12-2018

Randomization in Multiple Baseline Designs

Lissa Bonnstetter
lissabonnstetter@gmail.com

Randomization in Multiple Baseline Designs

by

Lissa N. Bonnstetter

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Applied Behavior Analysis

December, 2018

Thesis Committee:
Benjamin Witts, Chairperson
Kimberly Schulze
Michele Traub

Abstract

In empirical work, randomization is seen as an effective means of controlling for potential biases. This thesis reviews the *Journal of Applied Behavior Analysis* (JABA) multiple baseline design articles from 1968-2017 to identify if and how articles randomize their participants, phases, trials, settings, or other factors. In this review, we suggest the requirement of randomization in multiple baseline design articles that is determined prior to the start of baseline to improve internal validity.

Table of Contents

Chapter 1: Introduction

While great strides have been made in applied behavior-analytic research, continued enhancements to research designs and methods must be made to develop more credible research to further the field. In research, one way to enhance credibility is through the systematic elimination of bias.

The Oxford English Dictionary defines bias as "something that causes inclination for or against something" ("bias", 1989). In research, bias occurs when a researcher knowingly or unknowingly sets research parameters such that a particular outcome is more likely given those adjustments. In other scientific fields, a common protection against bias is through the use of randomization; however, while randomization is the core subject of this paper, I will first begin discussing current research methods and potential bias that can affect their results.

Bias and Research Methods

In reporting research methods, presenting the entirety of the procedures is essential. The reason for complete documentation is to permit other researchers to accurately replicate others' work. Details like the participants, materials, designs, and procedures used should be explained, including details about who was chosen to participate, why they were chosen, basic characteristics of participants, assignment to conditions, materials involved, the type of design used, and the procedure technicalities (see, e.g., VandenBos, 2010, section 2.06).

Careful planning of research methods increases the likelihood of sound results. Superior research methodologies consistently incorporate two qualities: the methods are appropriate for the questions being asked and the methods aim to systematically eliminate bias and confounding variables (Viera & Bangdiwala, 2007). Both bias and confounding variables affect the soundness of a study, but they are different. For example, in a between-subjects design for an attending-to-task intervention, one student is observed in the morning and one is observed in the afternoon. The participant in the morning is showing that the intervention is ineffective while the participant in the afternoon shows the intervention works. However, this difference might have nothing to do with the intervention. The morning participant may be from a low-income family and does not get an adequate breakfast. Hunger can have an effect on a person's behaviors, like attending, making it a confounding variable because the hunger affects the dependent variable, but it has not been accounted for in the research methods (Kleinman et al., 1998). If instead we knew about this confounding variable but still arranged it such that the low-income participant was run in the morning, this would be a bias.

It is important to recognize confounding variables in research, but as this paper is concentrated on bias we are going to focus on the different kinds of bias and their effects on research. Research is filled with potential sources of bias that can be difficult to avoid, particularly if the researcher is unaware of the possible bias that can alter the results of the research. Bias can be avoided if researchers take the appropriate measures in the method of their research to reduce the risks. Bias can occur in data collection, analysis, or interpretation, or during the publication phase (Podsakoff, Mackenzie, Lee, & Podsakoff, 2003)

Bias in data analysis can occur when data are fabricated, data are tested in a way that manipulates the results, and by rejecting data that do not support the hypothesis of the study (Podsakoff et al., 2003). For example, if two participants responded positively to an intervention, while the third and forth did not, removing those participant's data would create a bias in favor of the intervention. Bias in data interpretation can develop by inaccurately displaying data in a way that is in favor of the researchers hypothesis.

Publication bias comes from the research that journals choose to accept. If scientific journals only choose to accept research with results that are positive or in favor or a particular intervention, then the journal would not be accurately representing the research on those interventions, thus creating a bias (Shadish, Zelinsky, Vevea, & Kratochwill, 2016; Sham & Smith, 2014).

Bias in data collection can come from unsuccessfully collecting a sample of people to accurately represent a population. Inaccurately representing a population can occur through selection bias, which happens when predetermined randomization of participants does not occur. Volunteer bias can also occur when a sample of a population is collected in a way that will cause results that do not represent the population. For example, if participants are self-selected, when researchers ask for volunteers to share their opinions on a specific topic, the general population is not accurately represented. Other examples of bias in data collection are survivor bias and misclassification bias (Podsakoff et al., 2003).

In behavior analytic research, a protection against bias in data collection is interobserver agreement, which (arguably) measures the accuracy of data collection by using multiple independent observers, then comparing their agreement data (Watkins & Pacheco, 2000). While interobserver agreement protects against the inaccurate collection of data, it does not protects from bias that are developed before the experiment begins, such as design choices.

Current research standards in a variety of scientific fields aim to limit bias through research design. Behavior analysts tend to report single case studies more frequently than other research designs. Single case designs are considered the gold standard of research design over group designs because they collect in depth detail, are able to experiment on rare cases or specific populations, and often aim to protect from bias in research.

Chapter 2: Single Case Designs

Single case designs have assisted in the growth of behavior analytic research in literature on autism, education, and school based interventions, to list a few areas, by using individuals as their own control to demonstrate the effects of an intervention on a person's behavior. Single subject designs are able to accommodate small samples on specific conditions, allowing participants to be their own control, and the flexibility to alter and individualize conditions as needed (Smidt, Durham, Duffy, & Frank, 1996). The most commonly used single subject designs are comparison designs, reversal designs, and multiple baseline designs (Miller, 2006).

As the goal of single case design research is to demonstrate control over the dependent variable, a common strategy adopted in many studies is to turn the independent variable on and off, so to speak, in a withdrawal fashion; similarly, the relation between the independent variable and dependent variables can be reversed in a reversal design, or parametric variations can be explored. For example, through a withdrawal design a researcher could display a functional relation between positive reinforcement and on-task behaviors by systematically removing the intervention to observe a decrease in on-task behaviors. If on-task behaviors increase when the intervention begins again, then there is evidence of a functional relation.

While these and other designs provide evidence of functional relations, their use becomes inappropriate or unethical when the changes in behavior cannot be reversed (e.g., when you can't unlearn something) or when the behavior under question is dangerous (see Barlow, Nock, & Hersen, 2009). There might be times where stakeholders do not consent to the removal of a treatment once it has been enacted if the behavior being reduced is considered aversive. At other times, carryover effects from the intervention would make withdrawal or reversal designs difficult to demonstrate control. As an alternative, the best methodological option could be multiple baseline designs, as they permit demonstrations of functional relations without the requirement of removing the intervention. Two other alternatives to withdrawal or reversal designs are changing criterion designs or alternating treatment designs, but multiple baseline designs are much more commonly used in current behavior analytic research (Barlow et al., 2009).[1]

Multiple baseline designs demonstrate experimenter control through the use of staggered introductions of the intervention (Barlow et al., 2009; Christ, 2007; Hall, Cristler, Cranston, & Tucker, 1970). Each baseline and introduction is considered an A-B design, where the addition of a treatment phase for one participant should be displayed in a change or rate for one participant, while the other participants should not change (Barlow et al., 2009). Typically, multiple baseline designs start each behavior, setting, or participant in baseline until there is stability in the data or until a predetermined number of baseline data points have been collected. Once the data is stable, the intervention is introduced in temporal sequence. Instead of using withdrawal or reversal to demonstrate a functional relation, multiple baseline designs use temporal sequencing to rule out the likelihood that history, maturation, or other variable may be the cause of the change in behavior (Barlow et al., 2009). If behavior changes in only the baseline in which treatment is introduced, then it is suggested that the participants, behaviors, or settings are independent of each other. If each behavior changes subsequently only when introduced to the intervention, then a functional relation has been demonstrated. By using

[1] Throughout this paper, we treat all research designs as independent, although research designs can be combined.

multiple baselines to demonstrate the control of an intervention, the soundness of the study is strengthened. It is recommended to use at least three baselines, but the strength of study increases with the number of baselines that are used (Barlow & Hersen, 1973; Kazdin & Kopel, 1975; Wolf & Risley, 1971).

Multiple baseline designs across participants allow the observer to view the effects of the treatment for different participants. By moving each participant into the intervention at staggered times, the researcher is able to rule out alternative explanations related to timing. If each participant's behaviors are stable in baseline, but changes with the introduction of the intervention, we can be confident that the change in behavior was due to the introduction of the intervention. For example, Piazza and Fisher (1991) measured the total duration of sleep in a multiple baseline across participants design. In the study, a response cost intervention was implemented for four participants. Control was demonstrated by an increased duration of sleep for all participants only following the introduction of the intervention.

In multiple baseline designs across behaviors, one participant is studied, and the intervention is applied to different behaviors. For example, an experiment by Bornstein, Bellack, and Hersen (1977) measured the behaviors of four students in a multiple baseline across behaviors study. The experiment examined the effects of social skills training on four behaviors: ratio of eye contact to speech duration, loudness of speech, number of responses, and assertiveness. The results of the study demonstrated that the intervention had an effect on all the measured behaviors for each participant after the intervention was introduced and through the intervention phase.

Multiple baseline designs across settings are implemented to assess the effect of an interventions across different environment. In a multiple baseline across settings article, the productivity of staff at three restaurants was measured (George, & Hopkins, 1989). After an intervention was implemented, there was an increase in productivity. By measuring the intervention across three separate settings at staggered times, the intervention is strengthens by intrasubject replication.

As described above, an important feature of multiple baseline design is the independence of baselines; however, if interdependence is expected or required, the addition of independent baselines can control for the expectation of interdependence. If interdependence is expected, the experimenter should be able to manipulate multiple behaviors or people by the application of an intervention. For example, Gresham and Gresham (1982) compared different group-oriented contingency systems in a modified reversal design to decrease disruptive behaviors in a classroom setting. The study found that the interdependent contingencies were effective in reducing a variety of disruptive behaviors, while an addition of an independent baseline was able to demonstrate the functional relation between the intervention and the change in behaviors.

When the design of research is more sound (i.e., reliable and accurate), the readers can be more confident in the results (Pannucci & Wilkins 2010). Bias, which occurs when errors are introduced into testing by knowingly or unknowingly encouraging one outcome to occur, threatens the soundness of a study (Pannucci & Wilkins 2010). Aspects of single case designs are intended to prevent bias from seeping into the methods of research.

Bias and Multiple Baseline Designs

Single case designs have been preferred by behavior analysts because the functional relation between the intervention and the dependent variable is displayed through the changes in behavior over separate behaviors, settings, or participants (Barlow & Hersen, 1984). Additionally, single case designs attempt to control for the effect of extraneous events by showing that the change in behavior comes only when the intervention is introduced. Simply by comparing an individual's baseline behavior to their behavior in treatment, individual differences (between participants) are ruled out and functional relations can be identified (Miller, 2006). By using multiple subjects, the generality of the treatment outcome can be assessed; meaning that if an intervention demonstrates effective behavior change over multiple participants then the intervention is likely responsible for creating the behavior change (Miller, 2006; Ryan, Ormond, Imwold, & Rotunda, 2002). Successful replication of results across participants/behaviors/ settings also confirms the effectiveness and validity of the study (Miller, 2006). Single subject designs are able to accommodate small samples on specific conditions, allowing participants to be their own control, and the flexibility to alter and individualize conditions as needed (Smidt et al., 1996).

As explained above, elements of single case designs protect the validity and accuracy of the research in ways that other research does not, but it does not protect from biases created by the researcher. For example, interobserver agreement simply suggests the observers probably saw the same thing, which is hardly useful when the research methods have already introduced bias in other ways.

There are other ways, perhaps more subtle, that bias can seep into experimental design. It has been documented that the experimenter's interpretation biases can seep into experimental design through their expectations of their research. For example, if a researcher purposely assigns sequencing of participants to better produce the desired results, they have developed a bias. In this example, the use of randomization on the sequencing of participants could remove the bias.

Research using group design has an advantage over single case designs in their historical attempts through the mechanism of randomization. Randomization has the potential to strengthen the internal validity, or the soundness of the experiment, and thus the external validity (e.g., Kratochwill & Levin, 2010)

Randomization

A thorough study, one with external and internal validity, has the potential to be meaningful for the larger community, but a poorly conducted study could never be of use to the larger community because it has no utility. Randomization is a protective factor for the soundness of research, where bias compromises the soundness of the study. Randomization assists in making research more sound and consists of using systems of randomization—when the researcher invites chance (e.g., electronic means, rolling dice, drawing elements from a hat) to arrange variables of interest in an experiment rather than having the researcher determine their arrangement.

Kratochwill and Levin (2010) discussed the relation between randomized controlled trials (RCTs), which aim to minimize threats to an experiment's internal validity through randomization of participants, and applied behavior analytic research. In traditional group design research in other fields of science, the gold standard has been the RCT. The process of RCT could be a great benefit to behavior analytic research. As Kratochwill and Levin argued, RCT's typically randomize across groups, but it is possible to arrange a single case design in a way that will satisfy the requirements of a RCT, which include the randomization of recipients, time, settings, and outcome variables. The application of RCTs in applied behavior analysis has been limited, as RCT often require groups of participants, aggregated data, and the use of inferential statistics, all of which are not common in single case design research (Kratochwill & Levin, 2010). There are several admirable features of RCTs that make their strategies of interest to behavioral research. More significantly, there is nothing essential in single case designs that preclude them from the addition RCT features. RCTs help to eliminate selection bias, which is the distortions of conclusions made from the results of an intervention when the selection of individuals is not randomized properly (Kratochwill & Levin, 2010). Selection bias can include determining when a participant will start or end an intervention or reporting only the most significant results to better support the desired conclusion.

There are several important ways that randomization can be included in multiple baseline design research, including randomization of treatment components, which also apply to other variables (e.g., assignment of trials, tasks, conditions). Of particular importance to the MBD, and similar to the rationale provided in RCTs, is the randomization of participants (or settings, or behaviors) across baselines. Why randomization of assignment is important is not difficult to see; imagine a MBD across participants in which an intervention is aimed at improving some academic task. The researcher in this hypothetical example has three participants; one who learns quickly, one who learns slowly, and one who is somewhat fast to learn, but not as fast as most.

Assigning the fast learner to the shortest baseline would reduce the possibility of the participant improving their performance in the absence of the intervention, perhaps figuring out the task simply by exposure. Likewise, assigning the slowest learner to the longest baseline might not suffer this concern, as the likelihood of improving performance for that participant is relatively lower.

An additional important and unique randomization component in MBD is the randomization of baseline lengths when baseline lengths are predetermined. For example, if a study employs five baselines and it is expected that all participants must have at least three opportunities at baseline, then a series of baseline lengths from three to X can be selected. For our purposes, let us say that between 3 and 12 baseline sessions are selected, as we do not want to extend past 12 sessions for our intervention assessment. In this case, we would randomly select five session lengths from our 10 session lengths. Again we can see why randomization might be important by imagining an intervention in which it is possible to improve performance in baseline; a researcher might be motivated to keep baseline as short as possible, and randomization could help prevent that (assuming longer baseline lengths are included in the selection process).

A requirement of multiple baseline designs should be that randomization of the variable of concern is determined prior to the start of baseline and should be mentioned in the methods. Kratochwill and Levin (2010) explained that in the majority of multiple baseline designs, there is no discussion of the assignment or order of the intervention regarding the timing of intervention or order of participants. To have more valid data, readers need to know that randomization of orders of tasks, participants, etc., is determined prior to the start of baseline. If randomization is not determined prior to the start of baseline, there is a risk that the order of certain variables in the experiment may have been determined to better demonstrate effectiveness of an intervention. In an effort to see how well Kratochwil and Levin’s recommendations for randomization are represented in our literature, we completed an analysis of literature.

Chapter 3: Method

Materials

For this research, every simple multiple baseline design article in the *Journal of Applied Behavior Analysis* (JABA) from 1968-2017 was subject to analysis in this paper. Through the examination of each article in JABA, we isolated each simple multiple baseline study and outlined each article by documenting the authors, the type of multiple baseline design used, the category of randomization used, and the quotation from the article that mentioned randomization, if applicable.

The criteria for the simple multiple baseline design articles chosen was that the articles did not involve design variations of multiple baseline designs. For example, multiple baseline designs with embedded combinations of reversal and/or withdrawal designs were not included in this research. Any articles that were not simple or experimental were removed.

Procedure

JABA has been identified as the flagship of behavior analytic research and while the articles in JABA are not representative of all multiple baseline design articles, we argued that it would give us a fair representation of current practices.

All multiple baseline design articles were extracted from every issue of JABA up through 2017. After retaining the articles that met the criteria, each article was opened as a searchable pdf. There were two ways we searched the articles; when digitized articles were available we used the search function, but for older research articles, we read and identified keywords by hand. Any mention of randomization, baseline assignment, length, and other relevant baseline requirements was recorded. The results were examined to identify the number of multiple baseline design articles that included predetermined randomization. The results were also examined to determine if predetermined randomization could be applied to increase the soundness of the study.

Inter-rater reliability was determined by having a second reader identify any randomization, baseline assignment, length, and other relevant baseline requirements for articles in JABA from 2010 and 2015. An a priori decision was made to discontinue inter-rater data collection is 100% agreement with the primary research was achieved. The data were compared to the original researcher and found the data collected matched to 100%. A third reader was also recruited, but missed quotations and keywords the original researcher and second reader found. The third observer did not find anything that was missed by the researcher, which was an error on the part of the reader, so the third observer’s data were removed.

Chapter 4: Results and Discussion

We identified 375 articles from JABA that met the guidelines for simple multiple baseline design studies. Of those articles, 265 did not mention any type of randomization in their research methods, displayed in Figures 1 and 2. Of the 110 multiple baseline articles that contained randomization, 79 randomized tasks, 2 randomized settings, and 9 randomized participants. Furthermore, only three articles reviewed mentioned predetermined randomization. Twenty-four other articles mentioned randomization in other categories, predominately in IOA.

We found that majority of articles did not discuss randomization in any section of the methods. When randomization was mentioned, it was not indicated at what point the randomization occurred or how the condition was randomized. For instance, Scherrer and Wilder (2008) randomized participants, but it is unclear at what point in the research process the participants were randomized. Potentially, there may be bias in how the participants were chosen or the order they were placed in, which was prior to the randomization that was used in the study.

Peters and Thompson (2015) examined the effects of behavior skills training on participants' ability to ask a question or change the topic when a person's body posture indicated they were uninterested. In this study, two variables were randomized prior to the start of the experiment: how the experimenter behaved in each trial (interested or uninterested), and the consequence for responding correctly. Peters and Thompson used predetermined randomization, which reduces bias that can occur during data collection; however, the sequencing order of the participants who moved into the intervention phase was not randomized. This allows the experimenter to determine when each participant moves into the intervention phase, which can be arranged to display results that are in favor of the desired result.

Spellman, Whiting, and Dixon (2015) observed the effects of a behavior skills training procedure on gambling skills (card counting accuracy and chips won or lost) in a multiple baseline across participants design. Once stability was observed in the baseline phase, Spellman et al. randomly selected a participant to advance to the training phase. While randomly choosing a participant to move from baseline to the intervention was not predetermined, the order of participants moving to intervention was not up to the experimenter, making it less likely that bias from the experimenter could alter the data based on the sequencing of participants.

In a multiple baseline design across participant groups, Dixon and Vargo (2017) evaluated the effects of behavior skills training on responding during lockdown drills. Prior to the study, participants were randomly separated into three groups. In this experiment, using predetermined randomization eliminated the potential of the experimenter assigning participants into groups to create favorable results; however, it is unclear how the participants were chosen initially.

Wallace, Doney, Mintz-Resudek, and Tarbox (2004) examined instructional workshops on training educators to conduct functional analyses. The study did not include randomization in any part their methods; however, it was the only study we found that mentioned the lack of randomization as a limitation. In their discussion, Wallace et al. explained that the lack of randomization of the selection of participants may have produced a group of highly motivated participants which is not representative of typical teachers and school psychologists, creating a selection bias.

Based on the analysis of research in this paper, it is clear that the field of behavior analysis does not use randomization to protect against bias from the researcher, nor does it consider the lack of randomization to be a limitation of research. Perhaps this is due to limited understanding in the purpose of predetermined randomization in behavior analysis. It is apparent that researchers use randomization, but not it a way that is most effective for protecting against bias from the researcher. Adding randomization to other variables through the methods of research is great, but it does not mean that the research is free of bias, because all potential sources of bias should be accounted for.

As behavior analytic research continues to improve, the development of sound research should be a priority. The publication bias in behavior analytic research urges researchers to develop articles with favorable results (Sham & Smith, 2014). Because of this, researchers may knowingly or unknowingly be arranging variables in ways that will produce favorable results to increase their chances of developing published research. While the publication bias is an issue that needs to be resolved, using predetermined randomization can be a protection against the researcher bias that may be a result of the publication bias. For example, a researcher might arrange participants in a way that will demonstrate favorable results based on the participants' skills to have a better chance of being published. If the participants are randomized prior to the start of the study, the researcher is not able to create a bias in the arrangement of the participants.

Through this research, I am arguing for the requirement of predetermined randomization in multiple baseline design articles to be clarified for publication. As observed in the data collected through this research, there are some protections against bias through randomization in the methodology of research, but it is not a requirement of published behavior analytic research. A simple solution to this problem is to incorporate predetermined randomization either through computer randomization or manually using someone who is unrelated to the study and who is unaware of the potential results of the study. Adding this step to the research process will easily add valuable soundness to the study without taking much time away from the research. Much like IOA is a requirement of reliable research, identifying what was randomized and at what point in the research the randomization was applied should also be a requirement of research. If randomization is not used or is not determined prior to the start of the research, it should be listed as a limitation since bias from the researcher could have occurred.

References

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs. *Archives of General Psychiatry, 29*(3). doi:10.1001/archpsyc.1973.04200030017003

Barlow, D. H., & Hersen, M. (1984). Single case experimental designs: Strategies for studying behavior change. London: Pergamon Press.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Pearson/Allyn and Bacon.

"Bias." (1989) In *Oxford English online dictionary*. Retrieved October 11, 2018, from https://en.oxforddictionaries.com/definition/bias.

Bornstein, M. R., Bellack, A. S., & Hersen, M. (1977). Social-skills training for unassertive children: A multiple-baseline analysis. *Journal of Applied Behavior Analysis,10*(2), 183-195. doi:10.1901/jaba.1977.10-183

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools,44*(5), 451-459. doi:10.1002/pits.20237

Dickson, M. J., & Vargo, K. K. (2017). Training kindergarten students lockdown drill procedures using behavioral skills training. *Journal of Applied Behavior Analysis,50*(2), 407-412. doi:10.1002/jaba.369

George, J. T., & Hopkins, B. L. (1989). Multiple effects of performance-contingent pay for wait-persons. *Journal of Applied Behavior Analysis,22*(2), 131-141. doi:10.1901/jaba.1989.22-131

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *The Journal of Special Education,16*(1), 101-110. doi:10.1177/002246698201600110

Hall, R. V., Cristler, C., Cranston, S. S., & Tucker, B. (1970). Teachers and parents as researchers using multiple baseline designs. *Journal of Applied Behavior Analysis,3*(4), 247-255. doi:10.1901/jaba.1970.3-247

Kazdin, A. E., & Kopel, S. A. (1975). On resolving ambiguities of the multiple-baseline design: Problems and recommendations. *Behavior Therapy, 6*(5), 601-608. doi:10.1016/s0005-7894(75)80181-x

Kleinman, R. E., Murphy, J. M., Little, M., Pagano, M., Wehler, C. A., Regal, K., & Jellinek, M. S. (1998). Hunger in children in the United States: Potential behavioral and emotional correlates. *Pediatrics,101*(1). doi:10.1542/peds.101.1.e3

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*(2), 124-144. doi:10.1037/a0017736

Miller, L. K. (2006). *Principles of everyday behavior analysis*. Australia: Thomson/Wadsworth.

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and avoiding bias in research. *Plastic and Reconstructive Surgery,126*(2), 619-625. doi:10.1097/prs.0b013e3181de24bc

Peters, L. C., & Thompson, R. H. (2015). Teaching children with autism to respond to conversation partners' interest. *Journal of Applied Behavior Analysis,48*(3), 544-562. doi:10.1002/jaba.235

Piazza, C. C., & Fisher, W. (1991). A faded bedtime with response cost protocol for treatment of multiple sleep problems in children. *Journal of Applied Behavior Analysis,24*(1), 129-140. doi:10.1901/jaba.1991.24-129

Podsakoff, P. M., Mackenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. doi:10.1037/0021-9010.88.5.879

Ryan, S., Ormond, T., Imwold, C., & Rotunda, R. J. (2002). The effects of a public address system on the off-task behavior of elementary physical education students. *Journal of Applied Behavior Analysis, 35*(3), 305-308. doi:10.1901/jaba.2002.35-305

Scherrer, M. D., & Wilder, D. A. (2008). Training to increase safe tray carrying among cocktail servers. *Journal of Applied Behavior Analysis, 41*(1), 131-135. doi:10.1901/jaba.2008.41-131

Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49*(3), 656-673. doi:10.1002/jaba.308

Sham, E., & Smith, T. (2014). Publication bias in studies of an applied behavior-analytic intervention: An initial analysis. *Journal of Applied Behavior Analysis, 47*(3), 663-678. doi:10.1002/jaba.146

Smidt, G., Durham, D., Duffy, L., & Frank, T. (1996). Making a case for single case research. *Journal of Orthopedic & Sports Physical Therapy*. Retrieved October 1, 2018.

Speelman, R. C., Whiting, S. W., & Dixon, M. R. (2015). Using behavioral skills training and video rehearsal to teach blackjack skills. *Journal of Applied Behavior Analysis, 48*(3), 632-642. doi:10.1002/jaba.225

VandenBos, G. R. (Ed). (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Viera, A. J., & Bangdiwala, S. I. (2007). Eliminating bias in randomized controlled trials: Importance of allocation concealment and masking. *Family Medicine*, *39*(2), 132-137. Retrieved May 5, 2018.

Wallace, M. D., Doney, J. K., Mintz-Resudek, C. M., & Tarbox, R. S. (2004). Training educators to implement functional analyses. *Journal of Applied Behavior Analysis, 37*(1), 89-92. doi:10.1901/jaba.2004.37-89

Watkins, M., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education, 10*(4). Retrieved October 31, 2018.

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

Appendix

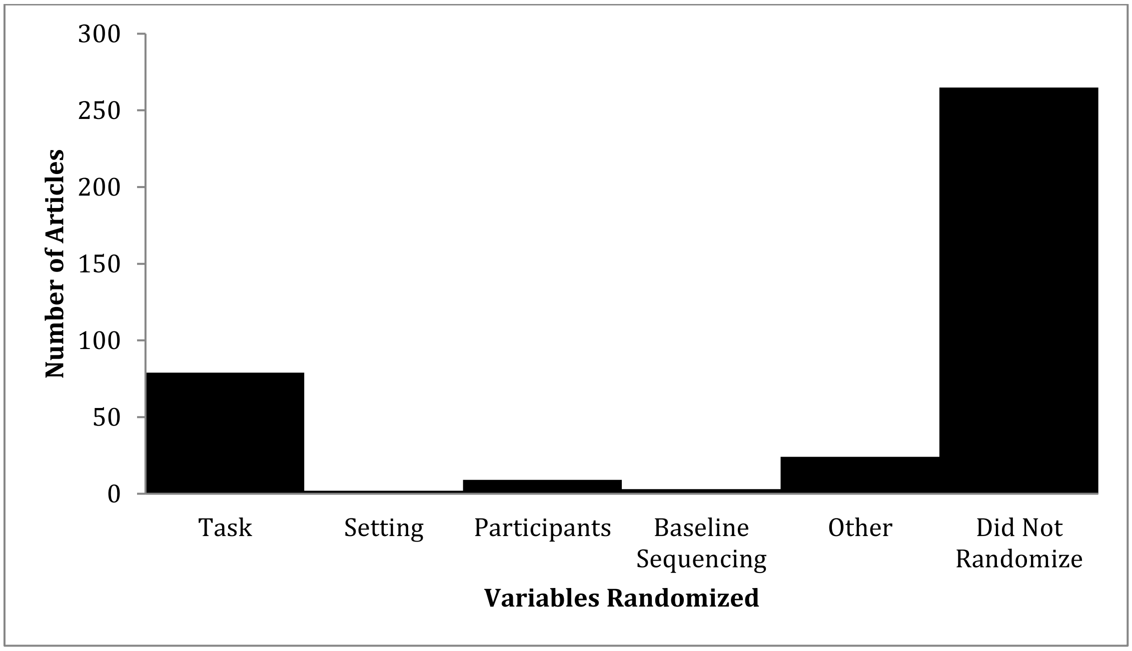

Figure 1. The total count of selected multiple baseline design articles that contained randomization of each variable.

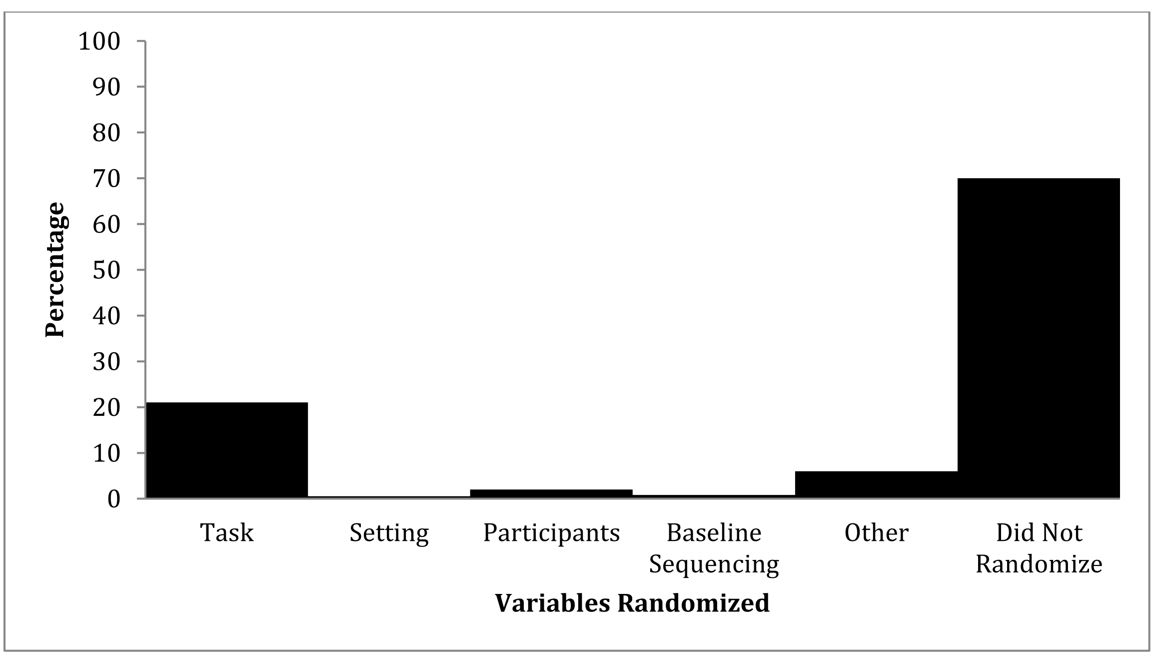

Figure 2. The percentage of selected multiple baseline design articles that randomized each variable.