Impact of a Priori Decision-Making and Response-Guided Decision-Making on Obtained Effect Sizes

By

Paige Bennett Eyler

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May 10, 2024

Nashville, Tennessee

Approved:

Jennifer Ledford, Ph.D.

Mary Louise Hemmeter, Ph.D.

Kristen Granger, Ph.D.

Katie Tuck, Ph.D.

I dedicate this dissertation to myself for the many years of studying, crying, swearing, and celebrating. It has been a wild ride.

ACKNOWLEDGMENTS

I want to thank my supervisor, Dr. Jennifer Ledford, for her invaluable guidance, encouragement, and unwavering support throughout my time at Vanderbilt University. Her expertise, feedback, and guidance have been instrumental in shaping who I am today.

I am also profoundly thankful to the members of my dissertation committee, Mary Louise Hemmeter, Kristen Granger, and Katie Tuck, for their insightful comments, suggestions, and scholarly contributions, which have enriched the quality and depth of this dissertation.

Special thanks are due to my husband and family for their boundless love, understanding, and encouragement during this journey. Finally, I would like to acknowledge my research team. Without them, this dissertation would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Single-case designs (SCD) are quantitative experiments where participants serve as their own control for evaluating the relationship between a dependent and independent variable (Ledford & Gast 2018). SCDs are increasingly more prevalent in research and are commonly used within special education for assessing the effectiveness of interventions tailored to individual needs (Shepley et al., 2022; Hammond & Gast, 2010). One of the many benefits of SCD research is that it can provide valuable insights for practitioners and researchers working in fields where individual variability is likely. Benefits of SCDs include the lack of a control group, which has both practical and social validity implications; the ability to draw causal relations given one or only a few participants, which is especially beneficial when low-incidence populations are of interest; and the ability to make data-based decisions, which is in line with most clinical and educational practitioners' needs (Ledford & Gast, 2018; Ledford et al., 2023; Shadish & Sullivan, 2011).

The most commonly used SCD is the multiple baseline design (MBD; Hammond & Gast, 2010; Ledford et al., 2019a; Pustejovsky et al., 2019; Shadish & Sullivan, 2011). This design is flexible because it can be used for reversible and non-reversible behaviors and does not require intervention withdrawal (Ledford et al., 2019b). Traditionally, data from multiple participants, behaviors, or contexts were (a) collected concurrently and (b) with varying baseline lengths to control for threats to internal validity visible via visual analysis (i.e., vertical analysis), with the most common design type being the *concurrent* multiple baseline across participants design (Bear et al., 1968).

Watson and Workman introduced the concept of a non-concurrent multiple baseline design in 1981 to increase the usefulness of MBD in applied settings. Non-concurrent variations of this design refer to designs in which data are collected from multiple participants, behaviors, or contexts without concurrent measurement (Watson & Workman, 1981). Non-concurrent multiple baseline designs were traditionally eschewed for concurrent versions (Ledford & Gast, 2018). However, recent arguments suggest these designs may be equally valid in at least some situations (Lanovaz & Turgeon, 2020; Ledford & Zimmerman, 2023; Slocum et al., 2022a). Specifically, the authors argue that historical standards are not necessarily grounded in empirical evidence (Lanovaz & Turgeon, 2020). With predetermined and random assignment (Watson & Workman, 1981), the probability of a coincidental event impacting participants across multiple tiers is minimal (Slocum et al., 2022a; Ledford & Zimmerman, 2023). Thus, understanding the nature of data collected in multiple baseline designs can impact the field.

Response Guided versus A Priori Decision-Making

When conducting SCDs that evaluate the effectiveness of an intervention (i.e., to answer a question about the effectiveness of an intervention versus a baseline condition), researchers must decide when to initiate the intervention (e.g., Barton et al., 2018; Golden et al., 2023; Ledford et al., 2017). One common way to determine when to intervene is through response-guided decision-making. Response-guided decision-making in SCD involves making data-driven condition adjustments based on ongoing data collection and analysis (Ledford & Gast, 2018). This type of decision-making is common in practice (i.e., when baseline data suggest the need for an intervention, a practitioner implements it; when intervention data suggest the need for a modification, the practitioner changes the intervention). However, in research, response-guided decisions about *when* to intervene are primarily concerned with *experimental* rather than *participant* needs. While practitioners may use baseline data to determine *if* a child needs treatment, the treatment begins once that need is established. When single-case designs are used, the need must also be established—but even after the need is established, *internal validity* must also be established. Thus, specific data patterns exhibited by the participant (e.g., increasing trends, variable data) and specific design requirements (e.g., responses by participants in other tiers of multiple baseline designs, meeting standards set by funding agencies; Horner et at., 2005; Ledford et al., 2023) influence when researchers choose to begin intervention when response guided decision making is used. Additionally, the utilization of response-guided decision-making for MBDs requires additional consideration. When to implement intervention depends on (a) how tier-specific baseline data look, (b) how the remaining tier's baseline data looks, (c) covariation between tiers, and (d) how quickly participants respond to intervention (Ledford & Gast, 2018).

Researchers can also determine when to intervene before study implementation (i.e., a priori decision-making) as an alternative to response-guided decision-making. Although a priori decision-making is a hallmark of group design research (e.g., randomized controlled trials), single-case researchers have historically not used this approach (Swan et al., 2020). Although uncommon, it has sometimes been used in the context of non-concurrent multiple baseline designs (Watson & Workman, 1981). However, it could theoretically be used in all single-case designs, including concurrent variations of the multiple baseline design. When utilizing an a priori approach in a single case multiple baseline design, researchers could randomly assign participants to varying conditions lengths rather than determining intervention initiation via visual analysis (Ledford & Zimmerman, 2023; Watson & Workman, 1981). Historically, single-case researchers have embraced response-guided decision-making as a strength of single-case design, allowing for flexibility in identifying functional relations and individualization of procedures (Ledford & Gast, 2018). However, response-guided decision-making could lead to biased outcome estimates (Ledford & Zimmerman, 203). On the other hand, a priori decision-making could lead to a reduced ability to identify a functional relation that exists (e.g., data are unstable or showing an increasing trend prior to intervention implementation).

Masked Analysis in SCD

In SCD, masked analysis, also known as blind analysis, refers to a procedure in which the individuals involved in assessing the outcomes or analyzing the data are kept unaware (masked or blind) of certain critical information (Ferron et al., 2017; Ledford & Gast, 2018). The purpose is to reduce bias and enhance the objectivity of the analysis. For example, an implementer doing their own visual analysis may be more likely to end the baseline phase and initiate the intervention phase due to preconceived notions about the intervention, the participants, or the outcomes. A visual analyst with less contextual information may make an objective determination. Masked analysis can occur during formative and summative visual analysis (e.g., Byun et al., 2017; Sallese et al., 2020). Formative visual analysis occurs throughout study implementation to identify ongoing behavior change and is critical to response-guided decision-making. In comparison, summative visual analysis occurs after the study implementation and is used to determine the presence of a functional relation. Thus, using masked analysis theoretically could improve both ongoing decision-making and functional relation determination. However, the extent to which bias is reduced via masked analysis has not been tested empirically.

Magnitude Determination in SCD

The impacts of using response guided versus a priori decision-making, and the impacts of using masked analysts, may be evaluated via not only potential impacts on visual analysis determinations, but also on the size of effects. Historically, single-case research findings have been assessed through visual analysis, a method that examines whether shifts in behavior can be attributed to the implementation of interventions (Ledford & Gast, 2018). Utilizing visual analysis, however, poses difficulties in quantifying the degree to which variations in independent and dependent variables relate to outcome fluctuations (i.e., magnitude of behavior change). In response to this difficulty, recent developments in the field of single-case research have incorporated effect size metrics to standardize the interpretation of these studies without replacing the visual analysis approach (Barton et al., 2017; WWC, 2020). Two commonly employed effect size metrics in single-case research are the log response ratio (LRR) and the within-case standardized mean difference (SMD; Chow et al., 2023). These metrics align nicely with SCD because they exhibit insensitivity to the number of sessions per phase and prove suitable for outcomes measured on a ratio scale, where zero signifies the absence of the behavior (Pustejovsky, 2018; Pustejovsky, 2019). These effect sizes are becoming increasingly popular as supplements to visual analysis.

Impacts of Design Decisions on Outcome Analyses

Analysis of single case design data, both in terms of visual analysis and supplemental statistical analysis, may be impacted by design decisions. For example, we found that effect sizes vary based on measurement and design type (Ledford et al., under review). Although not previously studied, the magnitude may also vary based on decisions about when to implement intervention (i.e., a priori or based on baseline data patterns) and whether the response of other participants is used to determine intervention implementation in different tiers (i.e., concurrent versus non-concurrent designs). There are benefits and drawbacks to each condition-determination strategy. For example, using response-guided decision-making may increase the potential risks for researcher bias and decrease the risk of a Type II error (Ledford & Zimmerman, 2020; Swan et al., 2020). The extent to which these things are likely to occur in practice has yet to be evaluated. However, using simulated data, Swan and colleagues suggest that response-guided designs may impact the inferences drawn (Swan et al., 2020). The extent to which these findings apply to "real" single-case data collected from human participants is unclear. However, from a practical standpoint, a priori decisions about baseline length may increase social validity and ethical use of SCD by limiting time spent in baseline and permitting researchers to identify beforehand the length of time for which participants will be expected to remain in baseline conditions. Thus, it is an essential area of study.

Purpose

Therefore, the following study aims to evaluate the impact of response-guided decision-making on obtained results compared to a priori decisions and gauge researcher perspectives regarding a priori decision-making within the context of SCDs. The following research questions guided this study:

> **Research Question 1:** Are obtained effect sizes for changes in social initiations larger for A-B comparisons when intervention implementation is determined based on response-guided decision-making relative to a priori decision-making?
>
> **Research Question 2:** Does the relation observed for A-B designs hold when calculating effect sizes for multiple baseline designs using response-guided decision-making versus a priori decision-making?
>
> **Research Question 3:** When presented with a brief survey, how do experts in the single case research design view non-concurrent multiple baseline designs that use a priori decision-making? Specifically, do experts (a) view a priori decision-making as beneficial

(b) consider non-concurrent multiple baseline designs to be appropriate, and (c) What perceived strengths and weaknesses of each method are reported?

CHAPTER 2

Method

This dissertation consists of two different studies. Study 1 was a quantitative study consisting of twelve SCDs. Study 2 consisted of a survey to evaluate expert perspectives regarding SCD practices.

Study 1: Design Comparisons

Study 1 consisted of two phases: (1) one where we conducted evaluations of an intervention for 10 participants using A-B comparisons; and (2) one where we conducted similar evaluations for 6 participants using two multiple baseline across participants designs.

Participants

All participants for Study 1 attended the same university-affiliated preschool. To be included in the study, participants had to

1) Be between 36 months and 6 years of age.
2) Have consistent school attendance.
3) Independently engage with a preferred toy for 3 to 5 minutes, according to parent report.

There were no diagnostic criteria for participation in the study. The principal investigator (PI) met with the preschool director and school Board Certified Behavior Analyst (BCBA) to identify potential participants. The PI worked with the school director and BCBA to identify eight potential dyads. Participants were paired into dyads based on similarities in language repertoire and observed social engagement with familiar adults (e.g., the school BCBA); that is, we attempted to create pairs of participants who were likely to respond similarly to intervention and baseline procedures (see descriptions below). Flyers, with a brief description of the study purpose and researcher contact information, were placed in the cubbies of all identified potential participants. All interested guardians were given additional information regarding the purpose of the study and the opportunity to meet with the PI to ask any questions they may have had. Consent was obtained for 16 participants across three different classrooms.

The participants' ages at the start of the study ranged from 3 to 5 years. Ethnicity, race, gender, and disability identity were determined by asking the participants' parents to report their child's preferred identification (See Appendix A). We received parent information for 15 of the 16 participants. Six of the included participants were identified as being a boy and nine of the included participants were identified as being a girl. Two of the included participants were identified as having a speech and language disability while 13 were identified as having no identified disabilities. Eleven participants were identified as being White non-Hispanic, one participant was identified as White with no ethnicity being reported, one was reported as Black non-Hispanic, and two were reported as being bi-racial with no ethnicity being reported (see Table 1 for additional details).

Table 1. *Included Participants*

Participant	Age	Gender	Disability	Race & Ethnicity	Dyad	Design
Silas	4	Boy	None	White non-Hispanic	A	RG AB
Charlotte	4	Girl	Speech & Lang	Black/AA non-Hispanic	A	AP AB
Neil	3	Boy	Speech & Lang	White non-Hispanic	B	RG AB
Demi	4	Girl	None	White non-Hispanic	B	AP AB
Jayden	4	Boy	None	White non-Hispanic	C	RG AB
Aiden	3	PNR	PNR	PNR	C	AP AB
Caleb	4	Boy	None	White non-Hispanic	D	RG AB
Evan	4	Boy	None	White	D	AP AB
Kaia	3	Girl	None	White non-Hispanic	E	RG AB
Maya	3	Girl	None	White non-Hispanic	E	AP AB
Priscilla	5	Girl	PNR	Bi-Racial (Asian + White)	F	RG MBP
Noelle	5	Girl	None	White non-Hispanic	F	AP MBP
Lyla	3	Girl	None	Bi-Racial (White + Indian)	G	RG MBP
Kinsley	3	Girl	None	White non-Hispanic	G	AP MBP
Otto	3	Boy	None	White non-Hispanic	H	RG MBP
Luna	3	Girl	None	White non-Hispanic	H	AP MBP

RG = Response guided. AP = A P=priori. MBP = Multiple baseline across participant. PNR = Parent did not report. AA = African American. Lang = Language.

Research Team

Five graduate students acted as primary implementers. The first author, a certified teacher and behavior analyst enrolled in a doctoral program in special education, trained four additional master's student implementers. The first author’s primary research interests involved improving the utilization of single case design methodology and improving student engagement in inclusive classrooms to promote more efficient instruction. Four of the five implementers identified as non-Hispanic White females, and one identified as an Asian-Korean female.

Six SCD experts acted as formative masked analysts. Three out of the 6 formative experts were BCBAs and enrolled in a doctoral program in special education at the time of the study. The remaining three formative experts were PhD-level behavior analysis with expertise in special education research. Five out of the 6 formative experts identified as Female, and one identified as Male. All 6 formative experts identified as White non-Hispanic.

Three SCD experts acted as summative masked analysts. Three SCD experts acted as summative masked analysts. All summative analysts held PhDs and were considered experts in the field of special education. One of the summative experts identified as a White non-Hispanic Male, one identified as a White non-Hispanic Female, and one as an Asian American Female.

Setting and Materials

For Phases 1 and 2, sessions occurred in a one-on-one format in a university-affiliated preschool. Sessions occurred outside the classroom in one of the three resource rooms available at the preschool. Sessions lasted approximately 5-10 min and were conducted once to twice daily. All sessions required the use of a Cannon camera for data collection purposes. Session data were coded via ProcoderDV™, a data collection software that facilitates the collection of observational data from digital media (Tapp, 2003). All graphs for formative and summative analysis were created in Microsoft Excel. A choice of commonly found toy sets (e.g., blocks, Play-Doh, trains) were available for each session (see Appendix B for a list of available toy sets). Stickers or stamps were offered to each participant at the end of each session, non-contingent on performance.

Response Definitions and Measurement

The primary dependent variable of interest was social initiations within the context of play. We selected social initiations as our dependent variable due to their inherent variability, which aligns well with standard effect size metrics (e.g., SMD, LRR; Chow et al., 2023; Pustejovsky, 2018; Shadish et al., 2014). Timed event data on participant social initiations were collected via ProcoderDV™ (Tapp, 2003).

Social initiations were defined as (a) gaining or directing attention, (b) showing or giving an object, (c) making a request, or (d) talking about the toy with a secondary indicatory being observed (e.g., looking at the researcher). A new initiation was scored if three or more seconds had passed since the previous initiation or any adult initiation.

Experimental Designs

Phase 1 consisted of two types of A-B comparisons—one that utilized response-guided decision-making and one in which the baseline lengths were predetermined. Similarly, Phase 2 consisted of two kinds of multiple-baseline across participant (MBP) designs—one that utilized response-guided decision-making in a concurrent design and one in which the baseline lengths were predetermined and randomly assigned (Ledford, 2018) in a non-concurrent design. Thus, we visually analyzed child behavior in the context of simple A-B comparisons (Phase 1) or multiple baseline comparisons (Phase 2). To analyze differences in effect sizes between groups (a priori and response guided), we compared groups by descriptively analyzing summative data (means).

Implementation Procedures for Phases 1 and 2

All sessions occurred one-on-one in a resource room. Each dyad was assigned a consistent implementer. For example, the first author implemented all sessions for Kaia and Maya (Dyad E), whereas a master's student implemented all sessions for Silas and Charlotte (Dyad A). All sessions started with the researcher presenting the child with two different toy sets and instructing them to pick which one they would like to play with during the session. All sessions concluded with the researcher offering the child a selection of stamps or stickers before returning to class.

Baseline. After the toy set selection was made, implementers informed participants that they would play by themselves for 5 min before playing together (e.g., "We're going to play with the magnet tiles by ourselves for 5 minutes, and when my timer goes off, we can play together"). All initiations by the child were responded to with a redirection to the toy set using a neutral tone (e.g., "Cool, right now I'm playing by myself"). All data were collected via video from the initial 5 min “play alone” portion. After the 5 min session, the implementer offered to play with the child for 3-8 min to decrease participant dissatisfaction and build rapport.

Intervention. The intervention for this study is the researcher's responsivity to participant play actions, loosely based on naturalistic developmental behavioral intervention strategies (Frey & Kaiser, 2011; Patel et al., 2016; Yoder & Warren, 2002). After the toy selection, implementers informed participants that they would play together for 5 min (e.g., "We get to play together for 5 minutes, and then you can pick a stamp before going back to class").

During intervention conditions, implementers followed the child's lead and were responsive to child play behaviors by:

1) *Imitating*—Repeating what the child vocalizes or engaging in the same motor play response as the child (e.g., both the child and the researcher push trains).
2) *Expanding*—Doing what the child does and then adding a related behavior (e.g., the child says “blue” while holding a train and the researcher says “blue train”); and
3) *Commenting*—The researcher makes a related comment on participant play actions or their play actions (e.g., "red train goes down").

Criteria for engaging in these responsivity behaviors was set to 20 percent or below for baseline and 80 percent or higher for intervention as measured by using 10s momentary time sampling intervals (see description below).

Interobserver Agreement (IOA)

For Phases 1 and 2, data were collected independently by two observers using video recording software. All coders attended an initial training session held by the first author and principal investigator. During these training sessions, all coders (1) engaged in didactic training with the PI to review the code book with examples and close non-examples, and (2) practiced coding a singular criterion-coded session alongside the PI. This in-vivo coding, alongside the PI, allowed coders to seek clarification on any ambiguous definitions before coding additional practice sessions.

Before coding study session data, all trainees achieved at least 80% accuracy with the PI across three additional practice criterion-coded sessions. Discrepancies were identified using the point-by-point method with a 6-second agreement time window (Yoder et al., 2018; see Appendix C for an example). IOA was calculated using the formula: (agreements / agreements +

disagreements) x 100 = IOA percentage. Reliability data were collected for at least 33% of sessions across all conditions and participants. Average reliability was 84% (range by child: 68 – 96%). Please see Table 3 for reliability data by participant.

Procedural Fidelity

An independent observer collected procedural fidelity (PF) data for at least 33% of sessions across all conditions and participants. PF data were collected via video using a 10-s momentary time sampling and event recording. Momentary time sampling measured the duration interventionists engaged in responsive play behaviors with the child. Adult responsiveness during baseline was intended to remain at or below twenty percent and at or above eighty percent for intervention. Other researcher behaviors (e.g., toy selection, post-baseline play, reinforcement, and responding to child initiations) were measured using event recording on a paper form (See Appendix D). Average PF was 97% (range by child: 86 – 100%). Please see Table 3 for reliability data by participant. PF and IOA data collection sessions were randomly selected via a random number generator.

Table 2. *Interobserver Agreement & Procedural Fidelity Data*

Participant	IOA *(Range)*	BSL Resp *(Range)*	INT Resp *(Range)*	PF *(Range)*	PF IOA *(Range)*
Silas	86% *(67 – 100%)*	18% *(13 – 23%)*	83% *(73 – 93%)*	97% *(86 – 100%)*	80% *(73 – 87%)*
Charlotte	90% *(78 – 100%)*	15% *(10 – 20%)*	87% *(77 – 100%)*	100%	95% *(93 – 97%)*
Neil	68% *(52 – 80%)*	8% *(3 – 13%)*	100%	86% *(79 – 100%)*	87% *(83 – 93%)*
Demi	77% *(65 – 89%)*	9% *(7 – 10%)*	95% *(93 – 97%)*	91% *(65 – 100%)*	93%
Jayden	84% *(67 – 92%)*	12% *(7 – 17%)*	88% *(87 – 90%)*	98% *(90 – 100%)*	84% *(77 – 90%)*
Aiden	87% *(85 – 88%)*	22% *(10 – 33%)*	82% *(77 – 90%)*	97% *(87 – 100%)*	90% *(83 – 97%)*
Caleb	77% *(69 – 88%)*	5% *(3 – 7%)*	77% *(74 – 80%)*	100%	90% *(83 – 97%)*
Evan	82% *(50 – 100%)*	10% *(3 – 17%)*	91% *(83 – 97%)*	100%	92% *(90 – 94%)*
Kaia	84% *(75 – 93%)*	9% *(7 – 10%)*	76% *(75 – 77%)*	100%	87% *(74 – 100%)*
Maya	83% *(67 – 91%)*	12% *(10 – 13%)*	85% *(80 – 90%)*	97% *(93 – 100%)*	93% *(91 – 94%)*
Priscilla	92% *(82 – 100%)*	2% *(0 – 3%)*	97% *(93 – 100%)*	100%	95% *(93 – 97%)*
Noelle	96% *(86 – 100%)*	2% *(0 – 3%)*	87% *(80 – 93%)*	100%	93% *(86 – 100%)*
Lyla	84% *(72 – 96%)*	9% *(0 – 17%)*	87% *(77-97%)*	100%	80% *(77 – 83%)*
Kinsley	92% *(80 – 100%)*	17% *(7 – 20%)*	90%	99% *(96 – 100%)*	82% *(80 – 83%)*
Otto	78% *(65 – 93%)*	10% *(7 – 13%)*	100%	99% *(94 – 100%)*	100%
Luna	86% *(79 – 93%)*	16% *(7 – 26%)*	95% *(90 – 100%)*	90% *(84 – 100%)*	85% *(73 – 97%)*
Average	84%	11%	89%	97%	89%

IOA = Interobserver Agreement. PF = Procedural Fidelity. BSL = Baseline. INT = Intervention. Resp = Responsivity.

Data Analysis

Phase 1: A-B Comparisons. Phase 1 included 10 A-B comparisons—five for each A-B comparison design type (response-guided or a priori). As previously stated, adults familiar with the included participants were consulted to help create equivalent dyads. Participants were randomly assigned to either a response-guided or a priori A-B design.

Response-Guided Decision-Making. Three experts in SCD who were unaware of our research questions (Ferron & Jones, 2006; Byun et al., 2017) were contacted and asked to consult on all experimental judgments about when to implement and discontinue intervention (see description below). Decisions were made by visually analyzing the data daily. When at least two of three expert raters indicated that data were stable in baseline or a therapeutic effect was observed in intervention, decisions on when to intervene and discontinue implementation were made.

A Priori Decision-Making. Before implementation, decisions about baseline and intervention length for the remaining Phase 1 participants were determined. The condition lengths were yoked across dyads; regardless of data patterns, the condition lengths assigned to participants in the a priori A-B designs matched their assigned partner in the response-guided A-B design.

Phase 2: MBD Comparisons. Phase 2 consisted of six participants—three participants for each MBP design. Similar to Phase 1, participants were randomly assigned to either a response-guided or a priori MBP design. Additionally, consistent with current standards and recommended practices (Ledford & Gast, 2018), participant dyads were randomly assigned to a tier.

Response-Guided Decision-Making. The baseline implementation occurred concurrently (i.e., simultaneously) across all three participants included in the response-guided MBP design. Similar to the response-guided A-B comparisons, three additional SCD experts who were not aware of our research questions (Ferron & Jones, 2006; Byun et al., 2017) made all experimental judgments about when to intervene and discontinue intervention by visually analyzing the data across all three tiers daily.

A Priori Decision-Making. Similar to the a priori A-B comparisons, condition lengths were predetermined based on the condition lengths from the response-guided MBP designs. Due to the randomization of participants to tiers and the decreased likelihood of shared history effects across participants (Ledford & Zimmerman, 2023), baseline implementation across participants did not occur concurrently (not simultaneously) for this MBP design.

Magnitude. All dependent variable data were analyzed using the SingleCaseES calculator (Pustejovsky et al., 2023) in R statistical software (R Core Team, 2021) to calculate the magnitude of behavior change between baseline and intervention conditions. The two effect sizes calculated for this study were within-case standard mean difference (SMD) and log response ratio (LRRi). SMD estimates the mean differences between baseline and intervention conditions relative to the pooled baseline variance, and LRRi estimates the proportionate change between baseline and intervention. We calculated all included comparisons' means and standard deviations to evaluate the differences between research practices (i.e., response-guided vs a priori).

Visual Analysis. In addition to the effect size metrics, visual analysis was used to determine the presence of a basic effect for Phase 1 and a functional relation for Phase 2. To determine summative effects, all data were analyzed by three additional experts, different from the three analysts who made formative decisions. Experts were told to evaluate data using visual analysis consistent with expert recommendations (Ledford & Gast, 2018; see Ledford et al., 2023 for description example)

Study 2: Expert Survey

Study 2 consisted of an online questionnaire used to gather information from single case design experts regarding current design implementation practices.

Participants

Field experts were contacted via email and sent a brief survey to evaluate perspectives on utilizing a priori decisions within single-case design. To be included in the survey data, participants must either have attended a single-case methodology conference, authored a single-case design study in the last two years, or been forwarded the survey link by a fellow researcher. Twenty-four SCD researchers responded to the survey. All respondents indicated that they held a master's or doctoral degree and had published a study that utilized SCD methodology within the past two years. Please see Table 2 for additional information regarding respondent demographics.

Table 3. *Survey Respondent Demographics*

Category	Reported Identification (%)	
Age (in years)		
	18 – 24	0 (*0%*)
	25 – 34	5 (*21%*)
	35 – 44	15 (*63%*)
	45 – 54	0 (*0%*)
	55 – 64	3 (*13%*)
	65 or older	1 (*4%*)
Race		
	White	20 (*83%*)
	Black/AA	0 (*0%*)
	Asian	3 (*13%*)
	NA/AN	0 (*0%*)
	ME	0 (*0%*)
	Bi/multiracial	1 (*5%*)
Ethnicity		
	Hispanic	1 (*4%*)
	Non-Hispanic	22 (*92%*)
	Other	1 (*4%*)
Highest Education Achieved		
	High School	0 (*0%*)
	Some College	0 (*0%*)
	Bachelors	0 (*0%*)
	Masters	6 (*25%*)
	PhD/EdD/MD	18 (*75%*)

Yrs = Years. AA = African American. NA = Native American. AN = Alaskan Native. ME = Middle Eastern. PhD = Doctor of Philosophy. EdD = Doctor of Education. MD = Medical Doctor.

Research Team and Positionality

The author (see description above) and their advisor, a Ph.D. level behavior analyst with expertise in SCD, coded all survey data. The advisor's research interests consisted of understanding the utilization of SCD and increasing its value and prevalence in special education, as well as using evidence-based instructional practices in early childhood contexts. Both researchers identified as White, non-Hispanic females and had collaborated on several previous quantitative and qualitative research projects.

Data Collection

The distributed survey consisted of the following questions: (1) Do you view a priori decision-making as beneficial to the field of SCD? (2) Do you view masked analysis as beneficial to the field? (3) Do you view non-concurrent designs as beneficial to the field? And (4) Do you view non-concurrent multiple baseline designs as an appropriate alternative to concurrent multiple baseline designs? And if so, under what conditions? In addition to open-ended responses, questions one, two, and three were accompanied by a zero to one hundred rating scale—where zero was representative of disadvantageous, fifty was representative of neutral, and one hundred was representative of advantageous. See Appendix E for an example of how survey questions were presented to the participants via REDCap® (Research Electronic Data Capture). REDCap® is a secure, web-based software platform engineered to facilitate data collection for research endeavors (Harris et al., 2009; Harris et al., 2019). Data were also collected on respondent demographics (see description above). All survey data were collected and managed via REDCap®.

Data Analysis

All survey data were coded using a grounded theory approach (Guest et al., 2013) to identify common themes across respondents. The first author (see description above) and her advisor (a doctoral-level behavior analyst with expertise in single-case design) independently reviewed all responses and produced a preliminary list of common themes across respondents for each of the four open-ended questions. These preliminary lists were then reviewed and discussed until a consensus was reached.

CHAPTER 3

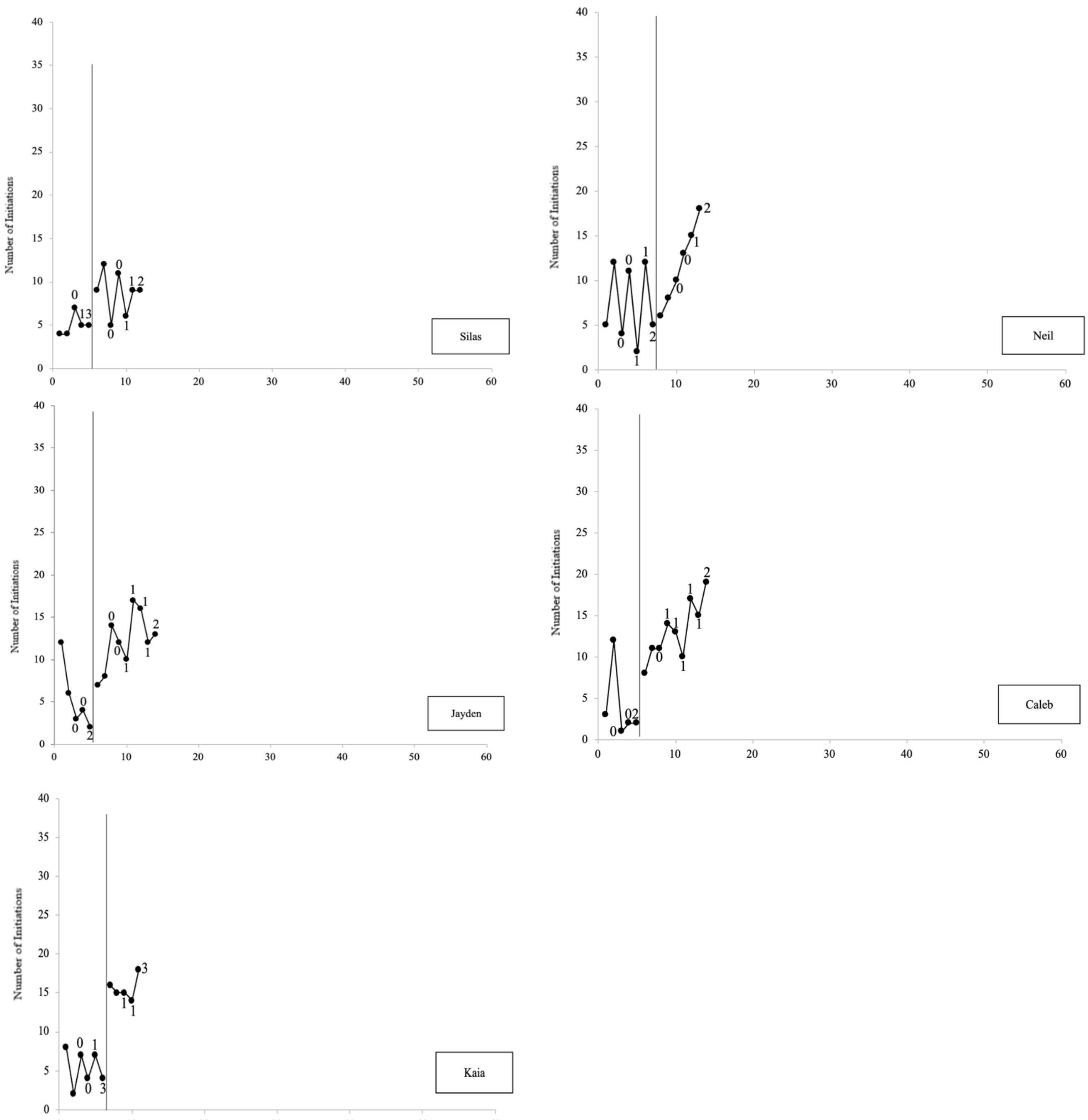
Results

Study 1: Design Comparisons

Phase 1: A-B Comparisons

Three formative masked analysts (see descriptions above) determined all condition lengths during the response-guided A-B designs. Baseline lengths ranged from five to seven sessions ($M = 6$), and intervention lengths ranged from five to nine sessions ($M = 7$). The three formative masked analysts agreed forty-six percent of the time on when to implement and discontinue intervention. Additionally, the formative masked analysts had total agreement on when to intervene for two (Silas and Kaia) out of the five participants. They had total agreement on when to discontinue intervention for one participant (Kaia). All masked analyst decisions for the response-guided A-B comparisons are depicted in Figure 1.

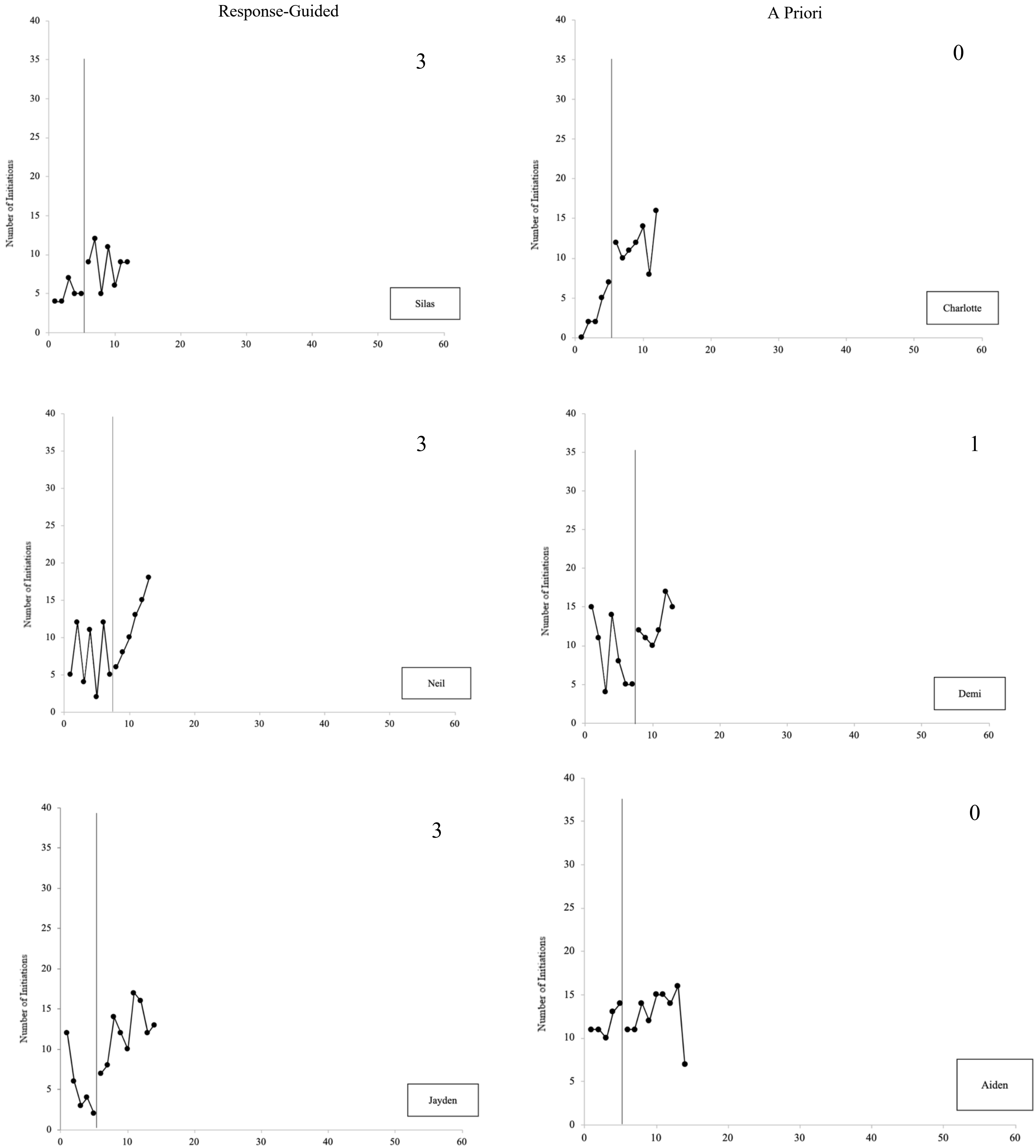
Figure 1. *Response-Guided Results for A-B Designs*



Note. The numbers represent how many of the experts voted to move conditions (baseline to intervention) or discontinue intervention implementation.

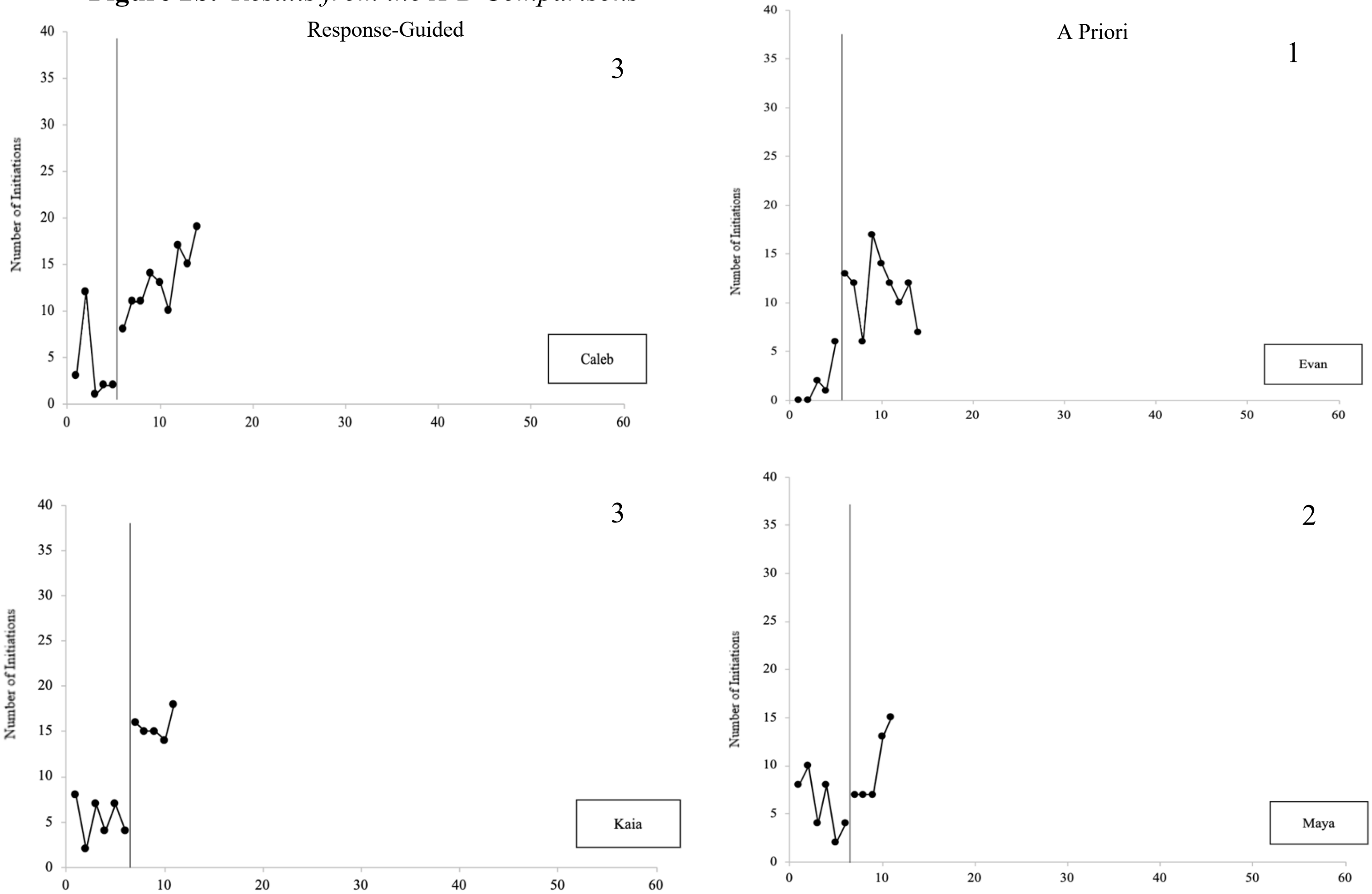
All children, except for Aiden and Neil, had an immediate and consistent level change following the implementation of the intervention (see Figures 2a and 2b). Data show an increasing trend throughout the intervention for Neil, and no demonstration of effect is observed for Aiden. Baseline data were consistently variable or had a decreasing trend for all participants except for Charlotte and Evan—both of whom were randomly assigned to a priori designs.

Figure 2. *Results from the A-B Comparisons*



Note. The number in the top right-hand corner represents how many of the summative experts stated that the design demonstrated an effect

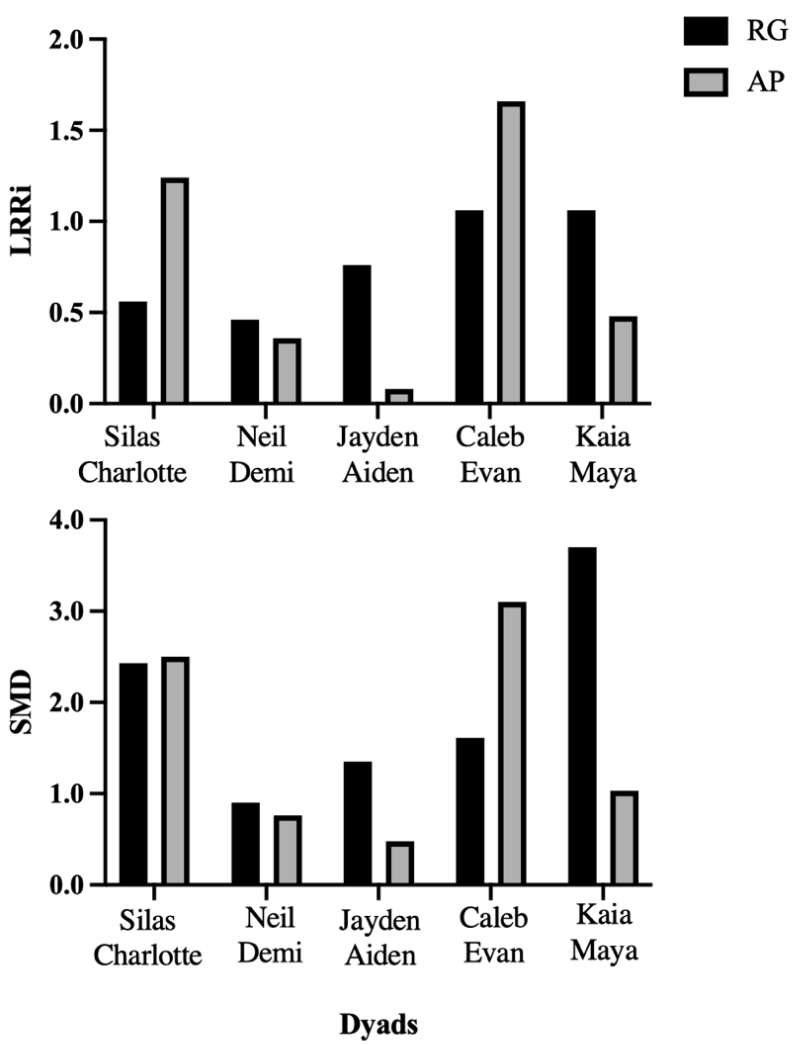
Figure 2b. *Results from the A-B Comparisons*



Note. The number in the top right-hand corner represents how many of the summative experts stated that the design demonstrated an effect.

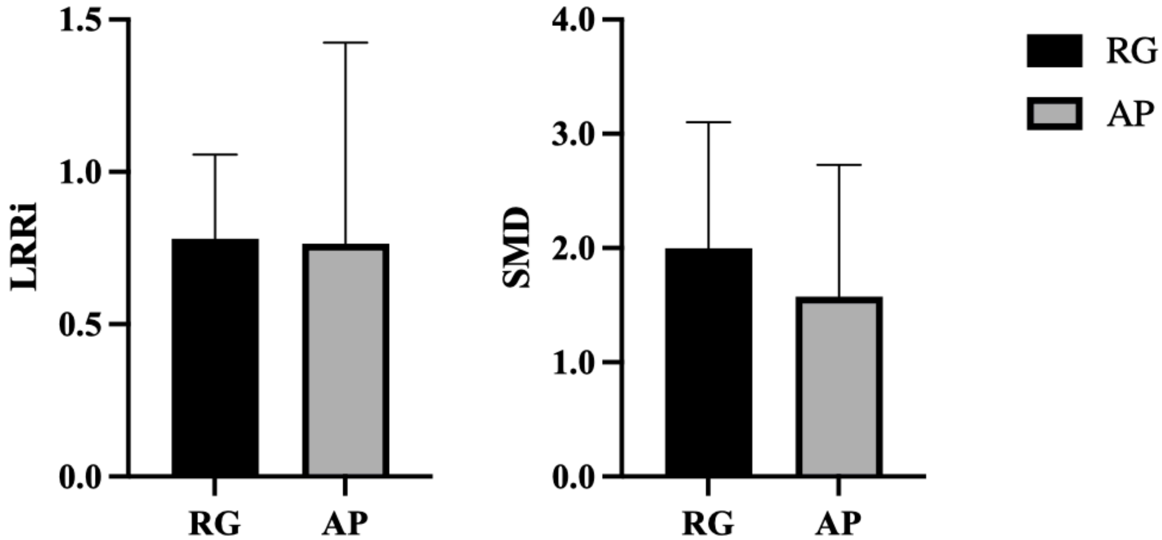
The three summative masked analysts identified six of ten designs as demonstrating an effect and agreed with each other seventy percent of the time. The mean value for response-guided A-B comparisons from Phase 1 was 0.78 (range: 0.46 – 1.06) for LRR and 1.99 (range: 0.9 – 3.7) for SMD. The mean value for a priori A-B comparisons from Phase 1 was 0.76 (range: 0.08 - 1.66) for LRRi and 1.57 (range: 0.48 - 3.1) for SMD. Effect size calculations across Phase 1 dyads are depicted in Figures 3 and 4.

Figure 3. *Phase 1 Effect Sizes Across Dyads*



LRRi = Log Response Ratio. SMD = Standard Mean Difference. RG = Response Guided. AP = A Priori.

Figure 4. *Phase 1 Mean Effect Sizes with Standard Deviations*

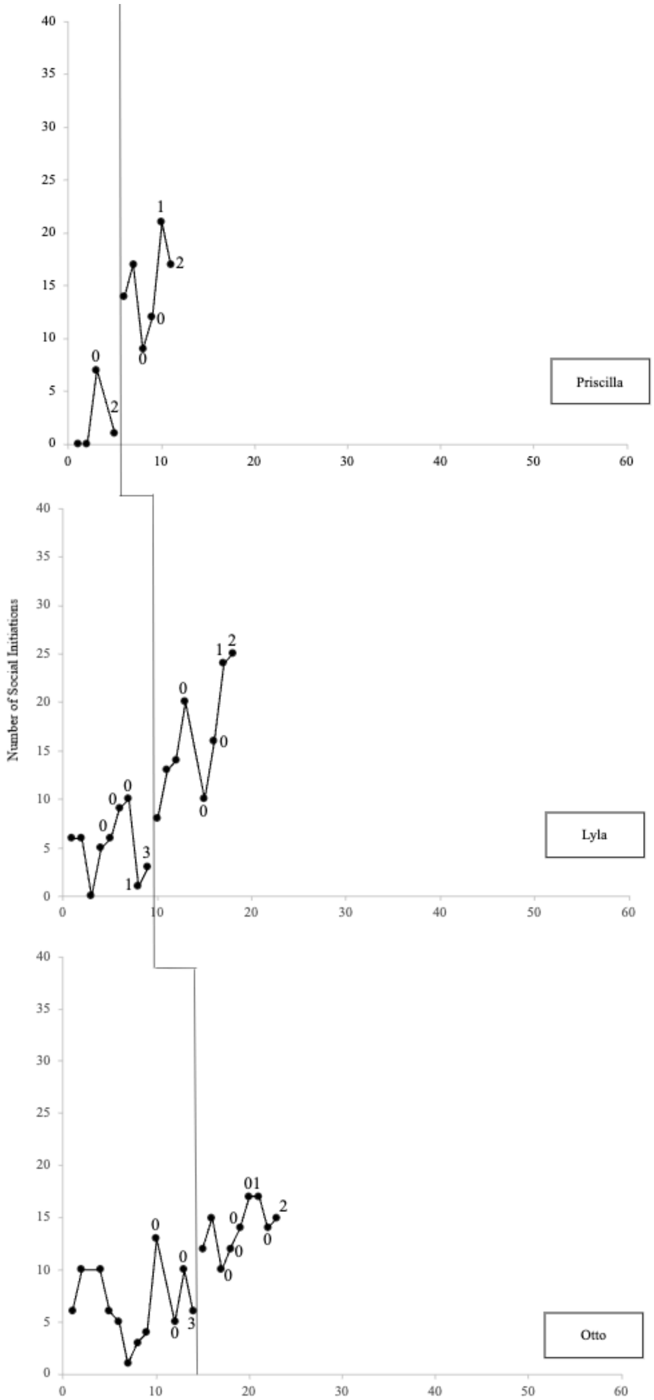


LRRi = Log Response Ratio. SMD = Standard Mean Difference. RG = Response Guided. AP = A Priori.

Phase 2: MBD Comparisons

Similar to Phase 1, three formative masked analysts (see description above) determined all condition lengths for the response-guided MBD. The baseline lengths consisted of four, nine, and twelve sessions (*M* = 8). The intervention lengths consisted of six, eight, and nine sessions (*M* = 8). The three formative masked analysts agreed seventy-eight percent of the time on when to implement and discontinue intervention. Additionally, the formative masked analysts had total agreement on when to intervene for two (Lyla and Otto) out of the three participants. The formative masked analysts did not reach total agreement on when to discontinue intervention for any of the participants. All masked analyst decisions for the response-guided MBD are depicted in Figure 5.

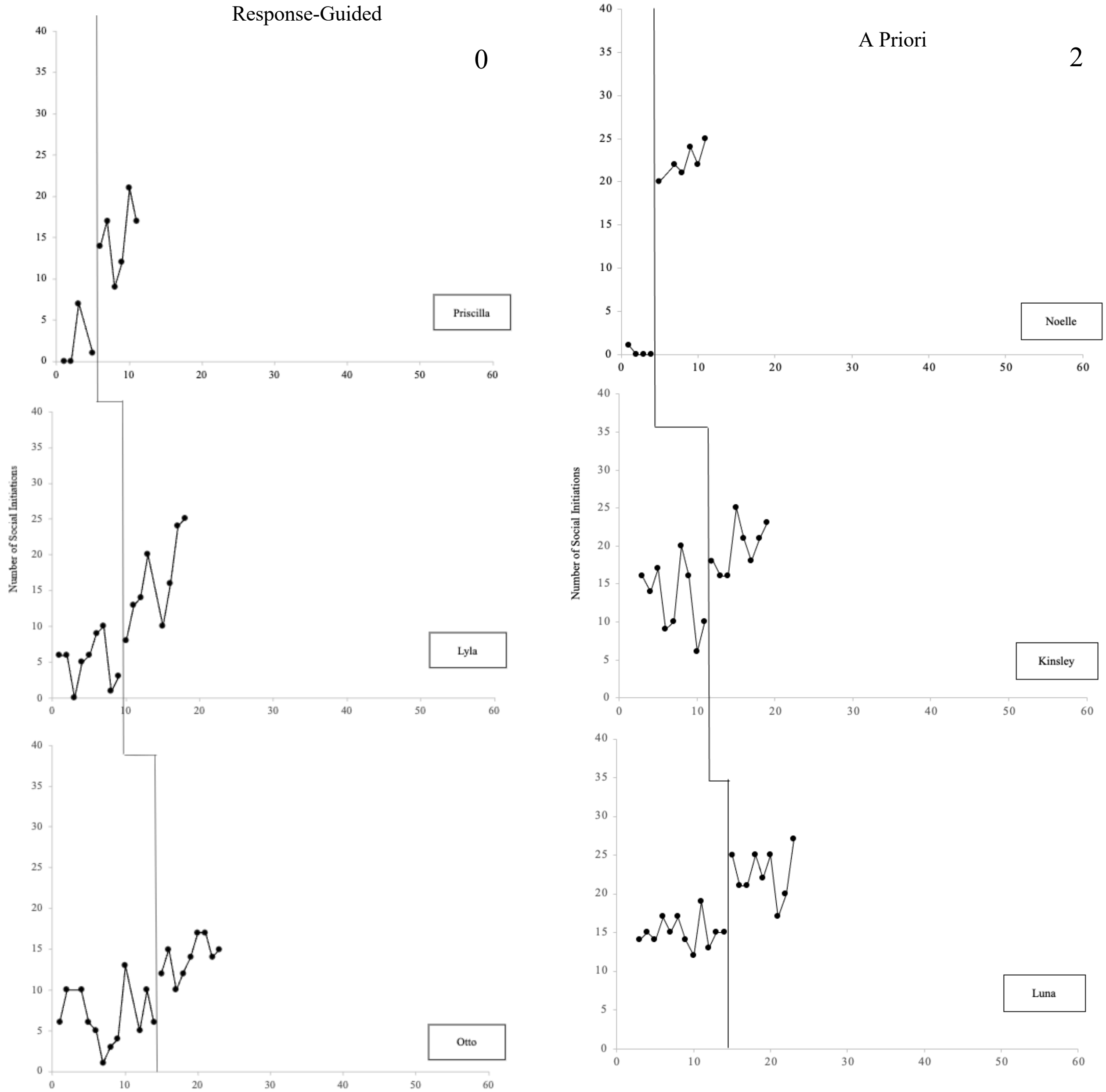
Figure 5. *Response-Guided Results for MBD*



Note. The numbers represent how many of the experts voted to move conditions (baseline to intervention) or discontinue intervention implementation.

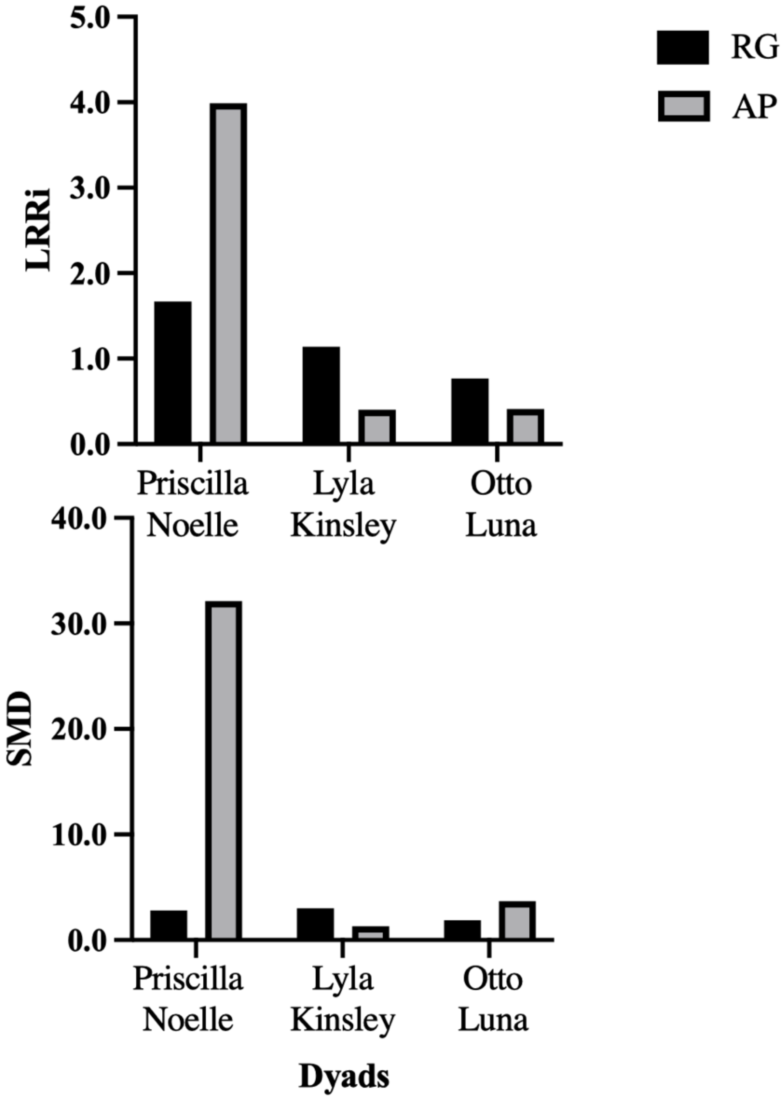
All children had an immediate and consistent level of change following the implementation of the intervention (see Figure 6). Baseline data were consistently variable or at floor levels for all participants. The three summative masked analysts identified the non-concurrent a priori MBD as having a functional relation and agreed with each other fifty percent of the time. None of the experts identified the response-guided design as having a functional relation. The mean value for the response-guided MBD from Phase 2 was 1.19 (range: 0.77 – 1.67) for LRR and 2.57 (range: 1.88 – 3.02) for SMD. The mean value for the non-concurrent a priori MBD design from Phase 2 was 1.60 (range: 0.40 – 3.99) for LRR and 12.37 (range: 1.31 – 32.12) for SMD. Effect size calculations across Phase 2 dyads are depicted in Figures 7 and 8.

Figure 6. *Phase 2 Multiple Baseline Designs*



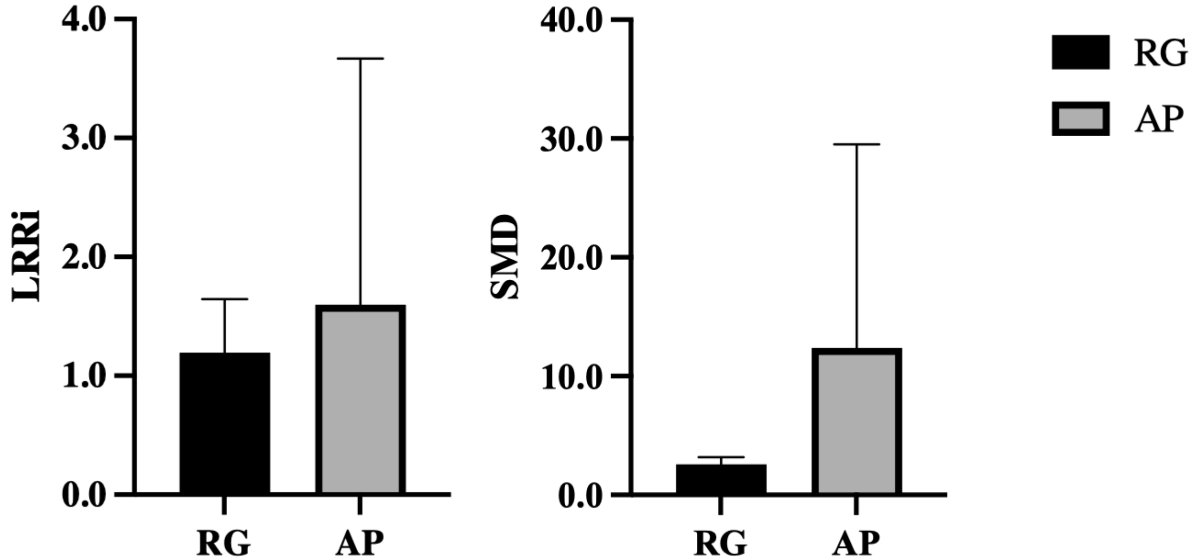
Note. The number in the top right-hand corner represents how many of the summative experts stated that there was a functional relation.

Figure 7. *Phase 2 Effect Sizes Across Dyads*



LRRi = Log Response Ratio. SMD = Standard Mean Difference. RG = Response Guided. AP = A Priori.

Figure 8. *Phase 2 Mean Effect Sizes with Standard Deviations*



LRRi = Log Response Ratio. SMD = Standard Mean Difference. RG = Response Guided. AP = A Priori.

Study 2: Expert Survey

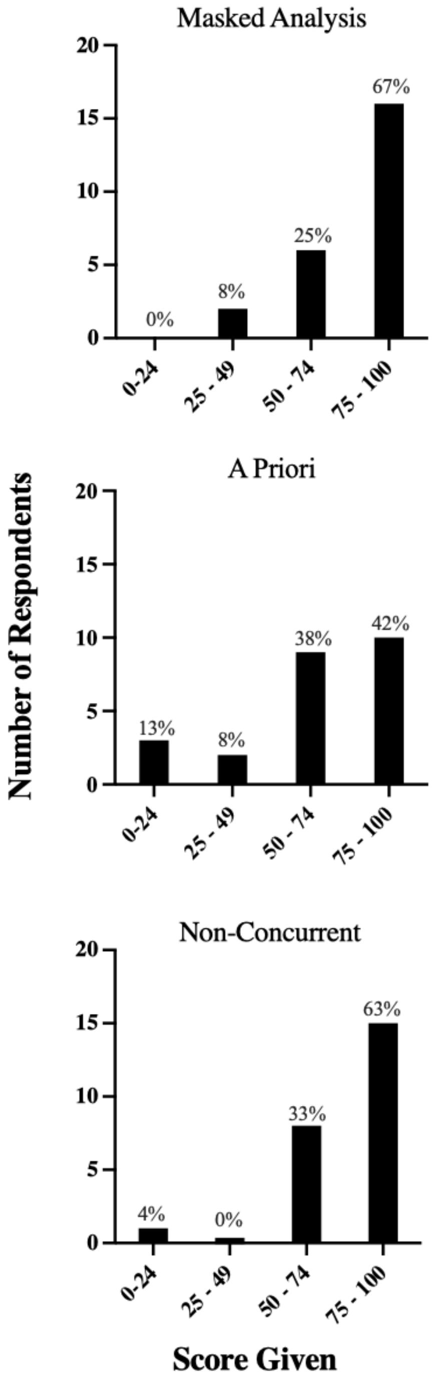
When asked to rate a priori, masked analysis, and non-concurrent designs on a scale of zero to one hundred, the distribution of scores was mixed with wide ranges across respondents. All quantitative survey data can be viewed in Table 4 and Figure 9. Qualitative results are divided by category and described below.

Table 4. *Quantitative Survey Data*

Question	Scale of 1 - 100
Do you view masked analysis as beneficial to the field of SCD?	Mean: 77.7 Median: 83 Range: 33 – 100
Do you view a priori decision-making as beneficial to the field of SCD?	Mean: 69.5 Median: 71.5 Range: 12 – 100
Do you view non-concurrent designs as beneficial to the field?	Mean: 77.9 Median: 81 Range: 18 – 100

SCD = Single-Case Design.

Figure 9. *Respondent Ratings of SCD Research Practices*



Note. Bars are representative of the number of respondents that gave scores within that quartile. With higher scores being associated with more positive views of the research practice within the field of SCD.

Formative and Summative Masked Analysis

Five distinct themes emerged from the respondents' responses regarding the utilization of masked analysis within the field of SCD:

1) Implementing masked analysis can be logistically challenging.
2) Utilizing masked analysis can limit researcher bias.
3) Engaging in masked analysis requires access to contextual information.
4) Utilizing masked analysis is not necessary for a rigorous study.
5) The usefulness of masked analysis can be dependent on the study context.

The majority of respondents (n = 16) acknowledged that the utilization of masked analysis within SCD can limit researcher bias when analyzing and reporting study results. As one researcher explained, "masked analysis provides the most unbiased mechanism to interpret the data." Although the usefulness of masked-analysis was widely acknowledged, respondents also noted some of the downsides to implementing masked-analysis within SCD research. Several respondents stated the expertise ("visual analysis is a skill, like many others, that requires practice and feedback to gain fluency") and study-specific knowledge ("knowledge of the intervention context and behaviors measured is required to meaningfully interpret results") were barriers to successfully implementing masked-analysis. Some respondents went as far as to state they did not view masked-analysis as necessary or even useful ("I think there are situations in which variability would be explained based on the type of behavior measured and if the person is entirely masked that could be problematic").

A Priori Decision Making

Nine distinct themes emerged from the respondents' responses regarding the utilization of a priori decision-making within the field of SCD:

1) There is a need for increased transparency within the field of SCD.
2) A priori decision-making can be beneficial when resources are constrained.
3) A priori decision-making can increase validity and decrease researcher bias.
4) A priori decision-making limits the flexibility of SCD research.
5) There can be challenges with data stability and variability when implementing a priori decision-making.
6) A priori decision-making can make it difficult to demonstrate experimental control.
7) A priori decision-making differs from the historical context of SCD.
8) The usefulness of a priori decision-making can depend on the study context.
9) The usefulness of a priori decision-making depends on study design.

Similar to masked-analysis, half of the respondents noted how the utilization of a priori decision-making within SCD could limit researcher bias. As one respondent noted, "a priori decision-making is beneficial because it removes possible threats to internal validity … we are biased to continue a phase until it improves so that we can claim a functional relation where one may not exist." However, ten respondents also noted how the utilization of a priori decision-making limits researchers' flexibility ("I see the value but also, SCD can be a great method to investigate individualized approaches, and at times it feels like going full a priori would limit the inherent flexibility of SCD").

Non-Concurrent Designs

Four distinct themes emerged from the respondents' responses regarding the usefulness of non-concurrent designs within the field of SCD:

1) Non-concurrent designs can be beneficial when resources are constrained.
2) The usefulness of non-concurrent designs depends on the study context.

3) Non-concurrent designs lack rigorous internal validity.
4) Non-concurrent designs have sufficient internal validity.

Over half of the respondents (*n* = 13) noted the usefulness of non-concurrent designs when resources are constrained (e.g., low incidence populations). One researcher noted, “I see nonconcurrent designs as beneficial based on researcher resources and participant resources. Most of the time, it is not feasible to recruit and implement intervention with a minimum of three participants simultaneously and rigorously.” Interestingly, respondents had differing opinions regarding the validity of non-concurrent designs. Four respondents stated their hesitancy regarding the validity of non-concurrent designs (“I have a hard time seeing how NC-MBDs are not just a series of independent AB designs”). In contrast, eight respondents stated how the think non-concurrent designs can be just as rigorous as concurrent designs (“there are also many reasons why nonconcurrent designs meet the same assumptions as concurrent designs”).

Seven additional themes emerged when respondents were asked if they viewed non-concurrent multiple baseline designs as an appropriate alternative to concurrent multiple baseline designs:

1) Hesitations about non-concurrent designs being less internally valid than concurrent designs.
2) Non-concurrent designs can be beneficial when resources are constrained.
3) The usefulness and rigor of a non-concurrent design depends on the study context.
4) More research is needed to validate non-concurrent designs as an appropriate alternative.
5) Non-concurrent designs and concurrent MB designs are equal regarding internal validity.
6) Non-concurrent designs are more socially valid and ethical than concurrent MB designs.
7) There is a need for increased transparency within the field of SCD.

The majority of respondents noted that nonconcurrent MB designs can be an appropriate alternative to concurrent MB designs ("under most conditions, I would say they are comparable and equally rigorous") when resources are constrained ("it can put less [of a] burden on the participant/client and conserve resources on the clinician/researcher") and when history threats are unlikely ("under conditions in which there would be no risk of cross-contamination or history effect across tiers"). Interestingly, although not a common theme regarding respondent opinion, two sources of scholarly work were consistently cited across both of the non-concurrent open-ended questions—Slocum et al. (2022) and Ledford and Zimmerman (2023).

CHAPTER 4

Discussion

As the use and impact of SCDs change over time, this research sought to understand current research practices and their impact on obtained outcomes. Specifically, it aimed to assess how response-guided decision-making affects outcomes compared to a priori decisions and to understand researchers' views on a priori decision-making in the context of SCDs.

Study 1: Design Comparisons

Results from Study 1 are variable across a priori and response-guided designs. This variability indicates that design decisions should be driven by context-dependent research questions and available resources rather than preconceived notions that selecting one design decision over the other will impact the results obtained. Interestingly, the percentage that the formative experts agreed was lower for the response-guided A-B designs than the response-guided MBD. This discrepancy in agreement may be attributed to the additional visual analysis required when implementing an intervention across tiers (Ledford & Gast, 2018). Additionally, none of the summative experts identified the response-guided MBD as having a functional relation. This lack of functional relation determination in the response-guided MBD may be attributed to the variability observed in the third tier. Lastly, effect sizes were larger for both top tiers in the MBD designs. This larger effect size in the first tiers may be attributed to the shorter baseline lengths or the age of the participants (both Priscilla and Noelle were five years old at the start of the study).

Study 2: Expert Survey

Quantitative responses from the survey indicate that there is some tension in the field regarding the implementation of a priori decision-making and non-concurrent designs. This polarization of opinions is highlighted in experts' evaluation of a priori decision-making on a scale of zero to one hundred. Three experts strongly evaluated a priori decisions as not beneficial to the field of special education with scores less than twenty-five. In direct contrast, six experts evaluated a priori decision-making as extremely beneficial to the field with perfect scores of one hundred. This variability in opinions highlights the recent push for more flexibility when evaluating the validity and acceptability of SCDs (Ledford et al., 2023).

Qualitative analysis across responses from all four open-ended questions highlights the field's belief that the selection of research procedures depends on the context of the study and available resources. Additionally, there was a repeated theme regarding the importance of transparency in reporting SCD procedures and study results. This belief in transparency corroborates some of the recent calls for more thorough reporting in the field of SCD (Ledford et al., 2023; Swan et al., 2020; Van Norman et al., 2023).

Recommendations and Future Directions

Our results suggest that researchers consider making design-specific decisions based on the context-dependent research questions and available resources. Additionally, researchers should report all design- and condition-specific decisions and the logic behind why they were made. Whenever the manuscript length is of concern, we strongly encourage researchers to include online supplemental materials (via journals or third-party repositories) to relay all research decisions to the relevant consumer. Results from this study provide several directions for future work. More research is needed to understand if these results are maintained across additional participants and dependent and independent variables. Additional qualitative research is needed better to understand current researcher perspectives regarding SCD best practices.

Limitations

The outcomes and recommendations presented from this study should be interpreted in light of several limitations. Results from Study 1 may not be generalizable to older populations or different dependent and independent variables. Additionally, Study 2 targeted highly experienced SCD experts, so the results may not be generalizable to the broader SCD researcher population.

Conclusions

This study evaluated the impact of response-guided decision-making on outcomes in comparison to a priori decisions and explored researchers' perspectives on pre-established decision-making within the framework of SCDs. Findings suggest that research decisions should primarily be influenced by design and resource-related factors rather than relying solely on the perceived advantages of response-guided decision-making. Future research is needed to better understand the generality of these results.

References

Barton, E. E., Ledford, J. R., Zimmerman, K. N., & Pokorski, E. A. (2018). Increasing the engagement and complexity of block play in young children. *Education and Treatment of Children*, *41*(2), 169-196.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91.

Byun, T. M., Hitchcock, E. R., & Ferron, J. (2017). Masked visual analysis: Minimizing type I error in visually guided single-case design for communication disorders. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1455-1466.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Guest, G., Namey, E. E., & Mitchell, M. L. (2013). *Collecting qualitative data. A field manual for applied research*. Thousand Oaks, CA: Sage.

Golden, A. K., Hemmeter, M. L., & Ledford, J. R. (2023). Evaluating the Effects of Training Plus Practice-Based Coaching Delivered Via Text Message on Teacher Use of Pyramid Model Practices. *Journal of Positive Behavior Interventions,* 10983007231172188.

Ferron, J., & Jones, P. K. (2006). Tests for the visual analysis of response-guided multiple baseline data. *The Journal of Experimental Education*, *75*(1), 66-81.

Ferron, J. M., Joo, S. H., & 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*(4), 701-716.

Frey, J. R., & Kaiser, A. P. (2011). The use of play expansions to increase the diversity and complexity of object play in young children with disabilities. *Topics in Early Childhood*

Special Education, 31, 99-111.

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

Harris, P., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J., (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support, *J Biomed Inform. 2009 Apr;42(2):377-81.*

Harris, P., Taylor, R., Minor, B., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Kirby, J., Duda, S. (2019) REDCap Consortium, The REDCap consortium: Building an international community of software partners, *J Biomed Inform. 2019 May 9 [doi: 10.1016/j.jbi.2019.103208]*

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.

Lanovaz, M. J., & Turgeon, S. (2020). How many tiers do we need? Type I errors and power in multiple baseline designs. *Perspectives on behavior science*, *43*(3), 605-616.

Ledford, J. R. (2018). No randomization? No problem: Experimental control and random assignment in single case research. *American Journal of Evaluation*, *39*(1), 71-90.

Ledford, J. R., Barton, E. E., Severini, K. E., Zimmerman, K. N., & Pokorski, E. A. (2019a). Visual display of graphic data in single case design studies. *Education and Training in Autism and Developmental Disabilities*, *54*(4), 315-327.

Ledford, J. R., Barton, E. E., Severini, K. E., & Zimmerman, K. N. (2019b). A primer on single case research designs: Contemporary use and analysis. *American Journal on Intellectual and Developmental Disabilities*, *124*(1), 35-56.

Ledford, J. R., Chazin, K. T., Lane, J. D., Zimmerman, K. N., Bennett, P. B., & Ayres, K. A. (2023, May). Single case analysis and review framework (SCARF). Retrieved from: http://ebip.vkcsites.org/scarfv2

Ledford, J. R., & Zimmerman, K. N. (2023). Rethinking rigor in multiple baseline and multiple probe designs. *Remedial and Special Education*, *44*(2), 154-167.

Ledford, J. R., Lambert, J. M., Pustejovsky, J. E., Zimmerman, K. N., Hollins, N., & Barton, E. E. (2023). Single-case-design research in special education: Next-generation guidelines and considerations. *Exceptional Children*, *89*(4), 379-396.

Ledford, J. R., Zimmerman, K. N., Chazin, K. T., Patel, N. M., Morales, V. A., & Bennett, B. P. (2017). Coaching paraprofessionals to promote engagement and social interactions during small group activities. *Journal of Behavioral Education*, *26*, 410-432.

Patel, N.M., Ledford, J.R., & Maupin, T.N. (2016). Responsive play interactions. In *Evidence based instructional practices for young children with autism and other disabilities.* Retrieved from http://ebip.vkcsites.org/responsive-play-interactions

Pustejovsky, J. E., Swan, D. M., & English, K. W. (2019). An examination of measurement procedures and characteristics of baseline outcome data in single-case research. *Behavior Modification*. Advance online publication. https://doi.

Sallese, M. R., & Vannest, K. J. (2020). The effects of a multicomponent self-monitoring intervention on the rates of pre-service teacher behavior-specific praise in a masked single-case experimental design. *Journal of Positive Behavior Interventions*, *22*(4), 207-219.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single- case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43*(4), 971–980. https://doi.org/10.3758/s13428-011-0111-y

Shepley, C., Shepley, S. B., & Spriggs, A. D. (2022). On the History of Single-Case Methodology: A Data-Based Analysis. *Journal of Behavioral Education*, 1-21. https://doi.org/10.1007/s10864-022-09477-2

Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., & Nichols, B. (2022a). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science*, *45*(3), 619-638. https://doi.org/10.1007/s40614-022-00326-1

Slocum, T. A., Joslyn, P. R., Nichols, B., & Pinkelman, S. E. (2022b). Revisiting an analysis of threats to internal validity in multiple baseline designs. *Perspectives on Behavior Science*, *45*(3), 681-694. https://doi.org/10.1007/s40614-022-00351-0

Swan, D. M., Pustejovsky, J. E., & Beretvas, S. N. (2020). The impact of response-guided designs on count outcomes in single-case experimental design baselines. *Evidence-Based Communication Assessment and Intervention*, *14*(1-2), 82-107.

Tapp, J. (2003). *ProcoderDV.* Nashille, TN: Vanderbilt Kennedy Center.

Van Norman, E. R., Klingbeil, D. A., Boorse, J., & Sturgell, A. K. (2023). A Summary of Relevant Demographic Characteristics of Multiple-Baseline Designs Targeting Academic *Skills. Remedial and Special Education*, 07419325231203343.

Yoder, P., Lloyd, B., & Symons, F. (2018). Observational measurement of behavior (2nd ed.). New York: Springer.

Yoder, P. J., & Warren, S. F. (2002). Effects of prelinguistic milieu teaching and parent responsivity education on dyads involving children with intellectual disabilities. *Journal* *of Speech, Language, and Hearing Research, 45*(6), 1158-1174.

Appendix A
Guardian Questionnaire

Participant: __________________ Date: _____________________________

Researcher: _________________ Relationship: ______________________

Study: _____________________

Participant Demographic Information

Birthday: ____

Gender: ____

Identified Race: ____

Identified Ethnicity (Hispanic / non-Hispanic): ____

Diagnosis/Disability status: ____

Participant Preferences

<u>Favorites</u>

Top three favorite activities: ____

Top three favorite toys/objects: ____

Other: ____

<u>Dislikes</u>

Activities: ____

Toys/objects: ____

Other: ____

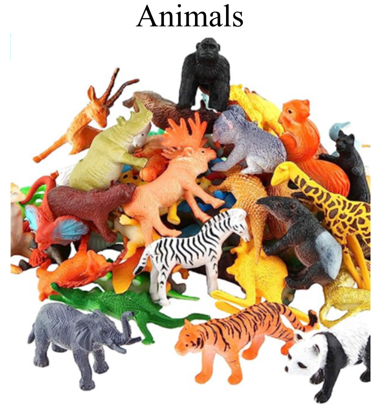
Participant Play Behaviors

Please view the attached flyer and list the type of object play ______________ most frequently engages in: ____

How long would you say _____________ can independently engage with a preferred toy set? (e.g., 6 minutes): ____

Appendix B

Examples of Available Toys

Larger Set	Accompanying Toys
Magnet Tiles 	Cars 
Play Doh 	Cookie Cutters 
Moon Sand 	Animals 

Appendix C

IOA Calculation for DV Example

	A	B	C	D	E
1	Coder 1			Coder 2	
2	00:52.2	initiation		00:52.7	initiation
3	01:32.3	initiation		01:33.1	initiation
4	01:50.5	initiation		02:26.6	initiation
5	02:25.8	initiation		02:56.0	initiation
6	02:55.5	initiation		03:31.4	initiation
7	03:30.4	initiation		03:40.4	initiation
8	03:38.5	initiation		04:20.0	initiation
9	04:19.0	initiation		04:58.4	initiation
10	04:57.9	initiation		05:07.5	initiation
11	05:05.7	initiation		05:14.4	initiation
12	05:14.0	initiation		05:29.6	initiation
13	05:28.7	initiation			
14					
15	Point by Point		0.916667		
16					
17					

Note. Discrepancies were identified using the point-by-point method with a 6-second agreement time window (Yoder et al., 2018).

Appendix D

Procedural Fidelity Count Data

Participant: ________________ Date: __________________________

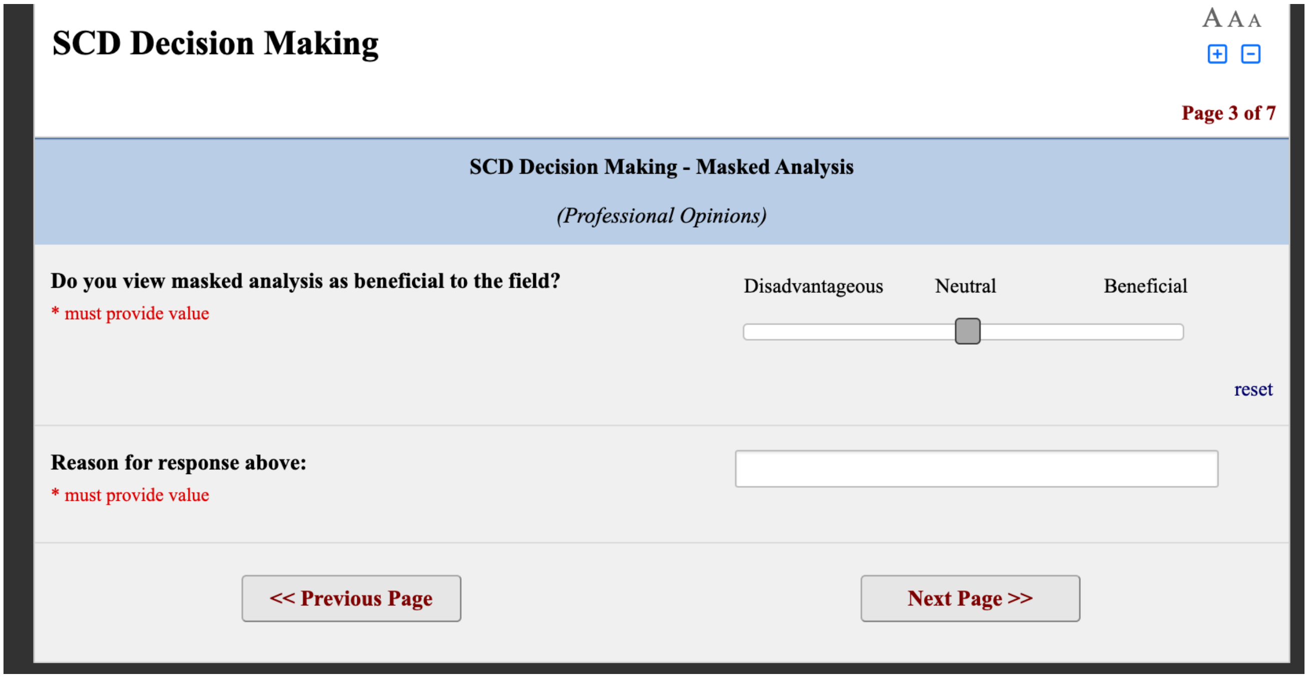
Implementer: _______________ Observer: _______________________

Session: ___________________ Condition: ______________________

Behavior	Observational Data	
1. Did the implementer offer two different play sets at the start of the session?	*Circle One* Yes	No
2. *If baseline*, did the implementer respond to child initiations with a redirection to the toy set using a neutral tone?	*Tally* Yes	No
3. *If baseline*, did the implementer play with the child for at least 3 minutes at the conclusion of the session, unless the child declined or indicated they were finished before 3 minutes was up?	*Circle One* Yes	No
4. Did the implementer offer the child a sticker or stamp at the conclusion of the session?	*Circle One* Yes	No
(Correct behaviors / Total behaviors observed) ______________ / ___________________ IOA = ______________________	*Count* Yes	No

Appendix E

Survey Format Example



SCD Decision Making

A A A

Page 3 of 7

SCD Decision Making - Masked Analysis

(Professional Opinions)

Do you view masked analysis as beneficial to the field?

* must provide value

Disadvantageous Neutral Beneficial

reset

Reason for response above:

* must provide value

<< Previous Page

Next Page >>

Note. Disadvantageous is represented by zero, neutral is represented by neutral, and beneficial is represented by a one hundred.