5.8** A record of talking out behavior of an educable mentally retarded student. Baseline, before experimental conditions. Contingent Teacher Attention, systematic ignoring of talking out and increased teacher attention to appropriate behavior. Baseline, reinstatement of teacher attention to talking out behavior. (Figure 2, p. 143, from: Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, T., & Porcia, E. (1971). The teacher as observer and experimenter in the modification of disputing and talking-out behaviors. *Journal of Applied Behavior Analysis*, 4, 141-149. Copyright 1971 by Society for the Experimental Analysis of Behavior, Inc. Reproduced by permission.)
A more recent example from the child literature is a study by Facon, Beghin and Riviere (2007) in which two visually impaired boys with a history of making inappropriate verbalizations were treated using planned ignoring of inappropriate verbalizations and differential reinforcement of appropriate verbalizations. The intervention was tested in each boy using an A-B-A-B design.

During all phases, the child was observed interacting with a therapist for 12 to 30 minute sessions. Each session was divided into 10 second intervals and the percentage of intervals during which appropriate and inappropriate verbalizations were made by the child were calculated (inter-observer agreement was 91%). During the baseline (A) phases, therapists responded to all verbalizations with a neutral response. During both treatment (B) phases, therapists ignored inappropriate verbalizations (i.e., did not speak until 10 seconds after the end of the verbalization) and responded to appropriate verbalizations with statements expressing strong interest in the child's statements.

The results of this study are presented in Figure 5.9. As shown for both children, inappropriate verbal responses occurred more frequently than appropriate responses during the initial baseline. During the intervention phase, this pattern changed significantly and appropriate verbalizations occurred much more frequently than inappropriate ones. The return to baseline showed an immediate reversal back to baseline levels, and the second intervention phase showed an immediate return of the intervention effects.

With respect to design considerations, we have here a very clear demonstration of the efficacy of the intervention on two occasions. As was the case in our prior example (R. V. Hall et al., 1971) baselines (especially the second) were shorter than treatment phases. However, in light of the immediate and dramatic improvements as a result of the intervention, the possible confound of time and length of adjacent phases does not apply in this analysis.

**A-B-A-B when phase change is not under complete experimental control**

As mentioned above, although the use of the A-B-A-B design allows the experimenter to draw strong inferences about treatment effects, it is not always clinically desirable or feasible to withdraw an effective intervention during the second baseline phase, particularly in the case of dangerous or harmful behaviors such as aggressive or self-injurious behaviors. Thus, it is possible that an experimenter begins an A-B-A-B design and is unable to withdraw treatment and must abandon this design in favor of an A-B design or an A-B-A-B with an abbreviated second baseline.

Another possibility, though, is that the subject him- or herself stops participating in the intervention during the first B phase, thus creating a quasi-experimental return to baseline. In such cases, the experimenter can flexibly continue with the investigation, but is now limited in the inferences that can be drawn given there is no longer complete experimental control over study procedures. As an example, Wallenstein and Nock (2007) tested the effects of aerobic
FIGURE 5.9 Percentage of intervals with appropriate and inappropriate verbal responses across experimental conditions for Baptiste and Paul. (Figure 1, p. 26, from: Facon, B., Beghin, M., & Riviere, V. (2007). The reinforcing effect of contingent attention on verbal perseverations of two children with severe visual impairment. *Journal of Behavior Therapy and Experimental Psychiatry*, 38 (1), 23–28.)

exercise on engagement in non-suicidal self-injury (NSSI; e.g., skin cutting) in a 26-year-old woman with a 13-year history of such behavior. In this study, the subject kept a daily log of self-injurious urges and behaviors across all phases. The study was planned as an A-B-A-B design, with treatment consisting of the subject engaging in a 60-minute aerobic exercise regimen (guided by an exercise video) three times per week and in response to any urges to engage in NSSI.
As presented in Figure 5.10, the subject reported a significant decrease in NSSI during the first, 5-week treatment (B) phase. Before the experimenters instructed the subject to discontinue use of the exercise regimen, the subject voluntarily decided to do so. During this quasi-experimental return to baseline (A), the subject reported a return of self-injurious urges and behaviors to initial baseline levels. This was followed by an instruction to re-introduce the exercise regimen (B), which was associated with a return of treatment effects. The quasi-experimental nature of this design precludes the experimenters from drawing causal inferences about the effects of the intervention, as unmeasured historical factor (e.g., break-up with significant other, problems in treatment) may have caused the subject to both stop exercising (i.e., begin the second A phase) and start engaging in NSSI once again. Nevertheless, these data provide some evidence for the effectiveness of this treatment approach and offer a basis for future experimental replications.

A-B-A-B with unexpected improvement in baseline

Difficulties can arise when interpreting the results of an A-B-A-B design if improvement occurs unexpectedly during baseline assessment. As an example, Epstein, Hersen, and Hemphill (1974) used an A-B-A-B design to study the
effects of feedback on frontalis muscle activity in a patient who had suffered from chronic headaches for 16 years. EMG recordings were obtained while the patient relaxed in a reclining chair in the experimental laboratory for 10 minutes following 10 minutes of adaptation during each of six baseline (A) sessions. During the six feedback (B) sessions, the patient's favorite music (prerecorded on tape) was automatically turned on whenever EMG activity decreased below a preset criterion level. In contrast, responses above that level turned off the music. The patient was instructed to "keep the music on" during the treatment phase. The intervention phase was followed by a six session return to baseline (A), which was itself followed by a six session return to feedback (B). The patient also was instructed to keep a record of the intensity of headache activity across all phases.

The results of this study are presented in Figure 5.11. As shown, EMG activity during baseline ranged from 28 to 50 seconds (mean = 39.18) per minute that contained integrated responses above the criterion microvolt level. Institution of feedback procedures resulted in decreased activity (mean = 23.18). Removal of feedback in the second baseline initially resulted in increased activity in Sessions 13–15. However, an unexplained but decreased trend was noted in the last half of that phase. This downward trend, to some extent, detracts from the interpretation that music feedback was the responsible agent of change during the first B phase. The importance of maintaining equal lengths of phases is highlighted here. If baseline measurement had been concluded on Day 15, an unequivocal interpretation (though probably erroneous) would have been made. However, despite the downward trend in baseline, mean data for this phase (30.25) were higher than for the previous feedback phase (23.18).

![Figure 5.11](image_url)
In the final phase, feedback resulted in a further decline that was generally maintained at low levels (mean = 14.98). Unfortunately, it is not clear whether this further decrease might have occurred naturally without the benefits of renewed introduction of feedback. Therefore, despite the presence of statistically significant differences between baseline and feedback phases and confirmation of EMG differences by self-reports of decreased headache intensity during feedback, the downward trend in the second baseline prevents a definitive interpretation of the controlling effects of the feedback procedure.

When the aforementioned data pattern results, it is recommended that variables that might have caused improvement during baseline be examined through additional experimental analyses. However, time limitations and pressing clinical needs of the patient or subject under study usually preclude such additional study. Therefore, the next best strategy involves a replication of the procedure with the same subject—or with additional subjects bearing the same kind of diagnosis (see chapter 10).

**A-B-A-B with monitoring of concurrent behaviors**

Most studies reporting on the use of the A-B-A-B design test the effects of an intervention on only one target behavior. However, a number of reports have highlighted the importance of monitoring concurrent (non-targeted) behaviors as well. This is of particular importance when side effects of treatment are possibly negative (see Sajwaj, Twardosz, & Burke, 1972). Kazdin (1973) has noted some of the potential advantages in monitoring the multiple effects of treatment on operant paradigms:

One initial advantage is that such assessment would permit the possibility of determining response generalization. If certain response frequencies are increased or decreased, it would be expected that other related operants would be influenced. It would be a desirable addition to determine generalization of beneficial response changes by looking at behavior related to the target response. In addition, changes in the frequency of responses might also correlate with topographical alterations. (p. 527)

The examination of potential collateral effects of an intervention of course should not be restricted to operant paradigms, but can be examined more generally whenever using experimental single-case designs.

As an example, Twardosz and Sajwaj (1972) used an A-B-A-B design to evaluate the efficacy of an intervention designed to increase sitting behavior in a 4-year-old, hyperactive, mentally retarded boy who was enrolled in an experimental preschool class. Beyond the assessment of the primary target behavior (sitting), the experimenters examined the effects of treatment on a variety of concurrent behaviors (posturing, walking, use of toys, and proximity of children). Observations were made during a free-play period (one-half hour) in which class members could choose their playmates and toys. During baseline (A),
the teacher gave the child instructions (as she did to all others in class) but did not prompt him to sit or praise him when he did. Institution of the sitting program (B) involved prompting the child (placing him in a chair with toys before him on the table), praising him for remaining seated and for evidencing other positive behaviors, and awarding him tokens (exchangeable for candy) for in-seat behavior. The sitting program was withdrawn in the third phase (A) and reinstated in the fourth phase (B).

The results of this study appear in Figure 5.12. Examination of the top part of the graph shows that the sitting program, with the exception of the last day in the first treatment phase, led to improvement over baseline conditions on both occasions. Continued examination of the figure reveals that posturing decreased during the sitting program, but walking remained at a consistent rate throughout all phases of study. Similarly, use of toys and proximity to children increased during administrations of the sitting program. In discussing their results, Twardosz and Sajwaj (1972) stated that:

This study . . . points out the desirability of measuring several child behaviors, although a modification procedure might focus on only one. In this way the preschool teacher can assess the efficacy of her program based upon changes in other behaviors as well as the behavior of immediate concern. (p. 77)

However, in the event that non-targeted behaviors remain unmodified or that deterioration occurs in others, additional behavioral techniques can then be applied (Sajwaj, Twardosz, & Burke, 1972). Under these circumstances it might be preferable to use a multiple baseline strategy in which attention to each behavior can be programmed in advance (see chapter 7).

A-B-A-B with no feedback to experimenter

As mentioned in Chapter 3, one of the main advantages of single-case experimental designs is that they provide the experimenter with ongoing feedback about the target behavior(s) and also allow the experimenter to flexibly modify the intervention over the course of the study. Thus, changes from one phase to the next are accomplished with the experimenter's full knowledge of prior results. Although these factors benefit the experimental clinician, they can introduce problems from a purely experimental standpoint. Critics of single-case experimental designs have suggested that the experimenter could introduce bias in the evaluation of the target behavior and in the actual application and withdrawal of specified techniques. One method of preventing such "bias" is to determine lengths of baseline and experimental phases on an a priori basis, while keeping the experimenter uninformed as to trends in the data during their collection. A problem with this approach, however, is that decisions regarding choice of baselines and those concerned with appropriate timing of institution and removal of therapeutic variables are left to chance rather than adjusted in order to display stability in the data.
FIGURE 5.12  Percentages of Tim’s sitting, posturing, walking, use of toys, and proximity to children during freeplay as a function of the teacher’s ignoring him when he did not obey a command to sit down. (Figure 1, p. 75, from: Twardosz, S., & Sajwaj, T. (1972). Multiple effects of a procedure to increase sitting in a hyperactive retarded boy. *Journal of Applied Behavior Analysis, 5*, 73–78. Copyright 1972 by Society for the Experimental Analysis of Behavior, Inc. Reproduced by permission.)
The use of an A-B-A-B design with a set number of sessions per phase no feedback to the experimenter (i.e., data were obtained from video recordings for all phases after the experiment ended) appears in a report Hersen, Miller, and Eisler (1973). The authors examined the effects of various conversational topics (non-alcohol and alcohol-related) on duration of loo and duration of speech in four heavy drinkers and their wives in ad libitum it actions videotaped in a television studio. Following 3 minutes of "warm-interaction, each couple was instructed to converse for 6 minutes (A phase) at any subject unrelated to the husband's drinking problem. Instructions repeated at 2-minute intervals over a two-way intercom from an adjoining room to ensure maintenance of the topic of conversation. In the next 6 min (B phase) the couple was instructed to converse only about the husband's drinking problem (instructions were repeated at 2-minute intervals). The last 12 min of interaction consisted of identical replications of the A and B phases.

The results of this study are presented in Figure 5.13 as mean data for four couples. As shown, speech duration data showed no trends across experimental phases for either husbands or wives. Similarly, duration of looking not vary across phases for husbands. However, duration of looking for wives significantly greater during alcohol- than non-alcohol-related segments of interaction. In the first non-alcohol phase, looking duration ranged from 26 to 43 seconds, with an upward trend in evidence. In the first alcohol phase (B), duration of looking ranged from 57 to 70 seconds, with a continuation of the upward linear trend. Reintroduction of the non-alcohol phase (A) resulted in a decrease in looking (38 to 45 seconds). In the final alcohol segment (B), looking once again increased, ranging from 62 to 70 seconds.

An analysis of these data does not allow for conclusions with respect to initial A and B phases inasmuch as the upward trend in A continued into B. However, the decreasing trend in the second A phase succeeded by the increase in the second B phase suggests that topic of conversation had a controlling influence on the wives' rates of looking. We might note here that if the experimenters were in position to monitor their results throughout all experimenter phases, the initial segment probably would have been extended until the wives' looking duration achieved stability in the form of a plateau. Then the second phase would have been introduced.

5.5. B-A-B DESIGN

An important variant of the A-B-A design is the B-A-B design, in which the first phase (B) typically involves the application of a treatment. In the second phase (A) the treatment is withdrawn and in the final phase (B) it is reinstated. Some investigators have introduced an abbreviated baseline session prior to the major B-A-B phases, resulting in the popular A-B-A-B design. Clinically, the B-A-B design is superior to the A-B-A design because the study begins and ends with the VOI.

administration of treatment. However, absence of an initial baseline phase precludes an analysis of the effects of treatment over the natural frequency of occurrence of the targeted behaviors under study. Therefore, the use of the more complete A-B-A-B design is preferred for assessment of singular therapeutic variables. Below we review two classic examples of the B-A-B design to illustrate this strategy.

B-A-B with group data

Ayllon and Azrin (1965) used the B-A-B strategy on a group basis in their evaluation of the effects of token economy on the work performance of 44 patients with psychosis. During the first 20 days (B phase) of the experiment, subjects were awarded tokens (exchangeable for a large variety of “backup” reinforcers) for engaging in work activities in their hospital inpatient unit. Over the next
20 days (A phase) subjects were given tokens on a non-contingent basis, regardless of their work performance. Each subject received tokens daily, based on the mean daily rate obtained in the initial B phase. In the last 20 days (second B phase) the contingency system was reinstated. Notably, this design also could be labeled a B-C-B design, as the middle phase is not a true measure of the natural frequency of occurrence of the target measure (see section 5.6).

The results of this study appear in Figure 5.14. During the first B phase, the entire group averaged about 45 hours of work per day. Removal of the contingency in the A phase resulted in a marked decrease to a level of one hour per day on Day 36. Reinstitution of the token reinforcement program in the second B phase led to an immediate increase in hours worked to a level approximating the first B phase. Thus, Ayllon and Azrin (1965) presented the first experimental demonstration of the controlling effects of token economy over work performance in state hospital psychiatric patients.

Importantly, when single-case experimental strategies are used on a group basis, it behooves the experimenter to show that a majority of those subjects exposed to and then withdrawn from treatment provide supporting evidence for its controlling effects. Individual data presented for selected subjects can be quite useful, particularly if data trends differ. Otherwise, difficulties inherent in
the traditional group comparison approach (e.g., averaging out of effects, effects due to a small minority while the majority remains unaffected by treatment) will be carried over to the experimental analysis procedure. In the current example, Ayllon and Azrin (1965) showed that 36 of their 44 subjects decreased their performance from contingent to non-contingent reinforcement. Conversely, 36 of 44 subjects increased their performance from non-contingent to contingent reinforcement. Eight subjects were totally unaffected by contingencies and maintained a zero level of performance in all phases.

Extending the previous point, virtually all of the designs discussed in this chapter can be used on a group basis. The overall strategy remains the same, but the specific procedures used and inferences permitted may differ on a case by case basis. For instance, Caruso and Kennedy (2004) used an A-B-A-B design to study the effect of a journal reviewer prompting strategy on the timeliness of reviews. Rather than following the same individual(s) over time, the authors observed the timeliness of journal article reviews during normal baseline phases (A) and during experimental phases (B) that consisted of an e-mail reminder to reviewers 7 days prior to the date the review was due. Results of this study revealed that prompting led to more on-time reviews and less variability in return times. The use of this design on a group basis provides compelling data about the effectiveness of the intervention; however, difficulties inherent in the use of group designs mentioned above remain.

**B-A-B from rogerian framework**

Although single-case experimental designs often are discussed within the context of the cognitive behavioral paradigm, these strategies can be easily and flexibly employed to study other psychological treatments. For instance, Truax and Carkhuff (1965) provided an earlier example of this in their well-known study on Rogerian therapy. These authors systematically examined the effects of high and low “therapeutic conditions” on the responses of 3 psychiatric patients during the course of initial 1-hour interviews. Each of the interviews consisted of the three 20-minute phases. In the first phase (B) the therapist was instructed to evidence high levels of “accurate empathy” and “unconditional positive warmth” in his interactions with the patient. In the following A phase the therapist experimentally lowered these conditions, and in the final phase (B) they were reinstated at a high level.

Each of the three interviews was audiotaped. From these audiotapes, five 3-minute segments for each phase were obtained and rerecorded on separate tapes. These tapes were then presented to raters (naive as to which phase the tape originated in) in random order. Ratings made on the basis of the Accurate Empathy Scale and the Unconditional Positive Regard Scale confirmed (graphically and statistically) that the therapist followed directions as indicated by the dictates of the experimental design (B-A-B). The effects of high and low therapeutic conditions were then assessed in terms of depth of the patient’s intrapersonal exploration.
Once again, 3-minute segments from the A and B phases were presented to “naive” raters in randomized order. These new ratings were made on the basis of the Truax Depth of Interpersonal Exploration Scale (reliability of raters per segment = .78).

Data on depth of intrapersonal exploration are plotted in Figure 5.15. Visual inspection of these data indicates that depth of intrapersonal exploration, despite considerable overlapping in adjacent phases, was somewhat lowered during the middle phase (A) for each of the three patients. Although these data are less than ideal (i.e., overlap between phases), this study illustrates that the controlling effects of therapist relational factors can be investigated systematically using the experimental analysis of behavior model.

5.6. A-B-C-B DESIGN

Another important variant of the A-B-A-B design is the A-B-C-B design, in which only the first two phases of experimentation consist of baseline and contingent reinforcement phases. In the third phase (C), instead of returning to baseline observation, reinforcement is administered in proportions equal to the preceding B phase but on a totally noncontingent basis. This phase controls for the added attention (“attention-placebo”) that a subject receives for being in a treatment condition and is analogous to the A₁ phase (placebo) used in drug evaluations (see chapter 6). In the final phase, contingent reinforcement procedures are reinstated. Thus the last three phases of study are identical to those in the B-A-B study used
by Ayllon and Azrin (1965) in the example described in section 5.5. In the A-B-C-B design the A and C phases are not comparable, inasmuch as experimental procedures differ. Therefore, the main experimental analysis is derived from the B-C-B portion of study. However, baseline observations are of some value, as the effects of B over A are suggested (here we have the limitations of the A-B analysis).

**A-B-C-B from the child literature**

Goldstein, Kaczmarek, Pennington, and Shafer (1992) examined the effects of a peer-mediated intervention on social interaction among 5 children diagnosed with autism. During all phases of the study observers coded audio- and videotaped interactions between five triads of children, each triad containing one autistic child and two non-autistic peers trained in facilitation strategies designed to increase social interaction behaviors performed by the target children.

During baseline (A phase) all children in each triad were instructed to talk and play with their friends. During the intervention phase (B), peers were instructed to use the facilitation strategies to increase social interaction behaviors performed by the target child in the triad. During the next experimental phase (C), peers were instructed to continue using their social interaction skills, but not to direct them only at the target child, and they were praised only for interactions with the other (non-autistic peer). Importantly, peers were allowed to interact with target children, but were only praised for interactions with non-target children. The authors note that they selected such an A-B-C-B design rather than a withdrawal (A-B-A-B) design in order to demonstrate that behavior change was related to the treatment variable itself, and not merely to the presence of any peer interaction. Finally, the last phase (B) in this design consisted of a return to the intervention in which peer facilitation strategies were directed at the target child.

The results from this study were presented individually for each target child. The results from one of those children are presented in Figure 5.16. As shown, the total number of communicative acts performed by the target child increased substantially during the peer intervention. This pattern was reversed during the reversal (C) phase, during which the frequency of social behaviors and total communicative behaviors directed toward the target children decreased to baseline (A) levels. Reintroduction of the peer intervention in the final phase (B) showed a return of the treatment effects, providing evidence of the controlling effects of the intervention. This pattern of results was observed in 4 of the 5 target children studied.

**A-B-C-B in a group application and follow-up**

An interesting application of the A-B-C-B design to a group of subjects was reported by Porterfield, Blunden, and Blewitt (1980). Subjects in this experimental analysis were "profoundly mentally handicapped" adults attending a center for those with mental retardation. The behavior targeted for modification
FIGURE 5.16 The frequency of social behavior demonstrated by target children. The circles indicate the total number of social behaviors, and the triangles indicate the total number of communicative acts. (Figure 2, p. 297, from: Goldstein, H., Kaczmarek, L., Pennington, R., & Shafer, K. (1992). Peer-mediated intervention: Attending to, commenting on, and acknowledging the behavior of preschoolers with autism. *Journal of Applied Behavior Analysis, 25* (2), 289–305.)
was participation in activities during a 1-hour period so designated during the 19 days of the study. Participation was defined by 12 separate activities and involved some of the following: watching television, dancing, responding to a verbal command, talking to another subject, and eating without assistance.

The baseline phase (A) lasted 3 days, with three staff members interacting with subjects in normal fashion. No specific instructions were given at this point. The B phase (room manager) lasted 5 days, with two staff members alternating for half-hour periods. Subjects in this condition were prompted and differentially reinforced for their participation. The C phase (no distraction) lasted 6 days and involved a maximum of two prompts to engage in activity, but subjects were not differentially reinforced. In the fourth phase (B) the room manager condition was reinstated. Then there was a 69-day follow-up period involving the room manager condition in the absence of the experimenter.

Data appear in Figure 5.17 and are presented as the percentage of subjects (i.e., trainees) engaged in activity. It is clear that baseline (A) functioning was poor, ranging from 25.7% to 37.9% participation. Introduction of the room manager (B) condition led to marked increases in participation (72.9% to 90.9%). However, when the no-distraction (C) condition was introduced, participation

![TRAINEE ENGAGEMENT](image)

**FIGURE 5.17** Percentage of trainees engaged during the activity hour for 19 days and follow-up days. (Figure 1, p. 236 from: Porterfield, J., Blunden, R., & Blewitt, E. (1980). Improving environments for profoundly handicapped adults: Using prompts and social attention to maintain high group engagement. *Behavior Modification, 4*, 225–241. Copyright 1980 by Sage Publications. Reproduced by permission.)
decreased to near baseline levels (21.5% to 48.0%). When the room management condition was reintroduced, in the second B phase, level of participation once again increased to 84.7% to 88.1%. This second application of the room management condition clearly documented the controlling effects of the contingency. Furthermore, data in follow-up confirmed that participation could be maintained (71.5% to 91.1%) in the absence of experimental prompting.

There are two noteworthy features in this particular example of the A-B-C design. First, even though the A and C phases were technically dissimilar, they were functionally quite similar. That is, the data pattern was the same as would be observed in an A-B-A-B design. However, contrary to the A-B-A-B design where there are two instances of confirmation of the contingency, only the B-C portion of the design truly reflected the controlling aspects of the room management intervention. Second, by making the dependent measure the “percentage of trainees engaged,” the experimenters obviated the necessity of providing individual data. However, from a single-case perspective, data as to percentage time active for each trainee would be most welcome.