Methods of Research in Sport Sciences

Quantitative and Qualitative Approaches

Gershon Tenenbaum
Marcy P. Driscoll

Meyer & Meyer Sport

(24) Single Subject Research: Roots, Rational and Methodology

Eitan Eldar

> "Science is, of course, more than a set of attitudes. It is a search for order, for uniformities, for lawful relations among the events in nature. It begins, as we all begin, by observing single episodes, but it quickly passes on to the general rule, to scientific law. If we could not find some uniformity in the world, our conduct would remain haphazard and ineffective" (Skinner, 1969, p. 13).

Single subject research (SSR) is also referred to as single case, intensive, within-subject, repeated measures, and time series experimental designs. The roots of SSR methodology go back to the early work of B. F. Skinner. His initial behavioral research (1930-1937) led him to distinguish between respondent behavior and operant behavior. In a series of experiments, he made repeated observations and manipulations of a well-defined behavior of a single subject (rats and pigeons), in a controlled context. The concluding chapter of his first book "The Behavior of Organism" (Skinner, 1938), discusses the differences between group and SSR, calling for a further refinement of SSR, as the basic methodology for studying human behavior. "Tactics of Scientific Research – Evaluating Experimental Data in Psychology" (Sidman, 1960), is a highly-cited book, providing comprehensive and rich guidelines of how to do and how to judge research in experimental psychology. In his work, Sidman offers a direct and systematic replication for achieving reliability and generality of research findings.

The development of Applied Behavior Analysis (ABA) is associated with the seminal article "Current Dimensions of Applied Behavior Analysis" (Baer, Wolf, & Risley, 1968), published in the first issue of The Journal of Applied Behavior Analysis (JABA). This article sets the ground rules for SSR in applied settings. Another significant contribution is "Strategies and Tactics of Behavioral Research" (Johnston & Pennypacker, 1980, 1993). This book, which derives from the classic work of Sidman, presents a complete and coherent rationale and description of behavioral research methods. The 1993 edition is much more communicative than the first one, allowing students and researchers to relate the SSR methodology in a clear and sound theoretical foundation. Numerous experimental, technical, theoretical and philosophical SSR articles can be found in The Journal of Applied Behavior Analysis (JABA), The Journal of Behavioral Education, The Behavior Analyst, The Journal of Experimental Analysis of Behavior (JEAB), and others.

The methodology presented in this chapter has been implemented in various educational (mostly special education) settings. It is suggested here as an exceedingly appropriate paradigm for studying behavior in sport contexts. The hypothetical examples provided in this chapter are intended to assist in generalizing the methodology into the sport and exercise sciences.

This chapter describes the basic features of SSR as the foundation for its experimental reasoning. It compares SSR to group research, and evaluates its potential contribution to the study of behavior in sport contexts. The major judgment tool in SSR – VISUAL ANALYSIS of graphic data is described and explained, followed by a detailed description of the BASELINE LOGIC, guiding the construction of all single subject designs. Once the logic has been formulated, five major designs (WITHDRAWAL, MULTIPLE BASELINE, CHANGING CRITERION, MULTITREATMENT, and ALTERNATING TREATMENT) and their variations are described. Each design is discussed in terms of its implementation, the demonstration of experimental control, and advantages and limitations. The chapter concludes with a discussion of SSR potential and adaptability in sport contexts.

24.1 *Basic Features of Single Subject Research*

The SSR methodology corresponds to the clinical / empirical orientation of applied behaviour analysis. It is focusing on socially important behaviors, and is committed to their improvement. The basic features characterizing SSR are closely related to the following comprehensive "working" definition of behavior.

> The behavior of an organism is that portion of an organism's interaction with its environment that is characterized by detectable displacement in space through time of some part of the organism and that results in a measurable change in at least one aspect of the environment. (Johnston & Pennypacker, 1993, p. 23).

Some basic features, derived from the essence of applied behavior analysis, led to the development of the SSR methodology. These are:

1. **Behavior is emitted by individuals**. Groups are comprised of individuals. Each individual has a specific interaction with the environment. Therefore, it is the individual's behavior that is studied, and not an average behavior of a group to which this individual belongs.

2. **Behavior takes time, and therefore requires repeated measurement**. Every behavioral event has a measurable duration. The frequency of the behavior within a specific duration is also amenable to measurement. Because behavior changes over time, it is essential to measure it systematically and reliably. Measuring behavior over time is vital to SSR, because experimental control is presented with the same participant, and not by a comparison to other (control) individuals. Therefore, data are collected prior to implementation of the independent variable, and throughout the study. Maintaining a reliable data collection should guarantee stability of data, and control of variables, extraneous to the context of study.

3. **Behavior is determined by the environment**. The "deterministic law" applies to all behaviors. It is subjected to logic and to order, like any other scientific phenomenon. Therefore, behavior analysts look for "functional relations" between the behavior and environmental events. This empirical orientation contributed to the development of SSR, focusing on causality rather than description or correlation.

4. **Behavior is systematic.** It is possible to identify stability in data of repeated behaviors, which is fundamental to all SSR. Variability in data is caused by interaction with the environment and can be explained. If variability is detected, an experimental manipulation is conducted, to control for it.

5. **Behavior is flexible - calling for a creative investigation.** The basic core of SSR is quite simple. The designs are built as a combination of "building blocks", tailored to answer a specific research question. The "baseline logic" serves all designs, while creative judgment is required to adjust it to the specific research context.

6. **Individuals respond to similar stimuli in a similar way.** Therefore, the reliability and the external validity (generality) of research findings in SSR is established if replication of previous procedures, produce similar outcomes (Johnston & Pennypacker, 1993; Sidman, 1960).

24.2 *Single Subject Design versus Group Design*

The essence and the implementation of SSR and "group research" (GR) differ substantially. SSR apply when the main concern is with individuals rather than with groups. GR provides information representing many individuals. Such data is not accurate for any specific individual in that group. "A prediction of what the average individual will do is often with little or no value in dealing with a particular individual. The actuarial tables of life insurance companies are of no value to a physician in predicting the death or survival of a particular patient" (Skinner, 1953, p. 19). Similarly, data showing a substantial motivation improvement of 20 players in a soccer team may not indicate that the most valuable player for that team (e.g., the scorer) has also been improved, following the implementation of a specific motivation program. In fact it may be that this player's motivation has actually regressed due to this specific intervention.

In GR participants are randomly assigned to groups. The analysis of the data is based on the assumption that measures of central tendency for each group represent the larger group (i.e., population) from which it was drawn. Analyzing groups' averages may "mask" important information regarding specific participants. Such an analysis, based on combined scores, does not allow drawing conclusions about an individual member of the group. This problem may be crucial for the coach, who's tactical plans are concentrated in a specific player. Therefore, empirical strategy and tactic in sport should be based on individual data.

Both group and SSR are methods of investigating the effect of the manipulation of the independent variable on the dependent variables. However, there are some conspicuous differences between the two. These are:

1. The source for EXPERIMENTAL CONTROL in SSR lies in comparing research conditions of the same participant. GR compares the differences among groups of participants.

2. Internal validity in SSR is determined by the contribution of the intervention to the target behavior of the participant. The significance of change depends on the unique characteristics of each individual (social validity). In GR the criteria for experimental validity are much more rigid. The level of significance is based on fixed statistical measures, determined ahead of time.
3. Data analysis in SSR is based on visual analysis of graphic data. In GR, it is based on statistical inference.
4. The ongoing measurement of behavior in SSR, followed by a continuous graphic presentation, enables the researchers to detect fluctuations that are sensitive to specific timing or context. A post-test measure in GR can not detect such process characteristics.
5. The graphic process analysis in SSR may also yield serendipitous findings or phenomena. Although unrelated to the research question, such information may contribute significantly to the area of study.
6. External validity in SSR is determined by replicating the intervention in similar context and with similar participants. In SR, the sampling methods and the population they represent confirm the external validity.

The above comparison has emphasized the advantages of SSR in the study of behavior and sport context. However, the criteria for selecting the research methodology should not necessarily be philosophical or paradigm dependent. They should serve the research question in the most appropriate way. If we are interested in a certain population (e.g., the whole team), then GR should be used to study the effect of the intervention (e.g., the effect of motivation training on "team spirit").

24.3 *Visual Analysis of Graphic Data*

The visual analysis of graphic data is the major "judgement tool" of behavior analysts. It is possible to detect a behavior change of an individual or a small group without the employment of statistical tools. Most independent variables studied in ABA lead to a clear and abrupt change in the dependent variables. Such a change is amenable to visual inspection (Baer, 1977).

The following advantages of graphic display of behavioral data are described in various publications (Baer, 1977; Cooper, Heron, & Heward, 1987; Michael, 1974; Parsonson & Baer, 1978; Sidman, 1960; Skinner, 1956). Visual analysis is a conservative approach, requiring a notable and stable behavior change, in order to be accepted as an important one. However, the analysis is relatively simple and may be administered with most computer programs available at homes, clubs, universities, and schools. The presentation is compact - displaying a detailed summary of all conditions in the research, time spent in each one, the design employed, the independent and dependent variables, the relationships between them, and follow-up details. The familiar presentation mode allows participants to understand and follow the process. It may also serve as an intervention, due to the reactive effect of posted graphs (e.g., provide the athletes with a weekly graph of their performance). The direct, immediate and continuos contact with the graphic data enables an ongoing formative evaluation of the process (e.g., if data show a decrease in motivation when the assistant coach is leading the practice, a personal adaptation may be implemented).

GRAPHIC PRESENTAIONS are visual illustrations of verbal statements and numeric data. They show at a glance the whole process, allowing a holistic evaluation of the intervention. In SSR a series of measures collected at each phase represent the dependent variables. The more data points collected at each phase, the more representative (i.e., valid) these data are of the target behavior. The following graphic tendencies serve as important judgment tools for the visual analysis of graphic data:

1. **LEVEL.** The value of the data path on the vertical axis. The mean level line (has an equal amount of data points above and under) represent the magnitude of the data in each phase of the research.
2. **TREND.** The direction of the data path across time. Trend may increase, decrease, or show a zero slope (horizontal). Trend value depends upon the nature of the target behavior. (e.g., an increase in the athlete's anxiety may be positive or negative).
3. **STABILITY / VARIABLITY.** The range in data points in the same phase. The closer data points are distributed around the mean data path, the greater the stability. An opposite situation means a high variability.

The analysis of graphic data is done within and between conditions, while only one variable is changed at each condition. Each condition should include a minimum of three separate and preferably, consecutive data points. The more variable the data are, the longer the condition should last. A within-condition analysis is concerned with the stability of data. The between-condition analysis compares the level and trend of data. A convincing demonstration of experimental control should show large and abrupt change in level (e.g., from low to high), and a remarkable change in trend direction (e.g., from decreasing or horizontal, to increasing). An overlap in data paths of two adjacent conditions, and high variability, weaken the experimental control and call for a further analysis.

Baseline Logic for Internal Validity

The first phase in implementing Single Subject Methodology is collecting "BASELINE DATA". These data provide information about the target behavior under "natural conditions", with no planned intervention related to the ongoing research or clinical treatment. It is the repeated measurement of behavior in units of frequency, duration, latency, or magnitude. A baseline is not necessarily a situation of no treatment or learning – it is the absence of the independent variable that defines it as such. Baseline data are the result of objective observations. They are not based on subjective judgments of the clinician or the researcher.

BASELINE DATA serve two functions – descriptive and predictive.

1. **Description of the target behavior**. Graphic and anecdotal descriptions of behavior provide objective knowledge about the participant's performance, the antecedents for the target behavior and the consequences of maintaining it. Such information is essential for programming the intervention.
2. **Prediction of future performance**. Stable baseline data enable the researcher to predict future probabilities of the target behavior, if the intervention had not been applied. Such PREDICTION enables the clinician to plan an appropriate success criterion for the subject and provides the researcher with a measure against which to compare the intervention data.

Baseline Logic – Defined

Single subject research methodology is based on BASELINE LOGIC (Sidman, 1960). It requires the repeated measurement of behavior, under at least two conditions: baseline (A) and intervention (B). A measurable change in the behavior following the intervention, when compared to baseline data, may lead to the conclusion that it is the intervention that caused the change.

BASELINE LOGIC is not a research design in itself – it is a set of guidelines for planning and decision-making in order to achieve functional relationships (also referred to as causality / experimental control) between the independent variable and the dependent variable. Measuring behavior under two conditions (A & B) does not demonstrate functional relationships (the behavior caused the change). Various strategies have been developed to verify such an assumption, all based on extension and adaptation of the A-B paradigm.

Baseline Logic – Components

BASELINE LOGIC provides basic methodological principles demonstrating that the change in behavior is related to the independent variables and not to confounding / extraneous variables. Such variables are events or situations that are not controlled by the researcher, but may have an effect on the target behavior. It consists of four components that should be present in every design demonstrating causality: PREDICTION, INTERVENTION, VERIFICATION, and REPLICATION. Another component – FOLLOW UP, is included in most designs, in order to ensure generalization and maintenance of the behavior change. This chapter presents baseline logic as it appears in the withdrawal design. A similar logic applies to all other designs. Baseline logic components are presented in Figure 4.35.

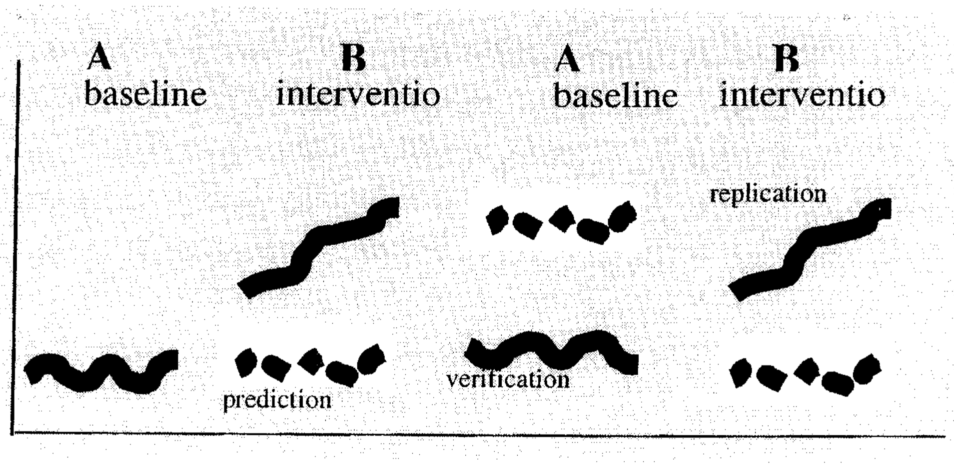

Figure 4.35 The Four Components of Baseline Logic – Baseline, Intervention, Verification and Replication, as they Appear in a Hypothetical ABAB Design. (The solid lines are actual baselines and interventions. The doted lines are predictions, if intervention had not been applied / removed).

PREDICTION is the assumption that baseline data will maintain level, trend and stability, if intervention is not applied. This is an extrapolation of a situation in which baseline condition would have remained unchanged. Intervention data are measured and compared against this extrapolation (Johnston & Pennypacker, 1993; Kazdin, 1982).

Effective Baseline Data

The following baseline characteristics are recommended for ensuring effective PREDICTION:

1. **Baseline length**. Baseline should include enough data points representing all aspects of the target behavior and not missing special events. On the other hand, it should not be too extended because it may withhold an effective intervention. Three data points are the minimal requirement for PREDICTION (Hersen & Barlow, 1976), but more information is usually required. If the target behavior is severe or risky, an immediate intervention should be implemented. A design with no initial baseline (e.g., B-A-B) should be employed in such a case.
2. **Baseline stability**. Baseline phase should be maintained until clear data STABILITY has been demonstrated (Baer, Wolf, & Risley, 1968). If data are variable the behavior should be further defined, and data collection reliability should be examined. Research in applied context leads to greater data VARIABILITY because some variables are difficult to control. Such variability should be treated as a natural part of individual functioning (Sidman, 1960). A baseline is considered stable if it does not include data points exceeding 50% of the average value of all points (Barlow & Hersen, 1984; Repp, 1983; Sidman, 1960). This is an arbitrary criterion that should be judged according to the research context.
3. **Baseline trend**. Data TREND is comprised of three or more data points, and can show increases, decreases, or zero trends. It is recommended to terminate a baseline with a deteriorating trend, or one that shows no change. Such data should support the need for intervention.

Intervention

The second phase in all single subject research designs includes measuring the behavior under intervention. The level and trend of data provide information about the effectiveness of the treatment. It is assumed that without INTERVENTION, baseline data would have remained unchanged (PREDICTION). It is also assumed that the intervention will have a positive effect on the target behavior.

Following a stable baseline, only one independent variable is administered. Intervention should be the least restrictive one for the participant, yet still effective. There are three possible outcomes for intervention:

1. **No change**. There are three methodological options for the researcher: a) continue with intervention, assuming a delay in the behavior change; b) modify the intervention; c) change the intervention.
2. **Improvement**. There are three options: a) maintain the successful intervention. This option (A-B) does not demonstrate causality, but it satisfies clinical judgment; b) apply intervention to other behaviors of the same subject (MULTIPLE BASELINE DESIGN); and c) withdraw the effective intervention or administer a placebo (WITHDRAWAL DESIGN). This is a powerful experimental tactic that is rarely implemented in educational settings because of the need to withdraw intervention.
3. **Deterioration.** In this case, intervention should be terminated, and an alternative one should be considered.

Verification

VERIFICATION provides assurance of the preceding PREDICTION, suggesting that without intervention, behavior would have remained unchanged. Verification contradicts the assumption that other variables, extraneous to the research design, are responsible for the change. One way of achieving verification is by withdrawing the intervention. If data reverse towards the previous baseline, it may be concluded that without intervention, no change would have occurred. Verification is based on the prediction that intervention effect would have been remained the same, if intervention had not been removed.

It should be noted that intervention withdrawal does not necessarily lead to a complete reversal to baseline. Some aspects of the behavior could already be acquired, causing a slow reversal.

Verification is strengthened if extraneous variables are controlled, and the reliability of data collection is improved. Verifying the prediction allows for claiming causality, but replicating the intervention effect increases the internal validity of the independent variable.

Replication within the Research

REPLICATION is a reintroduction of the independent variable. It reduces the probability that other extraneous variables caused the change in the dependent variables. The more an intervention be replicated, the more its believability will be strengthened (Baer, Wolf, & Risley, 1968). This principle is commensurate with single subject research methodology that is flexible and allows for adaptations.

It is recommended to extend the last intervention to ensure that data are stable and that the behavior has been acquired. Follow-up data, gathered after intervention is terminated, should provide information about the long-term effect of the independent variable.

Replication for Reliability and External Validity

Replication across studies is the repetition of a given experiment by the same experimenter, while keeping the independent variable and the participants constant (Sidman, 1960). This version of replications is called "DIRECT" because it repeats all aspects of the research. A series of successful direct replications, producing the same effect, strengthen the reliability of the results.

In SYSTEMATIC REPLICATION some aspects of the study are slightly and systematically varied (Cooper, Heron, & Heward, 1987). A successful systematic replication with different participants, with a slight change in the intervention or setting, should increase the reliability and the generality (external validity) of the results.

Direct replication is impossible in sport settings, because they are dynamic in nature. However, the systematic replication of an intervention may be an integral part of a training program. An example is having a sport psychologist "working" with an athlete on improving self-control. Once positive results are achieved, and experimental control (internal validity) is demonstrated, another study is initiated with a second athlete, then a third one, etc. The replication of effect with all participants, strengthen the generality of the results.

24.4 *A-B Designs*

A, B, and A-B designs do not demonstrate experimental control, but may contribute significantly to the professional's work. They present reliable data, enabling coaches and athletes to support their decisions with objective judgment.

"**A**" and "**B**" designs are also referred to as "case study" due to their descriptive nature. They describe the behavior prior to intervention (A), and during or after intervention has been employed (B). Unlike other case studies, based on narrative descriptions of a certain context or event, these designs present quantitative data. The graphic presentation of the data draws the level and trend of the target behavior, in a specific time frame and in a well-defined context.

A-B, also termed "simple time series design," is fundamental to all single subject designs. It presents behavior data in baseline phase and in the following intervention (first two phases in Figure 4.24). A-B is most suitable for evaluating educational and training goals that do not require experimental control. A-B is preferable to "A" because it measures the independent variable after presenting the behavior in its natural condition - with no intervention. It can also indicate that changes are not due to time (maturation), if an abrupt change, different than the stable baseline data path, is demonstrated with the presentation of intervention. This design is simple and therefore recommended highly for the practitioner who strives for success. It may demonstrate effective treatment, but cannot control for extraneous variables.

24.5 *Withdrawal Designs*

WITHDRAWAL (reversal) designs entail the programmed repeated introduction and withdrawal of the independent variable. Data comparison before, during, and after intervention enables the verification and replication of experimental control (as demonstrated in figure 1). The basic withdrawal design – ABAB, will be described here at length, followed by a short description of some withdrawal variations.

ABAB – Implementation

The following procedures are required for the implementation of ABAB:

1. A clear definition of the target behavior, and a description of other behaviors that could "react" to the planned intervention.
2. Collections of continuous baseline data, until clear and stable level and trend have been achieved.
3. Presentation of intervention following a deteriorating or "no change" baseline.
4. Continue intervention until the criterion for the target behavior has been achieved, or until a clear therapeutic trend has been demonstrated. Intervention phase length should be similar to baseline length, if possible.
5. Withdraw intervention when one of the two conditions (in #4) is met. Baseline 2 is similar to baseline 1, because there is no intervention, and data are collected in the same context.
6. Re-introduce the intervention (replication) after data in baseline 2 show no improvement or a deteriorating stable trend.
7. Continue intervention until therapeutic data are stable. It is recommended to collect follow-up data to ensure maintenance of the positive behavior change.

Experimental Control

WITHDRAWAL DESIGNS attain experimental control when data path shows improvement whenever intervention is introduced, and shows deterioration and return to baseline upon its removal. Every withdrawal and replication, followed by deterioration or improvement, strengthens the internal validity of the independent variable. This is a powerful strategy demonstrating causality with the same subject and with the same target behavior. BASELINE LOGIC components of withdrawal designs are described in the previous section (baseline logic).

An abrupt change detected immediately with the presentation of the intervention increases its internal validity. However, an abrupt change is not essential for claiming causality. In some cases, a gradual change is observed, because behavior has been acquired partially, or due to a carry-over effect of the intervention. In such cases, the stability of data plays a more significant role in determining experimental control.

Advantages

The major advantage of ABAB is the powerful experimental control it demonstrates. This is a simple design that is easy to explain and discuss. It is clear to the eye of the scientist, the practitioner, and the consumer.

The final phase of ABAB involves presentation of the independent variable. Ending a study with an effective treatment is warranted, enabling the maintenance of improvement. It also provides the researcher with extension possibilities for further study and refinement of the independent variable, or for comparison with other treatments.

Limitations

The limitations of ABAB are related to clinical and ethical, rather than experimental considerations. There is no sense in withdrawing an effective intervention prior to the acquisition of the target behavior. Such a conflict between clinical and experimental judgments should be considered when planning a withdrawal design.

Another limitation is related to acquired / irreversible behaviors (e.g., learning the correct kick in soccer). They do not lend themselves to withdrawal, and no reversal in data should be expected. An exception would be evidence with motivation problems interfering with the improvement of acquired behaviors. In such cases, the data trend should deteriorate upon withdrawal of the motivation program.

Variations of Withdrawal Designs

WITHDRAWAL DESIGNS are flexible to meet the needs of the participants and the context of the study. All variations adhere to the BASELINE LOGIC, presenting and withdrawing intervention.

ABA is the minimal withdrawal design demonstrating experimental control (the first three phases in Figure 16.35). This is an extension of an AB design, withdrawing intervention after stable data have been achieved. ABA is not frequently used in educational settings because it is terminated with a contratherapeutic change. It does not show a continuation of improvement as demonstrated in ABAB.

BAB design presents intervention at the first phase (no initial baseline), followed by withdrawal, and reintroduction of intervention. This design is applied when the target behavior is dangerous or unethical (e.g., a player may injure teammates during the practice), or when collecting baseline data is technically inappropriate (e.g., the training program has already started). It lacks the preliminary information about the target behavior in its "natural" context. However, it provides an immediate intervention, and is terminated with a continuation of treatment, enabling maintenance of the behavior change.

A "probe" phase in which an intervention is removed for a short period (e.g., one or two data points) may be applied when there is an abrupt deterioration in the target behavior. An immediate change in trend and level following reintroduction of intervention provides VERIFICATION of its effect. Probe data are taken when an extended withdrawal of intervention is either unethical or inappropriate (e.g., may jeopardize an already acquired skill).

Despite its limitations, the WITHDRAWAL DESIGN may be very effective in sport settings. There are many cases in which the withdrawal is an inevitable part of the process. For example, (a) The head coach is not present in some practices due to other duties (e.g., scouting); (b) in some games, certain players do not participate; (c) the athlete may try a certain diet or pre-competition arrangement. Such procedures, whether pre-planned or involuntary, may be included in a withdrawal design, enabling the evaluation of their contribution to the team / athlete.

24.6 *Multiple Baseline Designs*

The most frequently used design in educational settings is the multiple baseline (MB). This is a series of AB designs, in which BASELINE DATA are collected simultaneously, and intervention is introduced sequentially to each one of the target behaviors, participants or settings.

The MB design was formulated by Sidman (1960) and first introduced by Baer, Wolf and Risley (1968) as an alternative to reversal / withdrawal designs. It enables the measurement of a few dependent variables at the same time, when one is under experimental condition and the others are controlled. Figure 4.36 describes a hypothetical MB design across three different behaviors. It shows the BASELINE LOGIC components as they apply to this design.

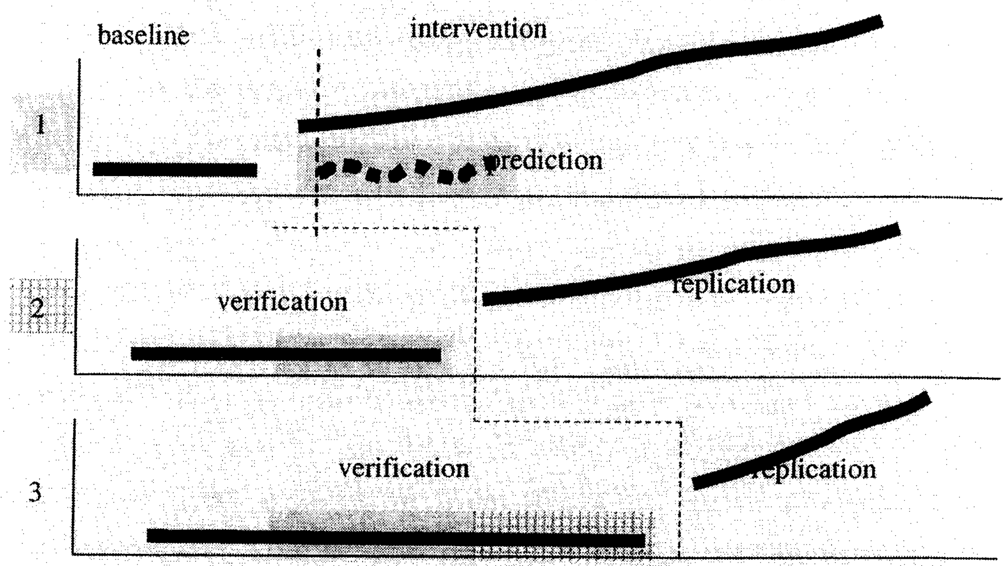

Figure 4.36 A Hypothetical Multiple Baseline Design. (The shaded area in tiers 2 & 3 serve as verification for the prediction for tier 1. The grid/shaded lines in tier 3 verify the prediction for tier 2).

Basic Forms

The basic arrangement of control and experimental conditions may be applied in three different forms:

1. **Multiple baselines across behaviors**. The effect of the independent variable is studied against different behaviors of the same participant or group. For example, the effect of a concentration training (independent variable) on the accuracy of passes, foul shots and assists in basketball (dependent variables).
2. **Multiple baselines across participants**. The effect of the independent variable is studied against different participants or groups, performing the same target behavior, in similar conditions. For example, the effect of a motivation program (independent variable) on the frequency of successful tackles (dependent variable) of three different soccer players, at the last 10 minutes of league games.
3. **Multiple baselines across settings**. The effect of the independent variable is studied in different settings in which the same behavior of the same participant/s is performed. For example, employing a self-control intervention for a "problematic" athlete in practices, competitions, and team gatherings.

Implementation

Each one of the three MB forms include two basic measures:

1. A simultaneous and continuous measure of baseline for all dependent variables.
2. A sequential presentation of intervention to all dependent variables.

The first step includes the collection of baseline data for all dependent variables, and setting criteria for a successful intervention. Once the baselines have been established, an intervention is applied to one variable, but not to the others, remaining under baseline condition. Then, rather than withdrawing intervention (as in withdrawal designs), it is applied to the next baseline, and so on. The following guidelines should be kept when implementing MB.

- The dependent variables should be FUNCTIONALLY INDEPENDENT. This means that there is no interaction among variables, and introducing the treatment to the first variable will not cause changes in the other unexposed variables. For example, changing the athlete's verbal inappropriate behavior may eliminate confrontations with referees. Therefore these are functionally dependent variables that are not amenable to the same MB design.
- The dependent variables should be FUNCTIONALLY SIMILAR so that they will respond to the same intervention. For example, cooperating with teammates and with the coach is functionally similar in that they may both be responsive to "social training". On the other hand, foul shooting and cooperation are not similar enough to respond to the same intervention (whether it is social or concentration training).
- The scale of measurement should be the same with all variables (e.g., percent of accuracy), to enable data analysis.
- Intervention should be introduced to stable baselines.
- Intervention should be introduced to the next variable, when the previous one has reached the target criterion, or when data trend is improving for three consecutive data points.
- Intervention condition is continued where applied, and baseline data are taken for all variables that were not exposed yet to treatment.
- It is recommended to vary the length of the unexposed baselines to control for extraneous variables such as "cyclic effect".
- Treatment is eventually applied to all variables.

Experimental Control

Experimental control is demonstrated when "where intervention is applied, change occurs; where it is not, change does not occur" (Horner & Baer, 1978, p. 189). It is the intention of the experimenter to show that each dependent variable is changing, in succession, only when the independent variable is introduced. Other unexposed baselines remain unchanged or vary just slightly (Baer et al., 1968).

Prediction. When baseline data are stable for the first variable, predictions are made that the level and trend will remain the same if all contextual variables will remain constant. A similar prediction is made with all other stable baselines.

Intervention. At this stage, intervention is implemented for baseline one, while all other baselines remain constant. If it is followed by a therapeutic change, it may be assumed that intervention is associated with the change in the dependent variable (as in AB design).

Verification. Continued baseline data of the other variables showing no change in stability, level, and trend, serve as verification for the prediction made for the previous baseline exposed to intervention. These baselines are still exposed to the same conditions under which the changed baseline had been measured. The fact that they have not changed, controls for time dependent extraneous variables such as specific / seasonal period of time, history and maturation. At the same time, only the dependent variables exposed to the intervention were changed.

Replication. The successive introduction of the intervention to unexposed baselines serves as replication of effect, if they are followed by a therapeutic change. No absolute rule can be given about the number of replications needed. Each additional one strengthens the internal validity and further controls for the "threat" of other extraneous variables. As with all other single subject designs, the decision about the number of replications is flexible and is left to the clinician and the researcher's judgment (Baer et. al., 1968). It has been demonstrated that a two-tier MB is sufficient enough to demonstrate causality (Cooper, Heron, & Heward, 1987). However, three or for tiers are frequently used in most MB designs.

Figure 4.36 shows the four components of baseline logic mentioned above, as they apply to the MB design.

Advantages

The MB design is empirically and practically oriented. It has the advantage of not withdrawing an effective treatment, thus supporting the maintenance of the behavior change. The prolonged data collection for several behaviors, participants, or settings, enables an empirical evaluation of the athlete or the program goals. Furthermore, the MB is consisted of a series of AB designs, providing ongoing information about the dependent variables before and after intervention.

The MB design enables the evaluation of generalization of behavior change. For example, the measurement of the same target behavior (e.g., concentration) in three different settings (individual practice, scrimmage, and competition) shows the degree of generalization attained.

Limitations

The MB design presents a weaker internal validity than in the withdrawal design. The reason lies in the VERIFICATION procedure. Verification in the MB design is achieved through other variables (participants, behaviors, or settings), remaining unchanged while the data for the variable under intervention show change. In withdrawal designs, verification is achieved within the same participant – if behavior data deteriorate when intervention is removed.

The MB design requires a delay in the delivery of intervention for the second and subsequent independent variables. Such a delay may not be appropriate if an immediate intervention is essential.

Other limitations are technical, relating to time required for implementing the MB design, and the interference that the prolonged baselines may cause to the participants and their surroundings.

24.7 *Variations to MB Designs*

Delayed MB. This variation has been constructed to overcome the limitation of prolonged baselines (Heward, 1978; Risley & Hart, 1968; Watson & Workman, 1981). While the first baseline data are being collected, other baselines are not yet measured. It is important to add the next baseline prior to implementation of intervention to the first baseline. This baseline overlap (a minimum of three data points) enables the

VERIFICATION element. If there is no such overlap, no causality may be claimed. Other baselines are added in the same delayed fashion.

The delay in measurement eliminates the need for extended baseline that may result in reactive effect (i.e., the observation procedure changes the target variables). This design may save resources in cases where they are limited. The delayed MB is flexible, enabling the addition of new variables, not available or relevant while the initial one was studied. If the first intervention is successful, other baselines, not presented initially, may be added, to strengthen the internal validity. Thus, a delayed MB has both advantages (i.e., flexibility and sensitivity to clinical procedures), and drawbacks (internal validity).

The limitations of this variation lay in the incomplete baselines. It may not be sufficient to collect only a few data points in each subsequent baseline, before the presentation of the intervention. Without an initial and concurrent measure of all variables, it may be difficult to provide an acceptable control for confounding variables such as history. The addition of delayed measurements of new variables to an already functioning MB should maintain a powerful internal validity, yet enable the required flexibility.

Multiple Probe Design. In this variation (Cuvo, 1979; Horner & Baer, 1978), data are not collected continuously on variables that have not yet been exposed to the independent variable. Probe measurements, interspersed throughout the unexposed baselines, are presumed to represent the level and trend of these variables. While intervention is presented to the first variable, probe data are intermittently collected for subsequent variables. It is advised to probe all variables when the first intervention reaches the desired criterion. Such a simultaneous measure increases internal validity. At least three consecutive data points are collected in the second tier, prior to the introduction of the second intervention. During the second intervention phase, intervention probes are conducted for the first variable, and baseline probes are taken for the third variable.

Both the probe and the delayed designs are suitable for coaches who wish to measure their athletes' progress, but lack the time and resources required for extended measurements. They may also reduce the intrusive effect of data collection during practices. Due to their functional similarities, both designs suffer similar limitations (see above).

Changing Criterion Design. The CHANGING CRITERION DESIGN (CCD) is ideal for sports contexts. It enables the athlete and the coach to evaluate gradual and systematic changes in performance, and therefore, to motivate progress. The CCD (Hartmann & Hall, 1976; Sidman, 1960) involves a successive change of the planned criterion for success (i.e., reinforcement contingencies), in graduated steps, from baseline to the desired terminal goal. Each phase of the design provides a baseline for the following one. Improvement in performance is presumed to occur concurrently with changes in the planned and stated task contingencies. Figure 4.37 describes the basic format of the CCD.

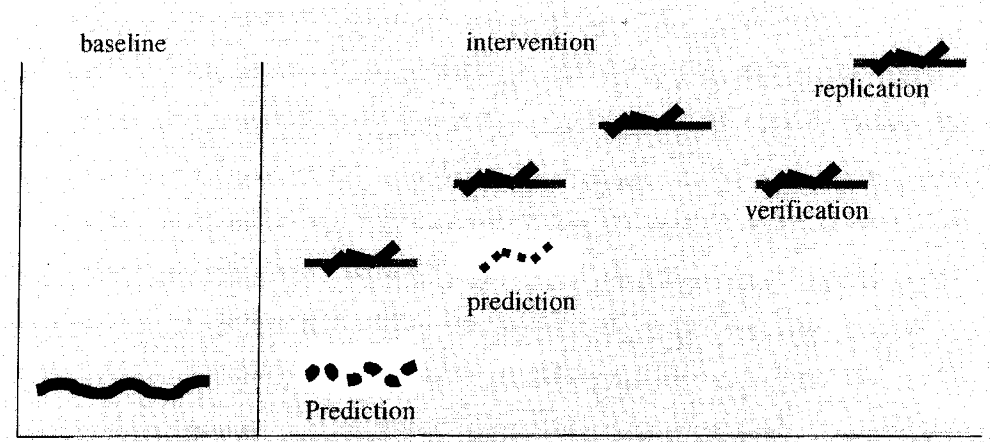

Figure 4.37 A Hypothetical Changing Criterion Design. (The dotted lines are predictions of data if intervention had not been applied).

Implementation

Implementing the CCD is comparable to operating a training program. Due to the applicable nature of this design, the following steps should be carefully executed, to maintain acceptable empirical outcomes:

1. A specific target behavior should be selected and behaviorally defined.
2. Baseline data should be taken to set a clear initial criterion for intervention, and to determine a desired criterion for terminating the intervention. Baseline data are taken here in the same manner used in all other single subject designs. Criterion changes must be large enough so they are visible, but not too large to frustrate the athlete and slow down the training program.

3. An intervention program is designed to meet the final criterion. Each intervention phase is defined clearly in terms of criterion, and the time in which it should be attained. These phases are the steps in the ladder the athlete should climb to reach the top rung. The criterion for each step is based on professional estimation of the athlete's ability, and knowledge of the target behavior. It is recommended that each successive performance step be based on the same difference as the previous one. For example, if baseline data show that only 10% of the athlete's moves are accurate, the criterion for the first phase may be 15% accuracy, for the second 20%, and so on.
4. After a stable baseline has been achieved, the first phase of intervention is presented, and so on. A minimum of four replications is recommended.
5. Each performance, which meets the set criterion for success, is followed by some sort of reinforcement to the athlete (e.g., a positive feedback in front of the whole team).
6. A few consecutive and successful data points are required to proceed from a given phase to a higher one. The minimal requirement is two consecutive successful measures, or two successful data points out of three attempts (e.g., the basketball player has two games in a row in which he/she had a frequency of 4 assists per 10 minutes).
7. Criterion for progress should be adjusted if reaching the criterion is too easy or too difficult.
8. The procedure is terminated when the target criterion has been achieved and maintained.

Experimental Control

The CCD is based on the match between the pre-established success criterion for each phase, and the actual value of the dependent variable at this phase. Experimental control is achieved when the level and the trend of the target behavior, vary concurrently with the manipulated change in the criterion and the contingent feedback it yields.

The change in the new phase should be immediate and should follow a stable level and trend in the preceding phase. The reason is that each phase serves as a baseline for the following one.

A short reversion to the previous criterion level may strengthen the experimental control. This can be done, by setting a lower requirement for reinforcement, after a higher one has already been achieved. A regression in performance in this case will verify that change in the performance was due to the intervention and not to other variables or to natural improvement. Varying the length of each phase should also control for extraneous variables and serve as VERIFICATION.

Advantages

The CCD is a positive practical procedure in that it supports a gradual improvement of performance to achieve experimental control. The CCD is tailored for complex behaviors and for extended acquisition periods. The sub-phases of this design are parallel to the phases in shaping (Cooper, Heron, & Heward, 1987; Skinner, 1953), which arrange successive approximations to the behavior, by repeatedly reinforcing minor improvements towards the terminal performance. It also supports behavior change programs, based on gradual change (e.g., force development).

The CCD is ideal for evaluating motivational programs designed to improve the performance of an individual athlete or a team. It can be used with accelerating (e.g., increasing running distance) and decelerating programs (e.g., weight loss) due to its stepwise graphic display.

Limitations

The CCD does not demonstrate a powerful experimental control. This is an extended procedure that may not be appropriate when an abrupt and immediate change is essential (e.g., when preparing for an upcoming competition). The CCD is appropriate only for evaluating already acquired behaviors.

24.8 Comparative Designs

Single subject research designs are mostly associated with studying a single participant and a single independent variable. This section describes another group of designs, comparing the effect of two or more independent variables, across one or more participants or behaviors. These designs are termed "COMPARATIVE", because they assess the relative effectiveness of a few interventions.

Multitreatment Design

The MULTITREATMENT DESIGN (MTD) is an extension of the withdrawal design, using the same baseline logic. Unlike the withdrawal design - presenting and withdrawing the same intervention, the MTD (Kratochwill, 1978) presents and removes different independent variables. The MTD may serve coaches who are considering different training methods, and are looking for empirical validation of the preferred one. Figure 4.38 presents an illustration of the MTD using hypothetical data.

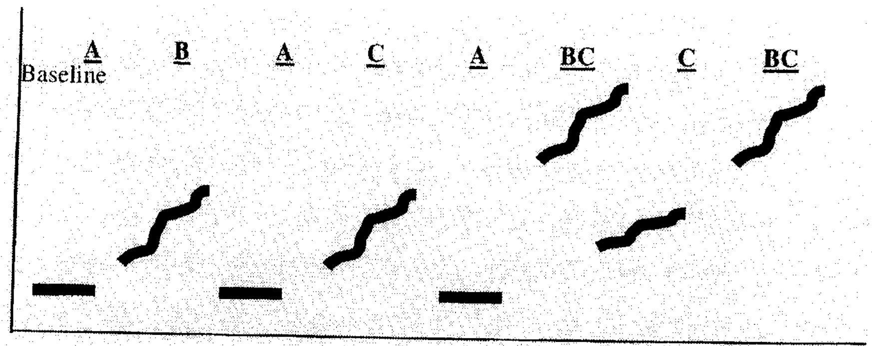

Figure 4.38 A Hypothetical Multitreatment Design Comparing Three Different Independent Variables (B, C, BC). Implementation.

As in withdrawal design, the MTD requires baseline measures of the target behavior to assess the athlete's present level of performance. Once a stable baseline is achieved, the independent variables are sequentially presented with a return to baseline following each presentation. It is important to determine in advance the order in which intervention will be introduced. If several participants take part in the study, the presentation order should be counterbalanced, to control for order effect.

The presentation of the first intervention is actually an ABA design. Then, the next intervention (C) is presented and reversed, with a notation A-B-A-C-A. Additional interventions may be introduced in the same manner. It is important to terminate the study with the most effective intervention (e.g., A-B-A-C-A-D-A-D).

It is also possible to evaluate a package of a few interventions and to compare their effect to other interventions. For example, two training methods (B & C) that show a moderate effect, may be combined (BC). This package is then compared to baseline condition (A-B-A-C-A-BC-A-BC), or to the separate treatments (A-B-A-C-A-BC-B-BC-C-BC). This variation stresses the flexibility of SSR, enabling modification of the design to meet the superior outcomes.

Experimental Control

Experimental control is demonstrated when the intervention is followed by a change in the dependent variable (see description for withdrawal design). Since there are a few compared interventions, they may be compared to each other (e.g., A-B-C-B-C) and not to a mediating baseline (e.g., A-B-A-C-A-C). In this case, one intervention (the less effective) serves as an active baseline for the other (the more effective). This presentation enables a direct comparison of the two interventions. It should be noted though, that only adjacent conditions might be compared. Non-consecutive conditions are subject to sequence effect and other interference, caused by the other conditions and by time elapsed between them.

External validity is achieved by replicating the intervention effect in different settings, with different behaviors and with different participants.

Advantages and Limitations

The MTD compares a few interventions in simple and clear fashion. It resembles the didactic logic of designing training programs, while attempting to reach the most effective procedure. Historical and maturation threats are substantial for this design because it introduces different interventions in a time lagged sequence. Another threat to internal validity is a carryover from one condition to another, and sequence effect, masking the pure contribution of each intervention.

24.9 *Alternating Treatment Design*

The ALTERNATING TREATMENT DESIGN (ATD) presents different interventions in alternation, and compares their effect on the dependent variable. The ATD (Barlow & Hayes, 1979) has also been referred to as multiple schedule design, alternating condition design, multi element baseline design, and simultaneous treatment design.

The most widely used definitions of ATD involve rapidly alternating two or more treatments or conditions, with the same person – within or between observational sessions (Barlow & Hersen, 1984). Most studies reported in the literature as a simultaneous treatment design have employed the same above definition of ATD. Figure 14.24.5 describes the prototype of the ATD.

Only one study (Browning, 1967) implemented a pure simultaneous presentation of treatments. It involves the concurrent presence of two or more treatments available at the same time for the participant's choice. The participant selects the most appropriate intervention at each given session. Therefore, this variation is actually measuring treatment preference rather than its effectiveness.

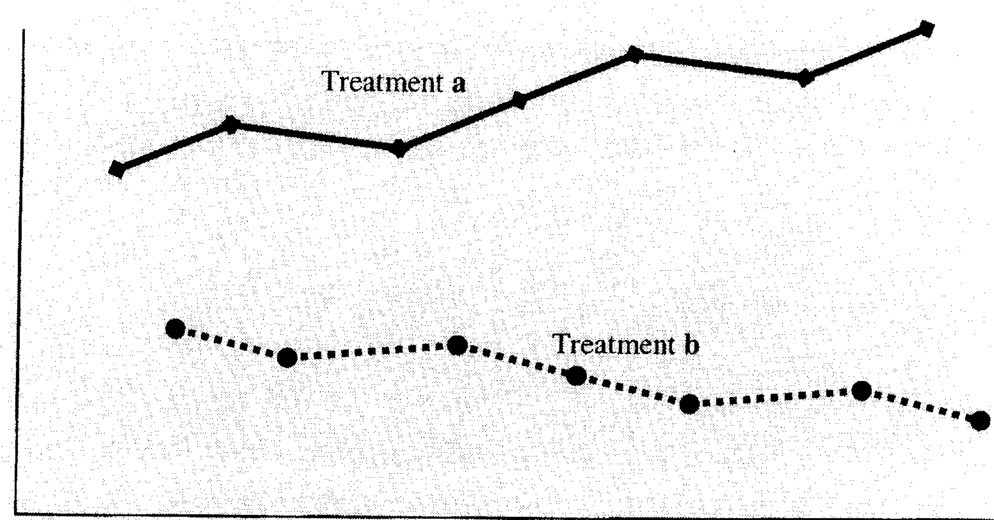

Figure 4.39 A Hypothetical Alternating Treatment Design, (comparing two different independent variables (A, B), with no baseline).

Implementation

Baseline data provide the researcher with clear and accurate "pictures" of the participant's ability. Target interventions are selected and scheduled for presentation prior to initiation of the study. The essence of this design is the "across time counterbalancing". This means that the different interventions are presented one after the other, with random variations in their order. For example, if two interventions (B & C) are involved, the ATD notation may look like this: A-B-C-C-B-C-B-B-C-B--. Interventions are alternated rapidly, on the same day or at every session.

Data for each independent variable are plotted on the same graph, creating clear and different data trends, one for each intervention. Data points for each intervention are represented by different symbols or colors. The connection of data points forms a clear and distinguishable curve for each variable. Unlike other single subject designs, the ATD does not require stability in data in order to introduce a new condition, because all trends are plotted concurrently.

All interventions are presented an equal number of times until a specific intervention demonstrates a clear superiority over the others. The final phase of the design continues the favorite intervention, terminating with therapeutic data.

Experimental Control

Graphs are analyzed for vertical differences between data trends for each intervention. Experimental control is concluded only if data trends are separated from one another. The larger the gap between the data trends, the more powerful is the experimental control. In case of an overlap between the data trends, no experimental control may be claimed. Data trends may be compared among themselves, with no need for baseline measures. However, comparing several interventions to a no-treatment condition may strengthen its credibility.

The rapid alternation of intervention controls for time and sequencing threats, is strengthening the ATD internal validity. External validity is achieved, as in all SSR designs, by replicating the intervention effect across participants, behaviors and settings.

Advantages

The ATD is an effective way to determine the relative advantage of a certain intervention over others. Although baseline data are recommended, interventions may be compared while skipping the baseline phase. This may be a significant advantage when an immediate intervention is required, or when collecting BASELINE DATA is intrusive or expensive.

Unlike the MTD, all interventions are presented in the same time frame, eliminating sequencing limitations associated with this design. The ATD is flexible and reactive, enabling the early detection, maintenance and follow-up of the more effective intervention.

Limitations

Due to the rapid alternations, a high level of proficiency is required for a reliable presentation of interventions. This highly technical expertise is essential because implementing the ATD is not natural to sport and educational contexts. It is uncommon to apply a few interventions, the same time, in order to attain the same goal.

The numerous alternations may interfere with the maintenance of behavior change in the presence of a certain intervention. The reason is the short exposure to each intervention, which is soon discontinued, to allow presentation of a different one. This problem is addressed in last phase of the ATD, maintaining the most effective treatment. However, relatively weak interventions may not "survive" the alternation period successfully.

Conclusions

This chapter introduced SSR designs that can be used in sport settings. WITHDRAWAL designs study causality by repeatedly producing the effect of the independent variable with the same athlete and on the same target behavior. MULTIPLE BASELINE and its variations evaluate the effect of an intervention by staggering its introduction into three or more baselines data series, across behaviors, participants, or settings. CHANGING CRITERION design evaluates experimental control by shifting the performance standards across time, in a step-by-step manner. COMPARATIVE DESIGNS measure the relative effectiveness of two or more interventions, by introducing them over the same time period.

The products of SSR are defined as "an increased ability to control the phenomenon of interest" (Johnston & Pennypacker, 1993, p. 13). SSR goes beyond the description of a process or the demonstration of correlation. Its main interest is showing functional relationships between measures of the dependent variables (e.g., athlete's motivation) and measures of the independent variable (e.g., motivation training intervention). This empirical/interventionist orientation allows the coach or the sport psychologist to assume accountability over the athlete's / team progress, ruling out the possibility that other extraneous variables were responsible for that change. Strategic and tactical decisions of athletes, coaches, and administrators, based on empirical findings should improve coaching and athletes' performance and add to the sport and exercise sciences accumulating knowledge.

SSR methodology is ideal for sport settings. Studying every athlete in individual sports is unique, requiring specific attention and a well-tailored methodology. Participants in team sports are also characterized by unique specialization. Studying each at a time, should enable the sport and exercise scientist and the professional to assess specific interventions and to recommend the suitable one.

The graphic presentation of the accumulated data may be used for an ongoing formative evaluation. It should be valuable for the coach and for the athlete to observe the data and to analyze performance accordingly, without waiting for the summative evaluation. This is powerful information, having a reactive power in itself.

Although still in its first steps, SSR is welcomed and published in quite a few journals. Let us hope that the use of this methodology in sport and exercise sciences will grow, adding to the body of knowledge in these fields, and to a high level of performance.

References

Baer, D. M. (1977). Reviewer's comment: Just because it's reliable doesn't mean that you use it. *Journal of Applied Behavior Analysis, 10*, 117-119.

Baer, D.M., Wolf, M.M., and Risley, T. R. (1968). Current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97.

Barlow, D.H., and Hersen, M. (1984). *Single Case Experimental Designs: Strategies for Studying Behavior Change* (2nd ed.). New York: Pergamon Press.

Barlow, D.H., and Hayes, S. C. (1979) Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*, 199-210.

Browning, R. M. (1967). A same – subject design for simultaneous comparison of three reinforcement contingencies. *Behavior Research and Therapy, 5*, 237-243.

Cooper, J.O., Heron, T. E., and Heward, W. L. (1987*). Applied Behavior Analysis*. Columbus, OH: Merill Publishing Company.

Cuvo, A. J. (1979). Multiple-baseline design in instructional research: Pitfalls of measurement and procedural advantages. *American Journal of Mental Deficiency, 84*, 219-228.

Hartmann, D. P., and Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis, 9*, 527-532.

Hersen, M. H., and Barlow, D.H. (1976). *Single Case Experimental Designs: Strategies for Studying Behavior Change*. New York: Pergamon Press.

Heward, W. L. (1978). *The Delayed Multiple Baseline Design*. Paper presented at the Fourth Annual Convention of the Association for Behavior Analysis, Chicasgo.

Horner, R. D., and Baer, D.M. (1978). Multiple-probe technique: A variation on the multiple baseline design. *Journal of Applied Behavior Analysis, 11*, 189-196.

Johnston, J. M., and Pennypacker, H. S. (1993). *Strategies and Tactics of Human Behavioral Research*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Kazdin, A. E. (1982). Single – Case Research Designs. New York: Oxford University Press.

Kratochwill, T. R. (Ed).(1978). *Single Subject Research – Strategies for Evaluating Change*. New York: Academic Press.

Michael, J. (1974). Stratistical inference for individual organism research: Mixed blessing or curse? Journal of *Applied Behavior Analysis, 7*, 647-653.

Parsonson, B. S., and Baer, D. M.(1978) The analysis and presentation of graphic data. In T.R. Kratochwill (Ed.) *Single Subject Research-Strategies for Evaluating Change*. New-York: Academic Press.

Repp, A. C. (1983). *Teaching the Mentally Retarded.* Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Risley, T.R., and Hart, B.M. (1968). Developing correspondence between the non-verbal and verbal behavior of preschool children. *Journal of Applied Behavior Analysis, 1*, 267-281.

Sidman, M. (1960). *Tactics of Scientific Research*. New York: Basic Books.

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental analysis*. New York: Appleton-Century.

Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.

Skinner, B. F. (1956). A Case history in scientific method. *American Psychologist*, 11, 221-223

Skinner, B. F. (1969). *Contingencies of Reinforcement: A Theoretical Analysis*. New York: Appleton-Century-Crofts.

Watson, P. J., and Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry, 12,* 257-259.