

Rapid Communication

A Brief Overview of Within-Subject Experimental Design Logic for Individuals with ASD

Cox DJ*

Department of Psychology, University of Florida, USA

***Corresponding author:** David J Cox, Department of Psychology, University of Florida, 945 Center Drive, Gainesville FL 32611, USA**Received:** May 10, 2016; **Accepted:** June 27, 2016;**Published:** June 28, 2016**Abstract**

Many individuals with ASD receive Applied Behavior Analysis (ABA) services. Most researchers and practicing, Board Certified Behavior Analysts (BCBAs) use within-subject designs to determine the effect of an intervention on behavior. This article outlines the logic for demonstrating experimental control from within-subject research designs for researchers and professionals currently unfamiliar with how to analyze and interpret these designs. Those not familiar with these research designs will be able to better understand and critique within-subjects research from behavior analytic journals, converse more easily with their behavior analytic colleagues, and use these designs to validate their own work with individual clients.

Keywords: ABA; Within-Subject Design; Evidence-Based Practices

Introduction

The Center for Disease Control estimates approximately 1 in 68 children are identified with Autism Spectrum Disorders (ASD) [1]. Many children with ASD require structured intervention to learn communicative, social, and daily living skills they may not learn as independently or rapidly as typically developing peers. The breadth of skill deficits demonstrated by individuals with ASD often requires remedial services provided by a variety of healthcare professions. Some of these professionals include Speech-and-Language Therapists (SLPs), Occupational Therapists (OTs), special education teachers, psychologists, and Board Certified Behavior Analysts (BCBAs).

Developing and maintaining collaborative relationships across disciplines will likely increase the overall consistency of interventions for individuals with ASD. As interdisciplinary collaboration continues to increase and be required by funding agencies, professionals face the difficulty of establishing and maintaining professional relationships between disciplines. These disciplines often differ widely in terminology, methods, and underlying theoretical assumptions. These differences have led some to discuss how collaboration might be accomplished without abandoning the theoretical foundations underlying a profession [2] through language common to all disciplines [3].

Since the author is most familiar with ABA, this brief overview covers experimental methods that are used by BCBAs and behavior analysts in clinical and research ABA settings. The goal is to briefly discuss the logic underlying various within-subject designs [4]. Discussing the rationale by which socially significant effects are demonstrated should allow the reader to understand how and why claims of effectiveness are made by researchers who employ these methods. This should, in turn, allow individuals previously unfamiliar with within-subject experimental designs to (a) feel more comfortable reading and interpreting behavior analytic research that uses within-subject designs, (b) better communicate with BCBA colleagues who use these designs within their practice, and (c) use these methods

in their own practice with individual clients where experimental evidence of the effect of an intervention is warranted.

The intended audience of this manuscript are clinicians and researchers who are primarily familiar with statistical techniques for data analysis. The first section of this manuscript discusses how the number of observations are often defined. The purpose of this section is to highlight the differences between obtaining large Ns across group designs versus within-subject designs, as well as the strengths and weaknesses of both designs relative to external and internal validity. The second section discusses the difference between statistical and visual analysis of data. The purpose of this section is to highlight differences between ABA research and common psychological research. Finally, the third section discusses the concepts of prediction, verification, and replication that are used for analyzing intervention effectiveness in three common within-subject empirical designs. The purpose of the third section is to provide examples and a description of how to analyze data from these designs to determine intervention effectiveness.

N as Number of Observations

Nearly all published research discusses the total N from the experiment. N in these contexts refers to the total number of observations made by researchers. In many situations N and number of participants may be equivalent, but this is not always so. For example, N would be equivalent for a researcher who chooses to record 1 observation on 100 participants, 10 observations on 10 participants, or 100 observations on 1 participant. Within-subject designs often use the latter approach across several subjects (see the *Journal of Applied Behavior Analysis* for a multitude of examples). Just as in group design research, findings from within-subject designs are interpreted in light of the number of observations present within each phase of the experiment.

All experiments seek to provide valid findings and research is often critiqued in terms of external and internal validity. Obtaining one observation for a large number of participants would result in

greater external validity of findings and lower internal validity [5]. That is, the researcher would be able to make general statements about the influence of an independent variable on a dependent variable for that particular sample of individuals. However, the researcher would be unable to say much about any individual participant from the sample. Compared to within-subject designs, external validity is a strength of group designs. The comparative weakness of group designs is decreased understanding of the mechanisms of individual behavior change [6].

In contrast, recording 100 observations for 1 participant would result in greater internal validity of findings and less external validity. Every response emitted by an individual varies in some dimension from previous responses. It is impossible to emit the exact same response twice. The more observations collected for a participant, the greater the probability the observations collected accurately reflect the effect of an independent variable on the dependent variable for that particular individual. This allows the researcher to make more specific statements about the effect of an independent variable on a dependent variable for that specific participant [6]. This becomes a significant strength of within-subject designs - especially for clinicians interested in the effect of an intervention for a specific individual. However, these designs are less equipped to make claims about how results generalize to other participants within the larger population compared to group designs.

The most appropriate data collection strategy likely depends on the research question being asked and resources available to the researcher. ABA often opts for a large number of observations for each participant. Over time, the results of a large number of participants who are exposed to the same independent variable are sometimes aggregated and analyzed [7]. However, the majority of researchers publishing within ABA use a handful of participants in any one study (see *Journal of Applied Behavior Analysis* for examples).

Evaluating Data

Another difference between ABA and many other psychological fields is how data are analyzed. Most fields of psychology use inferential statistics to determine the effect of an intervention on a target behavior. ABA typically uses visual analysis to interpret data and not statistical techniques. Visual analysis of clinical intervention typically uses two criteria for determining the effectiveness of an intervention. The first criterion is clear demonstration of functional control of the behavior of interest (more on this below). The second criterion is the notion of social significance as opposed to statistical significance.

A socially significant change can be defined as a change in the occurrence of behavior to levels acceptable to the client and socially relevant stakeholders (e.g., caregivers, other therapists, teachers, funding agencies, etc. [8]. This differs from statistically significant change. Statistical significance refers to the probability that an observed difference between the experimental and control group occurred if the independent variable really had no effect. A demarcating example could be as follows: An individual is observed to strike their own head at a rate of 100 times per hour with a standard deviation of 10 strikes to the head. An intervention put into place that consistently results in a reduction of self-injury to 50 times per hour across all participants

exposed to the intervention with a standard deviation of 5. This reduction will surely be statistically significant. However, these results would not be socially significant as the individual is still engaging in self-injurious behavior at rates harmful to his or her overall health. Statistically, the intervention would have been deemed effective. Based on social significance, the intervention would still need a lot of improvement. This does not mean that all statistically significant differences are not socially significant. Only that social significance is a different criterion by which to judge change in behavior.

The use of visual analysis has been argued by some as bias on part of the individual interpreting the data which leads to inconsistent analysis [9]. This has led some to propose structured criteria for interpreting within-subject design graphs [10] as well as the use of statistical techniques for interpreting within-subject design data [11]. However, other researchers have observed relatively high consistency among those trained in the visual analysis of within-subject design graphs [12]. As such, visual analysis in most published research and clinical settings does not use structured interpretation criteria nor statistical techniques.

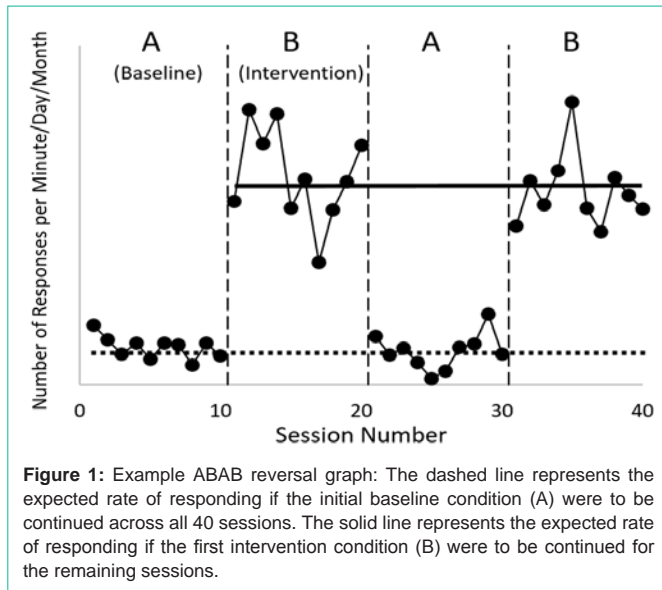
The second requirement for labeling an intervention as effective is functional control. Similar to group designs, demonstrating functional control using a within-subject design involves comparison of observations in a treatment condition to observations in a control condition. This control condition is often referred to as a baseline period and provides measurement of the behavior in the absence of the intervention. Observations are typically collected until stability in responding is achieved or the behavior is trending (i.e., increase or decreasing) in a direction opposite to the intended effect of the intervention. The overall level, variability, and trend during the baseline control condition is then compared to the level, variability, and trend during the intervention condition. If these comparisons can be verified and replicated within a subject, the intervention could be argued to have demonstrated functional control over responding. How verification and replication are demonstrated visually depends on the type of within-subject design used.

Common Within-Subject Designs

ABAB reversal design

Figure 1 shows an example ABAB reversal design. This design starts in a baseline/control condition until one of two things occur. First would be that stability in responding is observed. Stability can be defined as an absence of trend (behavior increasing or decreasing) and all measures fall within a small range of values (low variability) [11]. Second would be that the data path has a trend in the opposite direction to what is expected with the intervention. For example, if an intervention is aimed at decreasing self-injurious behavior, an increasing baseline trend would suggest the researcher could implement the intervention. This is because reduction in self-injurious behavior during the intervention would not be confounded with the baseline trend (Figure 1).

Following observation of a stable baseline (session 11 in Figure 1), the intervention would be introduced and implemented until stability in responding is achieved. If the intervention is believed to be potentially effective, the rate of responding should change from what would have been predicted had the baseline condition continued



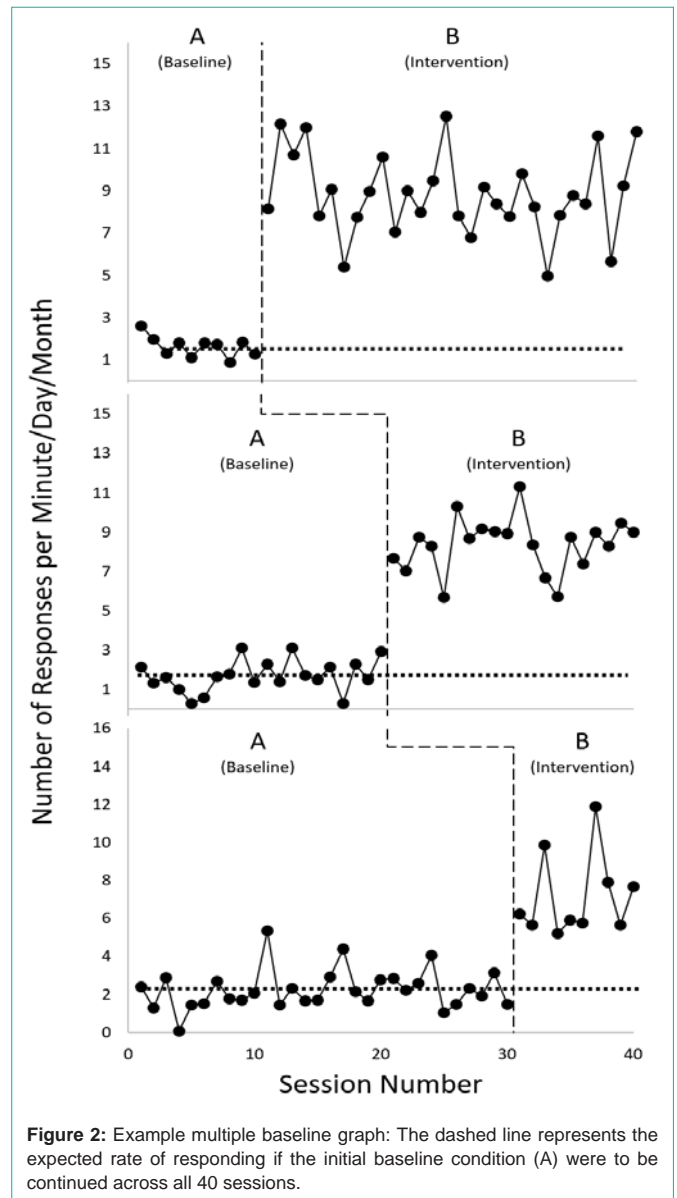
(dashed line in Figure 1). This is called a change in the level of behavior. The greater the change in level of responding and the more rapid the onset of level change, the more effective intervention is claimed to be. For example, an intervention that requires 10 observations before a small change in level of responding is observed would be considered a weak intervention. However, a near immediate change in level of responding combined with a large level change would indicate a much more effective intervention (Figure 1).

At this point, the design is an AB design and could be considered analogous to pre- post-test designs used in many group design studies. The primary exception is a larger number of observations for each individual participant in the example in Figure 1. Since potential confounding variables could have changed at the same time as implementation of the intervention, the next step in the ABAB reversal design is to remove the intervention (i.e., the first B phase) and go back to a baseline condition (second A phase - beginning at session 21 in Figure 1). If the intervention were responsible for the initial behavior change we would expect to observe responding reverse back to the original baseline level of responding (i.e., near the level indicated by the dashed line). That is, we would verify that the original baseline level would have continued if the intervention were not implemented [13].

Finally, to determine if the effect of the intervention can be replicated, the intervention is reintroduced. When the second B phase is implemented, responding should return to the rate observed in the first intervention condition (solid horizontal line in the final B phase in Figure 1). By replicating within the experiment, two important pieces of information are obtained [13]. First is that behavior change in the first B phase is less likely to have been the result of a confounding variable. Second, replication within an experiment demonstrates that the change in behavior can be made to occur again - the change is reliable.

Multiple baseline design

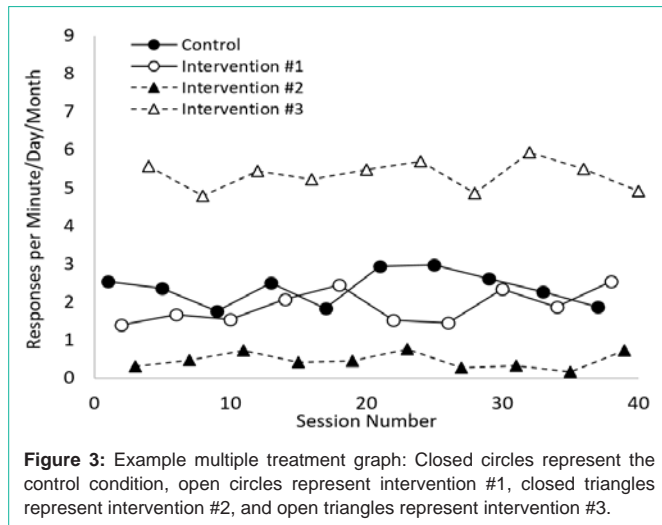
There are certain occasions where removing an intervention is not practical or not desirable. Examples include learning that



cannot be undone (e.g., learning to ride a bike), behavior that is too dangerous to return to baseline (e.g., aggression), or when limited time is of potential issue (e.g., increase in response that is first in a series of responses that build off one another). The multiple baseline is often used in these instances to verify and replicate the effects of an intervention across multiple participants [14,15], responses [16], or settings [17].

Figure 2 shows hypothetical data presented in a multiple baseline design. The plot shows data across multiple target responses. Again, each of the panels could be similar responses across different participants (i.e., multiple baseline across subjects), different responses emitted by the same participant (i.e., multiple baseline across responses), or the same response from the same participant across different settings (i.e., multiple baseline across settings such as clinic and home) (Figure 2).

Similar to reversal designs, the logic of multiple baseline designs



rests on verification and replication of the treatment effect with change from baseline to intervention occurring after stability is observed in the data path. Verification occurs when the intervention is implemented for the top response and the rate of responding remains unchanged for the second and third responses (sessions 11-20). A second verification of the intervention effect occurs when the intervention is implemented on the second response and responding remains unchanged for the third response (sessions 21-30). Replication occurs during two series of sessions. The first replication of the intervention effect occurs when a change from predicted levels of responding occurs for the second response when the intervention is implemented (sessions 21-40). The second replication of intervention effects occurs during sessions 31-40 of the third response when the intervention is implemented and behavior changes relative to what would have been predicted had baseline been continued.

Multiple treatment design

This design is often used to rapidly determine the influence of different interventions or components of an intervention on a response of interest. Readers familiar with behavior analysis will recognize this design from functional analyses of problem behavior [7,18]. In addition to assessment, multiple treatment designs have also been used for analyzing interventions [16]. As with both of the above methodologies, the multiple treatment design compares stable responding in a control condition to stable responding in intervention conditions (Figure 3).

Figure 3 presents hypothetical data for a multiple treatment design. The effect of each intervention on the target behavior is determined by comparing the data path for the intervention to the data path for the control condition. For example, intervention #1 resulted in comparable levels of responding to the control condition. This suggests intervention #1 had no impact on responding. This claim is made based on the high degree of overlap between data paths for both conditions. Interventions #2 and #3 resulted in a decrease and increase in the level of responding compared to the control condition, respectively. Depending on the purpose of the proposed interventions (i.e., to increase or decrease a given behavior), the data from Figure 3 would provide evidence of which intervention is most effective for that particular client.

Summary

Collaboration between disciplines can be enhanced by understanding the methods and reasoning by which professionals make statements regarding evidence-based practice. This article provided a brief overview of the logic behind analyzing within-subject experimental designs. These designs are commonly used by BCBAs and the empirical literature that comprises the science of ABA. Understanding the empirical designs presented here and the underlying analytical approach will allow professionals who are unfamiliar with within-subject designs to more easily evaluate behavior analytic literature and converse with behavior analytic colleagues about data they present. Finally, professionals who understand empirical rationale behind within-subject designs have another empirical tool they can use when either (a) group design studies are not feasible or (b) the professional is interested in understanding the effect of an intervention for one specific client in his or her research or clinical practice.

References

- Christensen DL, Baio J, Braun KV, Bilder D, Charles J, Constantino JN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network. 2016; 65: 1-23.
- Brodhead MT. Maintaining Professional Relationships in an Interdisciplinary Setting: Strategies for Navigating Nonbehavioral Treatment Recommendations for Individuals with Autism. *Behav Analysis Practice*. 2015; 8: 70-78.
- Cox DJ. From interdisciplinary to integrated care of the child with autism: the essential role for a code of ethics. *J Autism Dev Disord*. 2012; 42: 2729-2738.
- Roane HS, Ringdahl JE, Kelley ME, Glover AC. Single-Case Experimental Designs. Fisher WW, Piazza CC, Roane HS, editors. In: *Handbook of Applied Behavior Analysis*. The Guilford Press. 2011; 132-147.
- Johnston JM, Pennypacker HS. *Strategies and Tactics of Behavioral Research*: 2nd Edn. Erlbaum. 1993.
- Fisher WW, Groff RA, Roane HS. *Applied Behavior Analysis: History, Philosophy, and Basic Methods*. Fisher WW, Piazza CC, Roane HS, editors. In: *Handbook of Applied Behavior Analysis*. The Guilford Press. 2011; 3-13.
- Iwata BA, Pace GM, Dorsey MF, Zarcone JR, Vollmer TR, Smith RG, et al. The functions of self-injurious behavior: an experimental-epidemiological analysis. *J Appl Behav Anal*. 1994; 27: 215-240.
- Baer DM, Wolf MM, Risley TR. Some current dimensions of applied behavior analysis. *J Appl Behav Anal*. 1968; 1: 91-97.
- Deprospero A, Cohen S. Inconsistent visual analyses of intrasubject data. *J Appl Behav Anal*. 1979; 12: 573-579.
- Fisher WW, Kelley ME, Lomas JE. Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *J Appl Behav Anal*. 2003; 36: 387-406.
- Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. *Behav Ther*. 2011; 42: 284-299.
- Kahng S, Chung KM, Gutshall K, Pitts SC, Kao J, Girolami K. Consistent visual analyses of intrasubject data. *J Appl Behav Anal*. 2010; 43: 35-45.
- Cooper JO, Heron TE, Heward WL. *Analyzing Behavior Change: Basic Assumptions and Strategies*. In: *Applied Behavior Analysis*: 2nd Edn. Pearson Education Inc. 2007; 158-175.
- Beaulieu L, Hanley GP, Roberson AA. Effects of responding to a name and group call on preschoolers compliance. *J Appl Behav Anal*. 2012; 45: 685-707.
- Paden AR, Kodak T. The effects of reinforcement magnitude on skill

- acquisition for children with autism. *J Appl Behav Anal.* 2015; 48: 924-929.
16. Wacker DP, Berg WK, Berrie P, Swatta P. Generalization and maintenance of complex skills by severely handicapped adolescents following picture prompt training. *J Appl Behav Anal.* 1985; 18: 329-336.
17. Charlop MH, Trasowech JE. Increasing autistic children's daily spontaneous speech. *J Appl Behav Anal.* 1991; 24: 747-761.
18. Iwata BA, Dorsey MF, Slifer KJ, Bauman KE, Richman GS. Toward a functional analysis of self-injury. *J Appl Behav Anal.* 1994; 27: 197-209.