DOCUMENT RESUME

ED 201 092 EC 132 517

AUTHOR MacLeod, James C.: And Others
TITLE Single Subject Procedures: Applications for Educational Settings.
INSTITUTION Washington Univ., Seattle. Western States Technical Assistance Resource.
SPONS AGENCY Office of Special Education (ED), Washington, D.C.
PUB DATE Jul 80
CONTRACT 300-77-0508
NOTE 25p.: Handicapped Children's Early Education Program Series Paper No. 7.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Disabilities: Elementary Secondary Education: *Evaluation Methods: Program Evaluation: *Research Design: *Research Methodology: Student Evaluation
IDENTIFIERS *Single Subject Research Design

ABSTRACT

The monograph defines and discusses single subject research designs in special education. Advantages of this approach with measuring change in handicapped children are noted. Observational techniques (event recording, duration recording, interval recording, and time sampling) are described, along with three types of single subject evaluation procedures (A-B, reversal, and multiple baseline). A final section considers the presentation and interpretation of results. The authors conclude that the methodology of the single subject approach is simple and well suited to the individualized education program requirement of P.L. 94-142, the Education for All Handicapped Children Act. (CL)

Single Subject Procedures

Applications for Educational Settings

7*

*in a series of papers on topics of in[illegible]est to projects within the Handicapped Children's Early Educat[illegible] Program. Distributed by the **Western States Technical Assistance Reso[illegible]** Seattle, Washington.

WESTAR Coordinator for Series Paper
David [illegible]

WESTAR
Western States Technical Assistance Resource
345 N. Monmouth Avenue
Monmouth, Oregon 97361
503/838-1220, ext. 391

Reproduced and distributed under contract number 300-77-0518 from the Office of Special Education, U.S. Department of Education. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement should be inferred.

Single Subject Procedures

Applications for Educational Settings

James C. MacLeod
Judy Andrews
David N. Gr[illegible]

Child Neurology Clinic
Good Samaritan Hospital
Portland, Oregon

Distributed by WESTAR
July 1980

In view of demands for greater teacher accountability in the education of handicapped children, the use of empirically-based methods for documenting child change represents a valuable addition to educational programs. Since the relative impact of various instructional methods may not always be readily apparent, techniques for evaluating the impact of instruction on the individual child are necessary. Single subject evaluation procedures, the most appropriate technique to meet these needs, are particularly applicable to special education classrooms including handicapped preschool programs, resource rooms, and TMR and EMR programs.

Baer, Wolf, and Risley (1968) discuss single subject procedures in the context of "applied behavior analysis." This refers to the systematic study of the relationship between specific behaviors of the individual child and the specific techniques which are employed to change the child's behaviors. Lovitt (1975) describes five attributes of single subject evaluation procedures: (1) the direct measurement of the child's behavior, (2) the continuous measurement of the child's behavior on a daily or near daily basis, (3) an understanding of the idiosyncrasies of the child's behavior, (4) the demonstration of a functional relationship between the intervention and the child's behavior, and (5) interventions that are adequately described and therefore replicable. As such, single subject evaluation procedures are highly compatible with special education activities due to the emphasis placed upon the individual child.

Although group evaluation procedures have traditionally been used in the evaluation of treatment effectiveness, single subject evaluation procedures may be more appropriate when studying handicapped children for several reasons.

1. The variability with which the behavior of the individual child occurs may be documented. As such, fluctuations in the occurrence of the child's behavior may be accounted for on an ongoing basis.
2. The individual child serves as his own control. Therefore, any changes in the child's behavior may be evaluated within the context of the child's unique characteristics and behavioral repertoire. Single subject procedures show when the child's behavior has changed and also how it has changed. Further, single subject procedures document the relative speed with which the child acquires new skills and also the manner in which these skills were acquired.
3. The procedures grew out of problems frequently addressed by teachers. As such, the impact of instruction upon the individual child may be evaluated without disrupting ongoing educational activities. Single subject procedures are appropriately used in settings where the primary goal is to remediate the child's skill deficits or improve the child's social

functioning. Single subject procedures do not necessitate that changes occur in the intervention strategy for the child. Rather, they are meant to follow the course of treatment and instruction which may best serve the needs of the child.

4. In single subject procedures the relative impact of instruction is evaluated by visually analyzing the graphically presented data. Therefore, the use of statistical procedures are not necessary because the significance of the intervention is visually apparent.

While single subject evaluation procedures are useful in documenting individual sources of variance and the relative impact of intervention, there are certain limitations which must be considered. The degree to which one may generalize from the individual child to other children may be limited. As such, one cannot infer that an intervention which is effective with one child will be equally effective with another child. Therefore, Leittenberg (1973) advocates the use of single subject evaluation procedures for comparing the different interventions once their effectiveness has been established. Thus, single subject designs are useful techniques for the evaluation of specific instructional methods, but may not be sufficient for the comparison of different instructional methods which are already effective.

Group designs involve the comparison of two or more groups of children who have received different instructional methods. For example, an educator may be interested in determining the impact of one instructional method as compared with another. He may choose to divide a set of students into two groups, giving one group a particular instructional method while the other group receives another instructional method. Based upon some measure (for example. scores on a standardized test), the educator may compare the two groups to determine the relative effectiveness of the two instructional methods. Through the application of statistical procedures, the relative effectiveness of the methods can be identified.

However, objections to the use of group designs have been based upon a variety of criticisms (Bergin & Strupp. 1976; Hersen & Barlow, 1976). They include:

1. Group designs may require that a substantial number of children not participate in a given treatment condition. As such. the withholding of treatment from the children in the control group raises ethical questions.
2. Group designs require that a sufficiently large sample of handicapped children be included in each group. Due to the extreme variability of the skill levels and behaviors found among handicapped children and due to the differing needs of each child, it becomes difficult to identify homogeneous groups of children for studies.
3. Group designs may demonstrate statistically significant differences between groups, but the educational significance of the intervention for a particular child may be minimal (Bergin, 1966; Chassan, 1967; Sidman, 1960). The characteristics of the individual child are lost in group designs. As such, it becomes difficult for the practitioner to determine the extent to which the child in the classroom is similar to those children who exhibited change during the group evaluation procedure.

In special education settings, instruction may be different for each child depending upon his unique needs and attributes. Single subject evaluation procedures emphasize the child's uniqueness and focus on differences found in one child over time rather than differences between groups of children. Thus, single subject evaluation procedures may be more applicable to the study of handicapped children than are group procedures.

THE IMPLEMENTATION OF SINGLE SUBJECT DESIGNS

The implementation of single subject designs requires that three steps be followed:

1. The behavior or skill to be changed must be adequately defined.
2. The behavior must be measured over time and a baseline of the behavior must be obtained prior to intervention.
3. A treatment or instructional intervention must be initiated and the impact upon the baseline behavior must be monitored.

DEFINITION OF THE BEHAVIOR

The first step in implementing a single subject procedure is the selection and definition of a target behavior which the teacher or parent wishes to change. The selection of the target behavior is usually determined by the needs of the child and the goals of the teacher and parent. In addition, the target behavior that is selected must be both observable and measurable (Reynolds, 1968). For example, a child's behavior could be observed by seeing it or hearing it. The behavior may be measured by counting the number of times it occurs or by timing the length of each occurrence.

After selecting the target behavior, the behavior is defined. The definition of the target behavior must include both an objective description of the observable behavior and the procedures used to measure or record that behavior. The precision and accuracy with which the target behavior is defined can be checked by having two people simultaneously and independently observe and record the occurrence of the target behavior. If two observers agree on the occurrence of the target behavior while concurrently and independently observing the same child, then the definition of the target behavior is adequate. If two observers do not agree, then the definition of the behavior is not sufficient and must be more precisely and accurately defined. Precision can be obtained by debriefing the observers, noting any instances of disagreement, and rewriting the definitions to eliminate the discrepancies.

For example, suppose a teacher wants to modify a child's self-abusive behavior. First, the teacher must define the child's self-abusive behavior precisely and accurately in terms that are observable. The child may repeatedly hit himself on the head with his hand. For this child, self-abusive behavior may be defined as the number of times the child is observed hitting himself on his head with his hand during a 30-minute interval. The precision and accuracy of the definition can then be checked by having two independent observers simultaneously record the occurrence of the child's self-abusive behavior. It is important to stress that the idiosyncratic nature of the definition makes it highly unlikely that the same definition would be totally appropriate for different children manifesting self-abusive behaviors.

MEASUREMENT OF THE TARGET BEHAVIOR

After selecting and defining the target behavior, the level at which the behavior naturally occurs prior to intervention must be measured. This measure will serve as a standard or "baseline" against which the effectiveness of the intervention procedure can be evaluated.

The primary procedures used by educators to measure the target behavior are ones in which a person observes the child and records the occurrence of the behavior. These procedures are called observational techniques. However, automatic recording devices, such as teaching machines, which mechanically record a child's answers, and the direct measurement of a permanent or tangible product may also be used. For example, a teacher may directly measure a permanent product by counting the number of blocks a child stacks or the number of puzzle pieces properly placed.

There are several observational techniques which can be used in the classroom. These include: (a) event recording, (b) duration recording, (c) interval recording, and (d) time sampling. The selection of the appropriate observational technique is usually governed by the specific goals of the educator conducting the evaluation, the time and personnel available for the study, and the behavior which is to be observed.

EVENT RECORDING

An observer using event recording counts the number of times the target behavior occurs during a specified period of time. For example, a teacher may count the number of times a child gets out of his seat during a 30-minute observation session. Using this method, the frequency (number of times the behavior occurs during the 30-minute observation) and rate (average number of times the behavior occurs per minute) of the target behavior can be determined. This procedure can be utilized when there is a clear beginning and end to the behavior and when the length of each observation session is known. If the length of each observation session is the same, then both frequency and rate of the behavior may be used for comparisons between observation sessions. If the lengths of the observation sessions are not equal but are known, only the rate of the behavior may be used for comparisons between observation sessions.

DURATION RECORDING

Frequently the occurrence or non-occurrence of the target behavior is secondary to how long it lasts. In these situations, duration recording is an appropriate recording technique to use. A stop watch or other timing device may be used to measure the duration of the target behavior. For example, a teacher or observer may record the duration of a temper tantrum each time the behavior occurs. Total duration, average duration, and the percent of the observation session during which the behavior occurred can be determined using this method. Like event recording, this procedure requires that the target behavior have a fixed beginning and end. If an educator is interested in reporting only the total duration of the target behavior per session, then the length of each observation session must be identical. However, if an evaluator is interested in reporting the average duration or the percent of the session occupied by the target behavior, the length of each observation session can vary.

INTERVAL RECORDING

Interval recording is obtained by dividing the observation session into several separate but equal time periods. For example, a 5-minute observation session may be divided into twenty 15-second intervals. If the target behavior occurs at any time during an interval, it is noted as having occurred during that interval. For example, if the child is out of his seat at any time during a 15-second interval, a check mark is made in the appropriate interval on the observer's recording sheet. However, if a child is out of his seat more than once during the 15-second interval, only one check mark is made. The number and percent of intervals during which the child engages in the target behavior are reported. Since the procedure requires an observer's undivided attention, it may necessitate that someone in addition to the teacher is available to do the recording. Typically, the length of observation sessions is determined by noting the length of the behavior and setting the time period so that is is highly unlikely that more than one occurrence of the target behavior can occur within a given time interval.

TIME SAMPLING

Time sampling is one of the easiest observational techniques to use in an educational setting because it does not require an observer's undivided attention. Thus, a teacher can easily use time sampling to measure the target behavior while at the same time teaching a class. As with interval recording, the observation sessions are divided into a number of equal intervals. However, unlike interval recording, the child is observed only at the end of the interval. Therefore, the observer records whether or not the behavior is occurring at a specific instant. For example, a 3-hour session may be divided into twelve 15-minute intervals. At the end of each 15-minute interval, the observer looks at the child and records whether or not the behavior is occurring. For example, a teacher might look at a child every 15 minutes for three hours and record whether or not she is crying. Time sampling provides the teacher with the number and percent of intervals sampled during which a child engages in the target behavior. Like interval recording, time sampling can be used when the target behavior does not have a clear beginning and end. In addition, although the intervals within each session must be of equal length, the length of the sessions does not need to be equal if the percent of occurrence is reported.

INTEROBSERVER AGREEMENT

As previously noted, it is important that the target behavior is defined so that two observers simultaneously agree on the occurrence of behavior when they independently observe the same child. Thus, it is important that the observers are properly trained in the method of recording and in the definition of the target behavior. Although observers may initially observe the behavior in the same manner at the beginning of the study, their definition may independently "drift" and they may define the behavior differently by the end of the study (Johnson & Bolstad, 1973; O'Leary & Kent, 1973). As such, perceived changes or improvement in the target behavior may in fact be due to a change in how the behavior is measured. Consequently, several authors recommend training a minimum of two people to observe and record the behavior and periodically computing interobserver agreement throughout the evaluation (Bijou, Peterson & Ault, 1968; Johnson & Bolstad, 1973; McNamara & MacDonough, 1972; O'Leary & Kent, 1973; Reid, 1970; Reid & Demaster, 1972; Romanczyk, Kent, Diament & O'Leary, 1973).

Interobserver agreement can be computed between pairs of observers at any time throughout the course of the study. The most common method of assessing interobserver agreement has been that of percent agreement (Johnson & Bolstad, 1973; Kelly, 1977; Yelton, Wildman & Erickson, 1977). Percent agreement is calculated using the following formula:

$$\frac{\text{No. of Agreements}}{\text{No. of Agreements + No. of Disagreements}} \times 100 = \text{Percent Agreement}$$

Although percent agreement is an acceptable method for computing interobserver agreement, it does have several limitations (Johnson & Bolstad, 1973). Cohen's Kappa (1960) avoids the problems of percent agreement and has therefore been recommended for analysis of interobserver agreement (Hartmann, 1977; Yelton, Wildman & Erickson, 1977). For most educational evaluators, however, percent agreement is usually adequate and far simpler to compute. Generally, acceptable levels of interobserver agreement are above 75 percent for classroom behaviors (O'Leary & Becker, 1967; Hall, Lund & Jackson, 1968; Barrish, Saunders & Wolf, 1969). However, the level of agreement between observers may differ depending upon the type of behavior observed and the method of calculating reliability.

LENGTH OF THE BASELINE

The purpose of a baseline is to provide a standard against which the effectiveness of an intervention can be evaluated. Therefore, it is important not only to adequately define and obtain accurate measures of the behavior, but also to ensure that baseline measures are taken over a sufficient period of time.

Barlow and Hersen (1973) recommend that baseline data be gathered for a minimum of three sessions prior to beginning an intervention. At least three data points are necessary to reveal the presence of a pattern in the baseline data. For example, with a minimum of three points, an evaluator can determine if there is an upward, downward or stable trend in the data. An upward trend in the data during the baseline condition generally means that the occurrence of the target behavior is increasing over time (ascending baseline) and a downward trend in the data means that the occurrence of the target behavior is decreasing (descending baseline). In addition, at least three data points are necessary to give an investigator an idea of the variability of the target behavior during the baseline condition.

McNamara and McDonough (1972), Wolf and Risley (1971), and Baer, Wolf and Risley (1968) agree that baseline measurements continue until stability in the baseline data is attained. In other words, they contend that baseline data should be gathered until there is little change in the occurrence of the behavior between sessions. Therefore, if change in the behavior occurs after the intervention, it can be more easily attributed to the intervention. Sidman (1960) recommended that in a laboratory, under controlled conditions, rates of behavior should be within a 5 percent range of the mean. However, as Hersen and Barlow (1976) state, this is often not practical when doing applied research. For example, an investigator may not have the time to continue gathering pretreatment data until stability is attained. It is also difficult to isolate and therefore eliminate all sources of variance in applied settings.

Although conclusions regarding the effectiveness of an intervention must be made carefully, an instructional or treatment procedure is often implemented without a stable baseline. Several techniques for dealing with a variable baseline and the use of inferential statistics are available (see Hersen & Barlow, 1976). Instructional and treatment procedures are also sometimes started when a baseline is either ascending or descending. For example, if the intent of the teacher is to increase the occurrence of a child's behavior and the baseline is descending, the instructional procedure could be initiated. Since a change in the direction of the slope of the line is expected (i.e., descending to ascending), the effect of the procedure can be evaluated. However, if the intent of the teacher is to increase the occurrence of the specified behavior and the baseline is ascending, the intervention should not be initiated. The effectiveness of the instructional procedure would be difficult to assess as the slope of the line may not be expected to change. It would be difficult to attribute improvement to the procedure unless there was a marked change in the slope of the line.

INTRODUCTION OF THE INTERVENTION

After data have been gathered during the baseline condition, the intervention is introduced. The intervention may be an instructional or treatment technique which the educator believes will modify the occurrence of the target behavior. Data collection continues when the intervention is initiated and throughout the remainder of the evaluation. It is important that the data collection procedures are identical during the baseline and intervention conditions. Data must be gathered at the same time of day (Russell & Bernal, 1977) or during the same setting or class throughout the evaluation. For example, if a child is observed at 10:00 a.m. during the baseline condition, the child must also be observed at 10:00 a.m. after the introduction of the intervention, unless additional data collected at other times indicate that time of day is not a relevant factor.

It is also important that factors other than those part of the instructional intervention remain the same throughout the evaluation procedure. For example, it may not be wise to change a child's medication, classroom, or teacher during the evaluation unless these variables are part of the intervention process. If these extraneous factors are controlled and similar data collection procedures are followed throughout the evaluation, a change or improvement in the behavior following the intervention may be more easily attributed to the intervention.

A frequent error of educators using single subject procedures is the introduction of more than one type of instruction at a time in an attempt to assess the interactive effect of the two instructional methods (Hersen & Barlow, 1976). If more than one type of instruction at a time is introduced, it is difficult to tell which method contributed to changes in the target behavior. Therefore, only one factor at a time should be changed throughout the <u>entire</u> evaluation and only after stability has been obtained during the previous <u>treatment</u> condition.

SINGLE SUBJECT PROCEDURES

There are essentially three types of single subject evaluation procedures--A-B, reversal, and multiple baseline--each with advantages and disadvantages. All share a common goal of demonstrating the relative impact of a given intervention strategy upon a target behavior. All use the subject as his own control and all are appropriate for use in clinical/educational settings.

A-B DESIGN

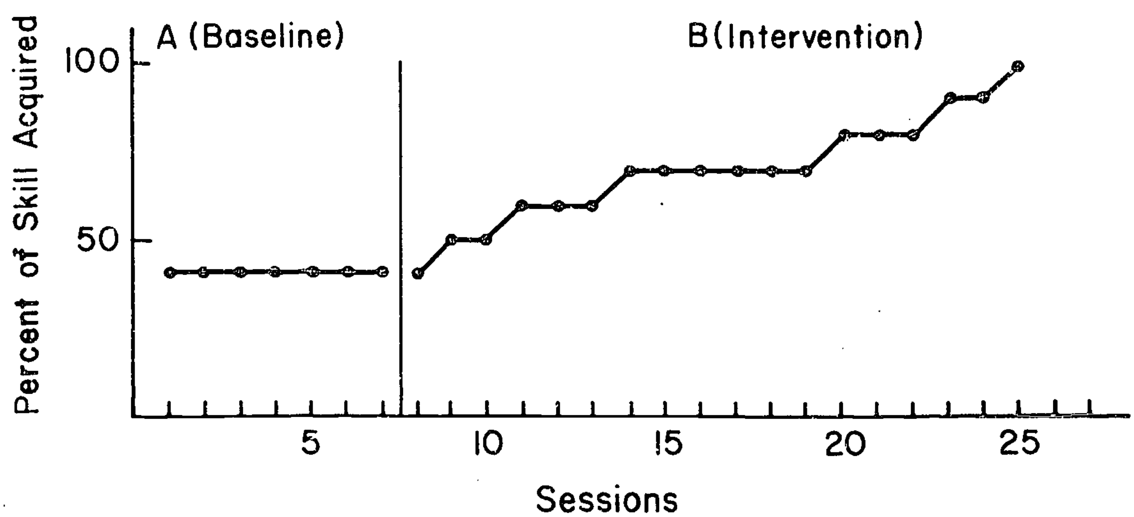
The simplest single subject procedure is the A-B design. With the A-B paradigm, the A represents a baseline condition and the B represents an intervention condition. The success of the A-B design is dependent upon obtaining an accurate baseline measure. Once the baseline measure is established, an intervention may be instituted. During the intervention, data are collected in the same manner as during the baseline condition. The impact of the intervention is determined by the extent to which the occurrence of the target behavior during intervention is different from the occurrence of the target behavior under baseline conditions. However, the A-B design does have certain limitations which must be considered (Risley & Wolf, 1972).

The major assumption of the A-B design is that the baseline measure is accurate and predictive of future behavior. It has been argued, however, that changes in the target behavior during the intervention phase may be due to factors external to the intervention (Wolf & Risley, 1971). For example, a child may receive instruction at home without the teacher's awareness, at the same time that an intervention is started at school. The child's learning might change to a greater degree than would have occurred with the classroom intervention alone. Therefore, when using an A-B design, one cannot assume that a change in the target behavior is always attributed to the intervention. Other factors (e.g., medication, illness) may be responsible for the observed change.

The extent to which external factors are controlled and precise baseline measures are established enhances the conclusions which may be drawn from the A-B procedure. The A-B procedure may be useful when instructional or treatment limitations prevent the withdrawal or intermittent application of the intervention. For example, an A-B procedure might be useful with a child who exhibits a high rate of self-destructive behavior. Due to the potential for physical harm to the child, it would not be appropriate to make unnecessary changes in the intervention once it had been initiated.

Figure 1 is an example of an A-B design used with a handicapped preschool child. As shown, the child's skill level remained stable during the baseline condition. Following the start of the instruction, changes occurred in the child's performance of the skill. A comparison of the child's performance during instruction with the child's performance during baseline reveals that the child acquired additional skills during instruction. Whether this increase is solely attributable to the instruction is not clear. It is conceivable that the skill increase might have occurred without the instruction. However, the fact that the behavior was stable during baseline may strengthen the teacher's confidence that the increase was due to the instruction.

FIGURE 1. Comparison of baseline and intervention conditions for a fine motor skill taught to a handicapped preschool child, utilizing the A-B design.



REVERSAL DESIGN

A more favorable single subject procedure is the A-B-A, or "reversal," design. In the A-B-A design, a baseline measure is obtained as in the A-B design. Following the baseline period (A), an intervention (B) is applied. The B phase is followed by a return to baseline or withdrawal (A) of the intervention. If the intervention has an impact, the target behavior might be expected to return to baseline levels during the reversal phase since the intervention which prompted the change in behavior is no longer in effect. Therefore, during the reversal condition, the extent to which the behavior "returns" or approximates the level of behavior during the baseline condition increases the teacher's certainty that the intervention was effective. If the behavior does not return to baseline levels during the reversal phase, several interpretations are possible: (1) the intervention was effective but the newly acquired behavior was not reversible, (2) the intervention was not effective and the observed change is attributable to some outside, uncontrolled variable(s), or (3) the intervention technique and other uncontrolled variables jointly contributed to the observed effect.

When the occurrence of the behavior during the intervention condition differs from the behavior during baseline and reversal conditions, the teacher's conclusions regarding the efficacy of the intervention are strengthened. As such, the inclusion of a reversal phase eliminates many of the objections raised to the A-B design. An observed change in the target behavior following the initiation of an intervention can be attributed to the intervention rather than to extraneous factors if the behavior returns to baseline levels during the reversal phase. However, the A-B-A procedure necessitates that the intervention be interrupted, an option which may not always be feasible or ethical.

Although there is no formula for determining the length of a given intervention condition, certain factors must be considered:

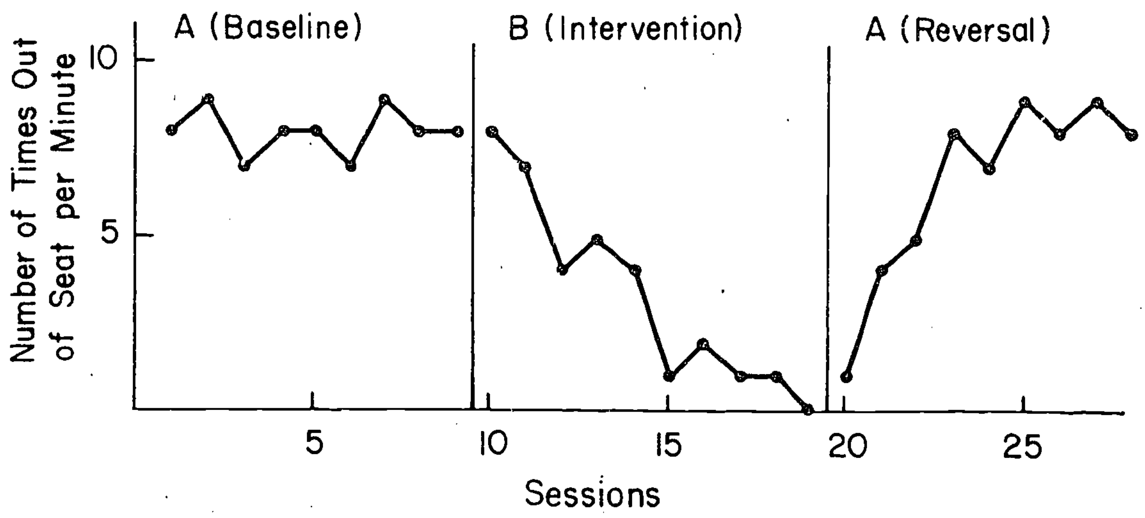
1. The intervention phase should continue until a degree of <u>stability</u> is obtained in the occurrence of the target behavior. As such, the data should show a steady improvement in the occurrence of the target behavior until the lack of a trend becomes apparent.
2. The intervention phase should be sufficiently long to assure that observed changes in the target behavior are not due to the <u>variability</u> with which the behavior naturally occurs. When there is a broad range of variability in the occurrence of the behavior, more data points are required to demonstrate the impact of the intervention. Therefore, the intervention should continue until the range of variability during the intervention phase does not significantly overlap the range of variability during baseline. This is particularly important during the later sessions of the intervention condition when the treatment effect should be most apparent. For example, a child got out of her seat an average of 18 times during a half-hour observation session. The range was from 14 to 24 occurrences during any given session. Following the application of a time-out procedure, the average decreased to 13 occurrences with a range of variability from 10 to 18 occurrences during a half-hour observation. Although the overall occurrence of the out-of-seat behavior had decreased, the range of overlap was too large (from 14 to 18 occurrences) to conclude that the time-out procedure was effective. As such, the intervention procedure continued until the range of overlap decreased further. When the range of overlap between baseline and intervention decreased to only two occurrences, the intervention procedure was terminated. Conclusions regarding the efficacy of the intervention can be made at this point since the overlap is minimal.

3. An equal number of data points should be included during each phase if possible (Barlow & Hersen, 1973). When baseline and intervention phases are of unequal lengths, observed changes may be due to extraneous factors associated with the unequal length of phases. For example, the child might receive additional attention during the longer phases, or greater effects due to maturation may occur during the longer phases. Thus, the length of the intervention condition is dependent upon the level at which the target behavior occurred during baseline, the variability of the behavior during baseline and intervention conditions, and the length of the baseline condition.

The A-B-A procedure may be an appropriate technique to use when documenting the impact of a behavior management intervention. However, the A-B-A procedure is not an appropriate technique for instructional interventions. Instructional interventions are directed toward increasing the child's ability to perform a given skill. As such, a return to baseline levels during a reversal condition may not be expected since increases in the child's ability to perform a skill may be relatively permanent.

Figure 2 shows an example of an A-B-A procedure used to document the impact of an intervention on the out-of-seat behavior of a handicapped preschool

FIGURE 2. An example of an A-B-A evaluation procedure used to document the impact of intervention on the out-of-seat behavior of a child.



child. As shown, there was a substantial decrease in the child's out-of-seat behavior during the intervention phase relative to the baseline phase. During the reversal condition, the child's out-of-seat behavior increased to the baseline level. Since the out-of-seat behavior decreased during the intervention condition and did not do so during the baseline and reversal conditions, one may infer that the observed decrease during the intervention phase was due to the intervention.

In most instances, an A-B-A-B procedure is more appropriate than the A-B-A procedure (Miller, 1973). With an A-B-A-B design, the intervention or treatment is reapplied after the reversal condition. The degree to which the two intervention conditions show a similar impact upon the target behavior, while differing from baseline and reversal conditions, increases one's confidence that the intervention was effective. With repeated intervention and reversal conditions showing similar results, evidence regarding the efficacy of the intervention is accrued. Thus, the greater the number of reversals, the greater one's assurance that the intervention is effective when the behavior repeatedly returns to baseline levels during the reversals.

The A-B-A-B procedure is regarded as more powerful than the A-B-A procedure, in that it allows for two (or more with repeated applications, A-B-A-B-A-B) opportunities to demonstrate the effectiveness of the intervention (Barlow & Hersen, 1973). An additional advantage of the A-B-A-B procedure is that the final condition is an intervention condition which may be continued following the evaluation. As such, the disadvantages associated with the withdrawal of the intervention are minimized.

The A-B-A-B procedure may be used with instructional interventions while the A-B-A procedure may not. Since skills which have been learned would not be expected to return to baseline during a reversal condition, the A-B-A procedure would offer little information regarding the impact of instruction. However, when the reversal condition is followed by a reapplication of the instructional intervention (as in the A-B-A-B design), comparison can be made between the child's skill gain during the baseline and reversal conditions with that during the intervention conditions. The degree to which the child demonstrates skill gains during instruction while the baseline and reversal phases remain stable provides evidence that the instruction contributed to the change.

As shown in Figure 3, the child's skill level increased during the instruction condition relative to the reversal condition. That the child's skill level increased <u>only</u> during the instruction condition suggests that the instruction was responsible <u>for</u> the observed change. The inclusion of the second instruction condition eliminates the objections which are found with the A-B-A design, i.e., that the intervention and reversal condition did not impact upon the target behavior and that any observed changes would have occurred with or without the intervention and reversal conditions. The degree to which the target behavior changes in a similar manner during the intervention condition while remaining stable during the baseline and reversal conditions increases one's certainty that the intervention had an impact.

Figure 4 shows an example of an A-B-A-B procedure used to decrease the occurrence of an inappropriate behavior. The degree to which the target behavior during the second intervention condition approximates the first intervention condition increases one's confidence that the intervention was effective. Thus, there are essentially two advantages to the A-B-A-B procedure over the A-B-A procedure for behavior management interventions: (1) the final condition is an intervention condition and (2) the second intervention serves as a replication of the first intervention which may provide additional information regarding its effectiveness.

An alternative to the A-B-A and A-B-A-B procedures is the B-A-B procedure. In the B-A-B procedure, the first condition is an intervention condition (B), followed by a non-intervention condition (A). The intervention condition (B) is then reapplied. The B-A-B procedure may be desirable when it is not possible to obtain

FIGURE 3. An A-B-A-B design used to document the impact of instruction upon the performance of a child's skill.

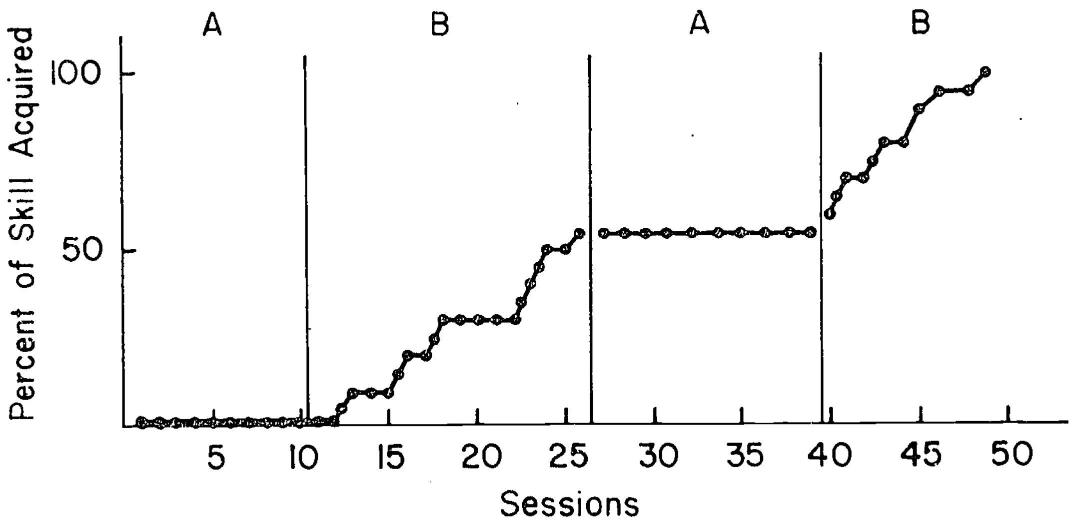
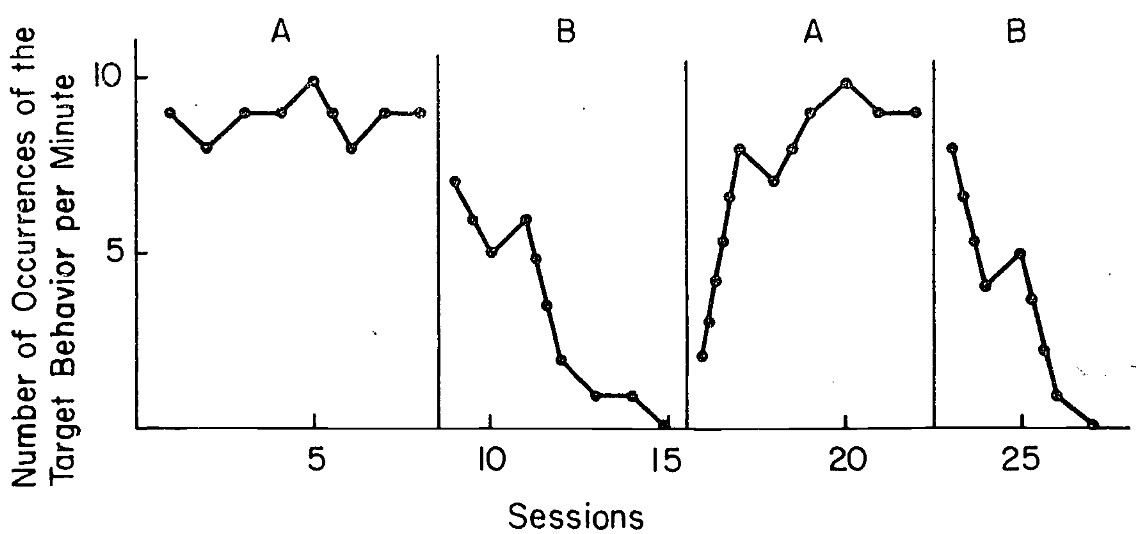


FIGURE 4. An example of an A-B-A-B evaluation procedure used to document the impact of a behavior management intervention.

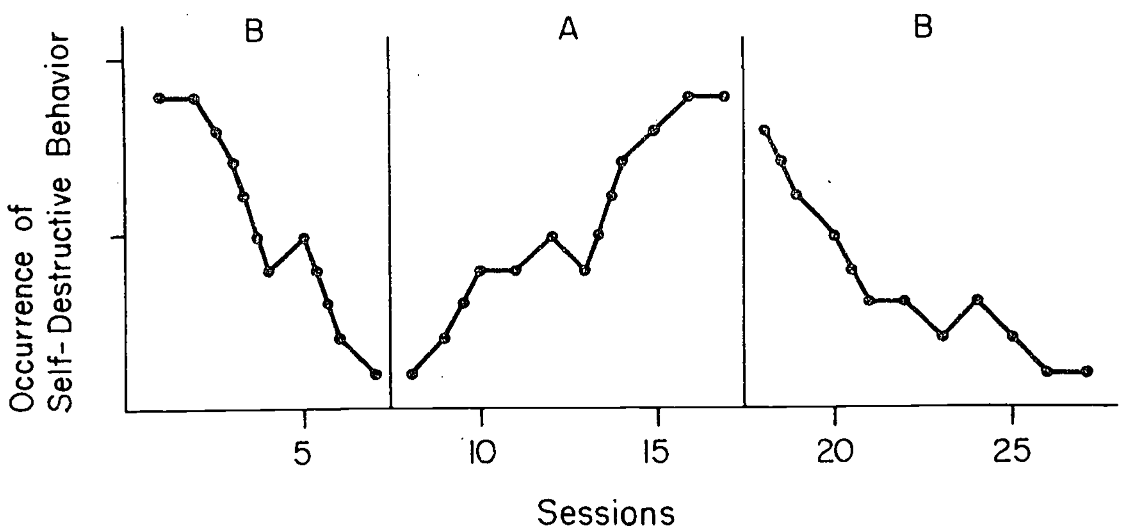


a baseline measure prior to the application of an intervention. For example, a child may exhibit a self-destructive behavior which is very damaging to his vision. In such an instance, it would not be appropriate to delay intervention for the purposes of collecting baseline data. The use of a B-A-B procedure would be appropriate.

In the absence of a baseline condition, however, it may not be possible to accurately determine the degree to which the intervention changed the occurrence of the target behavior from the level at which it would normally occur. An A-B-A-B procedure is more desirable when baseline measures may be obtained (Barlow & Hersen, 1973).

Figure 5 is an example of the B-A-B procedure used with a child exhibiting self-destructive behavior. As shown, the occurrence of the self-destructive behavior showed an overall decrease during the first intervention condition. This trend reversed during the non-intervention condition, returning to the level at which the target behavior occurred when the intervention was first initiated. During the second intervention condition, the occurrence of the target behavior again showed an overall decrease in a manner similar to that in the first intervention. The extent to which the occurrence of the target behavior reversed during the non-intervention condition relative to the intervention condition increases one's certainty that the intervention was effective.

FIGURE 5. An example of a B-A-B evaluation procedure used to document the impact of intervention on a child's self-destructive behavior.



MULTIPLE BASELINE PROCEDURE

A third, and somewhat more complex, single subject evaluation procedure is the multiple baseline technique (Baer, Wolf, & Risley, 1968). The multiple baseline technique is useful when the target behavior is potentially irreversible or when it is inappropriate to institute a reversal condition. Implementation of the multiple baseline technique involves a series of eight steps:

1. Identifying two or more target behaviors which are not expected to be related. Skills or behaviors which are related or similar, such that an intervention applied to one results in changes in the other, are said to be of the same response class. Behaviors or skills included in the multiple baseline technique must be independent of one another (i.e., different response classes). For example, if a parent designs a method to increase a child's verbal response to questions, one might also expect that the child's total verbal behavior will increase. Consequently, both behaviors should not be included in the same multiple baseline study.
2. Defining the target behaviors in a manner such that two or more observers may agree on the occurrence of the behaviors.
3. Obtaining data to support interobserver agreement on the occurrence of the target behaviors.
4. Initiating concurrent baseline collection procedures for each of the identified target behaviors.
5. Continuing the baseline measures until the absence of trends can be documented.
6. Implementing an intervention for one of the target behaviors while concurrent baseline measures continue for the remaining target behaviors (noting any change in the occurrence of the target behavior during intervention and any concurrent changes in the target behaviors still under baseline conditions).
7. Implementing an intervention for a second target behavior while concurrent baseline measures continue for any remaining untreated target behaviors. Changes in the occurrence of the second target behavior and concurrent changes in any of the other target behaviors should be noted.
8. Continuing to implement intervention in the same manner, one target behavior at a time, until all target behaviors are under intervention.

The strength of the multiple baseline technique is based upon two attributes: (1) each of the individual target behaviors follows an A-B procedure which allows for a comparison between the baseline and intervention conditions, and (2) control measures for the target behaviors under intervention are found in the concurrent baseline measures. Thus, the degree to which the occurrence of the individual target behaviors change during intervention relative to baseline suggests the intervention to be effective. The extent to which such changes occurred in the target behaviors under instruction, while the target behaviors under concurrent baseline conditions remained stable, further demonstrates the impact of the intervention.

The objections to the A-B procedure, that changes in the occurrence of the target behavior might have occurred with or without the intervention, are substantially decreased with the multiple baseline technique. If such changes would normally have occurred, one would expect similar changes in the concurrent baseline measures of the other target behaviors. Thus, when a change occurs for a behavior during intervention while the concurrent baselines remain stable, evidence is accrued that the intervention is contributing to the effect. This is the rationale which mandates that the selected target behaviors be independent of each other (i.e., different response classes). When target behaviors of similar response classes are utilized, an intervention applied to one of the behaviors might inadvertently impact upon the other. As such, the behaviors would co-vary and the control characteristics of the target behavior under baseline would be lost.

There are essentially three types of multiple baseline techniques (Hall, Cristler, Cranston, & Tucker, 1970), each of which follows the same basic procedures outlined above. The three types of multiple baseline techniques include:

1. Across Behaviors. This procedure involves the selection of two or more unrelated behaviors for each child. Intervention is applied to each in turn, with concurrent baseline conditions serving as control conditions for the intervention conditions. This procedure serves to evaluate the relative impact of intervention upon each of the target behaviors for a specific child.
2. Across Settings. The same target behavior is measured concurrently across two or more settings (e.g., home vs. school) for the same child. Intervention is applied in succession for each of the settings. Concurrent baseline measures across settings serve as control conditions for the intervention conditions. This procedure serves to evaluate the role of the environment in contributing to the occurrence of the target behavior.
3. Across Subjects. The same target behavior is measured concurrently across two or more children. One child receives intervention while concurrent baseline data are collected for the remaining children. Intervention is then applied in succession to the remaining children while concurrent baseline data are collected. This procedure controls for the influence of subject variables as they relate to the intervention.

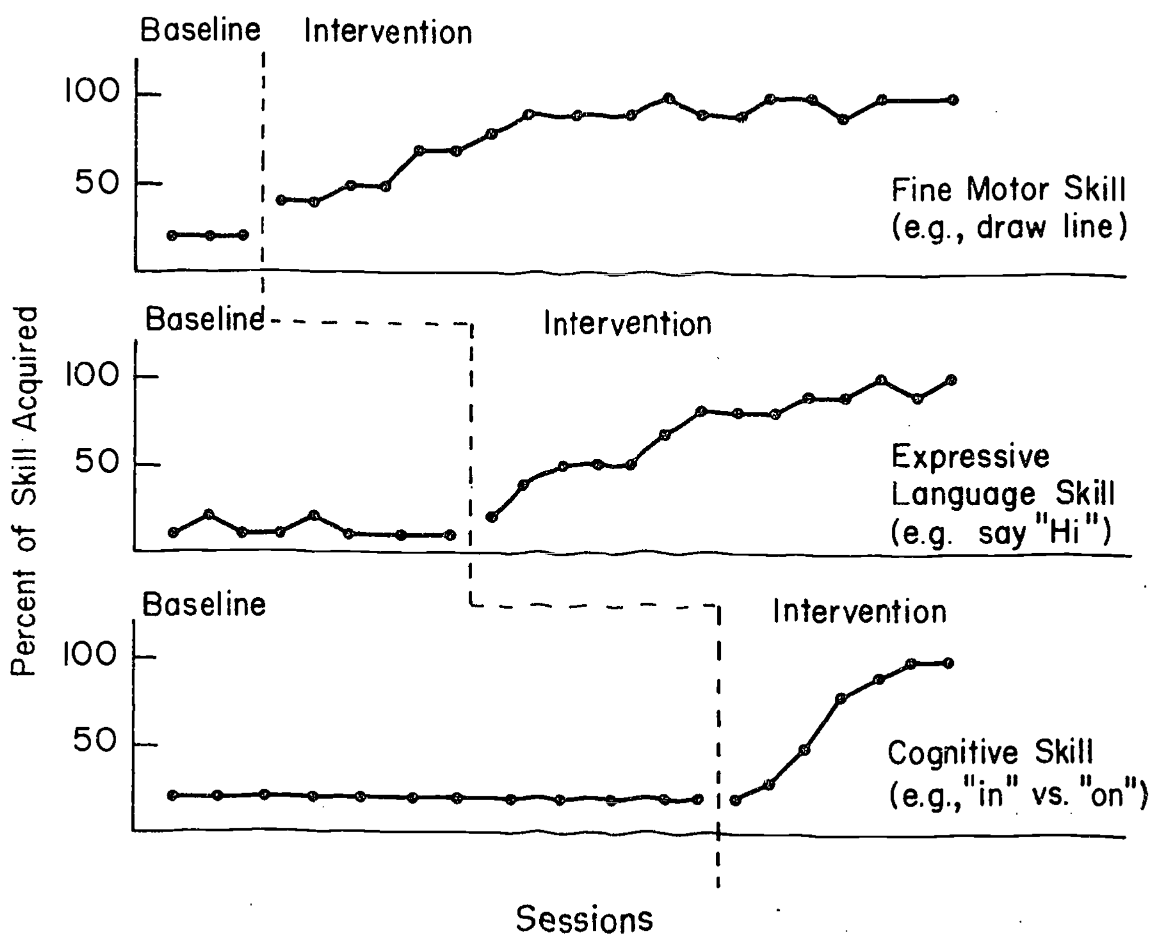
Multiple baseline techniques may be of particular value in educational settings in that a reversal condition is not necessary. Since instructional interventions are directed toward increasing the child's ability to perform a given skill in a relatively permanent manner, a return to baseline levels during a non-intervention condition may not be expected. Further, the multiple baseline technique may not be subject to the objections found in reversal evaluation procedures (that it may be inappropriate or unethical to withdraw intervention for a given target behavior). However, the multiple baseline technique does not demonstrate behavioral control as directly as reversal procedures since the impact of intervention is inferred from independent behaviors (Hersen & Barlow, 1976). Additional evidence may be obtained by including an A-B-A-B reversal procedure for one or more of the target behaviors in the multiple baseline technique (Kazdin & Kopel, 1975).

Figure 6 shows a simplified graphic representation of the multiple baseline design used to evaluate the impact of intervention with one child and three target behaviors (i.e., an across behaviors design). As shown, concurrent baseline measures were obtained for each of the three target behaviors. Intervention was applied to one of the target behaviors and a subsequent increase in the child's skill performance was noted. Concurrent baseline measures were also obtained for the remaining target behaviors while noting changes in their occurrence. Intervention was then applied to the second target behavior and subsequent increases in the child's performance of the skill were noted. A concurrent baseline was obtained for the third target behavior and change in its occurrence was noted.

Examination of Figure 6 shows that changes in the occurrence of the target behaviors are consistently greater during the intervention conditions than during the baseline conditions. The degree to which the occurrence of the behaviors changed during intervention relative to the respective baseline measure for each behavior (as in the A-B procedure) suggests the instruction to have been effective. The extent to which the occurrence of the target behaviors increased while the concurrent baseline measures remained comparatively stable further increases one's confidence that the intervention contributed to the observed increase.

The primary means of interpreting the results of the single subject design is through visual analysis of graphically presented data. Graphic presentation of the data allows the impact of an intervention technique to be easily demonstrated by illustrating that change in the target behavior occurred simultaneously with the initiation of the intervention. Although a variety of statistical procedures are available to aid in the interpretation of the data (see Gottman & Glass, 1978; Levin, Marasciulo & Hubert, 1978), visual appraisal of the data is most generally utilized. With visual analysis, the change in the behavior across sessions or conditions must be large enough to be visible to the eye. As Sidman (1960) noted, when using visual analysis to interpret data, interventions must produce "large effects" in order to be considered successful. Whereas a statistical comparison of behavior

FIGURE 6. Example of the multiple baseline technique used to document the impact of three instructional programs.



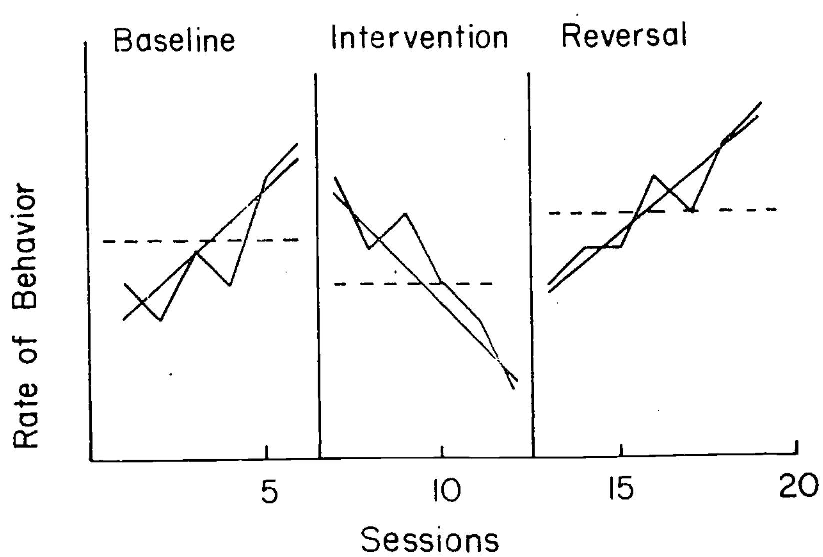
under baseline and trea ment conditions may indicate differences, a visual analysis of the data may t. Visual analysis is therefore the more conservative approach (Glass, Wilso Gottman, 1975; Parsonson & Baer, 1978). As Jones, Weinrott and Vaught (19 demonstrated, visual analysis results in fewer errors of incorrectly attributing c ange in behavior to the intervention. In addition, the clinical relevance of change in behavior across conditions which cannot be detected visually is questionable.

Data are usually graphed throughout the course of a single subject procedure. Thus, changes can be made in the intervention as the study proceeds, providing a degree of adaptability necessary in applied research. A standard format is used to graph behavioral data. The vertical axis (ordinate) is used to indicate the occurrence of the target behavior, while chronological time (the instructional session or the day) is plotted along the horizontal axis (abscissa) (see Figure 7). The occurrence of the target behavior may be expressed in the form of frequency, rate, duration, or percentage. The form of presentation depends upon the measurement method used to collect the data and the goals of the teacher.

Line graphs and bar graphs are frequently used in data presentation. In a line graph, discrete points are plotted and then joined with a line, as in Figure 7. Line graphs are appropriate when the time dimension is continuous and when the distances between points on the abscissa are equal. However, when the time between sessions is unknown, a bar graph is more appropriate (Parsonson & Baer, 1978). An example of a bar graph is in Figure 8. Additional information regarding the graphing of data from a single subject design is available in Cooper (1974), Koorland and Martin (1975), Meyers (1970), and Spear (1969).

Three properties of the data aid the instructor in an evaluation of the effectiveness of the in ention. These are: (a) trend in the target behavior;

FIGURE 7. An example of a line graph.

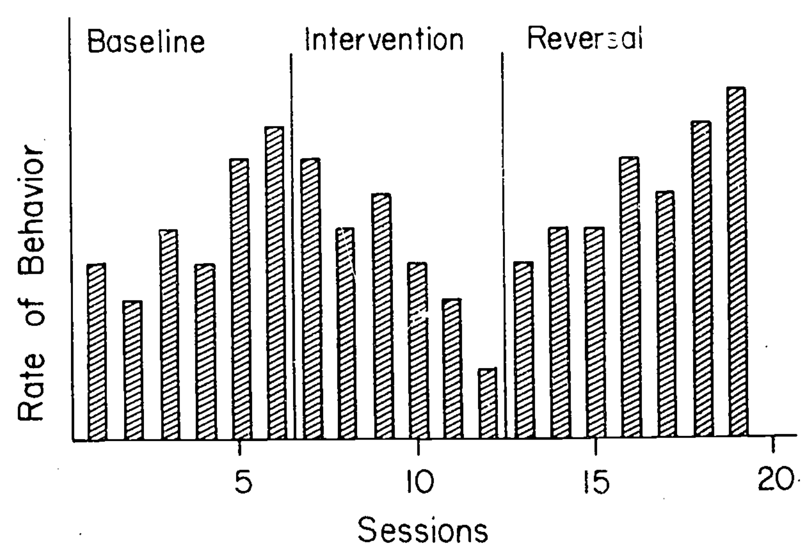


(b) the level of the behavior; and (c) the variability of the behavior. Visual analysis of the data includes determination of changes in trend, level and variability of the target behavior between conditions or phases.

TREND. A trend line, as represented by the solid line in Figure 7, represents the general trend of the data. Trend lines may be either drawn free hand or by using statistical methods (see Parsonson & Baer, 1978). If done carefully, the free-hand method is adequate for use by teachers in an evaluation of instructional techniques. A comparison of the slope of the trend lines during different conditions of the evaluation procedure yields information regarding the efficacy of the intervention. Both the amount and direction of the changes in trend provide information to the teacher. For example, during the baseline condition in Figure 7 there is an upward trend in the data. The slope of the trend line during the treatment condition changes dramatically, indicating that the intervention was effective. Changes of trend in the data within an experimental condition also provide information to the teacher regarding the efficacy of the intervention (Parsonson & Baer, 1978).

LEVEL. Changes in the level of the occurrence of the target behavior between conditions are also a good indication that the intervention is effective. All of the data points within a condition are averaged to provide an estimation of the level of behavior during a particular condition. The level of the target behavior is indicated by the dashed line in Figure 7. As can be seen from the data, there is a small but visible difference in the level of the target behavior between the baseline condition and the treatment condition and a larger difference between the treatment condition and the reversal condition. These differences, therefore, indicate that the intervention is effective.

FIGURE 8. An example of a bar graph.



VARIABILITY. A study of the variability between data points within and between conditions is also made by the teacher in analyzing the data. As previously mentioned, a stable baseline is preferred. Stable baselines usually provide a better basis for attributing change in the target behavior to the intervention. However, variable baseline data as compared with stable data during the treatment condition may be indicative that the treatment is effective. The elimination of variability may, in fact, be the goal of the treatment. Alternatively, an effective treatment may override any variability during the baseline condition. The occurrence of variability during the baseline condition may also provide the teacher with important information about extraneous factors affecting the occurrence of the target behavior. The identification and control of these factors could have important treatment ramifications.

SUMMARY

It has been argued that the single subject design can provide the means necessary to determine the impact of the treatment and instructional techniques used with handicapped children. One of the primary advantages of the single subject procedure is that it is relatively simple and analysis of the data does not require statistical sophistication. Thus, a teacher wishing to assess the efficacy of an instructional program needs only to follow a few steps. First, the behavior which the teacher wants to change must be selected and defined. The precision and accuracy of the definition can then be checked through analyzing the agreement between two people regarding the occurrence of the target behavior while simultaneously and independently observing the same child. Next, a baseline of the behavior is obtained. After the baseline is stable, the teacher can initiate the new instructional program. The behavior is measured continuously throughout the entire baseline and intervention process.

The teacher may select one of several designs to aid in the interpretation of the efficacy of the instructional method. The A-B-A design may not be used often since it may not always be appropriate or ethical to institute a reversal condition without reinstating the intervention. As such, the A-B-A-B and B-A-B designs may be more useful since they call for the intervention to be reinstated during the final phase. Since it is seldom appropriate to reverse classroom behaviors, as in a reversal design, most teachers will find multiple baseline designs to be the most appropriate. The B-A-B design may be chosen when it is not possible to obtain a baseline of the behavior prior to the intervention. This may occur when the evaluation procedure is started after the instructional program has begun or when it is necessary to begin an intervention immediately. Conversely, the multiple baseline design requires that baselines are obtained on several behaviors prior to initiating programs which may change these behaviors. This latter design is particularly applicable when a teacher wishes to change more than one behavior. The choice of the design is thus determined by the behavior to be modified, the instructional techniques to be evaluated, and the goals of the teacher.

Not only is the methodology of the single subject procedure simple, but it is also well suited to P.L. 94-142. The law requires that each child have an individualized education plan, based on the child's needs, that is continuously monitored and updated. The emphasis of the single subject design is on the individual. In addition, ongoing collection, graphing and interpretation of the data provide the teacher with continuous information regarding the child's progress so that the instructional plan can be changed when necessary.

REFERENCES

Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91-97.

Barlow, D. H., & Hersen, M. Single-case experimental designs: Uses in applied clinical research. Archives of General Psychiatry, 1973, 29, 319-325.

Barrish, H. H., Saunders, M., & Wolf, M. M. Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. Journal of Applied Behavior Analysis, 1969, 2, 119-124.

Bergin, A. E. Some implications of psychotherapy research for therapeutic practice. Journal of Abnormal Psychology, 1966, 71, 235-246.

Bergin, A. E., & Strupp, H. H. New directions in psychotherapy research. Journal of Abnormal Psychology, 1970, 76, 13-26.

Bijou, S. W., Peterson, R. F., & Ault, M. H. A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. Journal of Applied Behavior Analysis, 1968, 1, 175-191.

Chassan, J. B. Research design in clinical psychology and psychiatry. New York: Appleton-Century-Crofts, 1967.

Cohen, J. A coefficient of agreement for minimal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Cooper, J. O. Measurement and analysis of behavioral techniques. Columbus, Ohio: Charles E. Merrill, 1974.

Glass, G. V., Wilson, V. L., & Gottman, J. M. Design and analysis of time series experiments. Boulder: University of Colorado Press, 1975.

Gottman, J. M., & Glass, G. V. Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), Single-subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Hall, R. V., Cristler, C., Cranston, S. S., & Tucker, B. Teachers and parents as researchers using multiple baseline designs. Journal of Applied Behavior Analysis, 1970, 3, 247-255.

Hall, R. V., Lund, D., & Jackson, D. Effects of teacher attention on study behavior. Journal of Applied Behavior Analysis, 1968, 1, 1-12.

Hartmann, D. P. Consideration in the choice of interobserver reliability estimates. Journal of Applied Behavior Analysis, 1977, 10, 103-116.

Hersen, J., & Barlow, D. H. Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press, 1976.

Johnson, S. M., & Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Marsh (Eds.), Behavior change: Methodology, concepts and practice. Champaign, Illinois: Research Press, 1973.

Jones, R. R., Weinrott, M., & Vaught, R. S. Visual vs. statistical inference in operant research. In A. E. Kazdin (Chair), The use of statistics in N=1 research. Symposium presented at the Annual Convention of the American Psychological Association, Chicago, September, 1975.

Kazdin, A. E., & Kopel, S. A. On resolving ambiguities of the multiple-baseline design: Problems and recommendations. Behavior Therapy, 1975, 6, 601-608.

Kelly, M. B. A review of the observational data-collection and reliability procedures reported in the Journal of Applied Behavior Analysis. Journal of Applied Behavior Analysis, 1977, 10, 97-101.

Koorland, M. A., & Martin, M. B. Elementary principles and procedures of the standard behavior chart (3rd ed.). Gainsville, Florida: Learning Environments Inc., 1975.

Leittenberg, H. The use of single case methodology in psychotherapy research. Journal of Abnormal Psychology, 1973, 82, 87-101.

Levin, J. R., Marascuilo, L. A., & Hubert, L. J. Nonparametric randomization tests. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Lovitt, T. C. Characteristics of ABA, general recommendations, and methodological limitations. Journal of Learning Disabilities, 1975, 7, 432-443.

McNamara, J. R., & MacDonough, T. S. Some methodological considerations in the design and implementation of behavior therapy research. Behavior Therapy, 1972, 3, 361-378.

Meyers, C. H. Handbook of basic graphs: A modern approach. Belmont, California: Dickenson, 1970.

Miller, P. M. An experimental analysis of retention control training in the treatment of nocturnal enuresis in two institutionalized adolescents. Behavior Therapy, 1973, 4, 288-294.

O'Leary, K. D., & Becker, W. C. Behavior modification of an adjustment class: A token reinforcement program. Exceptional Children, 1967, 33, 637-642.

O'Leary, K. D., & Kent, R. N. Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, P. O. Davidson, & L. E. Acker (Eds.), Critical issues in research and practice. Champaign, Illinois: Research Press, 1973.

Parsonson, B. S., & Baer, D. M. The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change. New York: Academic Press, 1978.

Reid, J. B. Reliability assessment of observational data: A possible methodological problem. Child Development, 1970, 41, 1143-1150.

Reid, J. B., & Demaster, B. The efficacy of the spot-check procedure in maintaining the reliability of data collected by observers in quasi-natural settings: Two pilot studies. Oregon Research Bulletin, 1972, 12(8).

Reynolds, G. S. A primer of operant conditioning. Palo Alto, California: Scott, Foresman & Co., 1968.

Risley, T. R., & Wolf, M. M. Strategies for analyzing behavioral change over time. In J. Nesselroade & H. Reese (Eds.), Life-span developmental psychology: Methodological issues. New York: Academic Press, 1972.

Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.

Russell, M. B., & Bernal, M. E. Temporal and climatic variables in naturalistic observation. Journal of Applied Behavior Analysis, 1977, 10, 399-405.

Sidman M., Tactics of scientific research. New York: Basic Books. 1960.

Spear, M. E. Practical charting techniques. New York: McGraw-Hill, 1969.

Wolf, M. M., & Risley, T. R. Reinforcement: Applied research. In R. Glaser (Ed.), The nature of reinforcement. New York: Academic Press, 1971.

Yelton, A. R., Wildman, B. G., & Erickson, M. T. A probability-based formula for calculating interobserver agreement. Journal of Applied Behavior Analysis, 1977, 8, 1-12.