March 2023

Exploring Time-Varying Extraneous Variables Effects in Single-Case Studies

Ke Cheng
University of South Florida

Exploring Time-Varying Extraneous Variables Effects in Single-Case Studies

by

Ke Cheng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a concentration in Educational Measurement and Research
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: John Ferron, Ph.D.
Robert Dedrick, Ph.D.
Yi-hsin Chen, Ph.D.
Tony Tan, Ed.D.

Date of Approval:
March 21, 2023

Keywords: Single-case experiment design, SCD, extraneous variables, treatment effect

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The effect of time-varying extraneous variables has been studied in other statistical analyses such as using Kaplan–Meier or Cox regression analysis in survival analyses. Nonetheless, the effect of modeling versus not modeling individual specific time varying extraneous variables has not been explored in multiple-baseline single case designs through Monte Carlo simulation studies. Therefore, in my dissertation, I used simulation methods to explore for a variety of conditions (varying in the number of participants, number of observations per participant, type of extraneous variable effect, size of the true intervention effect) the impact of extraneous variables on bias and standard error of treatment effect estimates, as well as confidence interval coverage. I examined the degree to which bias, standard error, and confidence interval coverage are affected by including measures of the extraneous variables in the multilevel model used to estimate the average treatment effect.

Results showed that not modeling the extraneous variable effects led to substantial biases in the treatment effect estimates and 95% confidence intervals with coverage rates less than 50%. Modeling the extraneous variables led to unbiased effect estimates and confidence intervals for the treatment effect with 95% coverage rates.

Several limitation and implications are discussed in this dissertation. The simulation conditions as well as the outcomes could be expanded in future research. Also, different extraneous variable distributions can be modeled and tested after reviewing more single case design literature to identify other types of extraneous variable effects. Finally, methods for identifying and tracking changes in extraneous variables need to be developed and studied, so

that it is feasible to include these variables in the multilevel model used to estimate treatment effects in multiple-baseline studies.

CHAPTER ONE: INTRODUCTION

Single-case design (SCD), also known as single-subject research design, has a long history in behavioral sciences, and as evidenced by many reviews (Hammond & Gast, 2010; Shadish & Sullivan, 2011). It has been prominently used in the fields of education (Gage et al., 2018; Gast, 2005; Kennedy, 2005; Richards et al., 1999), psychology (Bailey& Burch, 2002; Johnson & Pennypacker, 2009; Kratochwill & Levin, 1992; Wheeler, 2017), and other disciplines. Based on Horner et al. (2005), over 45 professional journals incorporate SCD studies. Moreover, SCD has also been featured as an important methodology since the American Psychological Association (APA) recommended SCD alongside randomized controlled trials (RCTs) in their evidence standards (APA Divisions 12/53 and the APA Division 16) and has developed research-coding criteria for reviewing SCD research. In the education field, the Institute of Education Sciences (IES) has developed standards for review of SCDs, which is incorporated by the What Works Clearinghouse (WWC) (see Kratochwill et al., 2010, 2013, 2014), and the National Reading Panel (2000) has also developed standards for review. By and large, SCDs have been playing a significant and growing role in applied research.

In contrast to true experiments where the researcher randomly assigns participants to a control group and a treatment group, in a SCD, a participant is observed in both the control and treatment conditions. Basically, it involves collecting data repeatedly from the same subject (or participant) over short or long periods of time. Performance prior to the intervention (baseline phase) is compared to the performance after the intervention (treatment phase). In general, SCD employs one or more dependent variables that are repeatedly measured within and across

controlled conditions, and it is assessed for consistency throughout the experiment by frequent monitoring of interobserver agreement (e.g., the percentage of observational units in which independent observers agree) or an equivalent. The independent variable in a SCD research study is typically an intervention. The independent variables should be operationally defined to allow both valid interpretation of results and accurate replication of the procedures (Horner et al., 2005).

Gast and Ledford (2018) define research as "the systematic investigation and manipulation of variables in order to identify associations and understand the processes that occur in typical contexts" (p. 2). In other words, applied researchers strive to explore the relation between independent variables, the variables manipulated by researchers (interventions), and dependent variables or the outcome variables expected to be changed given the manipulation. To explore and understand the causal relations, researchers frequently utilize randomized control trials (RCTs), because causal inferences can be drawn making relatively weak assumptions (i.e., these designs have strong internal validity). Nonetheless, Kazdin (2011) argued that RCTs may not be feasible in some contexts, with some populations, and with some types of interventions. In addition, compared to the group design that assesses the average or mean level of performance and compares the change in group performance, the logic of SCDs is to measure individual participant's behavior repeatedly before, during, and after the intervention. The primary goals of a SCD are for each individual: a) to explore the causal relation between the intervention (independent variable) and a change in the dependent variable, b) to examine a single or a combination of treatment effects on the dependent variable, and c) to evaluate the relative effects of single or multiple independent variables on a dependent variable (Horner et al., 2005). By

using these types of design, researchers are able to investigate the changes in behaviors (trend) in the process as well as at the end of the intervention.

Problem Statement

The term, *internal validity*, refers to inferences about an observed causal relationship from A (baseline) to B (treatment) in which the independent variable(s) were manipulated (Shadish, Cook, & Campbell, 2002). To establish the internal validity of SCDs, one of the underlying assumptions is the effects of extraneous and confounding variables on the dependent variable should be controlled (see the "Criteria for Demonstrating Evidence of a Relation between an Independent and an Outcome Variable" in WWC, 2010). As such, only the independent variable (the treatment) substantially influences the dependent variable in the treatment phase. Nevertheless, it is impossible to avoid all threats to internal validity in any of the experimental and quasi-experimental designs. Shadish, Cook, and Campbell (2002) have defined the threats to internal validity as those other possible causes-reasons to think that the relationship between A and B is not causal, that it could have occurred even in the absence of the treatment, and that it could have led to the same outcomes that were observed for the treatment (p. 54). These external events, here referred to as extraneous variables, are defined as the variables that the researchers are not intentionally studying in the experiment, but that may impact the results, and confounding with the treatment. In the context of a single-case study, if all extraneous variables are held constant and the measurement is reliable, there should be little variability of the observations within a phase, and causal inferences are straight-forward (i.e., all observations would be close to the trend line). Here, the variability consists of stability, trend, shift in level, and outliers. The measurements of the variability are discussed explicitly in Chapter Two.

In applied SCD studies, however, substantial variability is often observed within a phase, including outliers, level change, and trend change, suggesting that not all extraneous variables are held constant over time. If extraneous variable effects are not distributed evenly across the phases, the treatment effect parameter estimates might be biased (e.g., an extraneous variable is distributed over time saying that it has a greater effect on treatment phase observations than baseline observations). For example, Bevan, Wittkowski, and Wells conducted a research study about reducing postpartum depression through metacognitive therapy in a multiple-baseline study in 2013. Two participants' data (participant B and D in the original article) were considered as unstable. The original quotes from the authors are

> Participant B's scores decreased during baseline, but this coincided with a time when her partner was on leave from work and she received much more support than usual. Her scores increased once this period passed; Participant D presented as severely depressed and expressed suicidal ideation during the assessment; therefore, a brief behavioral activation intervention was conducted to address the issue of risk, and this resulted in a decrease in baseline scores before MCT was introduced (p. 73).

Bevan, Wittkowski, and Wells have indicating that the absence of participant B's partner was an extraneous variable for participant B, while another treatment to treat the suicidal condition has been introduced for participant D. Therefore, these could be considered extraneous variables in this research and could bias the treatment effect estimation.

Most single-case researchers have addressed internal validity issues through either the design structure (such as using multiple-baseline design or reversal design) or the systematic replication of the effect within the course of the study (e.g., Hersen & Barlow, 1976; Horner et al., 2005; Kazdin, 2011; Kratochwill, 1978; Kratochwill & Levin, 1992). Researchers have also

addressed the internal validity issues by applying randomization techniques in structuring SCDs (Kratochwill & Levin, 2010). Although these techniques can strengthen the internal validity arguments, the effect estimation may still be biased by extraneous factors. The potential bias and standard error (SE) in the effect estimates may be reduced by statistical control over the extraneous factors (i.e., measuring these variables and including them as covariates in the effect estimation models).

Potential Extraneous Variables in Single-Case Designs

The What Works Clearinghouse (WWC) (see Kratochwill et al., 2010, 2013, 2014) has addressed nine possible threats to the internal validity in single-case research, including ambiguous temporal precedence, selection, history, maturation, statistical regression, attrition, testing, instrumentation, and additive and interactive effects of threats to internal validity. These threats normally are presented separately even though they are not totally independent in the context of experimental or quasi-experimental designs. In SCDs the greatest attention is given to threats that are suggestive of extraneous factors that many change over time (i.e., history, maturation, testing, and instrumentation). Extraneous variables in longitudinal studies are often summarized into two categories: time-independent variables and time-dependent (or time-varying) variables (Lalonde, 2015). A time-independent variable does not have within-subject variation, meaning that the variable is constant for an individual during the course of the study. A time-varying variable is a variable that involves both within and between-subject variation, meaning that the value of a covariate changes for an individual across time and can also change amongst subjects. Under the SCD context, accumulated effects from the time-varying extraneous variables may cause internal validity concerns. To this end, researchers need to take the time-varying extraneous variable into consideration when they are planning the study, so that

measures of these covariates can be made while they are doing the experiment. After the study is completed, the researcher should be able to debrief subjects to determine whether some other events may have influenced the dependent variable. Figure 2 demonstrates a graphical presentation of possible consequences for the occurrence of an extraneous variable on an AB design. If the extraneous variable occurred during the indicated time in Figure 2, and is not taken into account, the treatment effect might be biased under this circumstance.

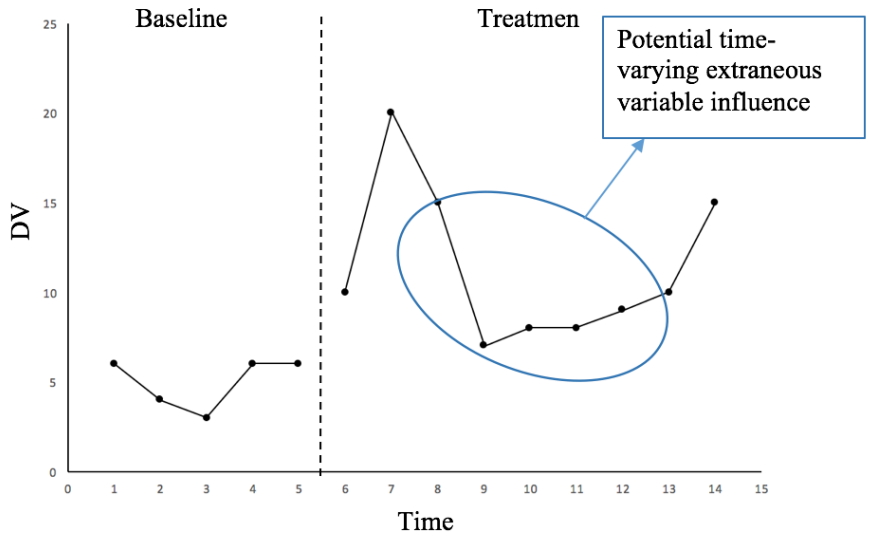


Figure 1. Graphical display of a time-varying extraneous variable effect of the outcome variable.

Purpose of This Study

As mentioned previously, extraneous variables that change over time may lead to biased treatment effect estimates. The effect of time-varying extraneous variables has been studied in other statistical analyses such as using Kaplan–Meier or Cox regression analysis in survival analyses. Nonetheless, the effect of modeling versus not modeling individual specific time varying extraneous variables has not been explored in SCD through Monte Carlo simulation studies. Therefore, in my dissertation, I used simulation methods to explore the impact of

extraneous variables on bias and standard error of treatment effect estimates, and to examine the degree to which bias and standard error can be reduced by including measures of the extraneous variables in the model used to estimate treatment effects. Multiple-baseline designs are focused on in this dissertation because they are the most common applied design in SCD research. It is hoped that this dissertation will inform the SCD researchers and practitioners about the potential influence of time-varying extraneous variables on treatment effect estimation, and illustrate how the measurement and inclusion of appropriate covariates could improve effect estimation, thereby helping the SCD researchers and practitioners to interpret the results of the treatment effect more precisely.

Significance of This Study

This dissertation will contribute to single-case applied researchers as well as methodologists who have the concern of extraneous variables that might influence the accuracy of the treatment effect estimation. In multiple baseline designs, extraneous variables might become noticeable if they have an effect simultaneously on the outcome variable. Therefore, this dissertation aims to provide a method for adjusting the model used to estimate effects from multiple-baseline designs to account for extraneous variables and to evaluate the appropriateness of the modified model.

Limitation

The conditions in this dissertation were chosen based on a review of published single-case literature, therefore the conclusions from this dissertation study can only be applied to studies with similar conditions. The interpretations of the conclusions need to be interpreted with caution when applied the conditions beyond this dissertation.

Definition of Terms

Between-Series Model. A statistical model where the subset of multiple-baseline study is used to compare between participants whose are in the baseline phase to those in the treatment phase.

Bias. A difference between a population value and an estimated value.

Confidence Interval Coverage. The proportion of replications in which 95% confidence intervals contain a population value.

Extraneous Variable. Any variable that researchers were not intending to investigate but which can potentially affect the outcomes of the research study as well as confounding with the treatment.

Fixed Effects. Regression coefficients which present the average effects across level-2 units in multilevel models.

Internal Validity. The inferences about an observed causal relationship from A (baseline) to B (treatment) in which the independent variable(s) were manipulated

Kenward-Roger. A method that adjusts degrees of freedom of the fixed effects for the small sample size conditions.

Level-1 Error. A residual or error from the predicted value to the observed value of observations within a level-1 unit in multilevel models.

Level-2 Error. Variability across level-2 units in multilevel models.

Multiple-Baseline Design. A type of single-case research design, which extends the AB design such that the baseline and treatment phases are established for multiple participants, multiple behaviors, or multiple settings.

Multilevel Modeling. A statistical model where nested data structure is considered for estimating parameters of the model. It allows researchers to have more than one level of the data structure.

Random Effects. The variabilities across level-2 units and level-1 units in multilevel models.

Relative Bias. Proportion of bias compared to the population parameter values (generating parameters).

Standard Error. The standard deviation of the sampling distribution or an estimate of that standard deviation.

Within-Series Model. Statistical models where multiple-baseline study observations are used to compare those are in the baseline phase to those in the treatment phase.

CHAPTER TWO: LITERATURE REVIEW

The literature review of this dissertation consists of five parts: a) single-case research design overview; b) an introduction of single-case research design types; c) analysis methods of single-case research designs concentrating on the effect size estimation methods; d) review of extraneous variables in single-case research designs; and e) review estimating effect size in single-case research designs when extraneous variables are present.

Single-Case Research Overview

Single-case research designs have been playing a major role in applied and clinical research including but not limited to educational and psychology fields. Since Kratochwill (1978) edited a volume of strategies for evaluating effects in single-subject research, the rapid proliferation of writing on single-subject research designs has been remarkable over the past 40 years. A considerable number of professional works across a variety of professional fields have been presented on both design and data analysis in the literature. For example, beyond the traditional discussion of the methodology within applied behavior analysis (e.g., Bailey, 1977; Johnson & Pennyparcker, 1980), there are presentations in clinical psychology (e.g., Barlow & Hersen, 1985; Johnson & Pennyparcker, 2009), social work (e.g., Fischer, 1978; Lane, Ledford & Gast, 2017), special education (e.g., Kenney, 2005), and communicative disorders (e.g., McReynolds & Keams, 1983). Therefore, the widening influence of single-case research methodology has a primary contribution to science. But more importantly, the application of this methodology has allowed a number of refinements of application to specific and unique

problems that may not have occurred without such diverse application across a variety of scholarly fields.

As described in chapter one, single-subject research design is a study of a unit (e.g., person) or multiple units (e.g., 4 to 8) and each unit serves as its own control. The study involves repeated measures from the same person over short or long periods of time. In general, the course of study is divided into at least two phases, and the participant is repeatedly measured under each phase over time. The phases are often designated by baseline and treatment (sometimes also denoted as A and B), and data from each phase are typically displayed on a graph. In a basic SCD, the dependent variable (displayed on the y-axis of the graph) is measured repeatedly over time (displayed by the x-axis) at regular intervals. Once a baseline of observation has been established - when a stable pattern emerges with at least three data points (WWC; Kratochwill et al., 2014), the intervention begins. The researcher continues to plot the observations while the intervention is implemented in the treatment phase (Figure 1).

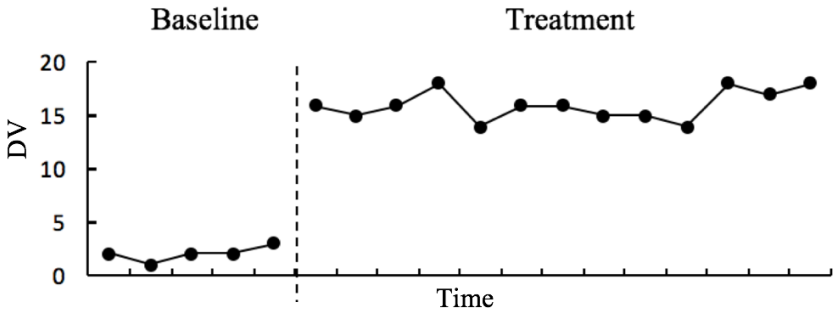


Figure 2. A basic single-subject design — AB design.

Type of Single-Case Research Designs

In single-case research design studies, several types of single-case research designs were

proposed previously, such as an AB design, a reversal (or withdrawal design), and a multiple-baseline design. These single-case research designs are discussed in the next sections.

AB Design

AB design is considered as the simplest SCD. AB design only consists of two phase – the baseline phase (denoted as A) and the treatment phase (denoted as B). The inference about the treatment effect can be made by comparing the difference of the dependent variable values between the baseline and treatment phases. Traditionally, randomization or replication of the baseline or intervention phases in the basic AB design will not be considered. As a result, AB designs have problems with internal validity and generalizability of results. The weakness of the AB design is in establishing causality because changes in outcome variables could be related to a variety of other factors, including maturation, experience, learning, and practice effects. However, Michiels and Onghena (2018) did a study of randomized AB designs. The results demonstrated the Type I error is under the control for the randomization tests, and the power of the randomized AB designs is sufficient for large treatment effects and large series lengths.

Reversal Design

The reversal design is an extension of the AB design (shown in Figure 3) as it includes alternation of the baseline and intervention phases. A baseline is established for the dependent variable and when steady state responding is reached, the treatment begins. Again, in the treatment phase (B phase), when the dependent variable reaches a steady state, the treatment will be removed by the researcher. By repeatedly measuring the dependent variable without any treatment, a second baseline is formed. This basic reversal design (ABA design) can also be extended with the reintroduction of the treatment (ABAB design), or another return to baseline (ABABA), and so on.

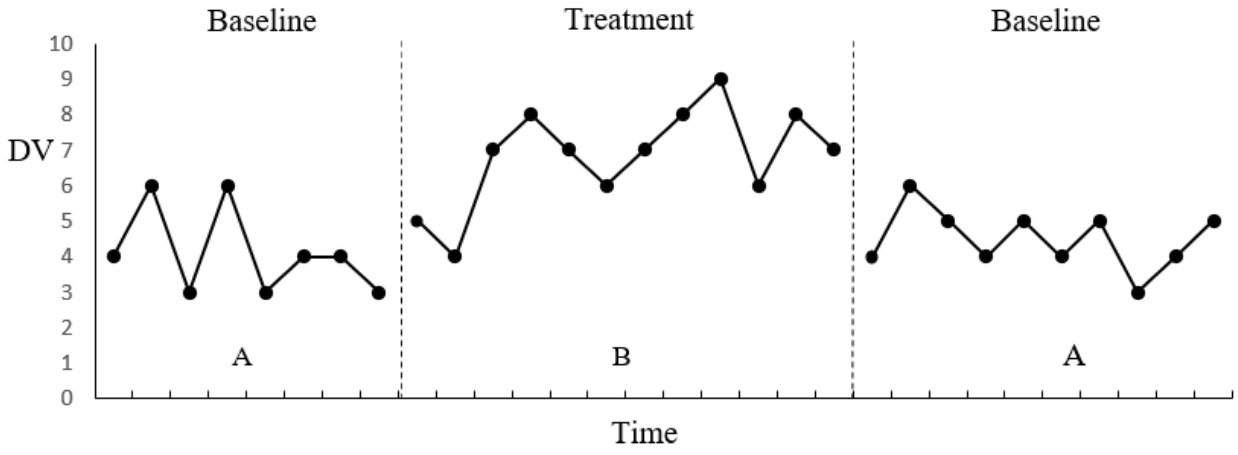


Figure 3. A reversal ABA design.

According to Lobo, Meoyaert, Cunha, and Babik (2018), incorporating at least four phases (e.g., ABAB, or ABABA) in the reversal design allows for a stronger determination of a causal relationship between the intervention and dependent variables. The relationship can be demonstrated across at least three different points in time – change in dependent variable from A to B, from B to A, and from A to B again. The other advantage for the reversal design is this design allows the incorporation of more than one intervention for each participant (e.g., multiple-treatment designs). These designs could allow researchers to study the effects of two different treatments on the dependent variable. However, challenges with including more than one treatment involve potential carry-over effects from earlier interventions and order effects that may impact the measured effectiveness of the interventions. Another concern of using reversal designs is researchers must determine that it is safe and ethical to withdraw the intervention, especially in cases where the intervention is effective and necessary.

Alternating Treatment Design

The alternating treatment design (ATD) consists of random alteration of two or more conditions such that each has an approximately equal probability of being present during each measurement occasion (Hains & Baer, 1989). The purpose of this design is to determine which treatment is more effective in changing one behavior. Logically, one condition cannot influence the performance of the dependent variable under other conditions. Therefore, Hains and Baer (1989) claims that "there is little reason to maintain a distinction in terminology between sequence, carry-over, and alternation effects. All that is of issue are sequence effects, sometimes in faster paced sequences, sometimes is slower paced sequences" (p. 60). Unfortunately, non-reversibility of effects would be a threat to internal validity, such as when learning a new word in one condition will result in the correct performance across all the conditions.

Multiple-Baseline Design

If reversal of the outcome variable it is not allowed (either unfeasible or unethical), multiple-baseline designs (MBD) will be used. According to Horner and Odor (2014), the multiple-baseline designs address the issue of how to study a dependent variable when it is not feasible or ethical to reverse the treatment effect. For example, if the treatment is for the child with autism, once the participant's behavior has improved, it may be impossible or unethical to remove the treatment which has been already implemented to influence the behavior. Another benefit of multiple-baseline designs is that the design structure may address the internal validity and ensure that the independent variable substantially influences the outcome variable through the replication of the intervention effect across subjects, settings, or behaviors. The staggered nature of introducing the intervention can eliminate some extraneous variable explanations for the behavior change (Morgan & Morgan, 2008).

In MBDs, four to eight participants will be normally involved, and each participant is randomly assigned to one of at least 3 baseline lengths. The treatment phases are staggered across time that createg different length of baseline phases across participants, behaviors, or settings. Figure 4 demonstrates the graphical display of a three participants multiple-baseline design.

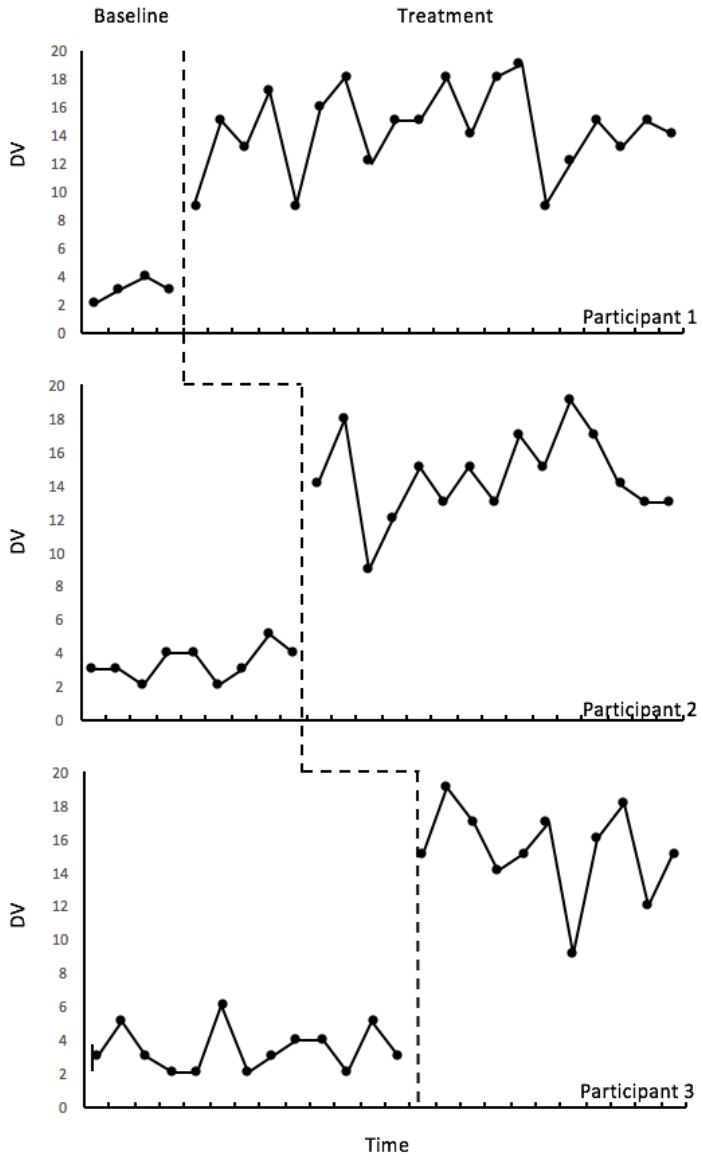


Figure 4. A graphical display of Multiple-baseline design.

Analysis Methods of Single-Case Design

Two main streams of SCD analysis methods are discussed in the follows: a) visually analyze the observations of participants, or b) statistical analysis of the observations.

Visual Analysis

Data from SCD have traditionally been analyzed visually and are graphed for each participant during a study with trend, level, and stability of data assessed within and between conditions. By examination of the immediacy of effect and overlap of data between the baseline and treatment phases, researchers are able to visually assess the treatment effect (Lobo, Meoyaert, Cunha, & Babik, 2018). Generally, when there is an immediate shift in level, and the data are stable over time in the baseline and treatment phase, it can be considered as a treatment effect. Moreover, several methods including training, structured criteria, and masked visual analysis have been suggested and those methods have shown improvement of the accuracy of visual analyses (Ferron, & Jones; 2006; Ferron, Joo, & Levin, 2016; Fisher, Kelley & Lomas; 2003). According to Kratochwill and et al. (2014), the primary reasons that visual analysis still plays a curtail role in analyzing single-case design are a) it is associated with the theoretical frame work of applied behavior analysis, b) when examining multiple SCD graphs, complex factors can be taken into account, and c) it is widely applied in the field of clinical practice since it normally focuses on the individual behavior change. However, visual analysis is not able to let researchers aggregate or synthesize the treatment effects across many single-case design studies (Kratochwill et al., 2014), and thus parametric statistics have been introduced to estimate effects.

Parametric Statistics

Although data from SCD have traditionally been analyzed visually, and visual analysis techniques have been recognized for a long time as effective and valuable (Michael, 1974), they

are still not adaptable to quantitatively weighted studies for meta-analytic purposes (Moeyaert et al., 2014). Statistical analyses such as single level regression analysis and multilevel modeling are beginning to be used more frequently, particularly as part of efforts to synthesize the results from multiple single-case studies.

Single-level Regression Analysis

Single-level regression analysis is one of the parametric statistical approaches for analyzing SCD datasets. The simplest model is written as:

$$Y_i = \beta_0 + \beta_1 * \text{phase} + e_i , \qquad e_i \sim N(0, \sigma^2), \qquad (1)$$

where Y_i is the continuous outcome variable at time *i*. β_0 is the average value of outcome variable in the baseline. β_1 is the estimated shift in level from the baseline to the treatment phases; this shift is also considered as the treatment effect. Phase is dummy coded as a value of 0 for the baseline phase, and 1 for the treatment phase. That is, when Phase = 0, the outcome value for the participant is $\beta_0 + e_i$; and when phase = 1 the outcome value for participant in the treatment phase is $\beta_0 + \beta_1 + e_i$, where e_i is the residual that represent the difference between the observed outcome value for participant and the predicted value given by the model.

If there are time trends, the model can be expanded to include a measure of time and its interaction with treatment:

$$Y_i = \beta_0 + \beta_1 * \text{phase} + \beta_2 * \text{time} + \beta_3 * \text{phase} * \text{time} + e_i , \qquad e_i \sim N(0, \sigma^2), \qquad (2)$$

where Y_i, phase, and e_i represent the same meaning as Equation 1. Time is typically centered such that 0 corresponds to the first measurement occasion in the intervention phase. As a consequence, β_0, is the expected baseline value for the participant projected for the beginning of intervention, and β_1 is the initial treatment effect for the participant (i.e., the difference between the projected baseline trajectory and the treatment phase trajectory at the beginning of

intervention). β_2 is the baseline slope for the participant, and β_3 is the change in slope between the baseline and treatment phases.

Multilevel Modeling

Multiple Baseline designs (MBD) have become the most frequently applied single-case research design (Shadish & Sullivan, 2011), and their use is continuing to grow (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013b). This design approach uses a varying time schedule that allows the researcher to determine if the application of treatment is truly influencing the change in behavior.

Within-Series Model. In the MBD setting, the treatment effect [Y_{tj} (B) – Y_{tj} (A)] can be estimated for each participant. In order to do that, researchers need to correctly specify the baseline phase model and treatment phase model for each participant, assume stability in the baseline, and calculate the difference between value of the dependent variable in the treatment phase at occasion *t* and estimate the value of the dependent variable that would have been obtained at that occasion had the participant stayed in the baseline phase (Wong, Wing, Steiner, Wong, & Cook, 2012). In typical MBD, observations are repeated measures within a participant and multiple participants are included in the study. The hierarchical structure can be analyzed with two-levels: the observations (level-1) are nested within participants (level-2). Therefore, if researchers assume there is not a time effect, the level-1 Model is written as:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * \text{phase} + e_{ij}, \qquad e_{ij} \sim N(0, \sigma_e^2), \tag{3.1}$$

where Y_{ij} is the continuous outcome variable for the j^{th} person at time *i*. β_{0j} and β_{1j} represents the intercept and the treatment effect respectively, and they vary across participants. Same as with Equation 1 and 2, the phase is dummy coded as values of 0 for the baseline phase, and 1 for the treatment phase. When Phase = 0, the outcome value for participant *j* is $\beta_{0j} + e_i$; and when Phase

= 1 the outcome value for participant *j* in the treatment phase is $\beta_{0j} + \beta_{1j} + e_{ij}$. e_{ij} is the residual term that is assumed homogeneous across phases, normally distributed with variance, σ^2, and to be independent or first-order autoregressive.

For the level-2 model, the variation of both β_{0j} and β_{1j} is described as:

$$\beta_{0j} = \theta_{00} + u_{0j} \quad u_{0j} \sim N(0, \sigma_{u0}^2), \qquad (3.2)$$

$$\beta_{1j} = \theta_{10} + u_{1j} \quad u_{1j} \sim N(0, \sigma_{u1}^2), \qquad (3.3)$$

where θ_{00} and θ_{10} are the average baseline and average treatment effect values across participants, and u_{0j} and u_{1j} are the student level errors (i.e., the deviations of the average baseline and treatment effects for participant *j* from the across participant average values). All the error terms from Equations 3.2 to 3.3 are assumed to be independent, and normally distributed. Similar to a single-level regression model, other predictors can be added into the level-1 or level-2 model. For example, if there is a trend in both baseline and treatment phase, *time* would be added in the level-1 equation and therefore the interaction of *time* and *phase* (time*phase) would also be estimated.

When adding a time effect, the outcome variable not only changed by phase, but also changed across time. Therefore, the level-1 model is:

$$Y_{ij} = \beta_{0j} + \beta_{1j}*\text{phase} + \beta_{2j}*\text{time} + \beta_{3j}*\text{phase}*\text{time} + e_{ij}, \quad e_{ij} \sim N(0, \sigma_e^2), \qquad (4.1)$$

where Y_{ij}, β_{0j}, β_{1j}, phase, and r_{ij} represent the same meaning as Equation 3.1. Time is centered such that time = 0 corresponds to the first measurement occasion in the intervention phase. As a consequence, β_{0j}, is the expected baseline value for person *j* projected to the beginning of intervention, and β_{1j} is the initial treatment effect for person *j* (i.e., the difference between the projected baseline trajectory and the treatment phase trajectory at the beginning of intervention).

β_{2j} is the baseline slope for the j^{th} person, and β_{3j} is the change in slope between the baseline and treatment phases.

The level-2 model is:

$$\beta_{0j} = \theta_{00} + u_{0j}, \quad u_{0j} \sim N(0, \sigma_{u0}^2), \tag{4.2}$$

$$\beta_{1j} = \theta_{10} + u_{1j}, \quad u_{1j} \sim N(0, \sigma_{u1}^2), \tag{4.3}$$

$$\beta_{2j} = \theta_{20} + u_{2j}, \quad u_{2j} \sim N(0, \sigma_{u2}^2), \tag{4.4}$$

$$\beta_{3j} = \theta_{30} + u_{3j}, \quad u_{3j} \sim N(0, \sigma_{u3}^2), \tag{4.5}$$

where θ_{00}, θ_{10}, θ_{20}, and θ_{30}, are the average values across participants of $\beta_{0j}, \beta_{1j}, \beta_{2j}$, and β_{3j}, respectively, and u_{0j}, u_{1j}, u_{2j}, and u_{3j}, are the participant-level errors, assumed to be distributed multivariate normal.

Combining level-1 and level-2 equations (equation 4.1 to 4.5) creates the final multilevel model as described as follows:

$$Y_{ij} = \theta_{00} + \theta_{10}*\text{phase} + \theta_{20}*\text{time} + \theta_{30}*\text{phase*time} + e_{ij} + u_{0j} + u_{1j}*\text{phase} + u_{2j}*\text{time} + u_{3j}*\text{phase*time}, \quad e_{ij} \sim N(0, \Sigma u^2), \tag{5}$$

where θ_{00}, θ_{10}, θ_{20}, and θ_{30} are the fixed effects, and u_{0j}, u_{1j}, u_{2j}, and u_{3j} are the random components of the model.

Between-Series Model. In MBD, the between-series model allows researchers to investigate the average causal effect (Ferron, Moeyaert, Van den Noortgate, & Beretvas, 2014). Between-series model allows researchers to avoid treatment effect bias that is caused by other factors than the treatment itself (strengthen the internal validity). Figure 5 demonstrates the example data set in which a MBD across three participants is used.

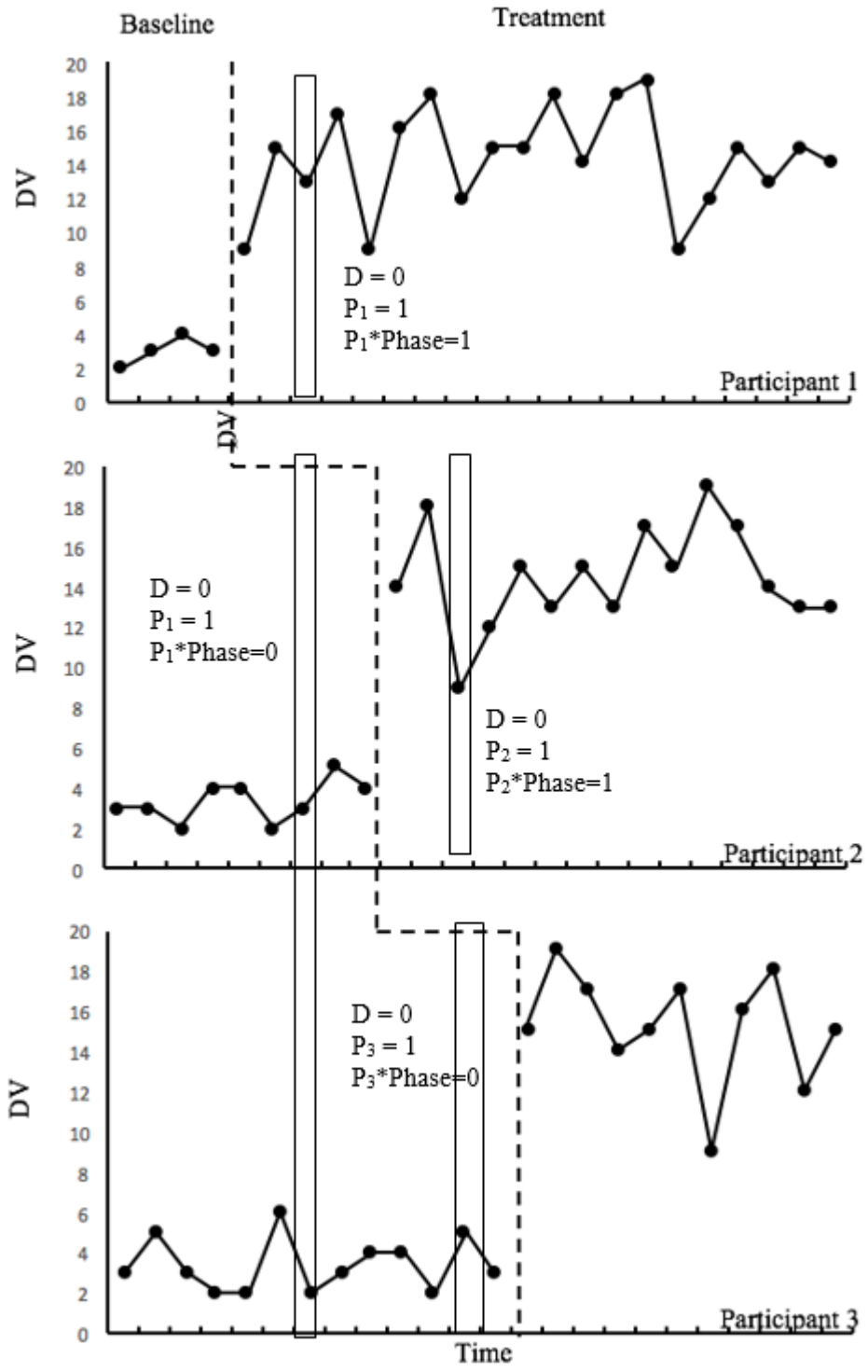


Figure 5. Graphical display of the data points for the between-series model.

The variable Phase is coded 0 for all baseline phase observations and 1 for all intervention phase observations. D_{ij} is created as a dichotomous variable that if the i^{th} observation for j^{th} participant is not in the enclosed box, then $D_{ij} = 0$, while $D_{ij} = 1$ means not 0. The between-series model utilizes dummy indicators P_1 to P_n, where n is one less than the number of interventions in the MBD. To be specific, P_{nij} is dummy coded 1 for i^{th} observation from j^{th} participant at the n^{th} box whereas $P_n = 0$, otherwise. Same as within-series model, phase is dummy coded 0 for baseline phase and 1 for treatment phase. To illustrate, see the example in Figure 4, which has three intervention start times. All the observations enclosed in the boxes are coded as P_1 or P_2 equals to 1. The enclosed observations three points after the first participant enters the intervention phase are coded such that $P_1 = 1$, whereas the enclosed observations three points after the participants 2 enters intervention are coded such that $P_2 = 1$. Therefore, the equation for the between-series model is:

$$Y_{ij} = \sum_{n=1}^{N} \left(\beta_n P_{nij} + \beta_{N+n} P_{nij} Phase_{ij}\right) + e_{ij} \qquad (6)$$

Note that the coefficient β_{N+n} is the between-series estimates of $E[Y_t(B) - Y_t(A)]$. For example, β_{N+1} is the treatment effect estimate after the first participant entered the treatment phase, while the other participants still stay in the baseline phase. Similarly, β_{N+2} is the between-series treatment effect estimate after the second participant entered the treatment phase, while the other participants are in the baseline phase. To estimate the common treatment effect averaging across the n time points (i.e., pooled estimate), the equation is written as:

$$Y_{ij} = \sum_{n=1}^{N} \left(\beta_n P_{nij} + \beta_{N+1} P_{nij} Phase_{ij}\right) + e_{ij} \qquad (7)$$

where β_{N+1} is the pooled treatment effect estimate across participants. Nonetheless, considering the time trend effect, observations in the baseline for each n^{th} time point are estimated separately.

Review of Extraneous Variables in Single-Case Design

Internal validity is an older topic in research design, but it is by no means outdated. To establish the internal validity of SCDs, one of the underlying assumptions is the effects of extraneous variables on dependent variable should be controlled. That being said, only the independent variable (the treatment) substantially influences the dependent variable in the treatment phase (Horner & Odom, 2014). Extraneous variables in longitudinal studies are often summarized into two categories: time-independent variables and time-dependent (or time-varying) variables (Lalone, 2015). The effect of time-dependent extraneous variables has been studied in other statistical analyses such as using Kaplan–Meier or Cox regression analysis in survival analyses, nonetheless, it has not been explored in SCD through simulation studies. To help methodologists better understand the features of potential time-varying extraneous variables in SCD, a systematic review of existing empirical SCD studies (including all types of SCDs) was conducted prior to this dissertation study. The articles published in peer-reviewed journals within nine years (from 2013 to 2021) were randomly selected. The searching criteria are introduced in the following sections.

Systematic Review Search Procedures and Selection Criteria

Search Strategy

A comprehensive search strategy of SCDs was developed prior to searching and identifying studies published in peer-reviewed journals that met the inclusion criteria. First, a computer-based search for the articles published in the major bibliographic databases was pre-selected (include a search of reference lists from relevant studies). The following primary key

terms and phrases appeared in the article: single-case designs, single-subject research designs, SCDs, alternating treatment designs, changing criterion designs, multiple-baseline designs, replicated single-case designs, and time-series designs. Given the fact that there was a considerable number of available articles (see Figure 6), articles published within nine years (from 2013 to 2021) were randomly selected. Specifically, 45 articles were randomly selected between 2013-2021, and there were a total of 45 articles reviewed in this review. The search was limited to studies published in the English language and those that appeared in peer-reviewed journals.

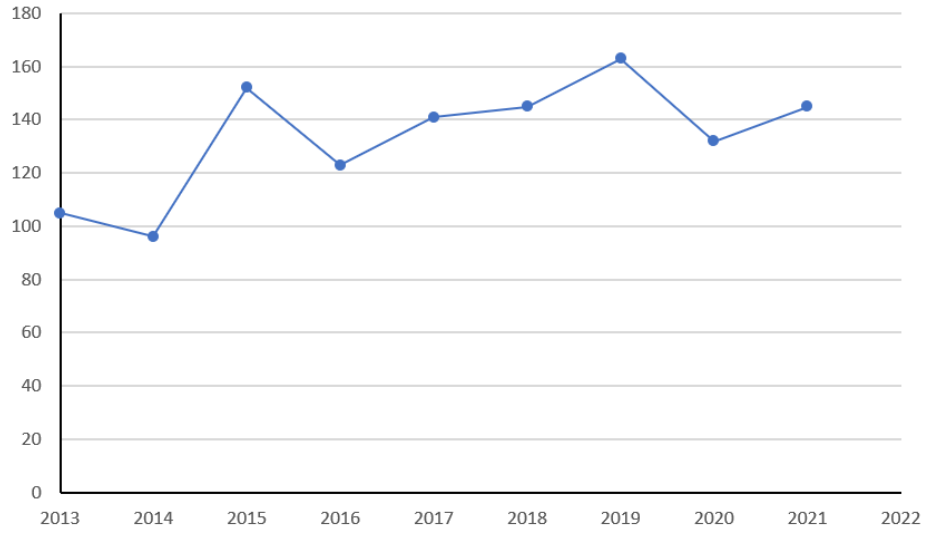


Figure 6. Number of single-case design studies published in PsychoINFO & Wiley Online Library

Study Selection

Four phases including study selection, screening, and coding were applied to select the highest number of applicable studies. Inclusion and exclusion were specifically stated in Table 1.

Phase 1 - an initial systematic review was conducted using PsycINFO and Wiley Online Library, which resulted in 1202 articles (year was customized from 2013 to 2021).

Phase 2 - titles and abstracts were screened: articles appearing to use a SCD from 2013 to 2021 retained (664). In this phase, studies that did not use a SCD or that did not explicitly state the treatment and measure of the outcome variable(s) were dropped.

Phase 3 - studies not pertaining to educational behaviors and social sciences were also dropped. There were a total 354 articles retained.

Phase 4 - each full-text article of the retained articles was carefully examined. The variables of my interest were also entered into the database.

Table 1. Inclusion and Exclusion Criterion for the Article Selection.

Inclusion Criteria
• Studies are published and available in the major bibliographic databases (include a search of reference lists from relevant studies) from 2013 to 2021. • The search is limited to English language. • Studies were selected considering one of the following terms: single-case designs; single-subject research designs; SCD; alternating treatment designs, changing criterion designs, multiple-baseline designs, replicated single-case designs, time-series designs. • The single-case model, treatment, and measure of the outcome variable are explicitly stated. • The single-case graphs are presented in the study. • Studies that meet the standards provided by What Works Clearinghouse (WWC; Kratochill et al., 2014)
Exclusion Criteria
• *The study meets the inclusion criteria but not applied in SCDs.* • *Studies are not pertinent to educational behaviors and social sciences.*

Of the 1202 original studies from 2013 to 2021, 664 studies were determined to apply SCDs under the context of educational behaviors and social sciences. Some studies were eliminated due to the structure of the SCDs was not in accord with the WWC standards. For example, reversal designs require a minimum of four phases (e.g., ABAB), and multiple baseline designs must demonstrate replication of the effect across at least three conditions (e.g., subjects,

settings, behaviors). For this review, 45 studies were selected. The journal sources of the 45 reviewed studies are summarized in Table 2.

Table 2. Journal Sources of Studies Included in the Systematic Review (N=45).

Journal Title	N
Journal of Applied Behavior Analysis	20
Behavioral Interventions	13
Journal of Counseling & Development	2
International Journal of Play Therapy	2
Journal of Positive Behavior Intervention	1
Journal of Research in Special Education Needs	1
Journal of Computer Assisted Learning	1
Learning Disabilities Research & Practice	1
Psychology in Schools	1
School Psychology International	1
Clinical Practice in Pediatric Psychology	1
Topics in Early Childhood Special Education	1

Coding Criteria Amplifications

The following variables were extracted and synthesized across studies:

Dependent Variable (behavior). In the SCD research, a dependent variable is the outcome or could be understood as the intended target behavior in the theory that should be changed given the intervention. In theory, the dependent variable should project stable and consistent patterns at the baseline and is continually observed during the intervention phase. According to Horner and Odom (2014), it is necessary to begin with the dependent variable when building a SCD study. Precisely defining the dependent variable(s) may help researchers to guide the selection of the independent variable(s), construct the relevant research question(s), and choose the appropriate measurement tools and procedures. In short, defining the dependent variable prior to the selection of research design may drive the whole research in the correct direction. Therefore, examining the dependent variable(s) in the selected studies may help researchers to better

understand the purpose of the research as well as the characteristic of the intervention. Moreover, the measurement of the dependent variable(s) was also tracked. Horner and Odom (2014) also argued that the selection of relevant and correct dependent variables will logically link the research question to the outcome variable. Hence, the decisions about how the dependent variables are measured are paramount for building a SCD.

Type of Single-case Designs. SCDs consist of multiple types of designs including alternating treatment designs, reversal designs, multiple-baseline designs, and combined designs. Different single-case designs influence the outcomes in different ways, and it is appropriate to affirm that a set of basic quality indicators is consistent across designs (Horner et al., 2005; Kratochwill et al., 2014). For example, due to a lack of replication in traditional AB designs, it seems difficult to draw a valid causal statement. The design selection will influence the internal validity, and different designs will provide different validity.

Characteristics of the Participant. The demographic information including age, gender, and ethnicity has been examined. This information could help in exploring the potential extraneous variable(s) which may cause the unstable trajectory of the datasets.

Dataset. The datasets from each article were extracted for the purpose of statistical analysis. To compare the datasets on the same scale, all the raw scores were transferred to z-scores.

Stability. The stability of the datasets in each article was examined because a lack of stability suggests the potential that an extraneous variable is affecting the outcome. Two methods were used to examine stability: 1) Cumming and Schoenfeld's stability criteria (1960), and 2) elements of the virtual visual analyst algorithm (Ferron & Joo, 2017). Cumming and Schoenfeld's (1960) criteria indicate that stability has been obtained if for 6 observed data

points, the difference between the mean of the first 3 observed data points and the mean of the second 3 observed data points is no greater than 5% of overall 6 observed data points.

Here, instability is defined as too much variability between the data at the beginning of the phase and at the end of the phase. To be specific, for 6 observed data points, the difference between the mean of the first 3 observed data points and the mean of the second 3 observed data points is no greater than 5% of overall 6 observed data points within a phase. The equation for calculating the stability that described in the above could be demonstrated as follows:

$$\left|\frac{\left(\frac{D1+D2+D3}{3}\right)-\left(\frac{D4+D5+D6}{3}\right)}{\left(\frac{D1+D2+D3+D4+D5+D6}{6}\right)}\right| \times 100$$

Equation 1

For example, if the 6 observed data points are D1=17, D2=18, D3=17, D4=19, D5=18, and D6=20. Thus, the stability for this 6 data points is:

$$\left|\frac{\left(\frac{17+18+17}{3}\right)-\left(\frac{19+18+20}{3}\right)}{\left(\frac{17+18+17+19+18+20}{6}\right)}\right| \times 100 = 9.19\%$$

Equation 2

This result, shown in the equation 2, indicates that the variation in the data between the first three data points and the last three data points is 9.19% of the mean of the 6 data points.

The judgment of whether or not the trajectory has a trend, outlier, or shift in level was determined and followed by Ferron, Joo, and Levin's (2017) virtual visual analyst (VVA) algorithm (see Table 3).

Table 3. Operational Parameters of the VVA Algorithm.

Decision	Operationalization
No consistent treatment* trend	Slope of least squares regression < .50 * SD
No trend at end of treatment*	Final 3 observations do not form an upward monotonic trend where the final 2 observations exceed the median of the segment
No outlier at end of treatment*	Final treatment* observation < M + 2 * SD
No within phase shift	$M_2 - M_1 < 1.5 * SD$

Note. *SD* is the standard deviation of the segment examined for stability; *M* is the mean of the segment being examined for stability; M_1 is the mean of the observation in the first half of a segment; M_2 is the mean of the observations in the last half of a segment.
*originally, the algorithm was designed to measure the baseline phase.

A Potential Time-varying Extraneous Variable is Suggested by the Researcher. Does the applied researcher mention the possibility of a potential time-varying extraneous variable? If the researcher mentioned the possibility of a potential time-varying extraneous variable in the experiment, was the potential time-varying extraneous variable identified?

Continuous or Categorical Data for the Potential Time-varying Extraneous Variable. This variable is an extended exploration of extraneous variables commented on by the researcher. If the researcher identified a potential time-varying extraneous variable, how was it distributed over time? Did it seem to impact multiple consecutive sessions indicating a continuous distribution? Or was it distributed in a more discrete or categorical fashion, in which each instance that the extraneous variable appeared it impacted a single session?

Results of the Systematic Review

In order to visualize the stability of each study under the same scale, the raw data were transformed to z-scores. The step included that the raw score subtracts the mean of the segment and then divided by the standard deviation of the segment [Z-score = (Raw score-

$Mean_{segment}$)/$SD_{segment}$].

Dependent Variables (behaviors) and Stability

As mentioned in the previous section, defining the dependent variable prior to the selection of the research design may drive the whole research to the right direction (Horder & Odom, 2014). To be specific, precisely defining the dependent variable(s) may help researchers to guide the selection of the independent variable(s), construct the relevant research question(s), and choose the appropriate measurement tools and procedures. Based on the review, 34 out of the 45 studies (76%) incorporated the dependent variable that pertains to the behavioral change; 6 out of the 45 studies (14%) were related to cognition improvement; and 5 out of the 45 studies (10%) were associated with other types of dependent variables (see Table 4 for detailed information).

Table 4. Descriptive Statistics of Dependent Variables.

		Stability			
	N	Variability[a]	Trend	Outlier	Shift in Level
DV					
Behavior Change	34 (76%)	30 (88.34%)	19 (55.88%)	19 (55.88%)	6 (23.68%)
Cognition Improvement	6 (14%)	3 (50%)	4 (66.67%)	0	4 (66.67%)
Other	5 (10%)	2 (40%)	2 (40%)	2 (40%)	1 (10%)

Note. DV = dependent variable; N = number of study.
[a]Variability between start and end of phase exceed Cumming and Schoenfeld's (1960) criteria.

Stability including variability between the values at the start and end of the phase, trends, outliers, and shifts in levels was analyzed along with the dependent variables. As presented in Table 4, based on statistical analysis, the datasets from 30 out of the 34 studies (88.34%) are not

stable under the behavior change; 19 out of the 34 studies (55.88%) have trends and outliers; and 6 out of the 34 studies (17.65%) have shifts in levels. Under the cognition improvement dependent variable, the datasets from 3 out of the 6 studies (50%) are not stable; 4 out of the 6 studies (66.67%) have shifts in levels and trend. For the variables which are categorized as "others", the datasets from 2 out of the 5 studies (40%) are not stable; 2 studies have trends, and 2 studies have outliers; 1 out of the 5 studies has shift in level. From the results in this section, it can be concluded that majority of the dependent variables are associated with behavior change. By analyzing the stability through the statistical criterion provided in the previous section, over half of the studies have data that are not considered stable.

Single-case Designs

Multiple baseline designs, reversal designs, alternating treatment designs, and mixed designs are the predominant single-case designs based on the review. Interobserver agreement across each design ranges from 95.81% to 99%, which indicates acceptable reliability and for observational datasets. The stability is discussed in the following section across the four SCDs.

Multiple-Baseline Design (64%). Twenty-nine out of the 45 studies (64%) utilized the multiple-baseline designs. Among them, 85% of the studies elaborated the advantages and rationales of applying the multiple-baseline designs. It can be concluded that the multiple-baseline designs allow researchers to test within-subject change across conditions, and it often involves multiple participants in a replication context. Within the multiple-baseline designs, the datasets from the 16 studies (52.8%) were examined by visual analysis; 2 studies (8.8%) were examined by statistical analysis; and only 1 study (2.9%) utilized both visual and statistical methods. Ten studies (29.4%) did not provide detailed information for the analysis method, which means the authors faiedl to identify as having used a visual or statistical analysis method.

Even these studies could be somehow inferred that visual analysis had been used, but it was not specified. Twenty out of the 29 studies (67.65%) were unstable; 14 out of the 29 studies (47.06%) had trends; 12 studies (41.18%) had outliers; and 4 studies (14.71%) had shifts in levels. Therefore, more than half of the studies were not stable judging by the given statistical analysis criteria.

Reversal design (22%). Ten out of the 45 studies (22%) applied reversal designs. Five out of the 10 studies (50%) utilized visual analysis. Five studies did not address the analysis method. Again, it might be inferred that visual analysis had been applied, but it was not described in detail. Based on Horner and Odom (2014), the situation of applying reversal designs often assumes that the dependent variables are reversible and expected to be discontinuous when the manipulation is not presented. Compared to the first baseline phase, the change of the dependent variables is evident as the independent variable (treatment) is first introduced. When the independent variable has been withdrawn, the dependent variable is expected to revert to the original baseline levels, and improves again when reinstating the independent variable. In other words, if the performance of the dependent variable remains at the same level even though the independent variable is withdrawn, the study becomes susceptible to internal validity concerns as the functional relationship between the dependent variable and independent variable fails to be demonstrated (Kratochwill & Levin, 2014). Moreover, the maintenance phase is often necessary to uphold the treatment effect in the reversal designs (Smith, 2012). Table 5 presents the detailed information regarding the stability under the reversal designs. From the review, datasets from 7 out of the 10 studies (70%) have large variability across time; 4 studies have trends; 5 studies have outliers; and 2 studies have shifts in levels. Even though the total number of the reversal

designs does not take up a large percentage of this review, more than half of the datasets from the 11 studies were not stable.

Alternating Treatment Design (9%). Four out of the 45 studies (8%) utilized alternating treatment designs, and all of them utilized the visual analysis method. Alternating treatment designs allow researchers to investigate the impact of two or more different independent variables on the dependent variable through alternating the introduction of the independent variables. It is paramount to hold other conditions constant throughout the study to ensure the internal validity and stability. Even though alternating treatment designs only take a small portion of this review, 3 out of 4 studies (75%) were determined to be unstable.

Combined design (5%). Two out of the 45 studies (6%) applied combined designs or mixed single-case designs. By analyzing the method of analysis, 1 study utilized visual analysis; 1 study utilized statistical analysis; and 1 study did not discuss the analysis method in detail. According to Kazdin (2011) and Kennedy (2005), the advantage of a combined design is the flexibility and the capacity to integrate the strengths of the various SCDs. Again, as alternating treatment design, combined designs also take a very small portion in this review. Among them, all of them were determined to be unstable.

Table 5. Descriptive Statistics of Stability Based on Single-Case Research Design

	Study	IOA	Methods of Analysis				Stability			
Research design	N	%	Visual	Statistical	Visual & statistical	Not Reported	Variability[a]	Trend	Outlier	Shift in Level
Multiple-baseline	29 (64%)	95.81	16 (52.84%)	2 (8.82%)	1 (2.94%)	10 (29.41%)	20 (67.65%)	14 (47.06%)	12 (41.18%)	4 (14.71%)
Reversal	10 (22%)	96.04	5 (50%)	0	0	5 (50%)	7 (70%)	4 (50%)	5 (50%)	2 (20%)
Alternating Treatment	4 (9%)	98	4 (100%)	0	0	0	3 (75%)	3 (75%)	3 (75%)	0
Combined	2 (5%)	99	1 (50%)	1 (50%)	0	1 (50%)	2 (100%)	2 (100%)	2 (100%)	1 (50%)

Note. *N* = number of study*; IOA* = interobserver agreement.
[a]Variability between start and end of phase exceed Cumming and Schoenfeld's (1960) criteria.

Potential Time-varying Extraneous Variables

In addition to many of the studies showing instability that may come from extraneous variables, in 10 out of the 45 studies (22.22%), extraneous variables were specifically suggested by the researchers in the results or limitation section. Table 6 provides the detailed information including the research purposes and the suggested potential extraneous variables. The graphs for those 10 studies are presented in Table 7 in the Appendix A. Direct quotes from the authors in each study were also presented in order to demonstrate the authentic expression. However, in these 10 studies, none of the authors specifically indicated the characteristics or the distribution of the potential extraneous variables. Nonetheless, from the direct quotes, the potential extraneous variable distributions could be sensed and built. For example, in Burns, et al.'s article (2015), the researchers investigated whether the conceptual intervention could help students that were struggling with the grade-appropriate centent in math. However, the results showed that there was no significant improvement of their math score. The authors stated in the

discussion section that "The study did not take into account the quality or focus of core instruction." Therefore, the "quality of the instruction" might be a potential time-varying extraneous variable that varied across time and participants. However, it was not tracked by the researchers during the experiment so it is hard to account for it in the estimation of the treatment effect. This dissertation aims to determine to what degree tracking and statistically accounting for extraneous variables may improve the estimation of treatment effects from single case studies.

Table 6. Potential Time-varying Extraneous Variables from the Reviewed Articles.

Research Purposes	Suggested Extraneous Variables
1. The effects associated with metacognitive therapy in postpartum depression.	"Participant B's scores decreased during baseline, but this coincided with a time when her partner was on leave from work and she received much more support than usual. Her scores increased once this period passed; Participant D presented as severely depressed and expressed suicidal ideation during the assessment; therefore, a brief behavioral activation intervention was conducted to address the issue of risk, and this resulted in a decrease in baseline scores before MCT was introduced" (Bevan, Wittkowski, & Wells, 2013, p. 73).
2. The impact of wellness-focused supervision on mental health counseling practicum students.	"Each person showed a decreased score during Weeks 9 or 10. One hypothesis for this low score is that there was a significant program challenge that occurred at that time" (Walen, Gage, & Lindo, 2016, p. 470).
3. Evaluation of increasing antecedent specificity in goal statements on adherence to positive behavior-management strategies.	"Varying activities across observation sessions may have been associated with differences in the saliency of discriminative stimuli for praise. For instance, opportunities to praise students while they were in the computer lab or quietly working on individual tasks may have been less salient than when students were encouraged to provide vocal responses during group instruction. The possibility remains that some activities (e.g., reading) present less salient discriminative stimuli for praise relative to others (e.g., math)" (Cohrs & Shriver, 2016, p. 777).
4. Using a conceptual understanding and procedural fluency heuristic to target math interventions.	"The study did not take into account the quality or focus of core instruction. The multiple-baseline design allowed for internally valid conclusions about the intervention, but the effect that classroom instruction also had on the data is unknown" (Burns, et al., 2015, p. 58).

Table 6. (Continued).

Research Purposes	Suggested Extraneous Variables
5. Evaluating the effects of on-task in a box as a class-wide intervention.	"Treatment integrity for Classroom C fell below 80% during implementation, which coincided with a decrease in class-wide on-task behavior. For sessions with low integrity, the teacher of Classroom C did not collect self- monitoring forms and did not provide reinforcement if the class met its goal. Therefore, the teacher was retrained on the intervention procedures using performance feedback" (Battaglia, Radley, & Ness, 2015, p. 752).
6. Initial investigation of nature-based, child-centered play therapy.	"P1: However, there was some variability that necessitates further discussion. Specially, during the 1st week of the baseline phase, P1's on-task behavior was uncharacteristically high in comparison with the other 2 weeks of the baseline phase. In seeking to explain this variability, the observers noted that P1 was focused on their presence during the 1st week of the baseline data, which may have influenced her behavior. She also exhibited a peak in total problem behavior during treatment Week 5 that may be explained by the presence of a substitute teacher in the classroom during classroom observations that week. There was also variability during treatment Week 7 and during the post intervention phase, which may be explained by her particular difficulty ending her relationship with the counselor and the opportunity to spend time in the natural play area. P2: In seeking to explain the variability in P2's scores, particularly during treatment Week 6, P2 was coughing frequently during the observations and appeared to be sick and irritable throughout the week. P3: The variability in the middle of treatment may be associated with having a substitute teacher during observation periods. P4: P4 had a significant family conflict that occurred during the middle of the treatment phase, which may account for the variability in scores. It is also noted that the post intervention phase occurred during the last month of the school year, which also may have in influenced the results" (Swank, Shin, Cabrita, Cheung, & Rivers, 2015, p. 446).

Table 6. (Continued).

Research Purposes	Suggested Extraneous Variables
7. Using video feedback to improve martial arts performance.	"There may have been some distractions from other students staying after class to practice other movements; Some minor and major injuries were incurred from doing other skills related to capoeira, but not as a result of the study, which may have impaired the performance of some skills being measured or have taken the participants out of the study altogether. As a result, intervention could not be evaluated for all three behaviors for three of the five participants" (BentitezSantiago & Miltenberger, 2016, p. 24).
8. Comparison of most-to-least to error correction for teaching receptive labelling for children diagnosed with autism.	"Mort had a previous history of engaging in non-compliance when provided corrective feedback. Anecdotally, throughout this study, Mort infrequently engaged in non-compliant behaviours; in addition, non-compliance was distributed evenly across both teaching conditions. However, his previous history may have accounted for quicker responding in the most-to-least condition for those targets he did master" (Leaf et al., 2016, p. 224).
9. Effects of therapy balls on children diagnosed with attention deficit hyperactivity disorder.	"Due to the various assignments/tasks that took place during independent seatwork time, data on classroom work samples (e.g., work completion and work accuracy) could not be collected consistently, which would have provided valuable information" (Taipalus, Hixson, Kanouse, Wyse, & Fursa, 2017, p. 425).
10. Effects of video modeling and feedback on mothers' implementation of peer-to-peer manding.	"One reason for the lack of functional control comes from Sam's data. His mands steadily increased and remained high but variable throughout baseline, and it was not until his mother entered the training phase that both his and her performances remained high and steady. This may have been because Emma (Sam's mother) began to set up the materials and play environment correctly (i.e. task analysis, steps 1–4) and intermittently used various prompts" (Madzharova, & Sturmey, 2015, p. 280).

Previous Research of Extraneous Variable Effects

Moeyert, Ferron, and Van der Noortgate (2013) did a simulation study to explore a method for adjusting the three-level model that include external events and evaluate the appropriateness of the modified model. In their study, a new approach was illustrated with real data sets. They included terms in the statistical model to account for external variables that have effects at the same point in time for all participants. Two scenarios were discussed in their study. In the first scenario, they assumed that a constant external event in a categorical format (i.e., the effect either does or does not affect each measurement occasion) and that this would affect all the participants at the same measurement occasions within a study, but the timing of the effect was randomly generated for the studies with a uniform distribution. In the second scenario, the effect of the external event was assumed in a decreasing magnitude format for all the participants in a study. The start points of the external events were randomly generated for the studies, and the average overall effect was the same for both scenarios. The results indicate that the treatment effects were biased when ignoring the external event effect, particularly when the number of measurement occasions and studies was small. A limitation of their method is that it only allows for the modeling of external events that simultaneously affect all participants. In this dissertation a method is developed to model extraneous variables that may differ from one participant to the next.

Summary

In this chapter, types of single-case research designs and their analysis methods are broadly reviewed. Based on the review, the multiple-baseline design is one of the most popular designs for single case practitioners. Multilevel modeling provides an appropriate method for estimating effects in multiple-baseline studies. Even though the design structure of the multiple-

baseline design may address the internal validity issue that ensures that the independent variable substantially influences the outcome variable through the intervention, there were many articles and studies indicating that the effect of extraneous variables might also influence the study results. A systematic review of the studies that applied single-case design methods published between 2013 to 2021 in peer-reviewed journals was done in this chapter. Based on the selection criteria, 45 articles were reviewed. The results from this review have again shown multiple-baseline design is the most applied design in single-case research setting. Another conclusion from the review is that the effect of extraneous variables might be a cause of variability and instability issues. Many authors have concluded this in their limitations section. Therefore, developing and evaluating an approach for modeling extraneous variable effects in single-case design is warranted.

CHAPTER THREE: METHODS

The methods section describes simulation design, data generation, fitting models, and parameter estimations of the simulation study. Two scenarios are introduced in this section.

Simulation Design

A Monte Carlo simulation study was conducted (using SAS 9.4 Institute, Inc., 2016) to empirically test the issues of modeling or not modeling the time-varying extraneous variables in Multiple Baseline studies (observations nested in participants). Four design factors were manipulated: number of observations in each participant (16 and 32), number of participants (4 and 8), and 2 extraneous variable factors intertwined within level-1errors: *Bernoulli* distribution and *Piecewise* distribution. Three effect sizes (0.6, 0.3, and 0) were also tested under each condition. Crossing all the simulation design factors results in a total of 2 (number of measurement occasions) x 2 (number of participants) x 2 (extraneous variable distributions) x 3 (effect sizes) = 24 simulation conditions. For each condition, 2000 replications are planned to be generated. The rationale of choosing the number of replications is based on the previous single-case simulation design studies (e.g. Ferron et al., 2009, 2010). The data generation was conducted using SAS/IML statement (SAS Institute, 2016). Two fitting models (miss-specified model and correct-specified model) was used to analyze the generated datasets, separately. The parameters were estimated using Kenward-Roger adjusted, and SAS MIXED statement was utilized in this procedure.

The convergence, effect estimates, treatment bias, treatment relative bias, standard errors, and 95% coverage of confidence intervals was compared across four fitting models for each dataset.

Design Factors

Shadish and Sullivan (2011) enumerated some parameters might influence on the quality of model estimation on the basis of their results of a thorough overview of 809 SSED studies (e.g. Alen, Grietens, & Van den Noortgate, 2009; Denis, Van der Noortgate, & Maes, 2011; Ferron et al.,2010; Wang, Cui, & Parrila, 2011). In this dissertation, two levels of measurement observations (16 and 32) and two levels of participants (4 and 8) were considered. These values are consistent with previous multilevel single case simulation designs (e.g. Ferron et al., 2009; Ferron, Farmer, & Owens, 2010; Moeyaert et al., 2014, 2016). Moreover, the level-1 errors were generated as first-order auto regressive with ρ equals to 0.2. The level-2 errors were assumed uncorrelated and distributed multivariate normal with the variance of 0. 25. [u_{ij} ~ N (0. 25)]. The average baseline value was set up at 0, and the treatment effect (i.e., the difference between the baseline level and the asymptote of the treatment phase) were set to 0.6, 0.3 and 0 for examine the different treatment effects. Two extraneous variable effects were also added to the treatment effect.

Extraneous Variable Distributions

The direct quotes indicating the extraneous variables were extracted from the reviewed articles and analyzed. Appendix A demonstrated the analysis results, and the observations at the time point that potentially contains the extraneous variables. Among ten studies, six studies contain the extraneous variables that are identified as dichotomous variables (Study 2, Study 5, Study 6, Study 7, Study 8, and Study 10). In these studies the extraneous factor was either present or absent at each occasion in time (i.e., it either does or does not impact the outcome). In contrast there were three studies where the extraneous variable varied continuously over time. (Study 3, Study 4, and Study 9). With an extraneous variable that was continuous the value could be high at one point in

time and then gradually decay over subsequent time points. One study includes both dichotomous and continuous extraneous variables (Study 1). Since only one study had both potential continuous and dichotomous extraneous variables that intertwine with the dependent variable, this case was excluded. Therefore, two methods of generating the effects of extraneous variables were investigated, one in which the extraneous factor was a binary variable, and one in which it was a continuous variable. In the next section the basic model that used to generate the data for the simulation was described. In the following section, there were descriptions of the method for generating extraneous variable effects, which were added to the data generated from the basic model.

Data Generation Basic Models

A Monte Carlo simulation study was conducted using SAS 9.4 (SAS Institute, Inc., 2016) with IML program. Data were planned to be generated assuming a two-level model (observations nested in participants).

The level-1 model that was used to generate data for person *j* is:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase + e_{ij} \qquad (8.1)$$

where Y_{ij} is the continuous outcome variable for the j^{th} person at time *i*. *Phase* is a treatment phase indicator variable coded 0 for baseline phase observations and 1 for treatment phase observations. β_{0j} represents the true average value of the baseline observations for person *j*, and β_{1j} is the treatment effect for person *j*. Figure 7 demonstrates the generated dataset for one person.

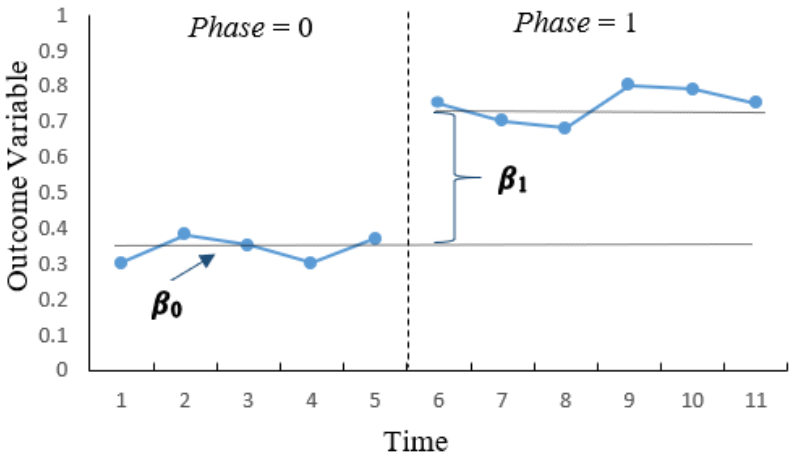


Figure 7. Demonstration of the level-1 data.

The level-2 model is:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (8.2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (8.3)$$

where γ_{00} and γ_{10} represent the cross person average value of β_{0j} and β_{1j}, whereas u_{0j} and u_{1j} are the person-level errors, assumed to be distributed multivariate normal. The level-1 error (e_{ij}) will be generated using a first order autoregressive, such that $e_{ij} = \rho\, e_{(i-1)j} + a_i$ and $a_i$is normally distributed with variance σ^2. The variance-covariance matrix for level-2 errors (u_{0j} and u_{1j}) will be generated as an uncorrelated diagonal matrix (covariances between pairs of regression coefficients will be set to zero). That is, $\boldsymbol{\Sigma u}$ = diag (τ_{00}, τ_{11}). *Time* will be centered for the purpose of modeling the extraneous variable effects in the treatment phase and the parameter values used for data generation (i.e., the values for γ_{00}, γ_{10}, ρ, σ^2, τ_{00}, and τ_{11}) will be defined when the basic model is expanded in the next section to include the extraneous variable effects.

The intervention starts points (i.e., the occasion corresponding to the first treatment observation) was varied across persons. For example, for four participants with 16 observations, the intervention started at time points 4, 6, 8, and 10, respectively. For four participants with 32

observations, the intervention started at time points 10, 14, 18, and 22, respectively. The intervention starts point pattern for the eight participants was set to be the same as for the four participant conditions, except the persons were paired. That is, with 16 observations, the intervention start point for persons 1 and 2 was set up at time point 4, the intervention start point for persons 3 and 4 was set up at time point 6, the intervention start point for persons 5 and 6 was set up at time point 8, and the intervention start point for person 7 and 8 was set up at time point 10. With 32 observations, the intervention start point for persons 1 and 2 was set up at time point 10, the intervention start point for persons 3 and 4 was set up at time point 14, the intervention start point for persons 5 and 6 was set up at time point 18, and the intervention start point for persons 7 and 8 was set up at time point 22. The extraneous variable effect would be introduced randomly at any time points in the treatment phase, then faded out over time as described in the next section.

Two Scenarios

As mentioned previously, the potential extraneous variables were categorized into two types: dichotomous variable and continuous variable. Therefore, two types of extraneous variable effects were generated and added through the addition of an extraneous variable effect to the basic data generation Equations 8.1 through 8.3.

First Scenario – Discrete Extraneous Variable

Data Generation. The extraneous variables in the first dataset were assumed as categorical format from a discrete probability distribution only. (In Study 2, 5, 6, 7, 8 and 10, the direct quotes regarding the extraneous variables are all indicating a special event occurs at certain measurement time points in which influence the outcome variable.) The distribution was determined as *Bernoulli* distribution.

Bernoulli Distribution. *Bernoulli* distribution is a type of discrete distribution which only have two possible outcomes labelled n = 0 (e.g. No) or n = 1 (e.g. Yes). The mean and variance for a *Bernoulli* random variable are $\mu = p$ and Var [X] = p (1 – p), where p denoted as the probability of having the extraneous variables, and (1 – p) demoted as the probability of not having the extraneous variables in the observation occasion. The extraneous variable will be randomly generated for each participant separately.

In this dissertation, the probability that the time point contains the extraneous variable (n = 1) is p and the probability for the time point that does not contain the extraneous variable (n = 0) is $q = 1 - p$, where $0 < p < 1$. Therefore, the probability density function for the *Bernoulli* distribution is:

$$P(n) = \begin{cases} 1 - p, & n = 0 \\ p, & n = 1 \end{cases}$$

and the variance for a *Bernoulli* random variable is:

$$\text{Var [X]} = p\,(1 - p).$$

For example, in Study 5, the direct quote from the authors is "Treatment integrity for Classroom C fell below 80% during implementation, which coincided with a decrease in class-wide on-task behavior. For sessions with low integrity, the teacher of Classroom C did not collect self-monitoring forms and did not provide reinforcement if the class met its goal. Therefore, the teacher was retrained on the intervention procedures using performance feedback" (Battaglia, Radley, & Ness, 2015, p. 752). For the Classroom C, there are totally 12 observations in the treatment phase, and 5 observations were considered as "no collection of self-monitoring forms and no reinforcement provided". Therefore, the probability that the effect mentioned above is 5/12 = 0.41, and the variance for this occasion mentioned above for Classroom C in the Study 5 is 0.41 (1 –

0.41) = 0.69. For the first extraneous variable situation in this dissertation, the probability of having an extraneous variable effect in the treatment phase will be set to 0.6 (60%).

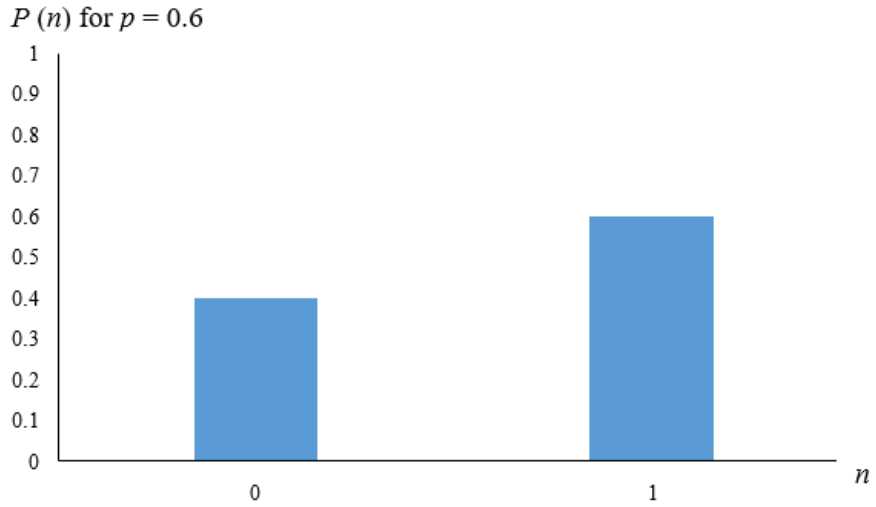


Figure 8. Bernoulli distribution with p = 0.6.

Based on the calculation demonstrated above, for four people with 16 observations, the intervention will enter at time point 4, 6, 8, and 10, respectively. Therefore, the number of data points in the treatment phase for each person are 13, 11, 9, and 7, respectively. As a result, the number of data points that contains extraneous variable in the treatment phase for each person are 8, 7, 5, and 4. For four people with 32 observations, the intervention was enter at time 10, 14, 18, and 22, respectively. Therefore, the number of data points in the treatment phase for each person were 23, 19, 15 and 11, and the number of data points that contains extraneous variable in the treatment phase for each person are 14, 11, 9, and 7. Same as four participants, eight participants had the exact number of data points that contains extraneous variable for 16 observations and 32 observations except the participants are paired as described above. The data points that contain the extraneous variable effect were randomly assigned in the treatment phase.

The level-1 equation is:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase + \beta_{2j} EV + e_{ij} \quad (11)$$

where β_{0j} represents the average value of the baseline observations, and β_{1j} is the treatment effect for person j. $Phase$ was dummy coded as a value of 0 for the baseline phase and 1 for the treatment phase. β_{2j} is the extraneous variable effect on the outcome variable Y_{ij}, and EV represents the extraneous variable which is coded as 0 for no extraneous variable effect and 1 for having extraneous variable effect. Specifically, EV was generated using UNIFORM function implemented in SAS/ IML. Note that if EV is greater than 0.4, the extraneous variable effect (0.2) would be added to the time points. The level-1 error (e_{ij}) will be generated using a first order autoregressive where $e_{ij} = \rho\, e_{(i-1)j} + a_i$. The variance of a_i (σ^2) will be set to .09and ρ will be set to .20.

The level-2 equations are:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (11.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (11.2)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (11.3)$$

where γ_{00} and γ_{10} represent the cross person average value of β_{0j} and β_{1j}. Since the baseline is not the research interest in this dissertation, β_{0j} was set up to 0. γ_{10}, the cross people average treatment effect, will be set up to 0.6, 0.3 and 0 for each condition. The extraneous variable effect, γ_{20} was set to 0.2. The level-2 errors (i.e., $u_{0j,\ 1j}$ and u_{2j}) were independently drawn from normal distributions with variances equal to 0.01 (i.e., $\tau_{00} = \tau_{11} = \tau_{22} = 0.01$).

Fitting Models. The corrected model and miss specified model were fitted to the generated datasets.

The Correct Model for the First Scenario. The equation for the model with discrete format extraneous variables is described as follows. It is equivalent to the first dataset generation model.

The level-1 equation is:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase + \beta_{2j} EV + e_{ij} \quad (12)$$

The level-2 equations is:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (12.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (12.2)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (12.3)$$

The level-1 error (e_{ij}) will be assumed first order autoregressive and the level-2 error structure will be assumed uncorrelated diagonal matrix.

Miss Specified Model for the First Scenario. The equation for the model without extraneous variables is described as follows.

Level-1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase + e_{ij} \quad (13)$$

Level-2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (13.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (13.2)$$

where the level-1 error variance was assumed first order autoregressive and homogeneous across phases and people, and the level-2 error structure was assumed an uncorrelated diagonal matrix.

Second Scenario – Continuous Extraneous Variable

Data Generation. By examine the description of the potential extraneous variables from Study 3, Study 4, and Study 9, the effect of the extraneous variable could vary continuously over time such that after the extraneous event occurred its effect may gradually decay over time leading to treatment phase trajectory that could potentially be modeled with a piecewise linear trajectory, such as shown in Figure 9.

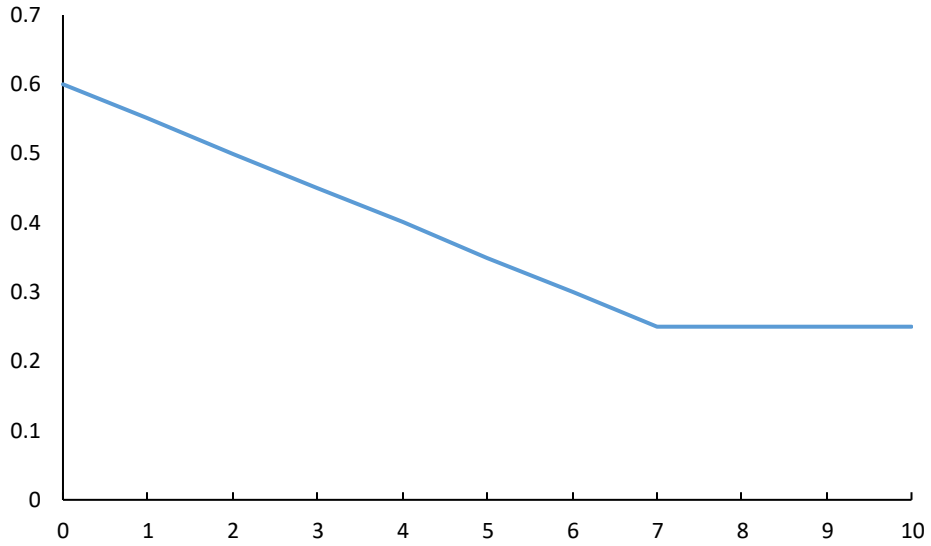


Figure 9. Piecewise trajectory to account for potential extraneous variables.

Piecewise Distribution. In some studies, a singular event may occur (e.g., a child receives a discipline referral and the impact of this event on the behavior may be greatest at the time of the event and then diminish over time. Or the extraneous factor may be a variable that impacts multiple consecutive days, such as the child getting sick or a change in the quality of instruction coinciding with a shift to a new instructional unit. In Study 4, the description for the potential extraneous variable from the original quote is

> "The study did not take into account the quality or focus of core instruction. The multiple-baseline design allowed for internally valid conclusions about the intervention,

but the effect that classroom instruction also had on the data is unknown" (Burns, et al., 2015, p. 58)."

Thus, the effect of the classroom instruction could be a continuous variable since it may vary across time, and the effectiveness of the extraneous variable may weaken with time. Because these sorts of effects may impact multiple consecutive observation sessions, and because there may be a pattern across time in the effects (such as when the effect weakens over time) they require a different effect generation model.

The level-1 data of the second scenario is generated as:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase + \beta_{2j} EV + \ _{ij} \quad (14)$$

where β_{0j} is defined as the average value of the baseline observations, β_{1j} is the treatment effect for person *j*, and *Phase* is defined as same in scenario one, with 0 for baseline and 1 for the treatment. *EV* is the extraneous variable and β_{2j} is the effect of the extraneous variable on the outcome variable Y_{ij}. The extraneous variable is introduced at a randomly selected time in the intervention phase for each person, and then will impact three consecutive observations. *EV* is coded 0 for all observations that are not impacted by the extraneous factor and coded 3, 2, and 1 for the three consecutive observations impacted by the extraneous factor. As for the first scenario, the level-1 error (e_{ij}) is generated using a first order autoregressive where $e_{ij} = \rho\, e_{(i-1)j} + a_i$. The variance of a_i (σ^2) is set to .09 and ρ is set to .20.

Also same as the first scenario, the intervention for four people with 16 observations start at time points 4, 6, 8, and 10, and the number of data points in the treatment phase for each person are 13, 11, 9, and 7, respectively. For four people with 32 observations, the intervention enters at time points 10, 14, 18, and 22, respectively. Therefore, the number of data points in the treatment phase for each person are 23, 19, 15 and 11.

The level-2 model was:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (14.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (14.2)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (14.3)$$

where γ_{00} and γ_{10} represent the cross person average value of β_{0j} and β_{1j}. Since the baseline is not the research interest in this dissertation, γ_{00} is set up to 0. γ_{10} is the cross people treatment effect and it is set up to 0.6. γ_{20} is the extraneous variable effect cross people, and it is set put to 0.1. u_{0j}, u_{1j}, and u_{2j} are the person-level errors, assumed distributed multivariate normal with the variance of 0. 01 [$u_{ij} \sim$ N (0, . 91); $u_{2j} \sim$ N (0, 0. 01); $u_{3j} \sim$ N (0, 0. 01)].

Fitting Models. The corrected model and miss specified model were fitted to the generated datasets.

The Correct Model for the Second Scenario. The equation for the model with continuous format extraneous variables is described as follows. It was equivalent to the second dataset generation model.

The level-1 equation was:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase + \beta_{2j} EV + {}_{ij} \quad (15)$$

The level-2 equations was:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (15.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (15.2)$$

$$EV = \gamma_{20} + u_{2j} \quad (15.3)$$

The level-1 error (e_{ij}) was assumed a first order autoregressive, and the level-2 error structure was assumed uncorrelated diagonal matrix.

Miss Specified Model for the Second Scenario. The equation for the model without extraneous variables was described as follows.

Level-1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase + e_{ij} \quad (16)$$

Level-2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (16.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (16.2)$$

where the level-1 error variance was assumed first order autoregressive and homogeneous across phases and people, and the level-2 error structure was assumed an uncorrelated diagonal matrix.

Parameter Estimation

For each simulated multiple-baseline study, ReML was applied to estimate parameters of both the correctly-specified model (i.e., the one including a term for the extraneous variable effect) and miss-pacified model (i.e., the one omitting the term for the extraneous variable effect), and the Kenward-Roger approach was used to compute adjusted standard error and degree of freedom to account for small sample size.

Simulation Outcomes

For each combination of sample size, series length, and extraneous variable scenario, 2000 multiple-baseline studies were simulated. Results were aggregated across the 2000 simulated data sets to estimate the convergence rate, the average effect estimate, bias in the effect estimate, relative bias in the effect estimate, and coverage of 95% confidence intervals. These results then were compared across the two extraneous variable scenarios as well as three treatment effect values.

Convergence Criteria

100% of the convergence occurred in every model. The convergence rates were calculated as proportion of replication in which estimations were reaching convergence.

Treatment Effect Estimates

The treatment effect estimates from the correct-specified model and the treatment effect estimates from the miss-specified model were averaged and compared across all the conditions and across the two scenarios.

Bias and Relative Bias

Bias was estimated as $\frac{\sum_{i=1}^{N\ of\ replication}(\widehat{\theta_i}-\theta)}{N\ of\ replication}$, where $\widehat{\theta_i}$ is the estimated treatment effect for the i^{th} simulated meta-analysis and θ is the set treatment effect (0.6, 0.3 and 0), and relative bias was estimated as $\frac{\sum_{i=1}^{N\ of\ replication}\left(\frac{\widehat{\theta_i}-\theta}{\theta}\right)}{N\ of\ replication}$.

Confidence Interval Coverage

Confidence interval coverage for each condition and analysis model was estimated as the proportion of the number of simulated meta-analyses where the 95% confidence interval for $\widehat{\theta_i}$ included θ.

Standard Error

Standard error is one of ways used in statistics to estimate the variability. It demonstrates as the sample mean which deviates from the generated population mean. The formula is given as: $SE = \frac{S}{\sqrt{n}}$.

Analysis of Dependent Variables

The ANOVA analyses are helping to detect the simulation condition effects and their effect sizes. Therefore, multi-way univariate ANOVAs were performed on bias, relative bias, standard errors, and coverage rates to analyze the variation in outcomes as a function of the simulation as well as the extraneous variable effects. Eta-square (η^2) were computed based on the ANOVA results to compare the effect sizes of the simulation conditions and extraneous variable effects. The calculation of Eta-square is using the proportion of variability of each dependent variable that associated with each simulation condition. The equation is calculated as the ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}):

$$\eta^2 = \frac{\text{SSeffect}}{\text{SStotal}}$$

The multi-way univariate ANOVA analyses and eta-square analyses were computed using PROC GLM in SAS. Cohen's (1992) effect size measures would be applied for the references, which are large: $\eta^2 \geq .15$, medium: $.06 < \eta^2 \leq .14$, and small: $\eta^2 \leq .06$.

CHAPTER FOUR: RESULTS

This chapter consists of the results and findings of this simulation study. This section reports each dependent variable of the study (i.e., relative bias, SE, and confidence interval coverage) across two extraneous variable scenarios (Bernoulli distribution and Piecewise distribution). Eta squared values (η^2) from the ANOVA tests of each dependent variable for each extraneous variable effect scenario provide an indication of the impact of the simulation design factors (i.e., number of participants, number of observations for each participant, size of the true intervention effect, and whether or not the model was correctly specified) as well as their interactions. The distribution of the simulation outcomes are sequentially reported and compared across conditions as well as across models that contains two types of extraneous variable effects. A 100% convergence rate was achieved across all replications, all design condition, each extraneous variable effect scenario, and for both correctly and incorrectly specified models. Note that the output was not examined for improper solutions (e.g., between case variance estimates of zero); rather simulation results were obtained by aggregating across all converged solutions. Table 10, 11, 12, and 13 in Appendix B provide a detailed summary of simulation results including each estimate of bias, relative bias, SE, and 95% confidence interval coverage. Results for extraneous variable scenario 1 (Bernoulli distribution) when the model is miss-specified to ignore the extraneous variable effect are provided in Table 10, whereas results from when the model is correctly specified to model the extraneous variable effect are provided in Table 11. Results from extraneous variable scenario 2 (piecewise distribution) are provided in Table 12 for the miss-specified model and in Table 13 for the correctly specified model.

Bias and Relative Bias for the Treatment Effect

The bias of the treatment effect was calculated by averaging the bias across replications. Relative bias was calculated by using the difference of the true value and estimated value divided by true value. To explore whether the simulation factors (number of participants, number of observations for each participant, size of the true intervention effect, and whether or not the model was correctly specified) had a substantial effect on the bias, several univariate ANOVA tests were conducted, one for each extraneous variable scenario. The η^2 value from the ANOVA tests are presented in Table 8 and Table 9 in Appendix B.

Extraneous Variable Scenario 1 (Bernoulli Distribution)

By examining the results from ANOVA analysis, it can be seen that all the variability in the bias estimates is explained by whether or not the model is correctly specified (η^2 =1.0), because when the model was misspecified (i.e., did not model the extraneous variable effect), the bias was .12 for each condition defined by the number of participants, number of observations per participant, and true effect size, whereas when the model was correctly specified to model the extraneous variable effect the bias was 0 across the conditions. As a consequence, all effects other the model specification explained no variability (η^2 = 0).

Figure 10 and Figure 11 shows the relative bias of miss-specified and correct specified estimators across the number of participants and number of observations for each participant. The box plot reveals the relative bias of the number of participants (4 and 8) and number of observations in each participant (16 and 32) for the treatment effect was distributed such that it was 0 for the correctly specified model and less than 5% bias of the population value for the missecified model.

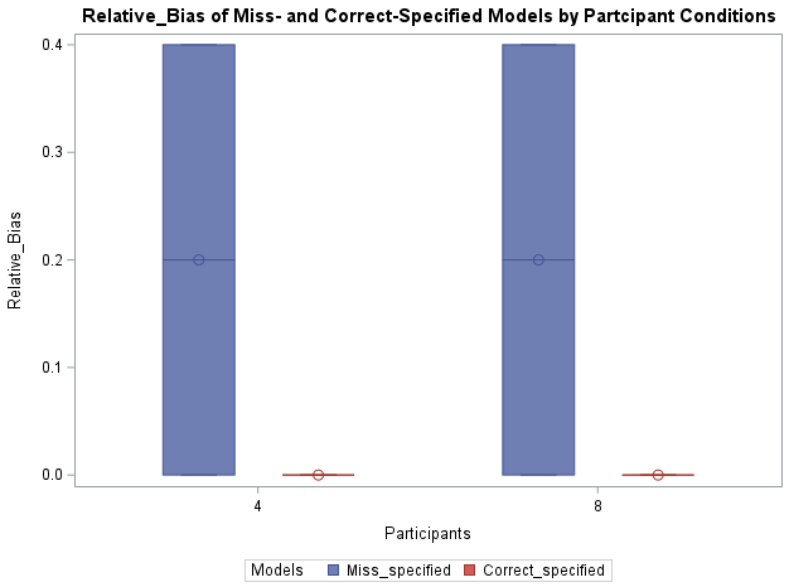


Figure 10. Box-plots: The relative bias of the miss-specified and correct-specified estimators by different participant conditions.

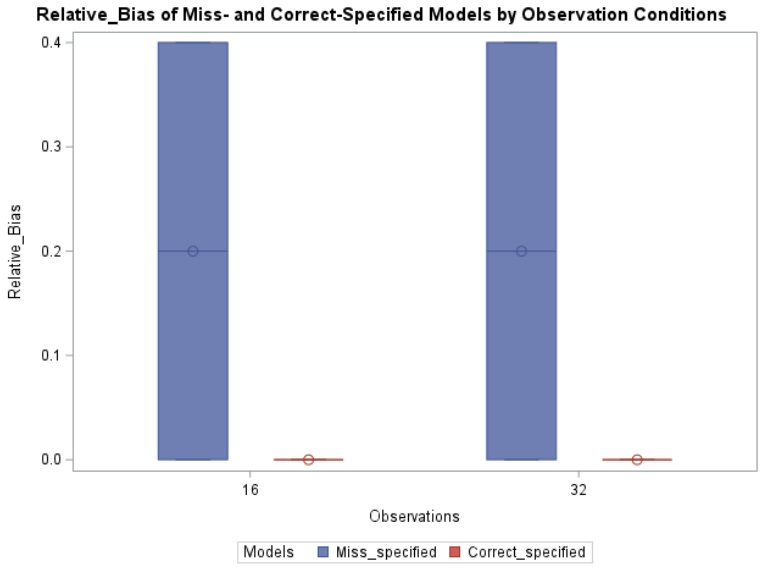


Figure 11. Box-plots: The relative bias of the miss-specified and correct-specified estimators by different observation conditions.

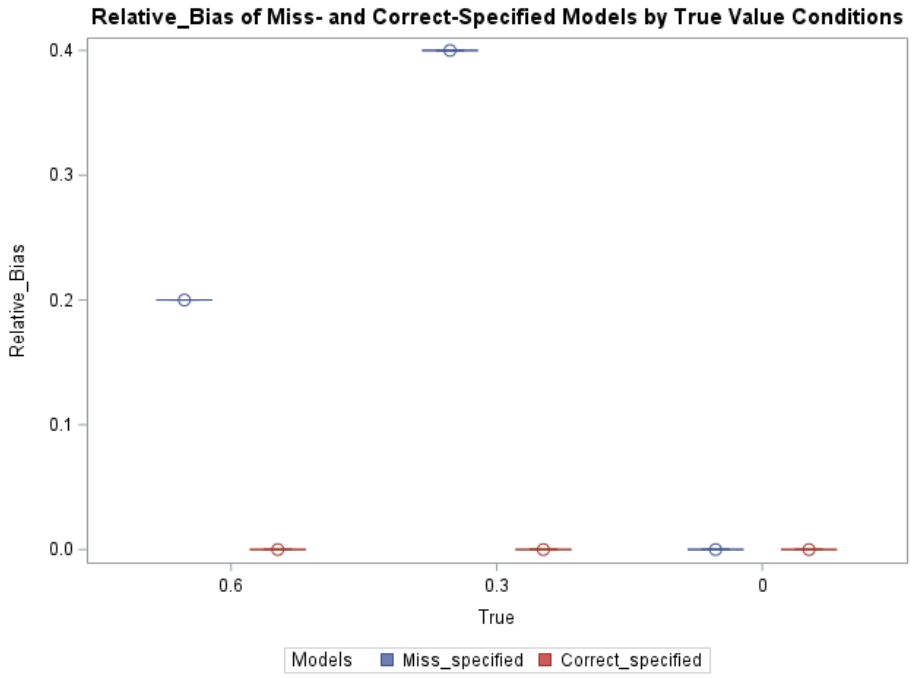


Figure 12. Box-plots: The relative bias of the miss-specified and correct-specified estimators by different true value conditions.

Figure 12 reveals that the relative bias was increased from 0.2 to 0.4 when the true value of the treatment effect switched from large effect to medium effect. Plus, the relative bias differences between miss-specified and correct-specified model were obviously observed across all the simulation condition. The parameters that used to calculate effect size were obtained from ANOVA analysis and given the medium effect size for the true value factor and large effect size for the extraneous variable effects. Therefore, it can be concluded that the miss-specified model with a medium effect size tend to have more bias than the miss-specified model with large effect size. This could be attributed to the size of the extraneous variable effect being a bigger proportion of the medium effect size.

Extraneous Variable Scenario 2 (Piecewise Distribution)

Same as the extraneous variable in Bernoulli distribution, the simulation condition including number of participants, number of observations in each participant, as well as the interactions of these variables were examined for both miss-specified and correct-specified models.

By examining the results from ANOVA analysis, the extraneous variable effect has a large effect on bias ($\eta^2 = 0.81$), and other factors including number of participants, number of observations, true value, as well as interactions of these factors shows small or no effects on bias estimates.

Figure 13 and Figure 14 shows the relative bias of miss-specified and correct specified estimators across number of participants and number of observations for each participant. The box-plot reveals the relative bias of the number of participants (4 and 8) and number of observations in each participant (16 and 32) for the treatment effect was distributed less than 10% bias of the population value across simulation conditions. Table 12 demonstrates the relative bias value as references.

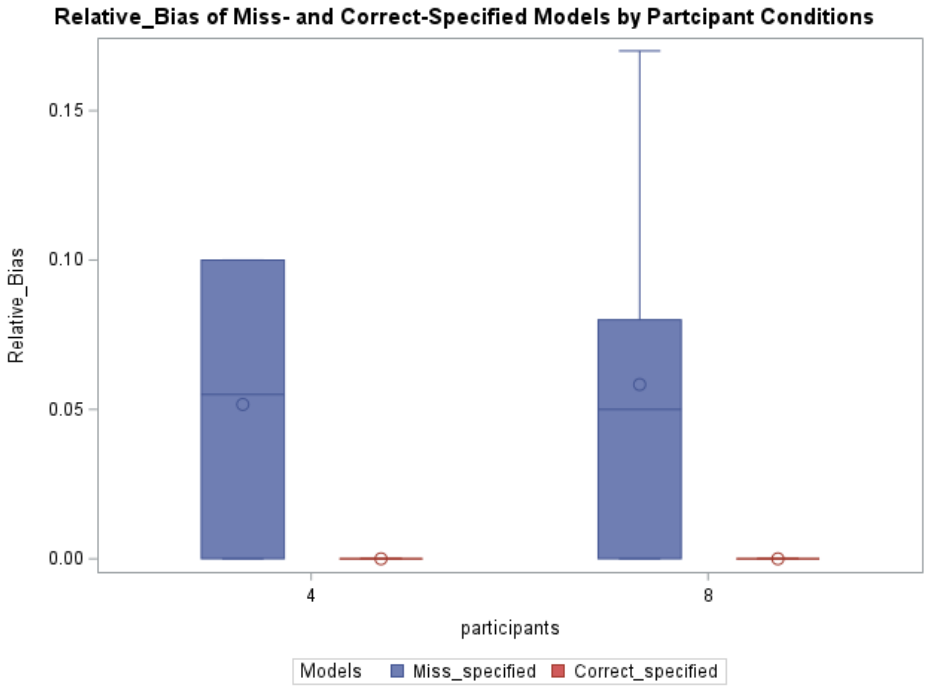


Figure 13. Box-plots: The relative bias of the miss-specified and correct-specified estimators by different participant conditions.

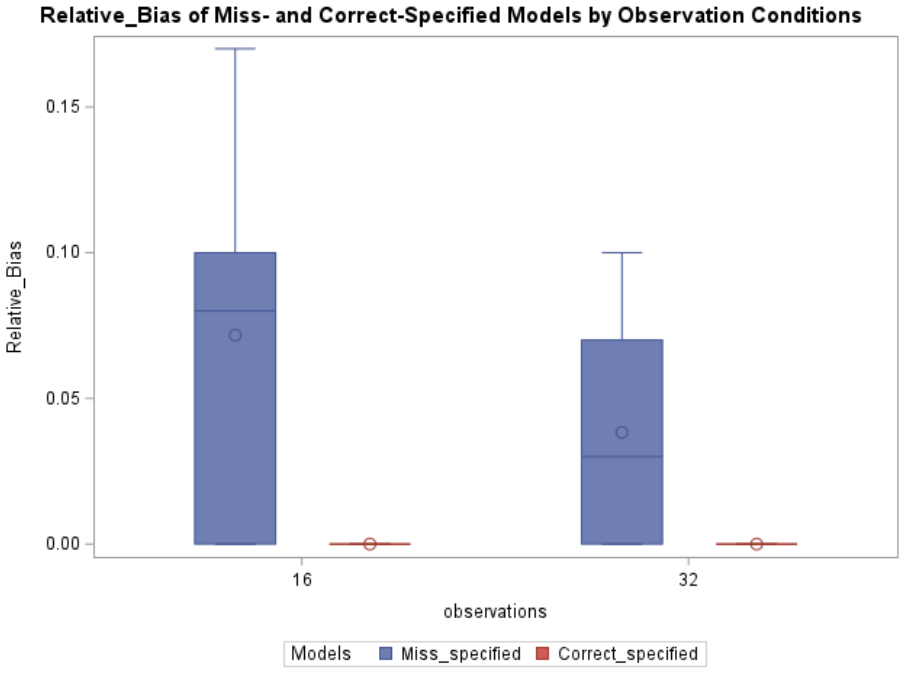


Figure 14. Box-plots: The relative bias of the miss-specified and correct-specified estimators by different observation conditions.

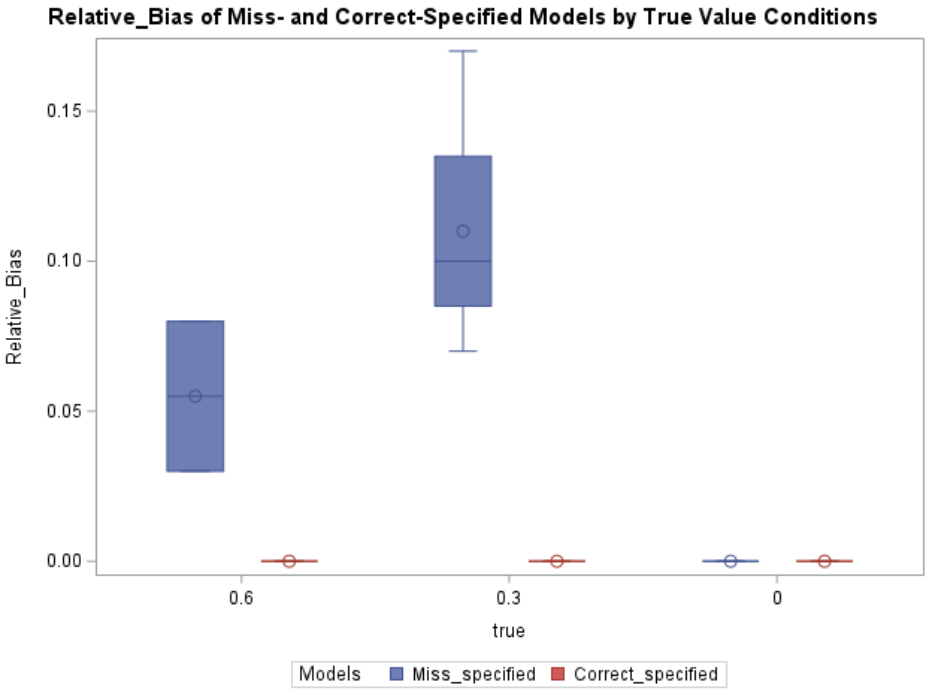


Figure 15. *Box-plots: The relative bias of the miss-specified and correct-specified estimators by different true value conditions.*

Figure 15 reveals that the relative bias increased from 0.05 to 0.10 when the true value of the treatment effect switched from large effect (0.6) to medium effect (0.3). Plus, the relative bias differences between miss-specified and correct-specified model were obviously observed across all the simulation condition. The parameters that used to calculate effect size were obtained from ANOVA analysis and given the medium effect size for the true value factor and for the extraneous variable effects. Therefore, the conclusion is consistent with the previous scenario, that the miss-specified model with a medium effect size tend to have more bias than the miss-specified model with a large effect size. Also, by including the extraneous variable effect in the model, the estimation of the treatment effect is much accurate than not including the extraneous variable effect in the analytic model.

SE for the Treatment Effect

Standard error (SE) is the standard deviation of the residuals. It tells how concentrated the data is around the line of best fit. In this dissertation, the SE also represents the variance across simulation replications. Table 8 and Table 9 in the Appendix B presents the complete SE results of the treatment effect across simulation conditions. Univariate ANOVA tests were also performed to detect simulation factors that have statistically significant effect on SE.

Extraneous Variable Scenario 1 (Bernoulli Distribution)

The ANOVA results reveal that number of participants, number of observations, and extraneous variable effect has large effect ($\eta^2 = 0.25$) on SE. Other factors and interaction effect have no effect or small effect on SE (results shown in Table 8).

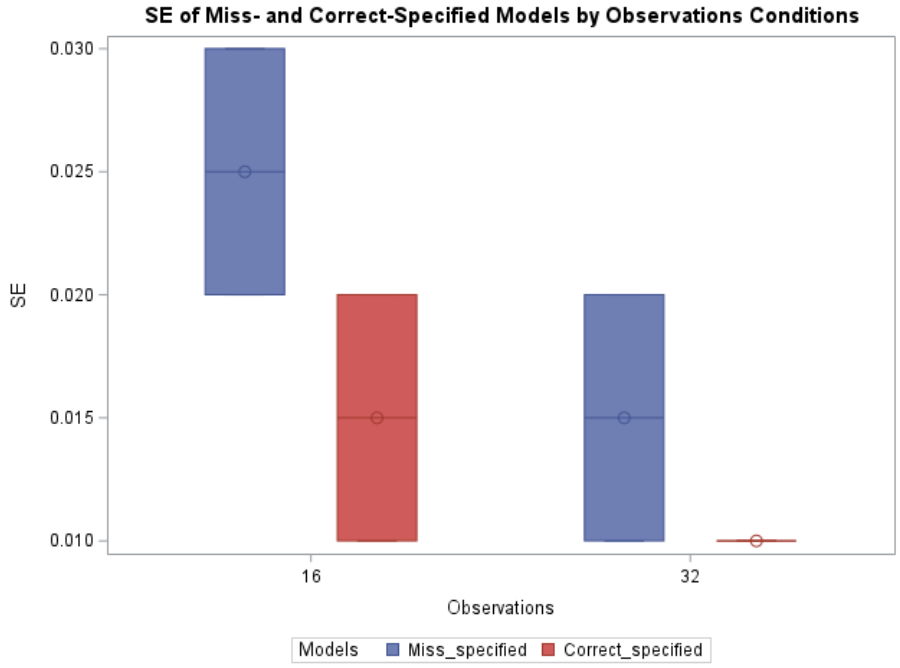


Figure 16. Box-plots: The SE of the miss-specified and correct-specified estimators by different number of observations conditions.

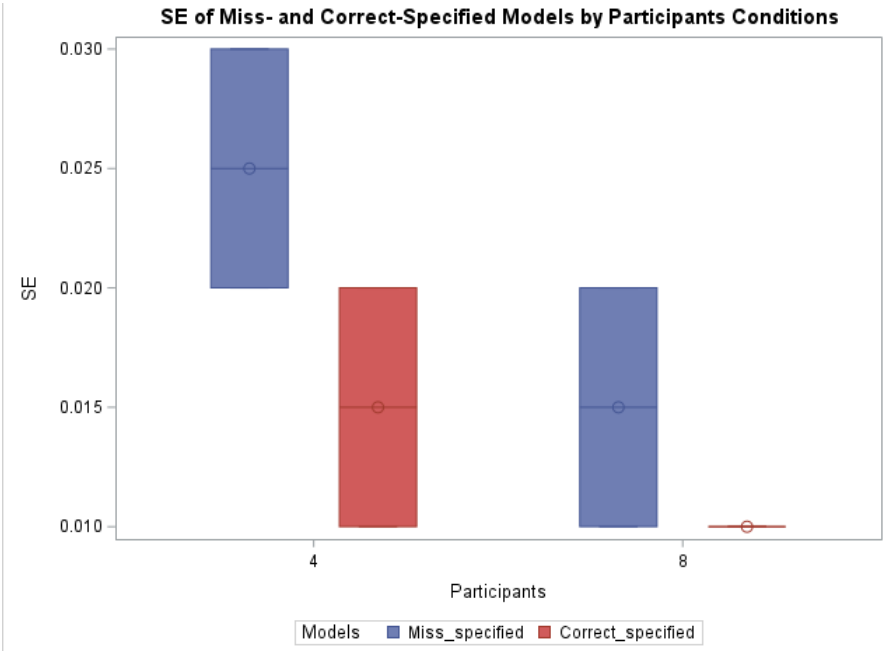


Figure 17. Box-plots: The SE of the miss-specified and correct-specified estimators by different number of participants conditions.

Figures 16 and 17 illustrate the SE differences between number of observations and number of participants as well as the differences between the miss-specified and correct-specified models. It can be concluded that for both miss-specified and correct-specified model, the smaller the number of observations nested within smaller number of participants (16 observations nested within 4 participants), the larger the SE. SE was dropped from 0.025 to 0.015 for 4 participants nested within 16 observations to 8 participants nested within 32 observations.

Extraneous Variable Scenario 2 (Piecewise Distribution)

The univariable ANOVA tests were also conducted for piecewise distribution conditions to determine the factors that might significantly influence the SE value. The results reveal that number of participants, number of observations, and the interaction of number of participants and number of observations has large effect ($\eta^2 = 0.33$) on SE, while other factors including true value and extraneous variable as well as the interaction of these factors has no effect on SE.

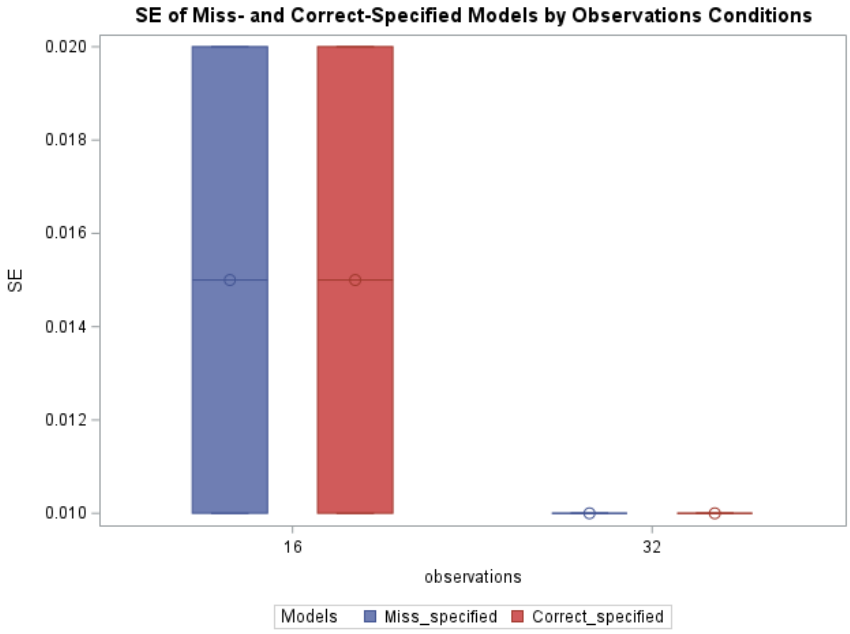


Figure 18. Box-plots: The SE of the miss-specified and correct-specified estimators by different number of observations conditions.

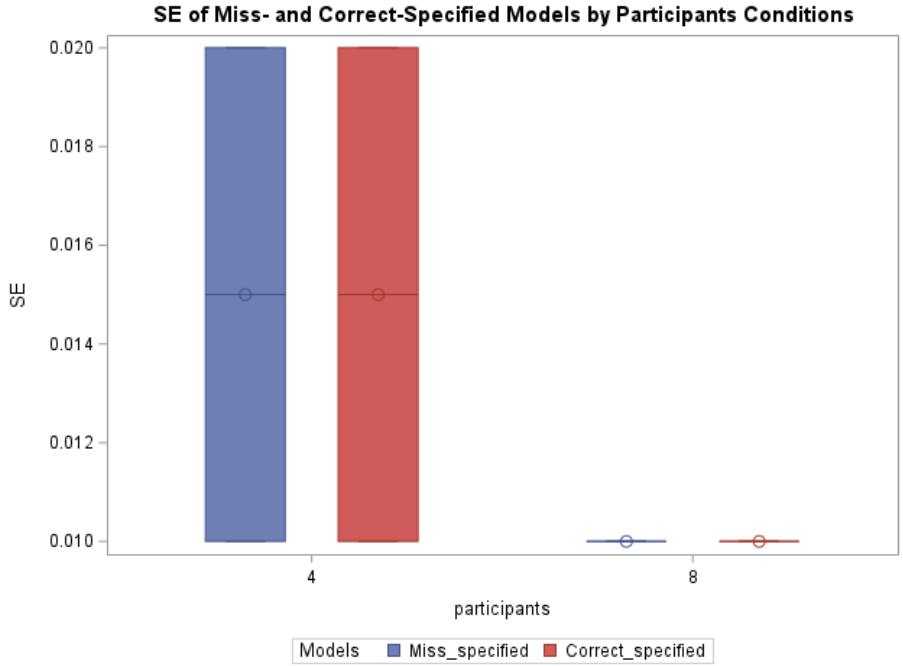


Figure 19. Box-plots: The SE of the miss-specified and correct-specified estimators by different number of participants conditions.

The box-plots from Figure 18 and 19 demonstrated when change of SE from small number of participants with small number of observations (16 observations nested within 4 participants) to large number of participants with large number of observations (32 observations nested within 8 participants). To be specific, the larger number of observations (32) and larger number of participants (8), the smaller the SE, that is, the SE dropped from 0.015 to 0 for 4 participants with 16 observations to 8 participants with 32 observations.

CI Coverage Rate for the Treatment Effect

The goal of the confidence interval coverage rate is 95% in this dissertation, which means the 95% of the time that the true value will fall into the confidence interval in each simulation replication. Still, univariate ANOVA tests were conducted to test the simulation factors that might statistically significant influence the on the CI coverage rate.

Extraneous Variable Scenario 1 (Bernoulli Distribution)

The only factor that influences the CI coverage rate is miss -and correct-specified models. Figure 20 demonstrates the for the correct-specified model, the CI coverage rate is 95% while for the miss-specified model, the CI coverage rate is around 2%. The effect size is 0.998 ($\eta^2 = 0.998$), which is large effect size.

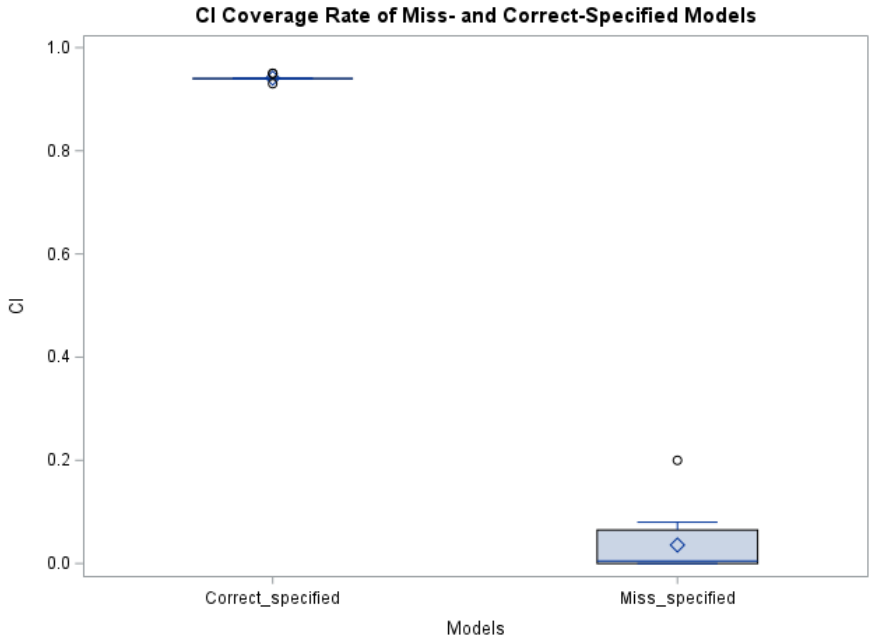


Figure 20. Box-plots: The CI coverage rate of the miss-specified and correct-specified models.

Extraneous Variable Scenario 2 (Piecewise Distribution)

Number of participants, number of observations, extraneous variable effect, the interaction of number of participants and extraneous variable effect, and the interaction of number of observations and extraneous variable effect are influencing the CI coverage rate. Figure 21 to Figure 25 demonstrate the 95% confidence interval coverage across different conditions mentioned above. Table 9 also demonstrate eta-squares respectively across each condition. Number of participants and the interaction of number of participants and extraneous variable effect has a medium effect ($\eta^2 = 0.08$) on 95% confidence interval coverage. Figure 21 shows the difference of 95% CI coverage across different number of participants. The confidence interval coverage is less with more number of participants (n = 16) than with less number of participants (n = 8), and the CI coverage difference is about 0.2. Figure 24 shows the CI differences between different number of participants but based on different models. It reveals

that for correct specified model, there is no difference of CI coverage between different number of participants, the CI coverage is all around 95%. However, under the miss-specified model condition, there are CI differences observed between different number of participants. Therefore, it can be concluded that only under miss-specified model condition, the number of participants influence the CI coverage, and with less number of participants (n = 4), the CI coverage (CI = 0.55) is larger comparing to the CI coverage (CI = 0.2) with more number of participants (n = 8).

Similar to number of participants, number of observations and the interaction of number of observations and extraneous variable effect are influencing the CI coverage. Figure 22 and Figure 25 reveals that only under miss-specified model condition, number of observations is influencing the CI coverage, and it has a small effect ($\eta^2 = 0.03$). The CI coverage (CI = 0.5 vs. CI = 0.25) is larger for a greater number of observations (n = 32) comparing to a smaller number of observations (n = 16).

Extraneous variable has a large effect ($\eta^2 = 0.78$) on the CI coverage. Figure 23 shows the difference of CI coverage between correct- and miss-specified models. The correct-specified model has a large CI coverage (CI = 0.95) than the miss- specified model (CI = 0.4).

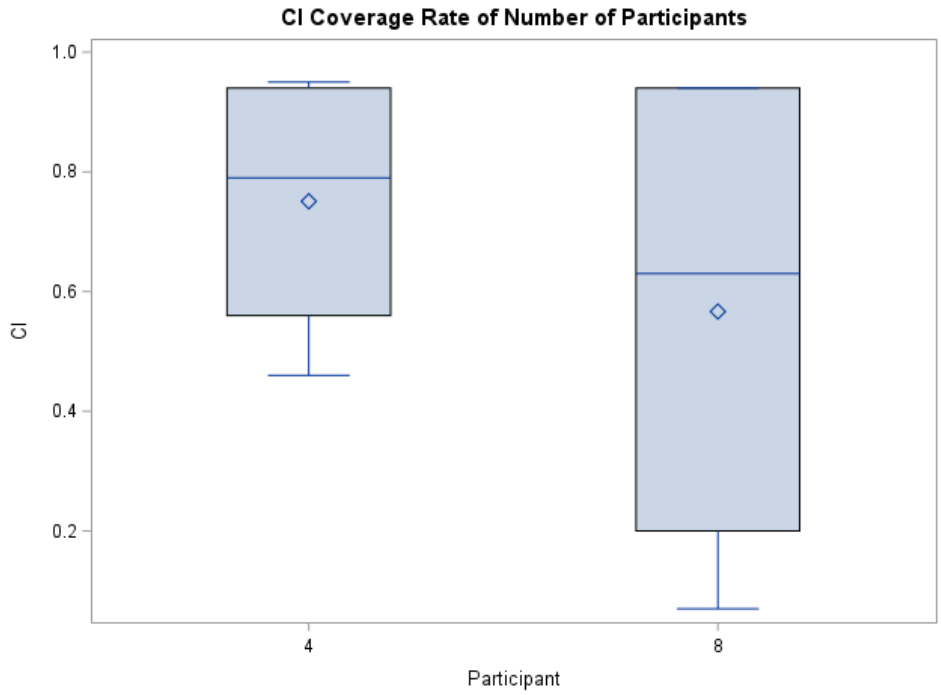


Figure 21. Box-plots: The CI coverage rate of number of participants.

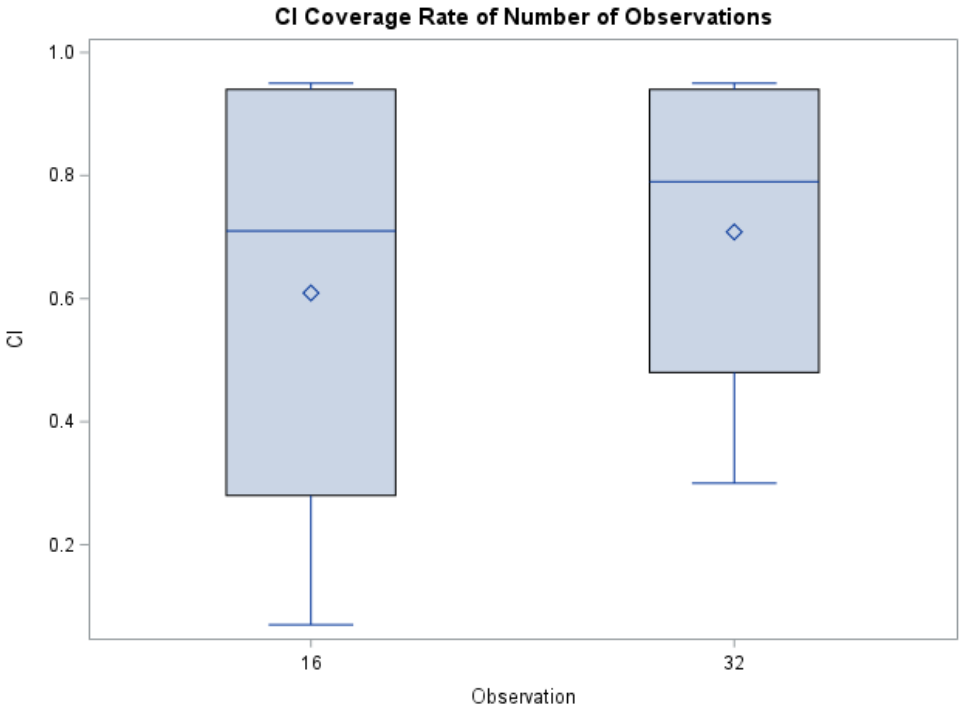


Figure 22. Box-plots: The CI coverage rate of number of observations.

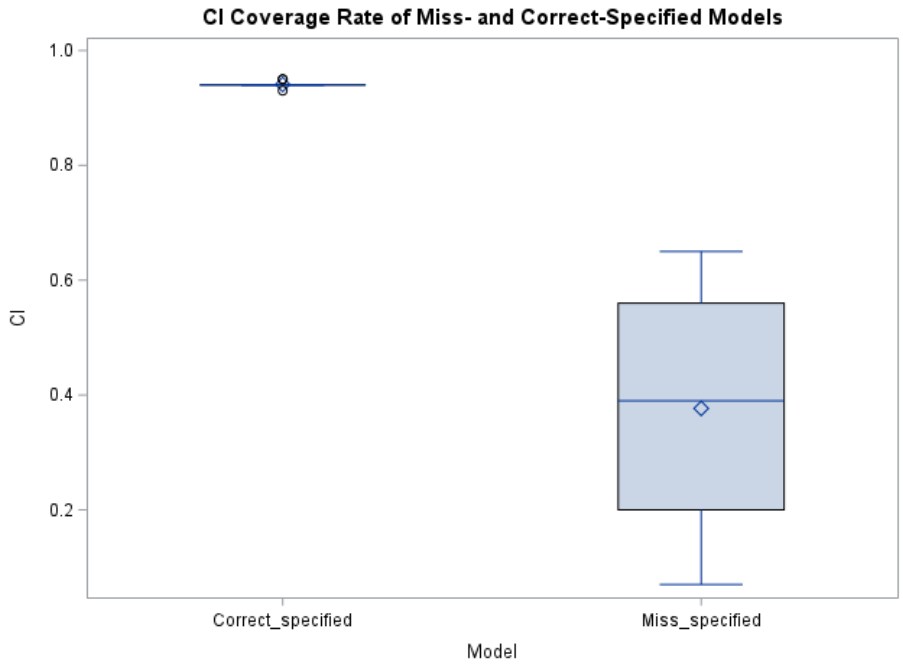


Figure 23. Box-plots: The CI coverage rate of correct- and miss- specified models.

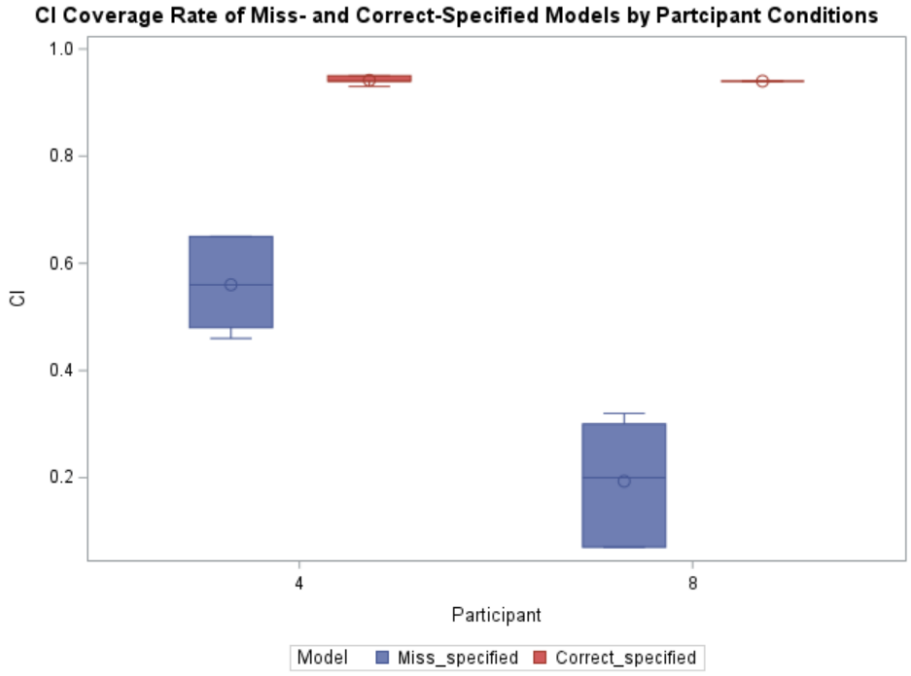


Figure 24. Box-plots: The CI coverage rate of correct- and miss- specified models by number of participant condition.

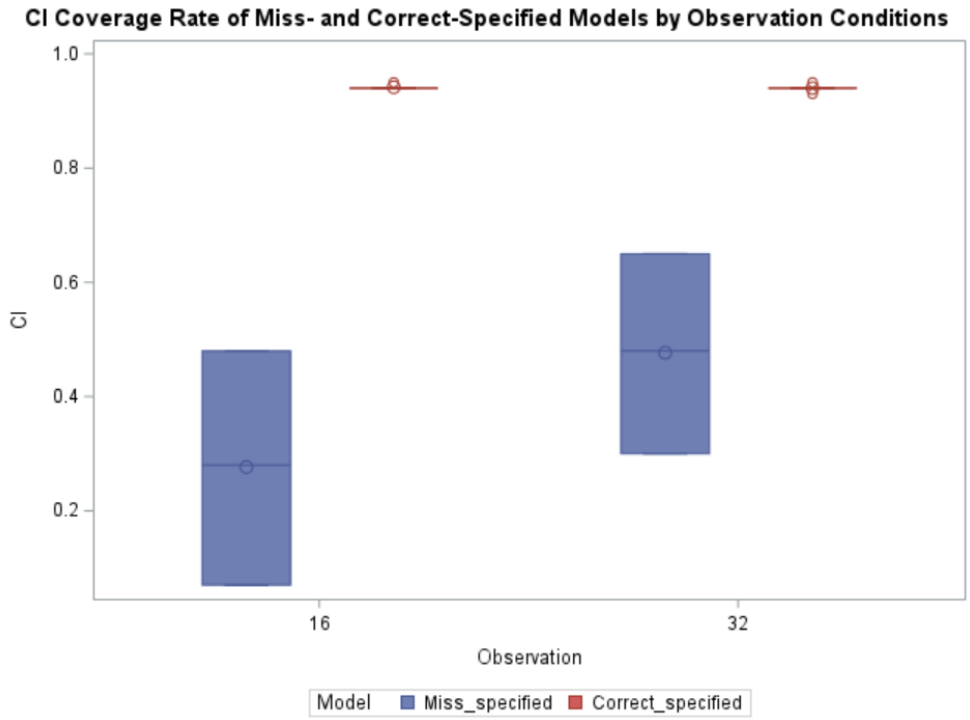


Figure 25. Box-plots: The CI coverage rate of correct- and miss- specified models by number of observation condition.

Summary of Results

Under bernoulli distribution condition, it can be concluded that the extraneous variable effect has large effect on Bias, SE, and CI coverage. Other conditions including number of participants, number of observations, and different true value, as well as their interactions has no effect on Bias and CI coverage. However, under miss-specified model condition, more number of participants (n = 8) with more number of observations (n = 32) has the smallest SE.

Under piecewise distribution condition, the extraneous variable effect has a large effect on Bias, SE, and CI coverage. Still, more number of participants with more number of observations causes less SE. While under miss-specified model condition, 4 participants with 32 observations produces better CI coverage.

CHAPTER FIVE: DISCUSSION

In this chapter, we discussed summary of this study, findings, implication, and application for the applied single-case researchers. Limitations and further research are also discussed at the end of this chapter.

Summary

This dissertation applied Monte Carlo simulation methods to explore the impact of extraneous variables on bias and standard error of treatment effect estimates, and to examine the degree to which bias and standard error can be reduced by including measures of the extraneous variables in the model used to estimate treatment effects. Two scenarios (two types of extraneous variable distributions – Bernoulli distribution and Piecewise distribution) were concluded and extracted from the empirical single-case research designed paper.

Data were generated using a two-level model (observation nested in participants). The observations for each participant in the baseline were generated as mean of 0 and the mean of the treatment effects were 0.6, 0.3 and 0 which represent large, medium, and no treatment effect. Level-1 errors were generated as autocorrelated (phi = 0.2) homogeneous across phases, while level-2 errors were generated under uncorrelated condition. Plus, the extraneous variable effect was also generated in the treatment phase. For the extraneous variable that consists in Bernoulli distribution in the treatment phase was considered as the first scenario, while the extraneous variable that consists in Piecewise distribution in the treatment phase was considered as the second scenario.

Two models for each scenario – miss-specified model, which is not including the extraneous variable effect, and correct-specified model, which is including the extraneous variable effect, were fit into the generated dataset. Parameter estimation accuracy and statistical inferences were systematically analyzed. Bias, relative bias, SE, and CI coverage rates were compared across simulation conditions by using ANOVA tests. The interactions of the simulation conditions were also examined through ANOVA tests.

Findings

Bias and Relative Bias of the Treatment Effect Estimate

The results of bias and relative bias for both Bernoulli and Piecewise distribution scenario were consistent. That is, only whether the model was correctly specified substantially explained the variation in bias estimates. When the model was correctly specified to model the extraneous variable effect, the bias was zero, but was non-zero when the extraneous variable effects were not modeled. Although previous research had not looked at adding terms to the model to estimate idiosyncratic extraneous variable effects, previous research had found that the fixed effects of multilevel models of single-case data were unbiased when models were correctly specified, including two level models without slopes (Ferron et al., 2009), three level models without slopes (Owens & Ferron, 2012), and three-level models with slopes (Moeyaert et al, 2014). In addition, one study that looked at modeling common, as opposed to idiosyncratic, external event effects found that modeling the effects removed the bias that was present when the effects were not modeled (Moeyaert et al., 2013). Furthermore, the biasing of the effect estimates to non-modeled extraneous variable effects was also found for some alternative types of idiosyncratic extraneous variable effects (Ferron et al., 2014). Thus, the bias results from this study are largely consistent with what may have been expected from previous research.

SE of the Treatment Effect Estimate

In Bernoulli distribution scenario, the interaction of observation factor and participant factor is statistically significant influence the SE of the treatment effect estimate. Plus, the extraneous variable effect also has a large influence on SE. Figure 16 and 17 demonstrated that the SE was smaller with large observations nested in large participants, and correct specified model has smaller SE than the miss-specified model. For Piecewise distribution, the results are the same as Bernoulli distribution, however, the extraneous variable effect factor does not influence the SE of the treatment effect.

CI Coverage Rate for the Treatment Effect

Only extraneous variable factor is influencing the CI coverage rate for the treatment effect in both Bernoulli and Piecewise distribution scenario. Figure 20 demonstrated that for miss-specified model in Bernoulli distribution, the CI coverage rate is only approximately 2% while for the correct-specified model, the CI coverage rate is around 95%. Also, the ANOVA results showed a large effect size for extraneous variable effect on CI coverage rate. For Piecewise distribution scenario, the CI coverage rate for miss-specified model is around 37% while the CI coverage rate for correct-specified model is 95% (Shown in Figure 21). The finding of coverage rates of 95% with correctly specified models is consistent with previous studies that did not simulate extraneous variable effects (Ferron et al., 2009; Moeyaert et al, 2014; Owens & Ferron, 2012), and adds to the support for the viability of applying multilevel models when estimating average treatment effects in single-case studies.

Limitations

This dissertation is only a start point of exploring extraneous variable effects in single-case design. Therefore, some limitations could be found in this dissertation. First, this simulation

study was conducted with only limited conditions. For example, only two types of extraneous variable (with Bernoulli distribution and Piecewise distribution) conditions were explored and modeled. Therefore, the interpretation of the conclusion needs to be interpreted with caution when applied the conditions beyond this dissertation. Future research could examine a wider range of single-case literature to uncover other extraneous variables with different distributions, such as a geometric distribution or a triangular distribution. Then the efficacy of modeling the different types of effects of these alternative extraneous variables could be examined. Also, more simulation conditions, including different numbers of observations, different numbers of participants, different levels of autocorrelation, the presence of baseline trends, and different types of single-case designs, could be examined to investigate the generalizability of the current findings.

In addition, this research was limited to the examination of effect estimation through multilevel models. Future research could examine the impact of modeling extraneous variable effects on other approaches to estimating effects (e.g., design comparable effect sizes (Shadish et al., 2014), within-case standardized mean differences (Gingerich, 1984), log response ratios (Pustejovsky, 2018), percent goal obtained indices (Ferron et al., 2020), etc.). Lastly, the results of this dissertation (values) were rounded to 2 decimals, which reduced the variability shown in the Boxplots of the standard errors.

Implications

Moeyaert et al., (2013) conducted a study to explore and present a method for adjusting the three-level model to external events that were common across cases and evaluates the appropriateness of the modified model. The results indicate that more bias observed in the treatment effects when ignoring a common external event, especially with a small number of

measurement occasions and studies. This result also corresponds to the results of this dissertation. In their study, two external events scenarios were also presented. In the first scenario, the pattern of external event was modeled as constantly influencing four measurement points for all the participants within a study. The external event in the second scenario was modeled similar as first scenario, the difference was the effects of the external event for these constant four measurement points were gradually faded. In this dissertation, the extraneous variable effects were modeled to be idiosyncratic (i.e., affecting a single participant, as opposed to all participants) and were distributed randomly across measurements occasions.

As mentioned in previous chapters, the influence of modeling and not modeling idiosyncratic extraneous variable effects that change over time has not been previously studied in a Monte Carlo simulation study. This dissertation aimed to contribute to the existing literature and demonstrates that the differences between modeling and not modeling the extraneous variable effect in terms of bias, relative bias, SE, and CI coverage rate. It is hoped that this dissertation will inform the SCD researchers and practitioners about the potential influence of time-varying extraneous variables on treatment effect estimation and illustrate how the measurement and inclusion of appropriate covariates could improve effect estimation, thereby helping the SCD researchers and practitioners to interpret the results of the treatment effect more precisely.

Secondly, for the single-case applied researchers as well as methodologists whose have the concerns of the accuracy of the treatment effect estimation that caused by extraneous variable effect, this dissertation tend to provide a guidance of how to modeling the extraneous variable when building analytic models. The extraneous variable can be modeled as an event that happened or did not happen in each observation (in a Bernoulli distribution) or can be modeled

as an event that has a relatively strong effect in couple of observations in the treatment phase, and then faded out as time passed by (Pricewise distribution). This information needs to be considered in the study designed, which is before the data collection. Therefore, for SCD practitioners, not only the outcome variable (dependent variable) needs to be tracked during the data collection, but also the extraneous events that happened in the treatment phases.

If potential extraneous variables may result from variation in how the treatment is delivered in a particular session, then the outcome variation related to the extraneous variables could be minimized by focusing on the fidelity of intervention delivery. Furthermore, monitoring treatment fidelity over time could provide the data needed to model fidelity related extraneous variables.

A more general approach to tracking potential extraneous variables would be to ask open ended questions after each session about whether other factors associated with the behavior being studied had changed since the last session. However, researchers need to be aware that these open-ended questions might influence or direct the outcomes. Therefore, researchers may need to be trained to avoid the consequences mentioned above. Another approach would be to develop a measurement tool, which included items for each factor that is known to be related to the behavior under investigation. For example, there could be items developed that are related to potential changes in medicine, health, sleep, other interventions or treatments, interpersonal relationships, etc., and then this measurement tool could be filled out each session. It would be helpful in future research to develop alternative approaches to gathering information on potential extraneous variables and then to investigate the efficiency and effectiveness of the methods. This dissertation found that if we were able to identify and measure extraneous variables, then we could improve our treatment effect estimation in multiple-baseline studies by including these

extraneous variables in our multilevel models. However, the advantages of more accurate treatment effect estimates can only be realized if we develop methods to tack and measure potential extraneous variables.

REFERENCES

American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents. (2008). *Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care.* Washington, DC: American Psychological Association.

Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Sage.

Burns, M. K., Walick, C., Simonson, G. R., Dominguez, L., Harelstad, L., Kincaid, A., & Nelson,

G. S. (2015). Using a conceptual understanding and procedural fluency heuristic to target math interventions with students in early elementary. *Learning Disabilities Research & Practice, 30*(2), 52-60.

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology, 21*, 397-414.

Cumming, W. W., & Schoenfeld, W. N. (1960). Behavior stability under extended exposure to a time correlated reinforcement contingency. *Society for the Experimental Analysis of Behavior,3*(1), 71-82.

Engel, R.J., & Schutt K. R. (2017). Chapter 8 single-subject design. *The practice of research in social work* (187-218). *Thousand Oaks,* CA: Sage.

Ferron, J., Goldstein, H., Olszewski, A., & Rohrer, L. (2020). Indexing effects in single-case experimental designs by estimating the percent of goal obtained. *Evidence-Based Communication Assessment and Intervention, 14*(1-2), 6-27.

Ferron, J., Joo S., & Levin J. R. (2017). A monte carlo evaluation of masked visual analysis in response-guided versus fixed-criteria multiple-baseline designs. *Journal of Applied Behavior Analysis, 50,* 701-716.

Ferron, J., Moeyaert, M., Van de Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: implication for design and analysis. *Psychological Methods, 19* (4), 493-510.

Gast, L. D., & Ledford, R. J. (2018). Research approaches in applied setting. In Ledford, R. J. & Gast, L. D. (Eds.), *Single case research methodology: Applications in special education and behavioral sciences,* 1-26. New York: Routledge.

Gingerich, W. J. (1984). Meta-Analysis of Applied Time-Series Data. *The Journal of Applied Behavioral Science*, *20*(1), 71–79.

Hains, A. H., & Baer, D. M. (1989). Interaction effects in multielement designs: Inevitable, desirable, and ignorable. *Journal of Applied Behavior Analysis*, *22(1),* 57-69.

Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983–2007. Education and Training in Autism and Developmental Disabilities, 45,187–202.

Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavior change.* New York: Pergamon.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71(2),* 165-179.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings (2nd ed.).* New York, NY: Oxford University Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Erlbaum.

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.

Kratochwill, T. R., & Levin, J. R. (2014). *Single-Case Intervention Research: Methodological and Statistical Advances*. Washington D.C.: American Psychological Association.

Lalonde, T. L. (2015). Modeling time-dependent covariates in longitudinal data analyses. *Innovative Statistical Methods for Public Health Data*, 57-79.

Lobo, M.A., Moeyaert, M., Cunha, A.B., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, 41, 187–197.

Michiels, B., & Onghena, P. (2018). Randomized single-case AB phase designs: Prospects and pitfalls. *Behavior research methods*, 1-23.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, T., & Van den Noortgate, W. (2013a). Modeling external events in the three-level analysis of multiple-baseline across participants designs: A simulation study. *Behavior Research Methods, 45*, 547–559.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013b). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research, 48*, 719–748.

Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2014). Three-level analysis of single-case experimental data: Empirical validation. The Journal of Experimental Education, 82, 1-21.

Owens, C. M., & Ferron, J. M. (2012). Synthesizing single-case studies: A Monte Carlo examination of a three-level meta-analytic model. *Behavior Research Methods*, 44, 795-805.

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology, 68*, 99-112.

Richard, S. B., Taylor, R., Ramasamy, R., & Richards, R. Y. (1999). Single-subject research: Applications in educational and clinical settings. *Belmont,* CA: Wadsworth.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*, 971–980.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, *52*(2), 123-147.

Wong, V. C., Wing, C., Steiner, P. M., Wong, M., & Cook, T. D. (2012). Research designs for program evaluation. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of psychology: Research methods in psychology, 2*, 316–341.

APPENDIX A: POTENTIAL CODING FOR EXTRANEOUS VARIABLES IN THE REVIEWED ARTICLES

Table 7. Potential Coding for the Extraneous Variables in the Reviewed Articles.

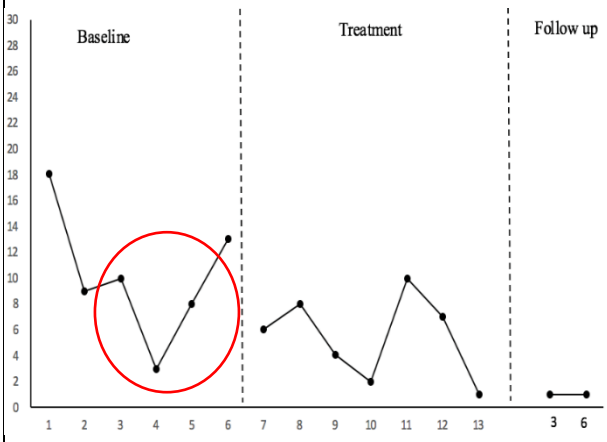
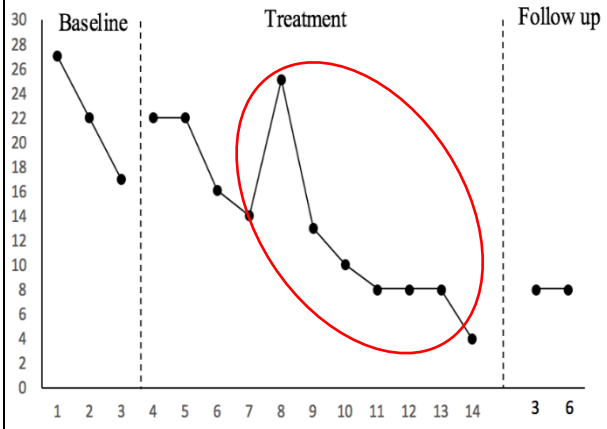
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
1. The effects associated with metacognitive therapy in postpartum depression.	"Participant B's scores decreased during baseline, but this coincided with a time when her partner was on leave from work and she received much more support than usual. Her scores increased once this period passed; Participant D presented as severely depressed and expressed suicidal ideation during the assessment;"(Bevan, Wittkowski, & Wells, 2013, p. 73).	The "partner on leave" could be coded as 0 with no leave and 1 with leave. A brief behavioral activation intervention could be treated as a continuous variable, that the effect of the behavioral activation intervention will be gradually faded.	 

Table 7.(Continued).

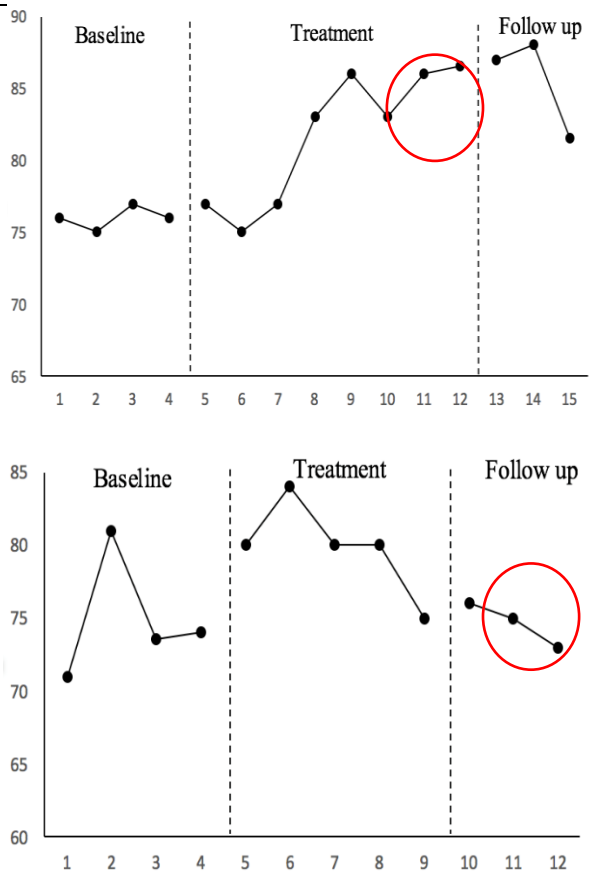
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
2. The impact of wellness-focused supervision on mental health counseling practicum students.	"Each person showed a decreased score during Weeks 9 or 10. One hypothesis for this low score is that there was a significant program challenge that occurred at that time" (Walen, Gage, & Lindo, 2016, p. 470).	A significant program challenge for week 9 or 10 could be coded as 1. Other weeks could be coded as 0.	

Table 7.(Continued).

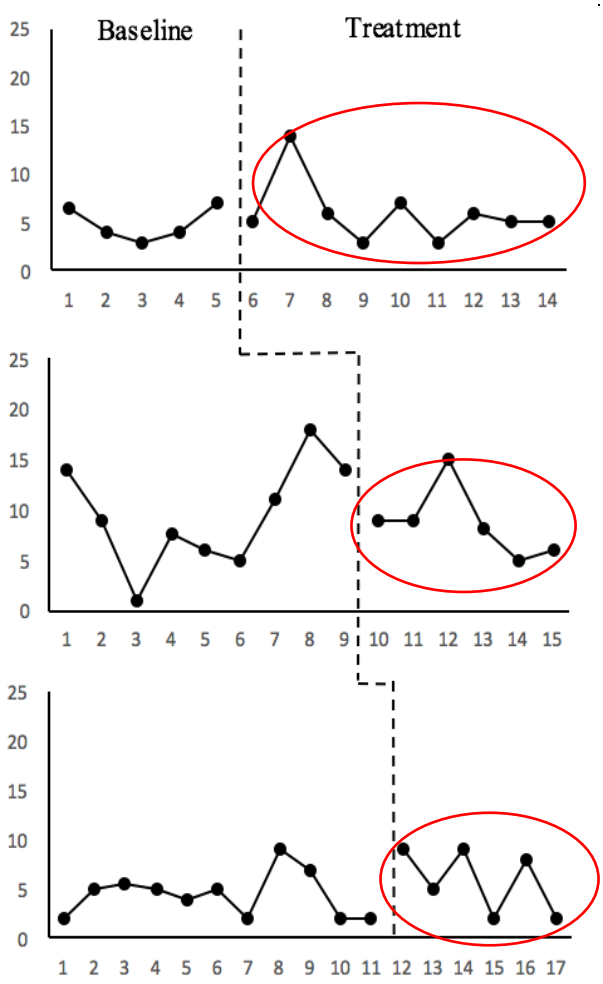
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
3. Evaluation of increasing antecedent specificity in goal statements on adherence to positive behavior-management strategies.	*"Varying activities across observation sessions may have been associated with differences in the saliency of discriminative stimuli for praise. For instance, opportunities to praise students while they were in the computer lab or quietly working on individual tasks may have been less salient than when students were encouraged to provide vocal responses during group instruction. The possibility remains that some activities (e.g., reading) present less salient discriminative stimuli for praise relative to others (e.g., math)" (Cohrs, & Shriver, 2016, p. 777).*	*The influence of varying activities (computer lab, quietly working on individual tasks, or group instruction) across sessions may influence the outcome variable. And it could be treated as continuous variable.*	

Table 7.(Continued).

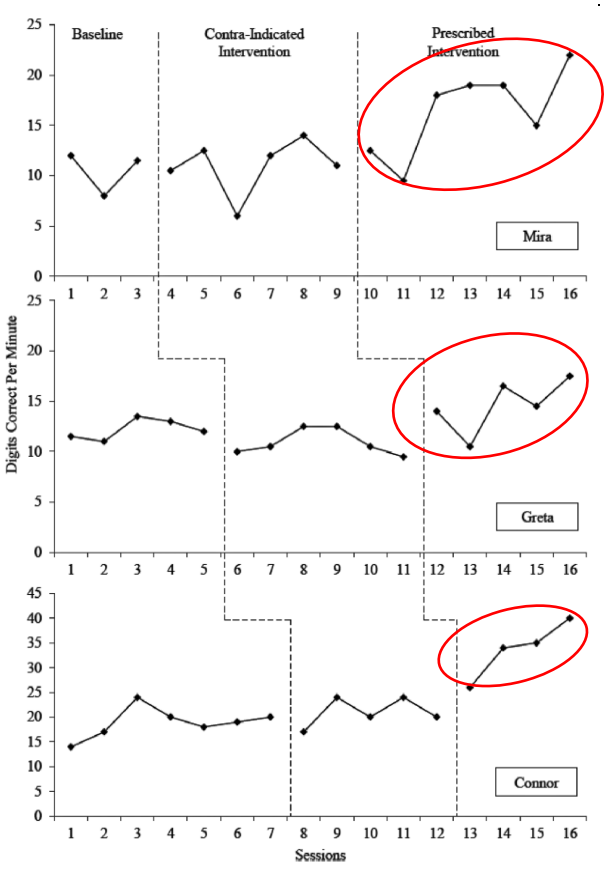
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
4. Using a conceptual understanding and procedural fluency heuristic to target math interventions.	"The study did not take into account the quality or focus of core instruction. The multiple-baseline design allowed for internally valid conclusions about the intervention, but the effect that classroom instruction also had on the data is unknown" (Burns, et al., 2015, p. 58).	The effect of the classroom instruction could be a continuous variable since it may vary across time.	

Table 7.(Continued).

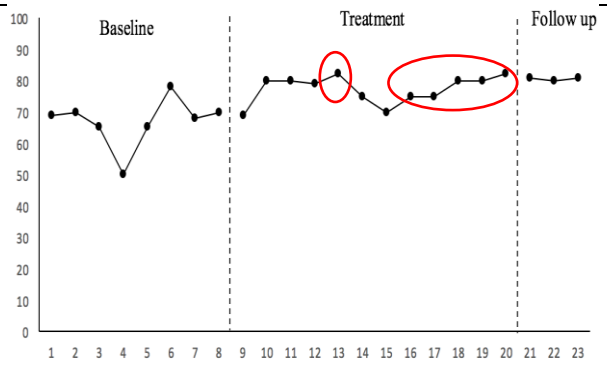
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
5. Evaluating the effects of on-task in a box as a class-wide intervention.	"Treatment integrity for Classroom C fell below 80% during implementation, which coincided with a decrease in class-wide on-task behavior. For sessions with low integrity, the teacher of Classroom C did not collect self-monitoring forms and did not provide reinforcement if the class met its goal. Therefore, the teacher was retrained on the intervention procedures using performance feedback" (Battaglia, Radley, & Ness, 2015, p. 752).	Teacher of classroom c did not collect self-monitoring forms and did not provide reinforcement on several sessions could be coded as 1, and other sessions could be coded 0.	

Table 7.(Continued).

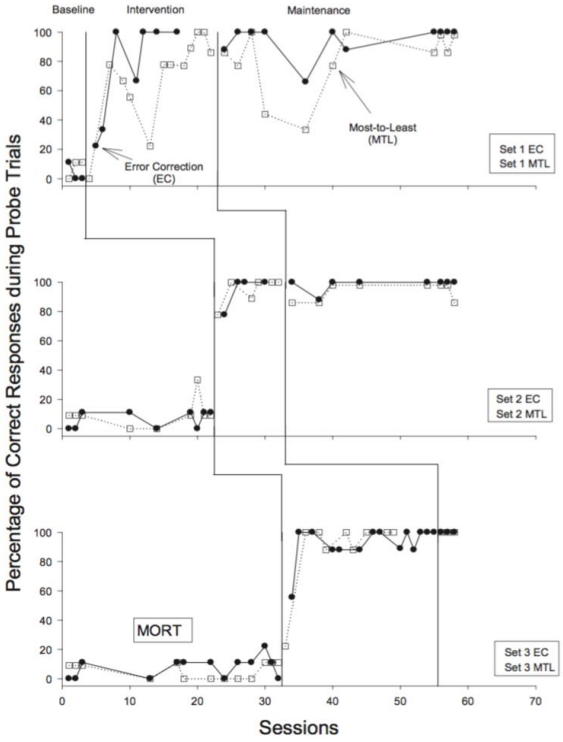
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
8. Comparison of most-to-least to error correction for teaching receptive labelling for children diagnosed with autism.	"Mort had a previous history of engaging in non-compliance when provided corrective feedback. Anecdotally, throughout this study, Mort infrequently engaged in non-compliant behaviours; in addition, non-compliance was distributed evenly across both teaching conditions. However, his previous history may have accounted for quicker responding in the most-to-least condition for those targets he did master" (Leaf et al., 2016, p. 224).	"Infrequently engaged in non-compliant behaviours" duing the observations could be coded as 1, and other obseration occations could be coded 0.	

Table 7.(Continued).

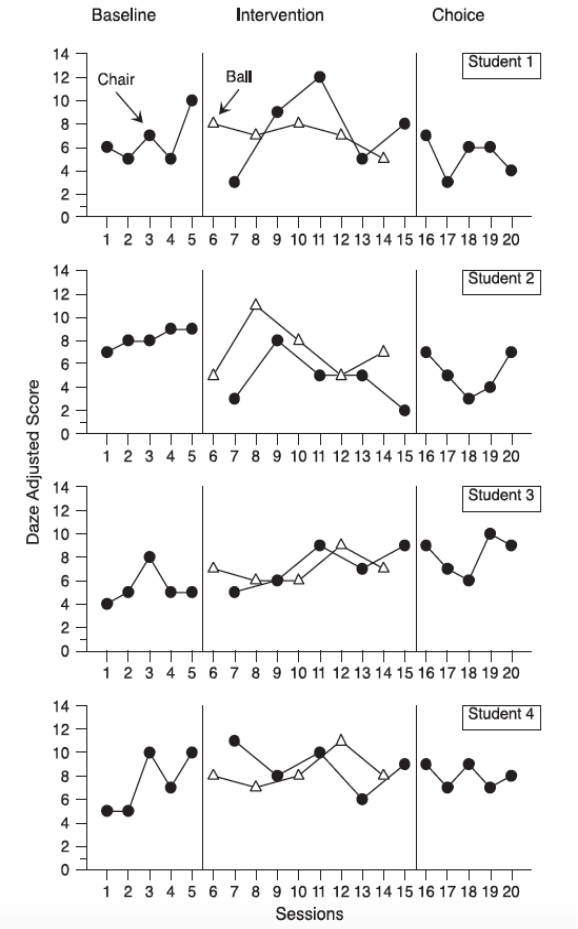
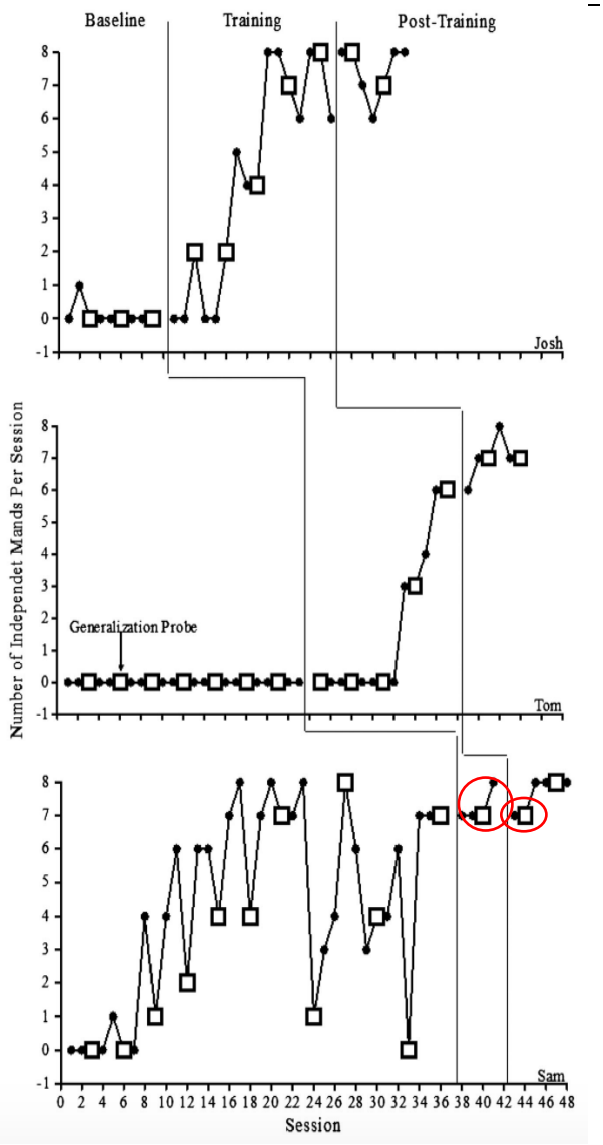
Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
9. Effects of therapy balls on children diagnosed with attention deficit hyperactivity disorder.	"Due to the various assignments/tasks that took place during independent seatwork time, data on classroom work samples (e.g., work completion and work accuracy) could not be collected consistently, which would have provided valuable information" (Taipalus, Hixson, Kanouse, Wyse, & Fursa, 2017, p. 425).	"Data on classroom work sample" (can be concluded as different classroom work tasks) could be a continuous variable and it could vary across person and across time since the participants might score differently at different observation occasions on their work sheets.	

Table 7.(Continued).

Articles	Suggested Extraneous Variables	Potential Coding for Extraneous variables	
10. Effects of video modeling and feedback on mothers' implementation of peer-to-peer manding.	"One reason for the lack of functional control comes from Sam's data. His mands steadily increased and remained high but variable throughout baseline, and it was not until his mother entered the training phase that both his and her performances remained high and steady. This may have been because Emma (Sam's mother) began to set up the materials and play environment correctly (i.e. task analysis, steps 1–4) and intermittently used various prompts" (Madzharova, & Sturmey, 2015, p. 280).	During the observation sessions, the observations that Emma (Sam's mom) incorrectly used materials could be coded as 1, and the observations that Emma correctly used the materials could be coded 0.	

APPENDIX B: SIMULATION RESULTS

Table 8. Eta-Squared for Bias, SE, and 95% Confidence Interval of the Treatment Effect Estimates for Bernoulli Distribution Condition.

Conditions	Eta-Square		
	Bias	SE	CI
Participant	0	0.29	0
Observation	0	0.29	0
Participant*Observation	0	0.03	0
True	0	0	0
Participant*True	0	0	0
Observation*True	0	0	0
Participant*Observation*True	0	0	0
Extraneous Variable Effect	1	0.29	0.998
Participant*Extraneous Variable Effect	0	0.03	0
Observation*Extraneous Variable Effect	0	0.03	0
Participant*Observation*Extraneous Variable Effect	0	0.03	0
True*Extraneous Variable Effect	0	0	0
Participant*True*Extraneous Variable Effect	0	0	0
Observation*True*Extraneous Variable Effect	0	0	0
Participant* Observation*True*Extraneous Variable Effect	0	0	0

Table 9. Eta-Squared for Bias of the Treatment Effect Estimates for Piecewise Distribution Condition.

Conditions	Eta-Square		
	Bias	SE	CI
Participant	0	0.33	0.08
Observation	0.03	0.33	0.02
Participant*Observation	0.03	0.33	0
True	0.02	0	0
Participant*True	0	0	0
Observation*True	0.01	0	0
Participant*Observation*True	0	0	0
Extraneous Variable Effect	0.81	0	0.78
Participant*Extraneous Variable Effect	0	0	0.08
Observation*Extraneous Variable Effect	0.03	0	0.03
Participant*Observation*Extraneous Variable Effect	0	0	0
True*Extraneous Variable Effect	0.02	0	0
Participant*True*Extraneous Variable Effect	0	0	0
Observation*True*Extraneous Variable Effect	0.01	0	0
Participant* Observation*True*Extraneous Variable Effect	0	0	0

Table 10. Bias, Relative Bias, Coverage, and Standard Errors Estimates for miss-specified model for Bernoulli Distribution Condition.

Conditions		True = 0.6	True = 0.3	(True = 0)	Avg.
4(16)	Effect Est (SE)	0.72(0.03)	0.42(0.03)	0.12(0.03)	
	Bias	.12	.12	.12	.12
	Relative Bias	0.2	0.4		
	95%CI	.06	.08	.07	.07
8(16)	Effect Est (SE)	0.72(0.02)	0.42(0.02)	0.12(0.02)	
	Bias	.12	.12	.12	.12
	Relative Bias	0.2	0.4		
	95%CI	.01	.01	.01	.01
4(32)	Effect Est (SE)	0.72(0.02)	0.42(0.02)	0.12(0.02)	
	Bias	.12	.12	.12	.12
	Relative Bias	0.2	0.4		
	95%CI	.01	.01	.01	.01
8(32)	Effect Est (SE)	0.72(0.01)	0.42(0.01)	0.12(0.01)	
	Bias	.12	.12	.12	.12
	Relative Bias	0.2	0.4		
	95%CI	0	0	0	0

Table 11. Bias, Relative Bias, Coverage, and Standard Errors Estimates for correct-specified model for Bernoulli Distribution Condition.

Conditions		True = 0.6	True = 0.3	(True = 0)	Avg.
4(16)	Effect Est (SE)	0.60 (0.02)	0.30 (0.02)	0 (0.02)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.95	.95
8(16)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.94	.94
4(32)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.93	.95	.94
8(32)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.94	.94

Table 12. Bias, Relative Bias, Coverage, and Standard Errors Estimates for miss-specified model for Piecewise Distribution Condition.

Conditions		True = 0.6	True = 0.3	(True = 0)	Avg.
4(16)	Effect Est (SE)	0.65 (0.02)	0.35(0.02)	0.05(0.02)	
	Bias	.05	.03	.05	.04
	Relative Bias	0.08	0.1		
	95%CI	.48	.46	.48	.47
8(16)	Effect Est (SE)	0.65 (0.01)	0.35(0.01)	0.05(0.01)	
	Bias	.05	.05	.05	.05
	Relative Bias	0.08	0.17		
	95%CI	.10	.07	.07	.08
4(32)	Effect Est (SE)	0.62 (0.01)	0.33(0.01)	0.03(0.01)	
	Bias	.02	.03	.05	.03
	Relative Bias	0.03	0.1		
	95%CI	.65	.64	.65	.65
8(32)	Effect Est (SE)	0.62 (0.01)	0.32(0.01)	0.03(0.01)	
	Bias	.02	.02	.05	.03
	Relative Bias	0.03	0.07		
	95%CI	.32	.30	.30	.31

Table 13. Bias, Relative Bias, Coverage, and Standard Errors Estimates for correct-specified model for Piecewise Distribution Condition.

Conditions		True = 0.6	True = 0.3	(True = 0)	Avg.
4(16)	Effect Est (SE)	0.60 (0.02)	0.30 (0.02)	0 (0.02)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.95	.95
8(16)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.94	.94
4(32)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.93	.95	.94
8(32)	Effect Est (SE)	0.60 (0.01)	0.30 (0.01)	0 (0.01)	
	Bias	0	0	0	0
	Relative Bias	0	0		
	95%CI	.94	.94	.94	.94

APPENDIX C: SAS CODE

```
*+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

The following simulation code is for burnolli distribution.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;
ods graphics off;

ods html close;

ods _all_ close;

proc printto log = junk;


proc printto log='C:\Users\tuq44546\Desktop\ke\diss\log.txt';

proc printto print = 'C:\Users\tuq44546\Desktop\ke\diss\output.txt';


%global _print_;

%let _print_ = off;

data j0;

input Estimate StdErr Lower Upper;

datalines;

. . . .

. . . .

;

data j00;

input Estimate StdErr Lower Upper;

datalines;

. . . .

. . . .

. . . .

;
```

```
%macro hlmsim (n, n1, n2, v2lev, phi, true, br);
%do i=1 %to &n;

dm 'odsresults; clear';


proc iml;

*++++++++++++++++++++++++++++++++++++++++++++++++++++++++
This part of the program creates the initial data set,
which contains the following variables:
IDlevel2: level 2 ID
time: potential level-1 predictor
phase: dichotomous level-1 predictor (0=baseline, 1=treatmetnt)
y: outcome
++++++++++++++++++++++++++++++++++++++++++++++++++++++++;
create j1 var{IDlevel2 time timec  phase EV y};


do ID=1 to &n2;
cut=0;
* cut = the last baseline observation;
if &n2=4 then do;
if &n1=16 & (ID = 1) then cut = 5;
if &n1=16 & (ID = 2) then cut = 7;
if &n1=16 & (ID = 3) then cut = 9;
if &n1=16 & (ID = 4) then cut = 11;
if &n1=32 & (ID = 1) then cut = 10;
if &n1=32 & (ID = 2) then cut = 14;
if &n1=32 & (ID = 3) then cut = 18;
if &n1=32 & (ID = 4) then cut = 22;
```

```
end;


if &n2=8 then do;

if &n1=16 & (ID = 1 | ID = 5) then cut = 5;

if &n1=16 & (ID = 2 | ID = 6) then cut = 7;

if &n1=16 & (ID = 3 | ID = 7) then cut = 9;

if &n1=16 & (ID = 4 | ID = 8) then cut = 11;

if &n1=32 & (ID = 1 | ID = 5) then cut = 10;

if &n1=32 & (ID = 2 | ID = 6) then cut = 14;

if &n1=32 & (ID = 3 | ID = 7) then cut = 18;

if &n1=32 & (ID = 4 | ID = 8) then cut = 22;

end;


IDlevel2=j(&n1,1,ID);

time=j(&n1,1,0);

timec=j(&n1,1,0); *centered time;

phase=j(&n1,1,0);

EV = j(&n1, 1, 0);



do ii=1 to &n1;
time[ii,1]=ii-1;
timec[ii,1]=ii-cut-1;


if ii > cut then phase[ii,1]=1;

end;

*creating level 1 error armasim: auto-correlated errors;

x=round(1000000*ranuni(0));

rr=armasim({1,&phi},0,0,.05,&n1,x);


r=rr;
```

```
*level 2 errors;

u0b=rannor(0);

*repeated the same person error for all the observation of a person;

u0=repeat(u0b*sqrt(&v2lev),&n1);*variation for baseline;

u1b=rannor(0);

u1=repeat(u1b*sqrt(&v2lev),&n1);*variation for treatment;

u2b=rannor(0);

u2=repeat(u2b*sqrt(&v2lev),&n1);*variation for EV variable;



*grand mean for intercept;

gamma00=0; *baseline avg = 0;

*treatment effect;
gamma10=&true;

gamma20=&br;

intercep=gamma00+u0;

effect=gamma10+u1;

*creating ev variable effect;
*generate uniform distribution as k, the numbers are from 0 to 1. if the
number <=.4 then ev=0,if the number >.4 then ev=0.2.
ev=0 for all baseline;


do iii=1 to &n1;
k=round(1000000*ranuni(0));
EV[iii,1]=uniform(k);
if EV[iii,1]<=.4 then EV[iii,1]=0;
if EV[iii,1]>.4 then EV[iii,1]=1;
if phase[iii,1]=0 then EV[iii,1]=0;
end;

EVb = EV*gamma20 + u2;

y = intercep + effect#phase + EVb +r;



*print intercep true phase r y;
```

```
*print IDlevel2 u0 u1 r gamma00 gamma10  intercep effect time timec phase cut
y ev;

append;

end; *closes the person loop;


close j1;

quit;

*****************************end of data
generation*****************************************************;

*this is generating the missing value for convergence, we assuming the value
are all missing, and if the estimates come out, it will replace the missing
value, otherwise,it will keep as missing;

data fixedpar1;

set j0;

data fixedpar2;

set j00;


* Model 1: the miss specified model;

proc mixed data =j1 covtest cl;

class idlevel2 ;

model y = phase / s cl alpha = .05 ddfm = kenward;

random int phase / sub = idlevel2; *vary across people;

repeated / sub = idlevel2 type=ar(1);

ods output solutionF=fixedpar1    /*F means the fixed effect;*/

(keep = Estimate StdErr Lower Upper);


Run;


* Model 2: the correct specified model;

proc mixed data =j1 covtest cl;
```

```
class idlevel2 ;

model y = phase EV/ s cl alpha = .05 ddfm = kenward;

random int phase EV/ sub = idlevel2; *vary across people;

repeated / sub = idlevel2  type=ar(1);

ods output solutionF=fixedpar2

(keep = Estimate StdErr Lower Upper);


Run;


%do j=1 %to 2;

      data fixed&j;

            set fixedpar&j;

            w&j = estimate; output;

            w&j = StdErr; output;

            w&j = lower; output;

            w&j = upper; output;

            drop estimate StdErr lower upper;

      run;

%end;

*transform the results to one column, and rename the column;


proc transpose data = fixed1
out = fixed1

(rename = (

col1=est1_int           col2=se1_int            col3=l1_int
      col4=u1_int

col5=est1_pha           col6=se1_pha            col7=l1_pha
      col8=u1_pha

));

run;

*proc print;

*run;
```

```
proc transpose data = fixed2
out = fixed2

(rename = (

col1=est2_int           col2=se2_int            col3=l2_int
      col4=u2_int

col5=est2_pha2          col6=se2_pha2           col7=l2_pha2
      col8=u2_pha2

col9=est2_EV            col10=se2_EV            col11=l2_EV
      col12=u2_EV
));

run;

*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

The following statements merge the output data sets resulting with one row

of data containing the point estimates, lower limit, upper limit, for

each fixed effect and variance component, for each DF method. For the fixed

effects, it also contains the DF. The data set is then appended with

a new row for each simulated data set.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;


%do j=1 %to 2;

      data all&j;
    set fixed&j;
    counter = &i;

      run;

%end;


%if &i = 1 %then %do;

      %do j=1 %to 2;

            data sumstat&j;
        set all&j;

            run;

      %end;
```

```
%end;


%else %do;

      %do j=1 %to 2;

            data sumstat&j;
        merge sumstat&j all&j;
        by counter;

            run;

      %end;

%end;

%end;*this end is for the very first do loop for &n;


*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
The following set of commands creates a series of indicator variables
based on whether the fixed effect parameter and the variance component
parameter falls between the lower and upper limit. It then computes
the width of the confidence interval for each effect for each DF method.
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;

*track the confidence interval coverage;
data sumstat1;
set sumstat1;
cov1_pha=0;
if (l1_pha <= gamma10) & (gamma10 <= u1_pha) then cov1_pha=1;
if l1_pha=. then cov1_pha=.;
wid1_pha =  u1_pha      - l1_pha;
run;



data sumstat2;
```

```
set sumstat2;

cov2_pha2=0;

if (l2_pha2 <= gamma10) & (gamma10 <= u2_pha2) then cov2_pha2=1;

if l2_pha2=. then cov2_pha2=.;

wid2_pha2          =     u2_pha2            -     l2_pha2;

run;

*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++++

Means are then calculated, giving estimates of bias in the fixed and variance

component effect estimates along with the average standard error,

the coverage probabilities for each effect,

and the average CI width for each effect.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++;

proc means noprint data = sumstat1;

var

est1_pha           se1_pha            cov1_pha           wid1_pha;

output out=meanstat1

mean =

est1_pha           se1_pha            cov1_pha           wid1_pha

n = n_sims

std (est1_pha)     = sest1_pha

;

run;



proc means noprint data = sumstat2;

var

est2_pha2          se2_pha2           cov2_pha2          wid2_pha2 ;

output out=meanstat2
```

```
mean =
est2_pha2          se2_pha2           cov2_pha2          wid2_pha2
n = n_sims
std (est2_pha2)    = sest2_pha2;
run;


ods listing;
%global _print_;
      %let _print_ = on;
%do j=1 %to 2;
      data meanstat&j;
            set meanstat&j;
            reps=&n;
            n1size=&n1;
            n2size=&n2;
            v2level=&v2lev;
            phi=-1*&phi;
            conv=n_sims/reps;
      run;
      proc print data=meanstat&j;
      run;
%end;


*for miss specified model;
data meanstat1;
      set meanstat1;
* file print;
      file "C:\Users\tuq44546\Desktop\ke\diss\miss-model.txt" mod lrecl=800;
```

```
      put @1(
est1_pha          se1_pha           cov1_pha          wid1_pha
n_sims
sest1_pha
reps
            n1size
            n2size
            v2level
            phi
            conv) (9.4);
run;

*for correct model;
data meanstat2;
      set meanstat2;
* file print;
      file "C:\Users\tuq44546\Desktop\ke\diss\correct-model.txt" mod
lrecl=800;
      put @1(
est2_pha2         se2_pha2          cov2_pha2         wid2_pha2
n_sims
sest2_pha2
reps
            n1size
            n2size
            v2level
            phi
            conv) (9.4);
run;
```

```
%mend;

%hlmsim(n=10,n1=32,n2=4,v2lev=0.0001,phi=-0.2, true=0.3, br=0.2);

run;
*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

The following simulation code is for piecewise distribution.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;
ods graphics off;

ods html close;

ods _all_ close;

*proc printto log = junk;


*proc printto log='';

*proc printto print = '';


%global _print_;

%let _print_ = off;

data j0;

input Estimate StdErr Lower Upper;

datalines;

. . . .

. . . .

;

%macro hlmsim (n, n1, n2, v2lev, phi, true);

%do i=1 %to &n;


dm 'odsresults; clear';


proc iml;


*++++++++++++++++++++++++++++++++++++++++++++++++++++++

This part of the program creates the initial data set,

which contains the following variables:
```

```
IDlevel2: level 2 ID

time: potential level-1 predictor

timec: centered time

phase: dichotomous level-1 predictor (0=baseline, 1=treatmetnt)

y: outcome

++++++++++++++++++++++++++++++++++++++++++++++++++++++++;

create j1 var{IDlevel2 time timec  phase EV y };


do ID=1 to &n2;

cut=0;

* cut = the last baseline observation;

if &n2=4 then do;

if &n1=16 & (ID = 1) then cut = 5;

if &n1=16 & (ID = 2) then cut = 7;

if &n1=16 & (ID = 3) then cut = 9;

if &n1=16 & (ID = 4) then cut = 11;

if &n1=32 & (ID = 1) then cut = 10;

if &n1=32 & (ID = 2) then cut = 14;

if &n1=32 & (ID = 3) then cut = 18;

if &n1=32 & (ID = 4) then cut = 22;

end;


if &n2=8 then do;

if &n1=16 & (ID = 1 | ID = 5) then cut = 5;

if &n1=16 & (ID = 2 | ID = 6) then cut = 7;

if &n1=16 & (ID = 3 | ID = 7) then cut = 9;

if &n1=16 & (ID = 4 | ID = 8) then cut = 11;

if &n1=32 & (ID = 1 | ID = 5) then cut = 10;
```

```
if &n1=32 & (ID = 2 | ID = 6) then cut = 14;

if &n1=32 & (ID = 3 | ID = 7) then cut = 18;

if &n1=32 & (ID = 4 | ID = 8) then cut = 22;

end;


IDlevel2=j(&n1,1,ID);

time=j(&n1,1,0);

timec=j(&n1,1,0); *centered time;

phase=j(&n1,1,0);

phase1=j(&n1,1,0);

phase2=j(&n1,1,0);

EV = j(&n1, 1, 0);


do ii=1 to &n1;
time[ii,1]=ii-1;
timec[ii,1]=(ii-cut)-1;
timec2[ii,1]=(ii-cut)-3;

****fix this part*********************************************;

if ii > cut then phase[ii,1]=1;

if ii > cut & ii < cut+3 then phase1[ii,1]=1;

if ii > cut & ii < cut+3 then phase2[ii,1]=1;

if ii > cut+2 then phase2[ii,1]=1;


*change the "cut" to the random number that generated from uniform
distribution;



if ii > cut then phase[ii,1]=1;

end;

*creating level 1 error armasim: auto-correlated errors;

x=round(1000000*ranuni(0));
```

```
rr=armasim({1,&phi},0,0,.05,&n1,x);

r=rr;



*level 2 errors;

u0b=rannor(0);

*repeated the same person error for all the observation of a person;

u0=repeat(u0b*sqrt(&v2lev),&n1);*variation for baseline;

u1b=rannor(0);

u1=repeat(u1b*sqrt(&v2lev),&n1);*variation for treatment;

u2b=rannor(0);

u2=repeat(u2b*sqrt(&v2lev),&n1);*variation for EV variable;


*grand mean for intercept;

gamma00=0; *baseline avg = 0;

*treatment effect;
gamma10=&true;


intercep=gamma00+u0;

effect=gamma10+u1;

EVb = EV+u2;

y = intercep + effect#phase + EVb +r;



*print intercep true phase r y;


*print IDlevel2 u0 u1 r gamma00 gamma10  intercep effect time timec phase cut
y ev;

append;

end; *closes the person loop;
```

```
close j1;

quit;

*****************************end of data
generation*************************************************;

*this is generating the missing value for convergence, we assuming the value
are all missing, and if the estimates come out, it will replace the missing
value, otherwise,it will keep as missing;

data fixedpar1;

set j0;

data fixedpar2;

set j0;


* Model 1: the miss specified model;

proc mixed data =j1 covtest cl;

class idlevel2 ;

model y = phase / s cl alpha = .05 ddfm = kenward;

random int phase / sub = idlevel2; *vary across people;

repeated / sub = idlevel2;

ods output solutionF=fixedpar1

(keep = Estimate StdErr Lower Upper);


Run;


* Model 2: the correct specified model;

proc mixed data =j1 covtest cl;

class idlevel2 ;

model y = phase EV/ s cl alpha = .05 ddfm = kenward;

random int phase EV/ sub = idlevel2; *vary across people;

repeated / sub = idlevel2;

ods output solutionF=fixedpar2
```

```
(keep = Estimate StdErr Lower Upper);

Run;

%do j=1 %to 2;
      data fixed&j;
            set fixedpar&j;
            w&j = estimate; output;
            w&j = StdErr; output;
            w&j = lower; output;
            w&j = upper; output;
            drop estimate StdErr lower upper;
      run;
%end;

*transform the results to one column, and rename the column;

proc transpose data = fixed1
out = fixed1

(rename = (

col1=est1_int           col2=se1_int            col3=l1_int
      col4=u1_int

col5=est1_pha           col6=se1_pha            col7=l1_pha
      col8=u1_pha

));

run;

*proc print;

*run;

proc transpose data = fixed2
out = fixed2

(rename = (

col1=est2_int           col2=se2_int            col3=l2_int
      col4=u2_int
```

```
col5=est2_pha2          col6=se2_pha2           col7=l2_pha2
      col8=u2_pha2

col9=est2_EV            col10=se2_EV            col11=l2_EV
      col12=u2_EV
));

run;

/*aggreage the generated datasets as dataset x*/
/*data x;*/
/*set x j1;*/
/*run;*/

*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

The following statements merge the output data sets resulting with one row

of data containing the point estimates, lower limit, upper limit, for

each fixed effect and variance component, for each DF method. For the fixed

effects, it also contains the DF. The data set is then appended with

a new row for each simulated data set.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;


%do j=1 %to 2;

      data all&j;
    set fixed&j;
    counter = &i;

      run;

%end;


%if &i = 1 %then %do;

      %do j=1 %to 2;

            data sumstat&j;
        set all&j;

            run;

      %end;

%end;


%else %do;
```

```
      %do j=1 %to 2;

           data sumstat&j;
        merge sumstat&j all&j;
        by counter;

           run;

      %end;

%end;

%end;*this end is for the very first do loop for &n;


*++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
The following set of commands creates a series of indicator variables
based on whether the fixed effect parameter and the variance component
parameter falls between the lower and upper limit. It then computes
the width of the confidence interval for each effect for each DF method.
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++;

*track the confidence interval coverage;
data sumstat1;
set sumstat1;
cov1_pha=0;
if (l1_pha <= gamma10) & (gamma10 <= u1_pha) then cov1_pha=1;
if l1_pha=. then cov1_pha=.;
wid1_pha =  u1_pha      - l1_pha;
run;



data sumstat2;
set sumstat2;
cov2_pha2=0;
```

```
if (l2_pha2 <= gamma10) & (gamma10 <= u2_pha2) then cov2_pha2=1;

if l2_pha2=. then cov2_pha2=.;

wid2_pha2          =     u2_pha2            -     l2_pha2;

run;

*+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++++

Means are then calculated, giving estimates of bias in the fixed and variance

component effect estimates along with the average standard error,

the coverage probabilities for each effect,

and the average CI width for each effect.

+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++;

proc means noprint data = sumstat1;

var

est1_pha           se1_pha            cov1_pha           wid1_pha;

output out=meanstat1

mean =

est1_pha           se1_pha            cov1_pha           wid1_pha

n = n_sims

std (est1_pha)     = sest1_pha

;

run;



proc means noprint data = sumstat2;

var

est2_pha2          se2_pha2           cov2_pha2          wid2_pha2 ;

output out=meanstat2

mean =

est2_pha2          se2_pha2           cov2_pha2          wid2_pha2
```

```
n = n_sims
std (est2_pha2)    = sest2_pha2;
run;


ods listing;
%global _print_;
      %let _print_ = on;
%do j=1 %to 2;
      data meanstat&j;
            set meanstat&j;
            reps=&n;
            n1size=&n1;
            n2size=&n2;
            v2level=&v2lev;
            phi=-1*&phi;
            conv=n_sims/reps;
      run;
      proc print data=meanstat&j;
      run;
%end;


*for miss specified model;
data meanstat1;
      set meanstat1;
* file print;
      file "C:\Users\tuq44546\Desktop\ke\diss\miss-model.txt" mod lrecl=800;
      put @1(
est1_pha          se1_pha           cov1_pha          wid1_pha
n_sims
```

```
sest1_pha

reps

            n1size

            n2size

            v2level

            phi

            conv) (9.4);

run;


*for correct model;

data meanstat2;

      set meanstat2;

* file print;

      file "C:\Users\tuq44546\Desktop\ke\diss\correct-model.txt" mod
lrecl=800;

      put @1(

est2_pha2         se2_pha2          cov2_pha2         wid2_pha2

n_sims

sest2_pha2

reps

            n1size

            n2size

            v2level

            phi

            conv) (9.4);

run;


%mend;

%hlmsim(n=5,n1=16,n2=4,v2lev=0.0001,phi=-0.2, true=0.6);
```

```
run;
```