Improving the Methodological, Analytical, and Cultural Impact of Behavior Analysis Via Utilization of Group Design Methods and Statistical Analyses

Mark R. Dixon[a,*], Zhihui Yi[b,] Amanda N. Chastain[c], Meredith T. Matthews[d]

[a,*] **Corresponding Author:** Mark R. Dixon, University of Illinois Chicago, USA.
E-mail: mrdixon@uic.edu
ORCID: https://orcid.org/0000-0002-0671-1183

[b] Zhihui Yi, University of Illinois Chicago, USA.
E-mail: zyi7@uic.edu

[c] Amanda N. Chastain, University of Illinois Chicago, USA.
E-mail: achast2@uic.edu

[d] Meredith T. Matthews, University of Illinois Chicago, USA.
E-mail: mmatth22@uic.edu

Abstract

The multi-decade debates within the field of behavior analysis as to the possible value and threat of group design methodology and statistical analyses on the purity of the field have weakened the discipline's maximal impact on the world. This paper rebukes the concerns and suggests that through such adoption behavior analysis may likely better achieve its world-changing ideals, and pragmatic initiatives. We begin with a historical trace of the current debate and describe the pros and cons to design/analysis inclusion, frame such matters within the context of contemporary issues which applied behavior analysts find themselves concerned, and ultimately put forward means by which broader, and perhaps more impactful, research questions can be asked and interpreted.

Keywords:

Group Design Methodology, Research Design, Statistics, Single-Subject Design

Introduction

One of the most distinguishable characteristics of the field of behavior analysis is its reliance on a research method approach entitled "single-subject design" (Cooper et al., 2007; DeRosa et al., 2019). This exploratory approach is not to be confused with a case study (Bolgar, 1965) whereby a single individual is studied in an uncontrolled manner and simply reported upon afterwards. Single-subject design (SSD) approaches to research questions instead systematically analyze the repeated effects of an independent variable on a dependent variable across one or more individual subjects (Cooper et al., 2007). This cluster of techniques are employed within the field because of the concerns that many of the field's founders had regarding the limitations of traditional research methods – often described as group or statistical research methods (Sidman, 1960; Skinner, 1956). In the early days of behavior analysis, B. F. Skinner himself recommended this departure from traditional psychological statistical methods and analysis because:

You cannot easily make a change in the conditions of an experiment when twenty-four apparatuses have to be altered. Any gain in rigor is more than matched by a loss of flexibility. We were forced to confine ourselves to processes which could be studied with the baselines already developed in earlier work. We could not move on to the discovery of other processes or even to a more refined analysis of those we were working with. No matter how significant might be the relations we actually demonstrated, our statistical Leviathan had swum aground (Skinner, 1956, pp. 113-114)

Essentially, Skinner was concerned that the analysis of the group did not provide meaningful analysis of any of the group members, and as such little could be discerned about the behavior of the single subject. And to Skinner, that single subject mattered as it was the level of analysis that appeared to be necessary for evaluating and changing behavior. Today the utilization of SSD reaches far beyond behavior analysis into other clinical and helping professions such as social work and health care (see Bloom et al., 2009 for a textbook length treatment).

There is no debate we wish to have over the utility of the SSD as a method within the field of behavior analysis, as the approach has been of great utility for decades in crafting a precise analysis of the controlling variables on an individual's behavior. Furthermore, we make no dismissal over clear discoveries that such a research tradition has allowed for in the field's history (e.g., functional analysis and intervention; see Beavers et al., 2013 for a review of literature of functional analysis). We do, however, have concern that the overreliance on SSD and omitting the inclusion of more traditional research designs has marginalized the impact the field is having on matters of great concern to behavior analysts. For our work to be taken more seriously beyond the walls of our own discipline, we may need to start writing, speaking, and describing our results in broader non-technical language, and look carefully at what is gained and lost by the specific research design chosen by the researcher. This approach is not meant to imply defeat or suggest minimizing of our research endeavors, but rather simply to greatly embrace the very core elements of what it is to be a pragmatic behaviorist – functional utility.

Beyond the initial proclamations by Skinner regarding research methods, many well-known scholars have spoken out against the incorporation of group designs into the field. In the late 1990s, a group of presenters at the annual Association for Behavior Analysis convention debated the adoption of these "non-behavioral" methods, and their views were eventually assembled within a special section of one of the field's peer-reviewed journals. In this special section of papers, a range of opinions are presented with repeated concerns being made about loss of understanding the behavior of the individual person

(Branch, 1999; Perone, 1999), and inferencing beyond the data (Ator, 1999; Branch, 1999; Davison, 1999), and the additional distraction that statistical analyses create (Perone, 1999) as a drift away from SSD could have for the field. Only an occasional pro-group design approach was presented (Crosbie, 1999), and statements as such hinted at a potential fractioning that might be underway even within the field of behavior analysis. More recent discussions on the superiority of SSDs have echoed similar sediments (Kyonka et al., 2019). Examples of group design methods and statistical analyses will occasionally appear in behavior analytic journals today (e.g., Dixon et al., 2022; Jang et al., 2012, Sutton et al., 2022; Silverman et al., 2007; Yi et al., 2022), yet they are a minority compared to the continued use of SSDs throughout. Even with such examples appearing within our own collection of scholarly journals, most behavior analytic textbooks fail to describe the rationale and usage of group methods and analyses for behavior analysts (Cooper 2007; Mayer et al, 2019; Sidman, 1960; yet see Belisle et al, 2021 as an exception) – thus potentially limiting an awareness of value and an understanding of how to construct research questions utilizing group designs.

The current social and political environment often has placed the field of behavior analysis in its crosshairs. We are increasingly being described as a field of insensitive determinants of client autonomy (Kirkham, 2017; McGill & Robinson, 2021), responsible for the development of alleged trauma in former clients (Kupferstein, 2018), insensitive to racial injustices (e.g., Čolić et al., 2022; Zarcone et al., 2019), and behind trends of interest that need to be more fully addressed and analyzed (e.g., DeFelice & Diller, 2019; Fontenot et al., 2019; Kornack et al., 2019; Morris et al., 2021; Wang et al., 2019). Even our most heavily dominated applied appendage – autistic care – is being challenged as non-effective (United States of America Department of Defense, 2021). Critics from within and beyond the field itself seem to believe that perhaps behavior analysis has not aged well in a fast-changing and culturally evolving society. We are clear of the risks that any field of inquiry may encounter when it too quickly drifts from historical roots because of modern themes and current interests. Yet on the other hand, a field which ignores the critique of itself by its own members suggests that a possible reappraisal may be indeed necessary. And furthermore, if the field's reaction to such criticism is with rhetoric and not data, increased dismissal of the utility and value of the overall field may result. Even those who speak up about creating change in the discipline (e.g., Jaramillo & Nohelty, 2022; Mathur & Rodriguez, 2022; Pritchett et al., 2022; Wright, 2019), doing better than in the past (e.g., Baires et al., 2022; Li et al., 2019), or improving inclusivity (e.g., Deochand & Costello, 2022; Levy et al., 2022; Lovelace et al., 2022) cannot and should not rest

after such assertions alone, but only after producing data by which to support such claims. We believe that only through data that change will occur at the magnitude of impact that appears desired, and most importantly the type of data that will yield the greatest change-making potential will be gathered using between group, large sample sized research designs and statistical analyses.

Many of the most impactful contributions in terms of scalability to improving the human condition have occurred when behavior analysts have adopted non-SSD approaches to demonstrating effects of the independent variable. One example involves the use of contingency management for the treatment of substance use disorders (e.g., Dunn et al., 2008; Higgins et al., 1994; Higgins et al., 1991; Ledgerwood et al., 2008). In many of these published studies, a comparison is made between groups of individuals assigned to either a traditional treatment condition in which participants receive drug treatment as usual (e.g., standard relapse prevention support, health risk education, group meetings, and individual counseling sessions), or a contingency management condition wherein abstinence behaviors resulted in payment in the form of vouchers exchangeable for community retail items (see Higgins et al., 2019 for a recent review of the literature). Related studies on analyzing the choice making of drug users have also centered around non-SSD methods and analyses (e.g., Heil et al., 2006; Nighbor et al., 2019; Thrailkill et al., 2022; Yoon et al., 2007). Another example of behavioral solutions that have been quite successful at achieving wide-scale acceptability and adoption is using relational framing techniques to treat mental and physical health conditions under the auspices of Acceptance and Commitment Therapy (Dixon et al., 2023). This treatment approach has almost exclusively utilized between-subjects research methods (see Twohig et al., 2007 as an outlier in its use of SSD), and gathered enough data to be deemed as effective enough for the World Health Organization (WHO, 2020) to distribute ACT self-help material in 21 languages "for anyone who experiences stress, wherever they live and whatever their circumstances" (p. 5). A final example comes from Positive Behavior Intervention and Supports (PBIS; Horner & Sugai, 2015) whereby social-culture and behavioral supports are implemented school-wide (Tier 1 and Tier 2) and at the level of the individual (Tier 3) to promote improved educational and social outcomes. Many documented successes of this work are presented with group designs speaking to comparisons made between non-PBIS exposed and PBIS exposed student groups of varying demographics, whereby the PBIS exposed students tend to fair better regarding social-emotional functioning, behavioral concerns, academic performance, bullying and peer rejection, and prosocial behavior (Bradshaw et al., 2010; Bradshaw et al., 2012; Horner et al., 2009; Waasdorp et

al., 2012). Many more examples of behavior analytic researchers who have stretched beyond the SSD research tradition can be found in the context of functional analysis (Kurtz et al., 2013), gambling (Habib & Dixon, 2010), and The Good Behavior Game (Joslyn et al., 2019). All of this is not to suggest that SSD themselves do not yield utility for better understanding of human behavior. We completely agree that SSD has a crucial role to play in the behavior analysis of today and tomorrow. However, we also must accept that in order to advance beyond our current limited impact we have made to changing the world through behavioral science (Dixon et al., 2018), that we should look carefully at what many of our most successful endeavors all appear to have in common – group research designs.

Single-subject research designs have limitations related to the generalizability of the study findings as well as the methodological constraints that limit the use of inferential statistical methods. Conclusions can hardly be made regarding a group or groups of subjects, as the baseline logic (Cooper et al., 2020) underneath most SSDs fundamentally focuses on inferencing the likelihood that a procedure is responsible for producing the observed changes at an individual level. This does not speak to the likelihood that a similar effect can be observed when such a procedure is applied to the population, the group of individuals upon whom behavior analysts wish to bring socially significant changes. To a certain degree this limitation is mitigated through systematic replications. Population-level inference is usually drawn by first randomly sampling the target population to create a group of subjects and then using the changes observed in the group under the procedure to make inferences about the likelihood of whether the population will respond in similar ways. Additionally, the field of applied behavior analysis (ABA) takes pride in its continued use of technical terminology which has largely allowed behavior analysts to effectively communicate with other behavior analysts. However, this technical language may be a barrier preventing behavior analysis to reach professionals from other disciplines (Becirevic et al., 2016). Historically, this insular vocabulary and research approach were the very reason the field crafted its own scientific journals (e.g., Journal of Applied Behavior Analysis, Journal of the Experimental Analysis of Behavior) to combat the inability to publish its research using SSDs in more traditional psychology journals that were departing from a behavioral tradition (Gollub, 2002).

All the positive features of group design and analysis do not imply that shortcomings of the approach fail to exist. Issues of sample representation, clinical/practical significance, effect sizes, maintenance, functional control, and generalization all remain ripe for continued debate. In conclusion, we believe that the

benefits of incorporating group designs and analyses outweigh the limitations – and as such, behavior analysts can advance further towards having a positive impact on the world by incorporating these sorts of methods into the means by which they speak to matters of interest, react to critics, and advocate for behavioral solutions to non-behavioral audiences.

Group Designs in Behavior Analytic Settings

In contrast to single-subject designs, quantitative studies using group designs use variables among participants within one group or across multiple groups to examine the impact of independent variables on dependent variables. In order to aggregate these data, various descriptive and inferential statistical methods are used to provide a relatively objective interpretation. There are many nuances in how group design studies are categorized and the different optimal statistical methods that go with them, such as complex mixed designs better analyzed using structural equation modeling (Duncan, 1969) and time series designs (Gottman et al., 1969). In this section, though, we primarily focus on three types of group design in their basic forms: within-subject group design, between-subject group design, and mixed group design. We also provide a brief description of the design, common statistical methods used, a sample question that can be studied using this design, and how this research method can be used to address some contemporary issues surrounding our field.

Within-Subject Group Design

The concept of a within group research design is that individual subjects are evaluated multiple times during the experiment. Such a design is more similar to an SSD than other sorts of group designs, whereby the individual subject is only examined once and compared to other subjects. Here in a within group design, a group of subjects may be exposed to one same independent variable multiple times, a range of levels of an independent variable, or a combination of variables. For example, 20 children with attention-deficit/hyperactivity disorder (ADHD) may be exposed to behavioral interventions for 8 weeks, and also medication for 8 weeks. The order of delivery of treatment may be randomized across the entire 20 children, and after exposure to both (i.e., 16 total weeks), an analysis could be made as to which sort of treatment was better in terms of outcomes on a dependent variable (e.g., performance, attention in class, parent reports of homework completion). Variations of this basic framework could include comparing low and high doses of ADHD medication, comparing behavioral treatment alone to behavioral treatment with medication, a period of no-treatment to that of behavioral treatment, or even low and high doses of behavioral interventions alone.

As these examples illustrate, in a within-subject group design, dependent variables are collected from participants within one same group assignment, sometimes across multiple time points. Participants are usually selected based on similar criteria (e.g., demographic compositions, existing conditions, exposure to similar interventions), and statistical methods are used to compare the relationships of group-level measures across multiple time points or to identify patterns and relationships among different variables. Using within-subject group design, behavior analysts craft purpose statements such as evaluating treatment progress over time for one group of participants, identifying relationships between outcome measures and potential predictors, and evaluating the extent to which measurements used are reliable and accurate. In contrast to SSD which heavily relies on visual inspection of the data obtained to determine effects of the intervention, group designs supplement the visual graphical differences that are plotted through the use of statistical procedures. Most commonly, these questions can be answered using the data obtained through repeated measure T-tests, simple linear regression, and correlation analysis. To comprehensively discuss these statistical methods is beyond the scope of the current paper, and interested readers are highly encouraged to reference statistics textbooks in psychology or education (e.g., Howell, 2012). Here we present examples with from published peer-reviewed articles to briefly discuss these study's conceptualization and statistical analysis process, as well as suggestions on how similar investigations can be made to address some emerging research questions and challenges the field faces.

Treatment Progress Overtime

Consider the following scenario: A behavior analyst is interested in evaluating whether a comprehensive treatment model (CTM) newly introduced at their clinic effectively produces measurable gains among a group of autistic learners. Although such a question can be examined by individually evaluating the outcome measures of each learner, it might be difficult to aggregate such data due to insurance reimbursement or comparisons across agencies when some learners are showing improvement in outcome measures, and some are showing decreases. Furthermore, with varying baseline measures, it is difficult, if not impossible, to control for the potential differing effect of the intervention on learners with varying abilities. Although interpretations can be made by categorizing learners into different groups and evaluating the change in the outcome measure's level, trend, and variability, as routinely done in visual analysis (Cooper et al., 2020), such interpretation is largely subjective in nature although some tools have been constructed to improve objectivity (Dowdy et al., 2022).

A within-subject group design will be a good fit in scenarios like this. In this case, the investigator will gather all learners' (i.e., population) or randomly select a subgroup of learners' (i.e., sample) data before and after the implementation of the CTM. Here we would compare each learner's follow-up measure against their baseline measures, and the most common will use repeated measure T-tests to detect whether there are statistically significant changes between the two time points. For example, Yi et al. (2022) investigated the impact of ABA service embedded within the student's Individualized Education Program (IEP) using data obtained from a public school. A repeated measure T-test showed a statistically significant difference between students' performance at the beginning and the end of the school year. Figure 1 illustrates the comparison being made in this scenario.

Figure 1.
Illustration for using within-subject group design to evaluate changes for a group of participants between baseline and follow-up.
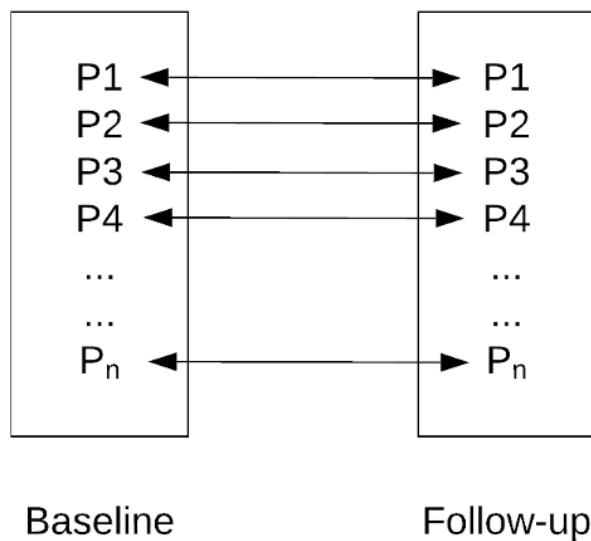
Within-subject designs using similar methods to evaluate participants' outcomes over time can have meaningful impact in the contemporary field of behavior analysis. There has been increasing attention in using behavior analytic principles to address social issues such as systematic racism (Shea et al., 2022), as well as critical reflections on the research and clinical practice of behavior analysis among the Black community (Čolić et al., 2022; Lovelace et al., 2022). Čolić et al. (2022) synthesized black caregivers' experience when it comes to autism care and provided specific accounts on racism manifested across its multiple stages. Čolić et al. provided multiple examples of how to address institutional racism and offered specific recommendations for ABA providers to combat racial bias. Sevon (2022) also proposed similar recommendations on increasing awareness on anti-Black racism. An important step the field should take

is to diligently listen to the voice from the community, consumers, and stakeholders and act accordingly to design and implement the behavior-changing system to dismantle these issues. Such endeavor can be strengthened by using within-subject research designs. For example, as an extension to Čolić et al., researchers might design and implement an intervention package among service providers consisting of awareness training on Black cultural values and intersectionality between Blackness and autism. They could subsequently incorporate behavior skill training on strengthening partnerships among stakeholders. A study can be conducted by first gathering qualitative and quantitative data on Black caregiver experience within behavior analytic settings. After the implementation of the intervention package among service providers, multiple waves of follow-up measures can be taken. Using within-subject group designs, statistical analysis can be done to detect whether the intervention package produced measurable improvements on Black caregivers' experience. Surely a single-subject approach could also be crafted here, however external validity is inherently reduced.

Predictors for Outcome

Another area where within-subject group design can be used is to identify predictors for certain outcome measures. Consider the following scenario: A behavior analyst who works at a local school district is interested in exploring whether there is a relationship between the amount of ABA service received and students' progress. In order to advocate for more resources devoted to ABA, the education team needs to reasonably demonstrate a relationship between the dosage of ABA and the amount of progress. In this situation, SSDs cannot easily answer this question. The outcome of ABA is often measured across multiple weeks or months. With alternating treatment designs, the short exposure to each dosage condition is not powerful enough to produce meaningful changes that can be detected. With reversal designs across multiple dosage levels, the potential confound of sequence effect and carryover effect is so large that it is very difficult to attribute measured gains to a specific condition. An alternative is to visually inspect a scatterplot with the treatment dosage on the x-axis and the amount of progress on the y-axis. However, when the number of learners is low and with high variability in outcome measures, it might be difficult to visually interpret the trend of the progress as the dosage increases along the x-axis.

A within-subject group design would be a good fit in this situation. Instead of relying on SSDs or visual analysis of the scatterplot, a linear regression analysis can be done to identify whether the dependent variable can be reliably predicted by a single or

multiple independent variables. For example, Yi et al. (2022) was interested in identifying factors that could predict participants' gain in school readiness skills. Using a cohort of 17 autistic students within a public school, the researcher conducted a simple linear regression on the dosage of PEAK-based instruction on participants' gain in Bracken School Readiness Scale (BSRA; Bracken, 2007). Results showed that the amount of PEAK-based instruction was a statistically significant predictor for their gain on BSRA, $F(1,14) = 5.31$, $p = .036$. The dosage of PEAK-based instruction accounted for 27.80% of the variance observed in BSRA gain.

Studies and analyses like this can deepen our understanding among several issues regarding the training for the next generation of behavior analysts and ensuring high quality of care. Several recent studies analyzed data released by the Behavior Analyst Certification Board (BACB) on certification outcomes among accredited and verified course sequences in this field (Dubuque & Kazemi, 2022; Matson & Konst, 2014). By comparing certification outcomes among multiple applicant characteristics (e.g., program mode and accreditation status), researchers reported trends in the number of applicants in the last decade and differences observed among applicants experienced different modes of learning (e.g., in person, remote, hybrid). An extension of this body of work would be to use within-subject group designs and to use analyses such as linear and logistic regression models in identifying the environmental factors that are mostly likely to impact the educational outcome. Researchers need to first gather more comprehensive applicant data, such as demographic information, social economic status, educational information (e.g., program mode, curriculum design, faculty-student ratio), fieldwork and supervision experience, and continue education. Researchers also need to collect more comprehensive outcome measures besides the board exam pass rate, such as consumer satisfaction, and apply appropriate statistical tests to identify predictors of these outcome variables. Similarly, the field has become increasing aware of staff burnout and its detrimental impact on the quality of care (Plantiveau et al., 2018). By using similar group design method, researchers can identify predictors of staff burnout and develop corresponding strategies to improve the quality of care.

Psychometric Properties

Behavior analysts have a long history in designing measurement systems for behavior changes overtime. Focusing on the individual's learning history, the field has long cautioned against standardized testing due to concerns on the inability to individualize the assessment's process, which might more accurately reflect on the behavior observed (Ayllon & Kelly, 1972; Koegel et al., 1997). As the field continues to evolve

and expand, concerns have been raised on the reliability of many assessments ABA providers use in clinical settings. The field of psychometrics studies the construction and application of assessment tools and an assessment's psychometric properties describe how well it measures what it claims to measure. Most commonly, researchers evaluate the instrument's validity and reliability, and, often time for a newly developed instrument, the extent to which the instrument yields similar outcome to established measures (Nunnally & Bernstein, 1994). In a systematic review conducted by Ackley et al. (2019), only four of the 18 ABA-based assessments reported data supporting its reliability. Evaluating psychometric properties of ABA-based assessment is not only a valid scientific objective, but also offers many benefits such as increasing the external validity of the field, simplifying assessment process, and increasing dissemination beyond behavior-analytic journals (Sutton et al., 2022). Issues like this cannot be answered by SSD, and within-subject designs can be useful in studying the psychometric properties of the instrument. For example, Sutton et al. (2022) evaluated the convergent validity and internal consistency of the PEAK Comprehensive Assessment. Lenoir et al. (2022) evaluated the convergent and age appropriateness of the Children's Psychological Flexibility Questionnaire.

Research studying psychometric properties is critically important to our field. An inaccurate or skewed measurement system is likely to render whatever conclusions made based on the observation or whatever progress captures via data collection invalid. This need is further amplified as the field keeps expressing interests on social issues, such as cultural humility and cultural responsiveness (Kolb et al., 2022; Wright, 2019). The behavior being measured can no longer be limited to simple operant classes. When a measurement system is developed and used to capture a group of behavior constituting a dynamic behavior system, the lack of psychometric studies on these instruments is concerning. Luckily, researchers are beginning to pay more attention during the development process and often seek to obtain feedback to revise their early draft. For example, Gatzunis et al. (2022) developed a Culturally Responsive Supervision Self-Assessment (CRSS) tool for supervisors to self-reflect on their cultural responsiveness during supervisions. Gatzunis et al. gathered feedback on CRSS and collected social validity data for its final form. An extension of their work would involve rigorous psychometric analysis of CRSS. Using within-subject group designs, researchers can extend this work by collecting responses among a large number of supervisors. CRSS's three domains can be verified by calculating the internal consistency among all items within the same domain. Furthermore, factor analysis can be used to examine whether the construct of the instrument correspond

to its theoretical underpinning, with the rationale being that items measuring the same domain should converge while items from different domains should be relatively independent. Test-retest reliability can be examined by administering the CRSS twice with one same group of participants and calculating the correlation between the two outcomes. Convergent validity can be examined by comparing the outcome of CRSS with established measures. Content validity can be examined by synthesizing the input from a group of subject matter experts. And most importantly, researchers can compare whether the self-reported CRSS outcome corresponds to perceptions from the supervisee.

Between-Subject Group Design

In contrast to within-subject group designs, between-subject group designs compare dependent variables collected between two or more groups of participants. Participants are usually categorized into multiple groups based on conditions and demographic characteristics, or are intentionally assigned to different groups which, later on, are exposed to different conditions (e.g., treatment options, waitlists). Here, the analysis primarily focuses on detecting the differences between the groups, which in turn, speaks to the impact independent variables have on dependent variables. Researchers can answer questions such as whether an added component of ACT can increase parental adherence to an online ABA caregiver training program (Yi & Dixon, 2021), comparing the efficacy of relational training procedures on intelligence (May & St. Cyr, 2021), and whether autistic individuals perform differently during skill assessments compared with neurotypical peers (Dixon et al., 2017). Most commonly, these questions are answered using independent sample T-tests or analysis of variance (ANOVA), depending on the number of groups. Figure 2 illustrates the comparison being made in this scenario.

Figure 2.
Illustration for using between-subject group design to compare differences among multiple groups.
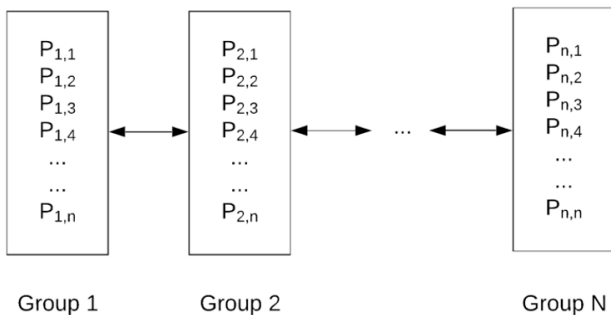
Compared with within-subject group designs, between-subject group designs have several methodological advantages and challenges. In a between-subject group design, one major concern is the inherited differences between the two groups. Suppose the researcher wants to compare two types of ABA intervention on participants skill gain across six months. Had two groups not being equal at the baseline condition, one can make the argument that any observed differences in skill gain might be attributed to the differing foundational learning skills between the two groups, rather than the different intervention. In within-subject group designs, differences among participants are less of a concern as participants are each compared against themselves, thus controlling for this difference. Another disadvantage for between-subject group designs is the requirement of the sample size. A between-subject group design using two groups of participants effectively double the number of participants required. This also leads to its potential insensitivity in detecting the treatment effect. More variability is inherently introduced with a larger number of participants. This variability often leads to smaller power in the statistical method used, decreasing its ability in detecting smaller changes. In other words, with all things being equal, studies using between-subject group designs need to produce a larger effect to avoid type-II errors.

At the same time, between-subject groups design also offer many methodological advantages. It is generally more flexible than within-subject group designs in the statistical methods used (Keppel, 1982), and can avoid sequence effects, which could be detrimental in certain within-subject design studies. When evaluating the differences across multiple treatment conditions using a within-subject research design, the order in which these conditions are exposed to participants might have a significant impact on the dependent variable. At the same time, a previously exposed condition might have carryover effects on the latter condition. This concern is similar to that of reversal designs used in SSDs. When using between-subject group designs, however, sequence effect is less of a concern since each group is independently exposed to its own condition.

Comparing Treatment Outcomes

A common application of between-subject group designs in behavior analytic settings is to compare outcomes of multiple treatment options. This is arguably one of the most important applied questions the field needs to answer: what works and what works better. Although SSDs such as alternating treatment designs and component analysis provide powerful demonstration on the impact of behavior-change procedures with an individual, and allows a direct comparison between different treatment conditions, aggregating such findings at a group level would better address the question on what is likely to work

when similar intervention is applied to a larger group of individuals.

For example Dixon et al. (2021) evaluated the impact of relational training procedures on participants' intelligence. The researcher randomly assigned a group of 17 autistic participants into two groups: a comprehensive ABA (C-ABA) group involving relational training procedures and a traditional ABA (T-ABA) group receiving instructions based on contingency-based learning and generalization, but not content that incorporated derived relational responding. In addition, 11 participants currently on the waiting list for ABA services served as a convenient waitlist control group. All participants' IQ was measured at baseline and after 12 weeks of intervention. A one-way ANOVA was conducted to compare the differences in participant's IQ change score. Results showed a statistically significant difference in participant's IQ change score among the three groups, $F(2,26) = 5.80$, $p = .008$. Post-hoc analysis indicated a statistically significant difference between participants in the C-ABA group and T-ABA group ($p = .042$), and participants in the C-ABA group and the waitlist control ($p = .009$). No statistically significant difference was detected between participants in the T-ABA group and those in the waitlist control ($p = .841$).

Studies like this using between-subject group designs are extremely important as functional utility is at the very core for a pragmatic behaviorist: questions concerning what works and what does not work to produce behavior change. With new intervention approaches and competing treatment options being developed daily, partitioners are ultimately tasked with providing the most appropriate and effective care. To answer the "what works" question requires using between-subject group designs. For example, there has been ongoing debate on optimal parameters of error correction and prompting strategies in discrete trial training. Yet often time, research remained at the individual level, with different studies reporting different outcomes. This speaks to the issue mentioned above, as it is difficult to synthesize outcomes from multiple or even a single study SSD with participants showing varying outcomes. An alternative would be to randomly assign participants into multiple groups, with each group exposed to one study condition. For example, participants in Group A will always receive errorless teaching procedures while participants in Group B will always receive least-to-most prompting. Researchers can then compare outcome measures between the two groups, such as trial to criteria, number of targets mastered, and social validity data. Often time, such analyses can be strengthened by introducing within-participants variables and this research design is called mixed group design.

Mixed Group Designs

Mixed group designs usually involve analyses that are conducted both at the within-subject level and at the between-subject level. They are often used in longitudinal studies involving multiple treatment conditions. Here participants are compared against their peers with different group assignments and against themselves across different timepoints. A wide range of statistical models can be used to detect the effect of group assignment and time, as well as to explore the interaction effect of the two independent variables. General linear models and mixed-ANOVA are often used in studies using group designs. Figure 3 illustrates the comparison being made in this scenario.

Figure 3.

Illustration for using a 2 (Time 1 VS Time 2) x N (Group 1 VS Group 2 VS ... VS Group N) mixed group design to compare differences among multiple groups at both timepoints.
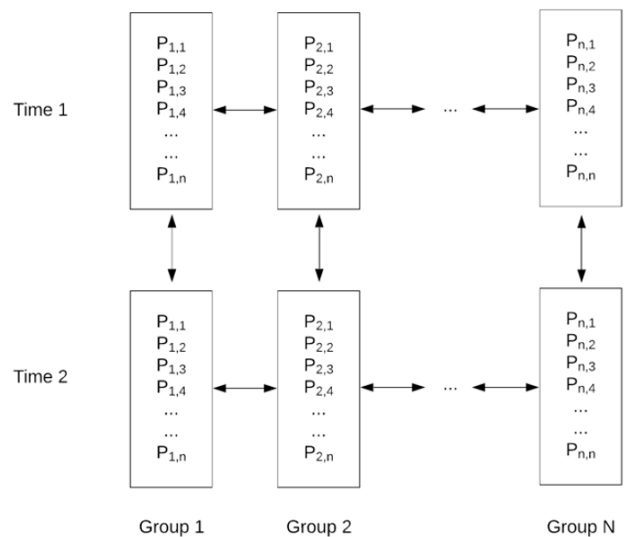
Randomized Controlled Trials

Among the few studies in the field of behavior analysis that utilize mixed group designs, the majority of them fall under the category of randomized controlled trials (RCTs). RCTs are widely accepted as a good standard in conducting casual analysis on the treatment's outcome with agreed upon procedural safeguards in maintain its internal and external validity (e.g., the CONSORT Statement; Schulz et al., 2010). In an RCT, eligible participants are randomly assigned into multiple groups of conditions with dependent measures captured throughout the study at various timepoints. During the analysis, researchers analyze the trajectory of dependent variables within each group, as well as comparing them among all groups at various timepoints.

For example, Sanders et al. (2020) conducted a RCT evaluating the impact of a rapid ABA assessment and treatment protocol among hospitalized autistic

children. Sanders et al. randomly assigned 36 eligible participants into two conditions. Those in the treatment group received a latency-based functional analysis and corresponding function-based behavior reduction plan. Those in the control condition received no active behavioral intervention. Participants clinical functioning, length of hospitalization, and perception from the medical team were evaluated before and after discharge. Results showed preliminary support of incorporating ABA procedures in in-patient hospital settings as those assigned to the treatment group demonstrating more improvements at a statistically significant level. In another example, a re-analysis of the Dixon et al. (2021) study was conducted using mixed-ANOVA to explore the potential interaction effect between group assignments (C-ABA VS T-ABA VS waitlist) and time (baseline BS follow-up; Yi et al., 2021). Results showed a statistically significant main effect of time and a statistically significant interaction effect. RCTs can be one of the most powerful tools for applied researchers, especially in addressing concerns on ABA's overall effectiveness (United States of America Department of Defense, 2021). A longitudinal RCT with multiple waves of data tracking participants overall development will provide strong evidence on the intervention's effectiveness or the lack of.

Conclusions

The exponential rise in the number of behavior analytic professionals signals an extremely bright future for the field of behavior analysis. The growth of the discipline alone is a metric of utility that our science has on saving the world around us (Dixon et al 2018). The time is ripe to couple this rise in popularity of the discipline with a rise of impact and verification that yes indeed – this field matters. We believe that a slight pivot from the reliance of SSDs to a greater adoption of group design methodology could produce great influences for our field to be taken seriously by outsiders. Training programs in behavior analysis should broaden their coursework in research designs to include some of methods noted here within. Clinicians should begin to more carefully examine how to optimize a blend of routine care with research techniques such as regression models, wait-list controls, and environmental comparison studies. Activists within the field wishing to champion a cause, should come forward with data – as such will more quickly alter cure idle hands and silent majorities. Our field has been defined as an enterprise deeply entrenched in pragmatic utilitarianism. Therefore, it is time to make peace with the pragmatic gains that can be accomplished via the occasional adoption of group designs into the field of behavior analysis at a more robust level than historically has occurred.

References

Ackley, M., Subramanian, J. W., Moore, J. W., Litten, S., Lundy, M. P., & Bishop, S. K. (2019). A review of language development protocols for individuals with autism. *Journal of Behavioral Education, 28*(3), 362-388. https://doi.org/10.1007/s10864-019-09327-8

Ator, N. A. (1999). Statistical inference in behavior analysis: Environmental determinants? *The Behavior Analyst, 22*(2), 93-97. https://doi.org/10.1007/BF03391985

Ayllon, T., & Kelly, K. (1972). Effects of reinforcement on standardized test performance. *Journal of Applied Behavior Analysis, 5*(4), 477-484. https://doi.org/10.1901/jaba.1972.5-477

Baires, N. A., Catrone, R., & May, B. K. (2022). On the Importance of listening and intercultural communication for actions against racism. *Behavior Analysis in Practice, 15*(4), 1042-1049. https://doi.org/10.1007/s40617-021-00629-w

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46*(1), 1-21. https://doi.org/10.1002/jaba.30

Becirevic, A., Critchfield, T. S., & Reed, D. D. (2016). On the social acceptability of behavior-analytic terms: Crowdsourced comparisons of lay and technical language. *The Behavior Analyst, 39*(2), 305-317. https://doi.org/10.1007/s40614-016-0067-4

Belisle, J., Stanley, C. R., & Dixon, M. R. (2021). *Research methods for the practicing behavior analyst*. Emergent Press LLC.

Bloom, M., Fischer, J., & Orme, J. G. (2009). *Evaluating practice : Guidelines for the accountable professional* (5th ed.). Allyn and Bacon.

Bolgar, H. (1965). The case study method. In B. B. Wolman (Ed.), *Handbook of clinical psychology*. McGraw-Hill.

Bracken, B. A. (2007). *Bracken School Readiness Assessment Third Edition*. The Psychological Corporation.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled rffectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12*(3), 133-148. https://doi.org/10.1177/1098300709334798

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130*(5), 1136-1145. https://doi.org/10.1542/peds.2012-0243

Branch, M. N. (1999). Statistical inference in behavior analysis: Some things significance testing does and does not do. *The Behavior Analyst, 22*(2), 87-92. https://doi.org/10.1007/BF03391984

Čolić, M., Araiba, S., Lovelace, T. S., & Dababnah, S. (2022). Black caregivers' perspectives on racism in ASD services: Toward culturally responsive ABA practice. *Behavior Analysis in Practice, 15*(4), 1032-1041. https://doi.org/10.1007/s40617-021-00577-5

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*. Pearson Education, Inc.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). Analyzing behavior change: Basic assumptions and strategies. In *Applied behavior analysis third edition*. Pearson Education.

Crosbie, J. (1999). Statistical inference in behavior analysis: Useful friend. *The Behavior Analyst, 22*(2), 105-108. https://doi.org/10.1007/BF03391987

Davison, M. (1999). Statistical inference in behavior analysis: Having my cake and eating it? *The Behavior Analyst, 22*(2), 99-103. https://doi.org/10.1007/BF03391986

DeFelice, K. A., & Diller, J. W. (2019). Intersectional feminism and behavior analysis. *Behavior Analysis in Practice, 12*(4), 831-838. https://doi.org/10.1007/s40617-019-00341-w

Deochand, N., & Costello, M. S. (2022). Building a social justice framework for cultural and linguistic diversity in ABA. *Behavior Analysis in Practice, 15*(3), 893-908. https://doi.org/10.1007/s40617-021-00659-4

DeRosa, N. M., Novak, M. D., Morley, A. J., & Roane, H. S. (2019). Comparing response blocking and response interruption/redirection on levels of motor stereotypy: Effects of data analysis procedures. *Journal of Applied Behavior Analysis, 52*(4), 1021-1033. https://doi.org/10.1002/jaba.644

Dixon, M. R., Belisle, J., Rehfeldt, R. A., & Root, W. B. (2018). Why we are still not acting to save the world: The upward challenge of a post-Skinnerian behavior science. *Perspectives on Behavior Science, 41*(1), 241-267. https://doi.org/10.1007/s40614-018-0162-9

Dixon, M. R., Hayes, S. C., & Belisle, J. (2023). *Acceptance and Commitment Therapy for Behavior Analysts: A practical guide from theory to treatment*. Taylor & Francis LTD.

Dixon, M. R., Paliliunas, D., Barron, B. F., Schmick, A. M., & Stanley, C. R. (2021). Randomized controlled trial evaluation of ABA content on IQ gains in children with autism. *Journal of Behavioral Education, 30*(3), 455-477. https://doi.org/10.1007/s10864-019-09344-7

Dixon, M. R., Paliliunas, D., Weber, J., & Schmick, A. M. (2022). A large-scale naturalistic evaluation of the AIM curriculum in a public-school setting. *Behavior Analysis in Practice, 15*(1), 156-170. https://doi.org/10.1007/s40617-021-00569-5

Dixon, M. R., Rowsey, K. E., Gunnarsson, K. F., Belisle, J., Stanley, C. R., & Daar, J. H. (2017). Normative sample of the PEAK relational training system: Generalization module with comparison to individuals with autism. *Journal of Behavioral Education, 26*(1), 101-122. https://doi.org/10.1007/s10864-016-9261-4

Dowdy, A., Jessel, J., Saini, V., & Peltier, C. (2022). Structured visual analysis of single-case experimental design data: Developments and technological advancements. *Journal of Applied Behavior Analysis*, *55*(2), 451-462. https://doi.org/10.1002/jaba.899

Dubuque, E. M., & Kazemi, E. (2022). An investigation of BCBA exam pass rates as a quality indicator of applied behavior analysis training programs. *Behavior Analysis in Practice, 15*(3), 909-923. https://doi.org/10.1007/s40617-021-00660-x

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin, 72*, 177-182. https://doi.org/10.1037/h0027876

Dunn, K. E., Sigmon, S. C., Thomas, C. S., Heil, S. H., & Higgins, S. T. (2008). Voucher-based contingent reinforcement of smoking abstinence among methadone-maintained patients: A pilot study. *Journal of Applied Behavior Analysis, 41*(4), 527-538. https://doi.org/10.1901/jaba.2008.41-527

Fontenot, B., Uwayo, M., Avendano, S. M., & Ross, D. (2019). A descriptive analysis of applied behavior analysis research with economically disadvantaged children. *Behavior Analysis in Practice, 12*(4), 782-794. https://doi.org/10.1007/s40617-019-00389-8

Gatzunis, K. S., Edwards, K. Y., Rodriguez Diaz, A., Conners, B. M., & Weiss, M. J. (2022). Cultural responsiveness framework in BCBA® supervision. *Behavior Analysis in Practice, 15*(4), 1373-1382. https://doi.org/10.1007/s40617-022-00688-7

Gollub, L. R. (2002). Between the waves: Harvard pigeon lab 1955–1960. *Journal of the Experimental Analysis of Behavior, 77*(3), 319-326. https://doi.org/10.1901/jeab.2002.77-319

Gottman, J. M., McFall, R. M., & Barnett, J. T. (1969). Design and analysis of research using time series. *Psychological Bulletin, 72*, 299-306. https://doi.org/10.1037/h0028021

Habib, R., & Dixon, M. R. (2010). Neurobehavioral evidence for the "near-miss" effect in pathological gamblers. *Journal of the Experimental Analysis of Behavior, 93*(3), 313-328. https://doi.org/10.1901/jeab.2010.93-313

Heil, S. H., Johnson, M. W., Higgins, S. T., & Bickel, W. K. (2006). Delay discounting in currently using and currently abstinent cocaine-dependent outpatients and non-drug-using matched controls. *Addictive Behaviors, 31*(7), 1290-1294. https://doi.org/10.1016/j.addbeh.2005.09.005

Higgins, S. T., Budney, A. J., Bickel, W. K., Foerg, F. E., Donham, R., & Badger, G. J. (1994). Incentives improve outcome in outpatient behavioral treatment of cocaine dependence. *Archives of general psychiatry, 51*(7), 568-576. https://doi.org/10.1001/archpsyc.1994.03950070060011

Higgins, S. T., Delaney, D. D., Budney, A. J., Bickel, W. K., Hughes, J. R., Foerg, F., & Fenwick, J. W. (1991). A behavioral approach to achieving initial cocaine abstinence. *American Journal of Psychiatry, 148*(9), 1218-1224. https://doi.org/10.1176/ajp.148.9.1218

Higgins, S. T., Kurti, A. N., & Davis, D. R. (2019). Coucher-based contingency management is efficacious but underutilized in treating addictions. *Perspectives on Behavior Science, 42*(3), 501-524. https://doi.org/10.1007/s40614-019-00216-z

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior Analysis in Practice, 8*(1), 80-85. https://doi.org/10.1007/s40617-015-0045-4

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*(3), 133-144. https://doi.org/10.1177/1098300709332067

Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.

Jaramillo, C., & Nohelty, K. (2022). Guidance for behavior analysts in addressing racial implicit bias. *Behavior Analysis in Practice, 15*(4), 1170-1183. https://doi.org/10.1007/s40617-021-00631-2

Jang, J., Dixon, D. R., Tarbox, J., Granpeesheh, D., Kornack, J., & de Nocker, Y. (2012). Randomized trial of an eLearning program for training family members of children with autism in the principles and procedures of applied behavior analysis. *Research in Autism Spectrum Disorders*, *6*(2), 852-856.

Joslyn, P. R., Donaldson, J. M., Austin, J. L., & Vollmer, T. R. (2019). The Good Behavior Game: A brief review. *Journal of Applied Behavior Analysis, 52*(3), 811-815. https://doi.org/10.1002/jaba.572

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Prentice-Hall.

Kirkham, P. (2017). "The line between intervention and abuse" – autism and applied behaviour analysis. *History of the Human Sciences, 30*(2), 107-126. https://doi.org/10.1177/0952695117702571

Koegel, L. K., Koegel, R. L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders, 27*(3), 233-243. https://doi.org/10.1023/A:1025894213424

Kolb, R. L., Robers, A. C., Brown, C., & McComas, J. J. (2022). Beyond cultural responsivity: Applied Behavior Analysis through a lens of cultural humility. In *Handbook of special education research* (Vol. I, pp. 144-157). Routledge.

Kornack, J., Cernius, A., & Persicke, A. (2019). The diversity Is in the details: Unintentional language discrimination in the practice of applied behavior analysis. *Behavior Analysis in Practice, 12*(4), 879-886. https://doi.org/10.1007/s40617-019-00377-y

Kupferstein, H. (2018). Evidence of increased PTSD symptoms in autistics exposed to applied behavior analysis. *Advances in Autism, 4*(1), 19-29. https://doi.org/10.1108/AIA-08-2017-0016

Kurtz, P. F., Fodstad, J. C., Huete, J. M., & Hagopian, L. P. (2013). Caregiver- and staff-conducted functional analysis outcomes: A summary of 52 cases. *Journal of Applied Behavior Analysis, 46*(4), 738-749. https://doi.org/10.1002/jaba.87

Kyonka, E. G. E., Mitchell, S. H., & Bizo, L. A. (2019). Beyond inference by eye: Statistical and graphing practices in JEAB, 1992-2017. *Journal of the Experimental Analysis of Behavior, 111*(2), 155-165. https://doi.org/10.1002/jeab.509

Ledgerwood, D. M., Alessi, S. M., Hanson, T., Godley, M. D., & Petry, N. M. (2008). Contingency management for attendance to group substance abuse treatment administered by clinicians in community clinics. *Journal of Applied Behavior Analysis, 41*(4), 517-526. https://doi.org/10.1901/jaba.2008.41-517

Lenoir, C., Hinman, J. M., Yi, Z., & Dixon, M. R. (2022). Further examination of the Children's Psychological Flexibility Questionnaire (CPFQ): Convergent validity and age appropriateness. *Advances in Neurodevelopmental Disorders, 6*(2), 224-233. https://doi.org/10.1007/s41252-022-00259-5

Levy, S., Siebold, A., Vaidya, J., Truchon, M.-M., Dettmering, J., & Mittelman, C. (2022). A look in the mirror: How the field of behavior analysis can become anti-racist. *Behavior Analysis in Practice, 15*(4), 1112-1125. https://doi.org/10.1007/s40617-021-00630-3

Li, A., Gravina, N., Pritchard, J. K., & Poling, A. (2019). The gender pay gap for behavior analysis faculty. *Behavior Analysis in Practice, 12*(4), 743-746. https://doi.org/10.1007/s40617-019-00347-4

Lovelace, T. S., Comis, M. P., Tabb, J. M., & Oshokoya, O. E. (2022). Missing from the narrative: A seven-decade scoping review of the inclusion of Black autistic women and girls in autism research. *Behavior Analysis in Practice, 15*(4), 1093-1105. https://doi.org/10.1007/s40617-021-00654-9

Mathur, S. K., & Rodriguez, K. A. (2022). Cultural responsiveness curriculum for behavior analysts: A meaningful step toward social justice. *Behavior Analysis in Practice, 15*(4), 1023-1031. https://doi.org/10.1007/s40617-021-00579-3

Matson, J. L., & Konst, M. J. (2014). Early intervention for autism: Who provides treatment and in what settings. *Research in Autism Spectrum Disorders, 8*(11), 1585-1590. https://doi.org/10.1016/j.rasd.2014.08.007

May, B. K., & St. Cyr, J. (2021). The impact of the PEAK curriculum on standardized measures of intelligence: A systems level randomized control trial. *Advances in Neurodevelopmental Disorders, 5*, 245-255. https://doi.org/10.1007/s41252-021-00199-6

Mayer, G., Sulzer Azaroff, B., & Wallace, M. D. (2019). *Behavior analysis for lasting change*. Sloan Publishing.

McGill, O., & Robinson, A. (2021). "Recalling hidden harms": Autistic experiences of childhood Applied Behavioural Analysis (ABA). *Advances in Autism, 7*(4), 269-282. https://doi.org/10.1108/AIA-04-2020-0025

Morris, C., Goetz, D. B., & Gabriele-Black, K. (2021). The treatment of LGBTQ+ individuals in behavior-analytic publications: A historical review. *Behavior Analysis in Practice, 14*(4), 1179-1190. https://doi.org/10.1007/s40617-020-00546-4

Nighbor, T. D., Zvorsky, I., Kurti, A. N., Skelly, J. M., Bickel, W. K., Reed, D. D., Naudé, G. P., & Higgins, S. T. (2019). Examining interrelationships between the Cigarette Purchase Task and delay discounting among pregnant women. *Journal of the Experimental Analysis of Behavior, 111*(3), 405-415. https://doi.org/10.1002/jeab.499

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Perone, M. (1999). Statistical inference in behavior analysis: Experimental control is better. *The Behavior Analyst, 22*(2), 109-116. https://doi.org/10.1007/BF03391988

Plantiveau, C., Dounavi, K., & Virués-Ortega, J. (2018). High levels of burnout among early-career board-certified behavior analysts with low collegial support in the work environment. *European Journal of Behavior Analysis, 19*(2), 195-207. https://doi.org/10.1080/15021149.2018.1438339

Pritchett, M., Ala'i-Rosales, S., Cruz, A. R., & Cihon, T. M. (2022). Social justice is the spirit and aim of an applied science of human behavior: Moving from colonial to participatory research practices. *Behavior Analysis in Practice, 15*(4), 1074-1092. https://doi.org/10.1007/s40617-021-00591-7

Sanders, K., Staubitz, J., Juárez, A. P., Marler, S., Browning, W., McDonnell, E., Altstein, L., Macklin, E. A., & Warren, Z. (2020). Addressing challenging behavior during hospitalizations for children with autism: A pilot applied behavior analysis randomized controlled trial. *Autism Research, 13*(7), 1072-1078.

Schulz, K. F., Altman, D. G., Moher, D., & the, C. G. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Trials, 11*(1), 32. https://doi.org/10.1186/1745-6215-11-32

Sevon, M. A. (2022). Schooling while Black: Analyzing the racial school discipline crisis for behavior analyst. *Behavior Analysis in Practice, 15*(4), 1247-1253. https://doi.org/10.1007/s40617-022-00695-8

Shea, P., Johnson, P., & Togade, D. (2022). Using relational frame theory to examine racial prejudice: A tool for educators and an appeal for future research. *Behavior Analysis in Practice.* https://doi.org/10.1007/s40617-022-00767-9

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* Basic Books.

Silverman, K., Wong, C. J., Needham, M., Diemer, K. N., Knealing, T., Crone-Todd, D., ... & Kolodner, K. (2007). A randomized trial of employment-based reinforcement of cocaine abstinence in injection drug users. *Journal of applied behavior analysis, 40*(3), 387-410.

Skinner, B. F. (1956). A case history in scientific method. *American psychologist, 11,* 221-233. https://doi.org/10.1037/h0047662

Sutton, A., Pikula, A., Yi, Z., & Dixon, M. R. (2022). Evaluating the convergent validity of the PEAK Comprehensive Assessment (PCA): Intelligence, behavior challenges, and autism symptom severity. *Journal of Developmental and Physical Disabilities, 34*(4), 549-570. https://doi.org/10.1007/s10882-021-09814-9

Thrailkill, E. A., DeSarno, M., & Higgins, S. T. (2022). Intersections between environmental reward availability, loss aversion, and delay discounting as potential risk factors for cigarette smoking and other substance use. *Preventive Medicine, 165,* 107270. https://doi.org/10.1016/j.ypmed.2022.107270

Twohig, M. P., Shoenberger, D., & Hayes, S. C. (2007). A preliminary investigation of acceptance and commitment therapy as a treatment for marijuana dependence in adults. *Journal of Applied Behavior Analysis, 40*(4), 619-632. https://doi.org/10.1901/jaba.2007.619-632

United States of America Department of Defense. (2021). *The Department of Defence comprehensive autism care demonstration auunal report 2021.* https://health.mil/Reference-Center/Reports/2021/12/03/Annual-Report-on-Autism-Care-Demonstration-Program-for-FY-21

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The Impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine, 166*(2), 149-156. https://doi.org/10.1001/archpediatrics.2011.755

Wang, Y., Kang, S., Ramirez, J., & Tarbox, J. (2019). Multilingual diversity in the field of applied behavior analysis and autism: A brief review and discussion of future directions. *Behavior Analysis in Practice, 12*(4), 795-804. https://doi.org/10.1007/s40617-019-00382-1

World Health Organization. (2020). *Doing what matters in time of stress: An illustrated guide.* https://www.who.int/publications/i/item/9789240003927

Wright, P. I. (2019). Cultural humility in the practice of applied behavior analysis. *Behavior Analysis in Practice, 12*(4), 805-809. https://doi.org/10.1007/s40617-019-00343-8

Yi, Z., & Dixon, M. R. (2021). Developing and enhancing adherence to a telehealth ABA parent training curriculum for caregivers of children with autism. *Behavior Analysis in Practice, 14*(1), 58-74. https://doi.org/10.1007/s40617-020-00464-5

Yi, Z., Koenig, J., & Dixon, M. R. (2022). Comparing low dosages of ABA treatment on children's treatment gains and school readiness. *Advances in Neurodevelopmental Disorders*. https://doi.org/10.1007/s41252-022-00296-0

Yi, Z., Schreiber, J. B., Paliliunas, D., Barron, B. F., & Dixon, M. R. (2021). P < .05 is in the eye of the beholder: A response to Beaujean and Farmer (2020). *Journal of Behavioral Education, 30*, 489-511. https://doi.org/10.1007/s10864-021-09435-4

Yoon, J. H., Higgins, S. T., Heil, S. H., Sugarbaker, R. J., Thomas, C. S., & Badger, G. J. (2007). Delay discounting predicts postpartum relapse to cigarette smoking among pregnant women. *Experimental and Clinical Psychopharmacology, 15*, 176-186. https://doi.org/10.1037/1064-1297.15.2.186

Zarcone, J., Brodhead, M., & Tarbox, J. (2019). Beyond a call to action: An introduction to the special issue on diversity and equity in the practice of behavior analysis. *Behavior Analysis in Practice, 12*(4), 741-742. https://doi.org/10.1007/s40617-019-00390-1