ASSESSMENT OF THE SOCIAL VALIDITY OF TREATMENT OUTCOMES

by

MELISSA HINES

(Under the Direction of Joel Ringdahl)

ABSTRACT

The current literature focusing on the social validity of the commonly used goal of an 80% reduction from baseline rates of responding is scarce to non-existent. In this study, we showed caregiver participants reduction exemplar videos with various levels of reduction from baseline rates of aggressive behavior. After each reduction exemplar video, the caregiver participants completed a rating scale in order to measure their acceptability of the behavior and the improvement from baseline. The data from this study showed mixed results both supporting and opposing the use of an 80% reduction in aggressive behavior as a socially valid goal for treatment outcome.

INDEX WORDS: Social validity, 80% reduction, Aggressive behavior, Acceptability, Treatment outcomes, Severity, Improvement, Applied behavior analysis

ASSESSMENT OF THE SOCIAL VALIDITY OF TREATMENT OUTCOMES

by

MELISSA HINES

Bachelor of Science, Kennesaw State University, 2020

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2024

ASSESSMENT OF THE SOCIAL VALIDITY OF TREATMENT OUTCOMES

by

MELISSA HINES

Major Professor:	Joel Ringdahl
Committee:	Jennifer Hodnett
	Kevin Ayres

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2024

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

ASSESSMENT OF THE SOCIAL VALIDITY OF TREATMENT OUTCOMES

INTRODUCTION

In recent years, social validity has become an increasingly important topic in every field of psychology, including Applied Behavioral Analysis (ABA). Social validity in ABA is defined as the extent to which the goals, treatments, or outcomes of an intervention are socially acceptable and important (Foster & Mash, 1999; Snodgrass et al., 2018; Wolf, 1978). Given that the clinical field involves direct client care, it is impacted by social validity. Behavior analysts can create and implement treatments with empirically supported evidence, decrease challenging behavior, and/or increase adaptive functioning. Regardless of clinical significance to practitioners, the most important part of a behavioral treatment is the impact of the change for direct and indirect consumers which includes client's caregivers (Schwartz & Baer, 1991; Wolf, 1978). If the behavior change is not socially valid or meaningful to caregivers, behavior has not been changed effectively (Baer et al., 1968; Foster & Mash, 1999).

In ABA that takes place in a clinical context, there are common outcomes behavior analysts strive for when implementing interventions. For practitioners focusing on behavior reduction, a common goal is to obtain an 80% reduction in target behaviors from baseline rates. Over the years, an 80% reduction in target behavior has become routinely used in behavior reduction as the indication of a successful intervention (Hanley et al., 2005; Hagopian et al., 2020). In a brief review of 34 research articles from the past five years published in the Journal of Applied Behavioral Analysis, conducted by the primary investigator, only 13 out of 34 articles

reported their reduction goal. Within these 13 articles, six reported their reduction goal was 80% and the other seven articles used other reduction goals ranging between 50% and 100% (contact first author for the full list of articles included, Table 1.1 shows the goals used by the seven articles in the review that did not use an 80% reduction as a measure of success).

Table 1.1

Articles that did not use an 80% Reduction

Article	Reported Percent Reduction for Success
Briggs et al., 2019	90%
Frank-Crawford et al., 2021	50%
Greer et al., 2023	85%
Rose & Beaulieu, 2019	90%
Stuesser & Roscoe, 2020	100%
Tsami & Lerman, 2020	100%
Verriden & Roscoe, 2019	70%

An example would be a child with baseline rates of challenging behavior equivalent to 60 instances per hour. For this child, an 80% decrease in challenging behavior would be 12 instances an hour. To behavior analysts, this change is significant, and if data is stable at this level, treatment may be concluded as successful. However, from the perspective of the caregivers

and family members, this may not be a significant change. This child's challenging behavior, at 12 instances an hour, could still be a crisis for this family. What if challenging behavior is so intense it causes significant injury to the child or family members? What if there is an infant or toddler being targeted by the child's challenging behavior? What if the level of destruction in the home causes significant financial troubles for the family? What if the baseline levels of challenging behavior are so high an 80% decrease is still over 100 instances an hour? In all these circumstances, the 80% reduction of challenging behavior may be a significant change for behavior analysts, but not for the consumers of the treatment.

Little to no research has been done to assess the social validity of establishing a goal to obtain an 80% reduction from baseline responding or evaluate if this should be a universal and routinely used measure of success in behavior reduction. The purpose of this study was to evaluate the goal of an 80% reduction in target behavior's social validity and acceptability by comparing it to other levels of behavior reduction.

CHAPTER 2

AN ANALYSIS OF THE SOCIAL VALIDITY OF THE 80% REDUCTION LINE METHOD

Participants

The experimenter recruited three caregiver participants through the waitlist (clients waiting for treatment), current clients (clients currently receiving treatment) and discharged clients (clients who have finished treatment) of an autism center. The inclusion criteria included speaking English as their first language and identifying as the primary caregiver of a child who has a diagnosis of a developmental disability. Additionally, the caregiver's child previously engaged or currently engages in aggressive behavior.

Setting

Study procedures took place at an autism center in the southeast. Upon arrival, caregivers were taken to a meeting room with a table and chairs. The experimenter sat next to the participants while delivering instructions and showing the videos. Materials included a laptop computer, pencil or pen, the reduction exemplar videos, and the social validity rating scale.

Materials

The experimenter recruited three individuals for the use of audiovisual recordings from their admission at the autism center. These audiovisual recordings were developed into the reduction exemplar videos being used in the study. The inclusion criteria for the individuals included that they had to have received treatment in the severe behavior program for aggression

at an autism center. Demographics and treatment information about clients included in the reduction exemplar videos are displayed in Table 2.1.

Table 2.1

Demographics for Individuals in Reduction Exemplar Videos

The acronyms being used are "H" for height, "W" for weight, "lbs" for pounds, "ASD" for autism spectrum disorder, "ADHD" for attention deficit hyperactivity disorder, "AGG" for aggression, "DIS" for disruption, "ELOPE" for elopement, "DRA" for differential reinforcement of alternative behavior, "MS" for multiple schedule, and "RIRD" for response interruption and redirection.

Individual	Gender	Age	Size (H; W)	Diagnosis	Target Behavior(s)	Treatment During Videos
1	Male	13	N/A; 141 lbs	ASD, ADHD	AGG, SIB, DIS, ELOPE	FCT+DRA+RIRD +Activity
2	Male	14	5'3"; 93 lbs	ASD	AGG, DIS	MS+Token Economy
3	Male	12	4'10"; 117 lbs	ASD	AGG, SIB	FCT+MS+ Punishment

For each of the three individuals, the baseline videos were developed to be representative of high levels of aggression clients may display before starting a behavioral intervention. The other three videos represented a 45-55% reduction, a 75-85% reduction, and a 100% reduction in aggression. Detailed information about each reduction exemplar video is included in Table 2.2.

Table 2.2

Description of Reduction Exemplar Videos

	Video	Rate Per Minute Aggression
Video Set 1 (Client 1)	Baseline	16.6
	45-55% Reduction	9
	75-85% Reduction	3.2
	100% Reduction	0
Video Set 2 (Client 2)	Baseline	9
	45-55% Reduction	4.4
	75-85% Reduction	1.8
	100% Reduction	0
Video Set 3 (Client 3)	Baseline	7.2
	45-55% Reduction	3.4
	75-85% Reduction	1.8
	100% Reduction	0

A range was used in place of a 50% reduction and an 80% reduction. This is because it was not feasible to find a 5-minute video that represented an exact 50 % or 80% reduction in aggressive behavior for the individuals used in the study.

For the individuals in the reduction exemplar videos, verbal consent was obtained from their caregivers by the experimenter to use audiovisual recordings to share with other study participants. Verbal consent was documented and retained. The videos were already recorded for their child's treatment purposes and in the possession of the autism center. So, consent was obtained verbally, over the phone, for the convenience of the consenting caregivers. Consent was not obtained from the clients themselves as they were all under the age of 18. Assent was not obtained because all recruited clients have high support needs and/or were unable to use complex

communication skills to understand assent and/or sign an assent form. They did not have any other participation in the study other than being present in the reduction exemplar videos shown to caregiver participants.

Social Validity Rating Scale

The social validity rating scale consisted of one question for the baseline videos and two questions for the videos that represent the client's behavior after intervention has started. The rating scale given after the baseline videos is included in Appendix A and the rating scale given after the reduction videos is also included in Appendix B. Question one was developed to evaluate the overall impact of the challenging behavior displayed in the video. More simply, question one is to evaluate the severity of the challenging behavior. Question two was developed to evaluate the acceptability of the change in aggressive behavior from video one, to each of the reduction videos.

The Clinical Global Impressions-Improvement Scale (CGI-I) and the Clinical Global Impressions-Severity Scale (CGI-S) informed the development of the rating scale used in this study (Guy, 1976). The CGI-I and CGI-S were selected in consideration of their relevance to the study. Specifically, these scales examine severity of behavior, impact of behavior, and outcome of intervention. (Guy, 1976; Toolan et al., 2022). Additionally, the CGI-I and CGI-S have empirical support for clinical application (Bearss, 2015; Lomas Mevers et al., 2020; Sheridan et al., 2021; Toolan et al., 2022). The questions were adapted to fit the current study.

Caregiver participants rated question one on a scale from 1 (Normal, i.e., as if the individual had no unique needs beyond typical needs of a child) to 7 (Extreme, i.e., impossible to perform regular activities of daily life; Toolan et al., 2022). A higher score by caregiver participants would indicate higher severity of the aggressive behavior. For question two,

caregivers rated from 1 (Very Much Improved, i.e., as if the individual had no unique needs beyond typical needs of a child/treatment is not needed), through 4 (No Change), to 7 (Very Much Worse; Bearss, 2015; Lomas Mevers et al., 2020; Sheridan et al., 2021). A lower score by caregiver participants would indicate greater improvement in the aggressive behavior. Descriptions were added to the Likert-type scales to help further the caregiver participant's understanding of the choices.

Data Collection

For this study, the data collection process was simple. For the reduction exemplar videos, the experimenter recorded frequency data for aggressive behavior on each video and converted it into a rate per minute. For the rating scale, the experimenter took the data from each caregiver participant and entered it into an excel document for analysis.

The average interobserver agreement (IOA) was 91% for the rates of aggression displayed in the reduction exemplar videos. Additionally, IOA data was collected on the experimenter data collection from the rating scale. For rating scale data collection, IOA was 100%. To promote procedural fidelity, a script was created to be read during each meeting with caregiver participants. A copy of the script can be found in Appendix C.

General Procedures

This study used a multi-element design and a survey. For caregiver participants, written consent was obtained during appointment one. Each caregiver participant was then assigned scheduled times where they met with an experimenter at the autism center, watched three different sets of four videos (12 total), and completed social validity rating scales. Each set of videos shows a different individual engaging in aggressive behavior (three different individuals total). The order of individuals in the videos and of the reduction exemplar videos themselves

was randomized for each participant. The description of the order of reduction exemplar videos for each participant is included in Table 2.3.

Table 2.3

Order of Reduction Exemplar Videos for Each Participant

	Client Order	Video #1	Video #2	Video #3	Video #4
Participant #1	Client 2	Baseline	75-85%	45-55%	100%
	Client 1	Baseline	75-85%	100%	45-55%
	Client 3	Baseline	75-85%	45-55%	100%
Participant #2	Client 2	Baseline	75-85%	45-55%	100%
	Client 1	Baseline	100%	75-85%	45-55%
	Client 3	Baseline	75-85%	45-55%	100%
Participant #3	Client 1	Baseline	45-55%	100%	75-85%
	Client 3	Baseline	45-55%	75-85%	100%
	Client 2	Baseline	75-85%	45-55%	100%

The data obtained by the social validity rating scale was then recorded and analyzed by the experimenter. After the last appointment, caregiver participants were debriefed on the purpose of the study.

Specific Procedures

Appointment One

Once caregiver participants arrived for their first appointment, they were taken to a room with a table and two chairs. The experimenter provided a written consent form for the caregiver participants and read the entire form. After confirming that they understood all the information within the written consent document and that they had no questions, the experimenter asked the caregiver participants if they would like to sign the consent form and participate in the study.

Once the form was signed by the caregiver participant, the experimenter scanned the document for secure storage, provided them with a copy and walked them out.

Appointment Two

Just as in appointment one, caregiver participants were taken to a room with a table and two chairs upon arrival to the autism center. Once seated the experimenter began by reading the first part of the script to the caregiver participants which describes the layout of the appointment, reminds them of the risks involved and that they can remove themselves from the study at any time, and asks them to refrain from speaking to the experimenter during the playing of the reduction exemplar videos.

Once the experimenter had the first set of reduction exemplar videos ready on the laptop, they read the next section of the script that explains the background of the first video, which contains the baseline rate in aggressive behavior for that individual. After watching the 5-min baseline video, the experimenter read the next section of the script and had the caregiver participant complete one rating scale question. After the completion of the first question, the procedure for viewing the other three reduction exemplar videos was as follows: the experimenter read the next section of the script and had the caregiver participants view their first of the other three reduction exemplar videos which represented a 45-55% reduction, a 75-85% reduction, or a 100% reduction in aggressive behavior from the baseline video. After observing this video, the experimenter read the next part of the script and had the caregiver participants fill out the second rating scale that included two questions. After the caregiver participants viewed all four of the reduction exemplar videos (baseline and the other three reduction exemplar videos) and filled out the provided rating scales, the experimenter collected the completed rating scales and walked them out.

Appointment Three

Just as in appointment one and two, caregiver participants were taken to a room with a table and two chairs upon arrival to the autism center. Once seated the experimenter began by reading the first part of the script to the caregiver participants which describes the layout of the appointment, reminds them of the risks involved and that they can remove themselves from the study at any time, and asks them to refrain from speaking to the experimenter during the playing of the reduction exemplar videos.

Once the experimenter had the second set of reduction exemplar videos ready on the laptop, they read the next section of the script that explains the background of the first video, which contains the baseline rate in aggressive behavior for the second individual. After watching the 5-min baseline video, the experimenter read the next section of the script and had the caregiver participant complete one rating scale question. After the completion of the first question, the procedure for viewing the other three reduction exemplar videos was as follows: the experimenter read the next section of the script and had the caregiver participants view their first of the other three reduction exemplar videos which represented a 45-55% reduction, a 75-85% reduction, or a 100% reduction in aggressive behavior from the baseline video. After observing this video, the experimenter read the next part of the script and had the caregiver participants fill out the second rating scale that included two questions. After the caregiver participants viewed all four of the reduction exemplar videos for the second individual (baseline and the other three reduction exemplar videos) and filled out the provided rating scales, the experimenter collected the completed rating scales and walked them out.

Appointment Four

Just as in appointment one, two, and three, caregiver participants were taken to a room with a table and two chairs upon arrival to the autism center. Once seated the experimenter began by reading the first part of the script to the caregiver participants which describes the layout of the appointment, reminds them of the risks involved and that they can remove themselves from the study at any time, and asks them to refrain from speaking to the experimenter during the playing of the reduction exemplar videos.

Once the experimenter had the third set of reduction exemplar videos ready on the laptop, they read the next section of the script that explains the background of the first video, which contains the baseline rate in aggressive behavior for the third individual. After watching the 5-minute baseline video, the experimenter read the next section of the script and had the caregiver participant complete one rating scale question. After the completion of the first question, the procedure for viewing the other three reduction exemplar videos was as follows: the experimenter read the next section of the script and had the caregiver participants view their first of the other three reduction exemplar videos which represented a 45-55% reduction, a 75-85% reduction, or a 100% reduction in aggressive behavior from the baseline video. After observing this video, the experimenter read the next part of the script and had the caregiver participants fill out the second rating scale that included two questions. After the caregiver participants viewed all four of the reduction exemplar videos for the third individual (baseline and the other three reduction exemplar videos) and filled out the provided rating scales, the experimenter collected the completed rating scales. After the conclusion of the fourth and final appointment, the experimenter asked caregivers if they would like more information on the purpose of the study,

provided this information if requested and asked the caregiver participants if they had any further questions before walking them out.

CHAPTER 3

AN ANALYSIS OF THE SOCIAL VALIDITY OF THE 80% REDUCTION LINE

RESULTS

Analyses of the data from the rating scales are included in Table 3.1 and are represented visually in Figure 3.1, Figure 3.2, Figure 3.3, Figure 3.4, Figure 3.5, Figure 3.6, and Figure 3.7.

Table 3.1

Rating Scale Results

Included in the table is the median score given for each question in all 4 categories of reduction exemplar video (baseline aggressive behavior, 45-55% reduction in aggressive behavior, 75-85% reduction in aggressive behavior, and 100% reduction in aggressive behavior).

	Rating Scale Question	Median Score
Baseline	Question 1	7
45-55% Reduction	Question 1	6
	Question 2	3
75-85% Reduction	Question 1	5
	Question 2	2
100% Reduction	Question 1	3
	Question 2	2

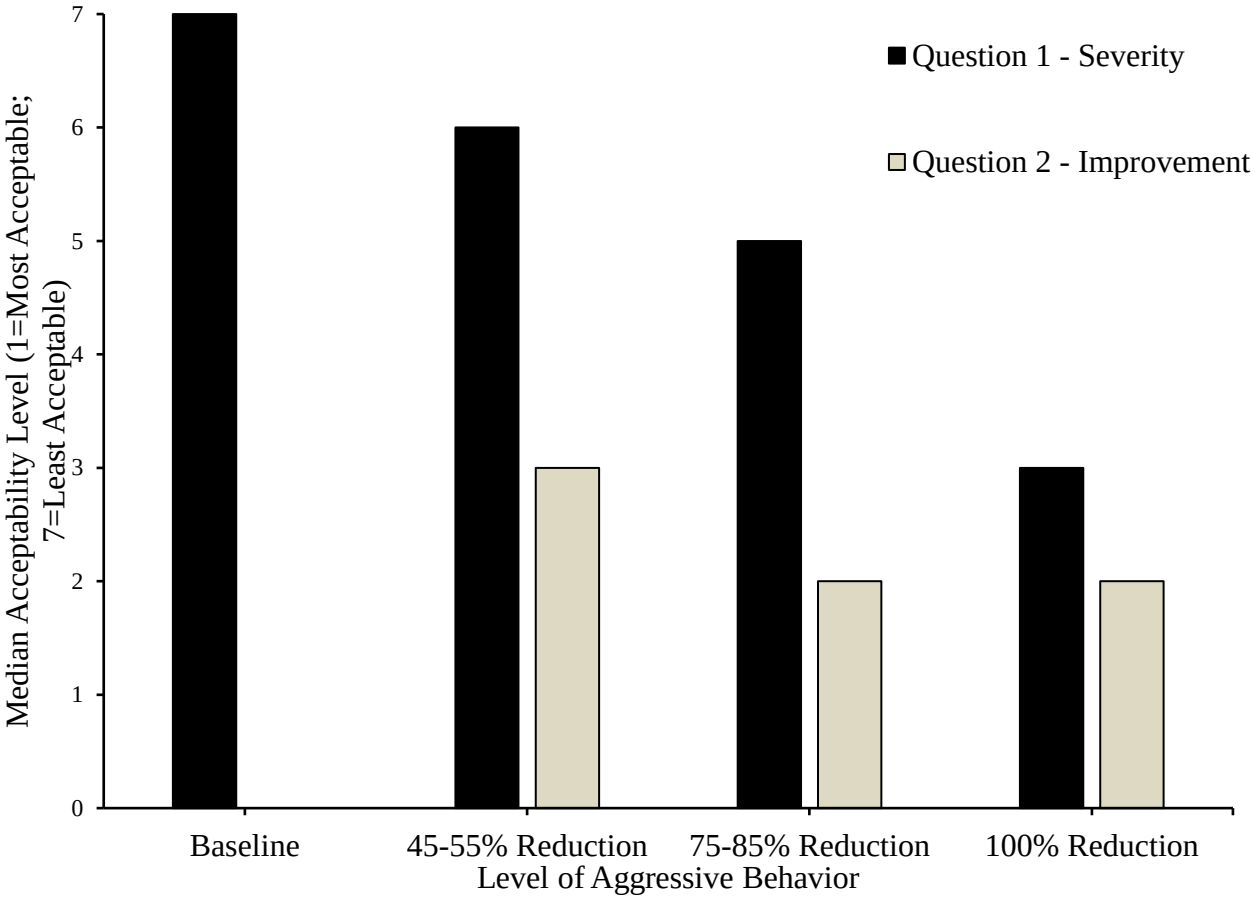

Figure 3.1

Median Severity and Improvement Ratings Based on Question and Level of Aggressive Behavior

Figure 3.1 shows caregiver's median acceptability scores. In this graph, the bars correspond to the different rating scale questions as identified in the legend. As with the CGI-I and CGI-S, a lower score corresponds to a higher level of acceptability or improvement (Sheridan et al., 2021). So, the lower the bar the more acceptable the aggressive behavior level and treatment outcome. The median rating for overall acceptability of aggressive behavior in the baseline video was 7 (Extreme, i.e., impossible to perform regular activities of daily life). For the 45-55% reduction, question one's median score was 6 (Severe) and question two's median score was 3 (Minimally Improved). For the 75-85% reduction, question one's median score was 5

(Marked) and question two's median score was 2 (Much Improved). For the 100% reduction, question one's median score was 3 (Mild) and question two's median score was 2 (Much Improved).

Participant One

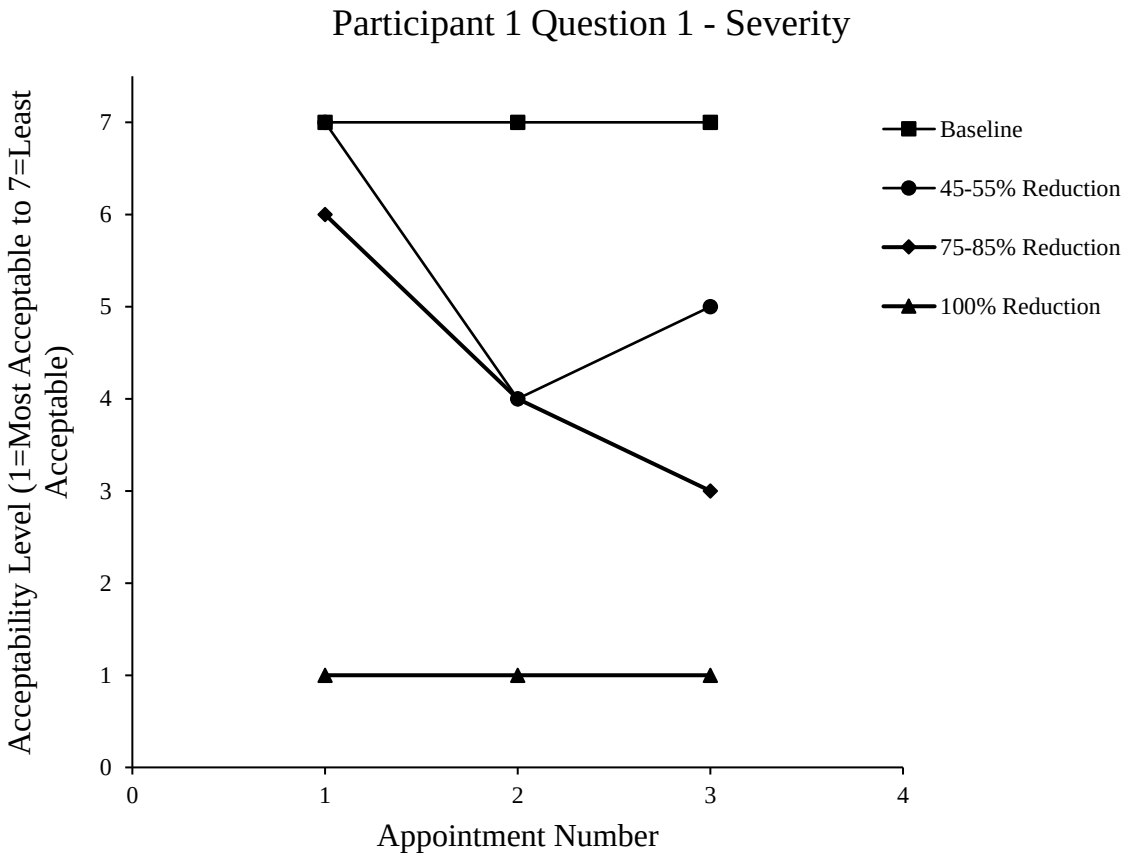

Figure 3.2

Median Severity Ratings for Each Video and Each Appointment for Participant One

Figure 3.2 shows participant one's median severity scores based on reduction level and appointment.

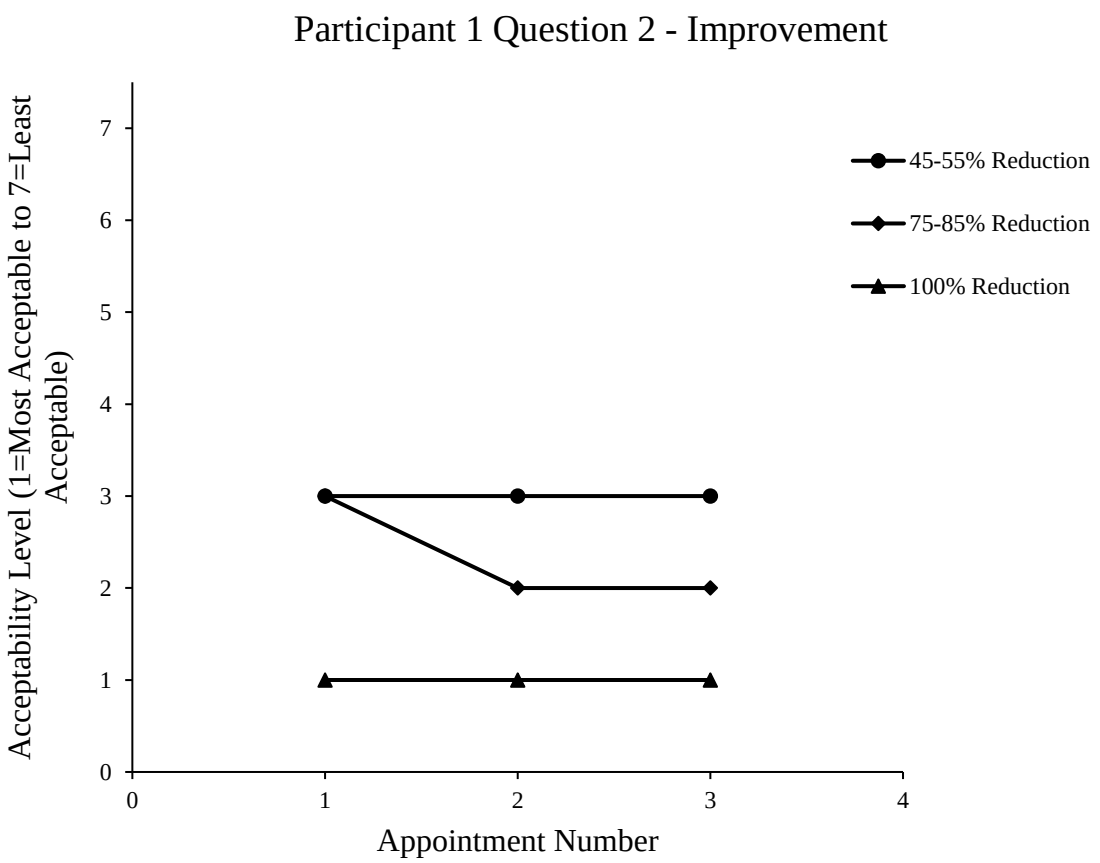

Figure 3.3

Median Improvement Ratings for Each Video and Each Appointment for Participant One

Figure 3.3 shows participant one's median improvement scores based on reduction level and appointment.

Participant Two

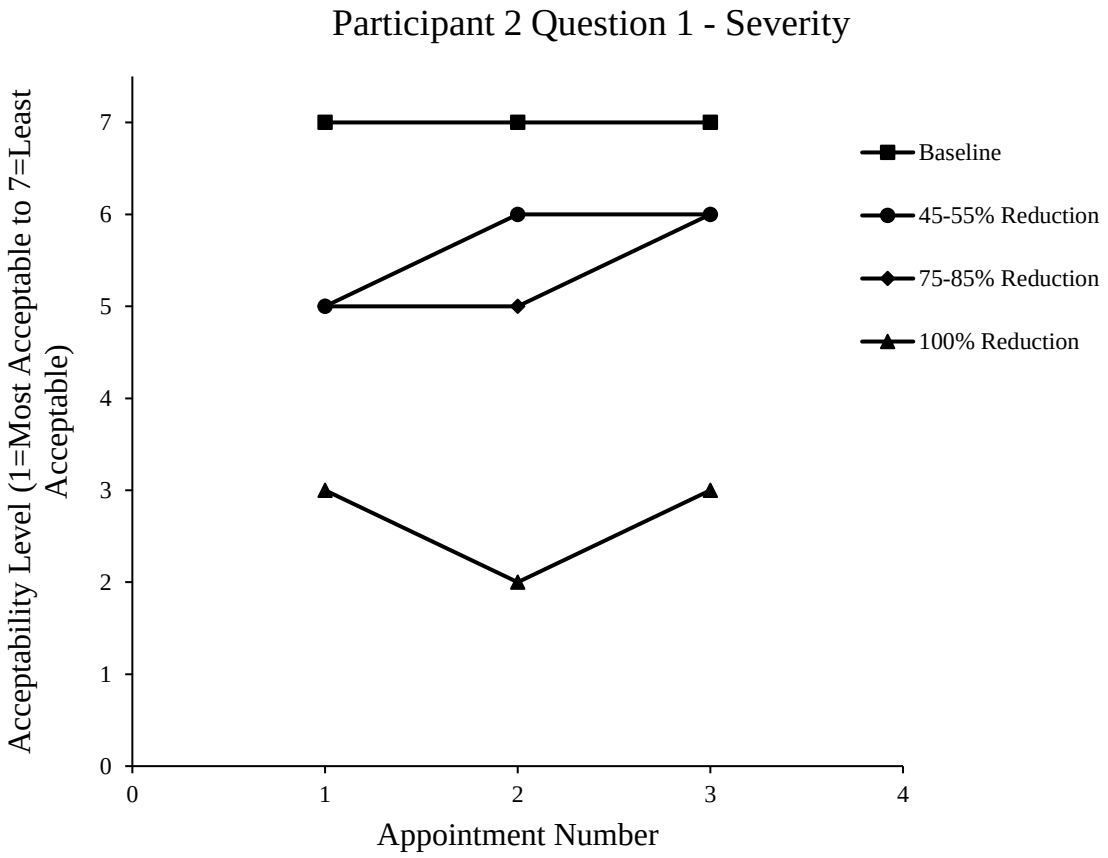

Figure 3.4

Median Severity Ratings for Each Video and Each Appointment for Participant Two

Figure 3.4 shows participant two's median severity scores based on reduction level and appointment.

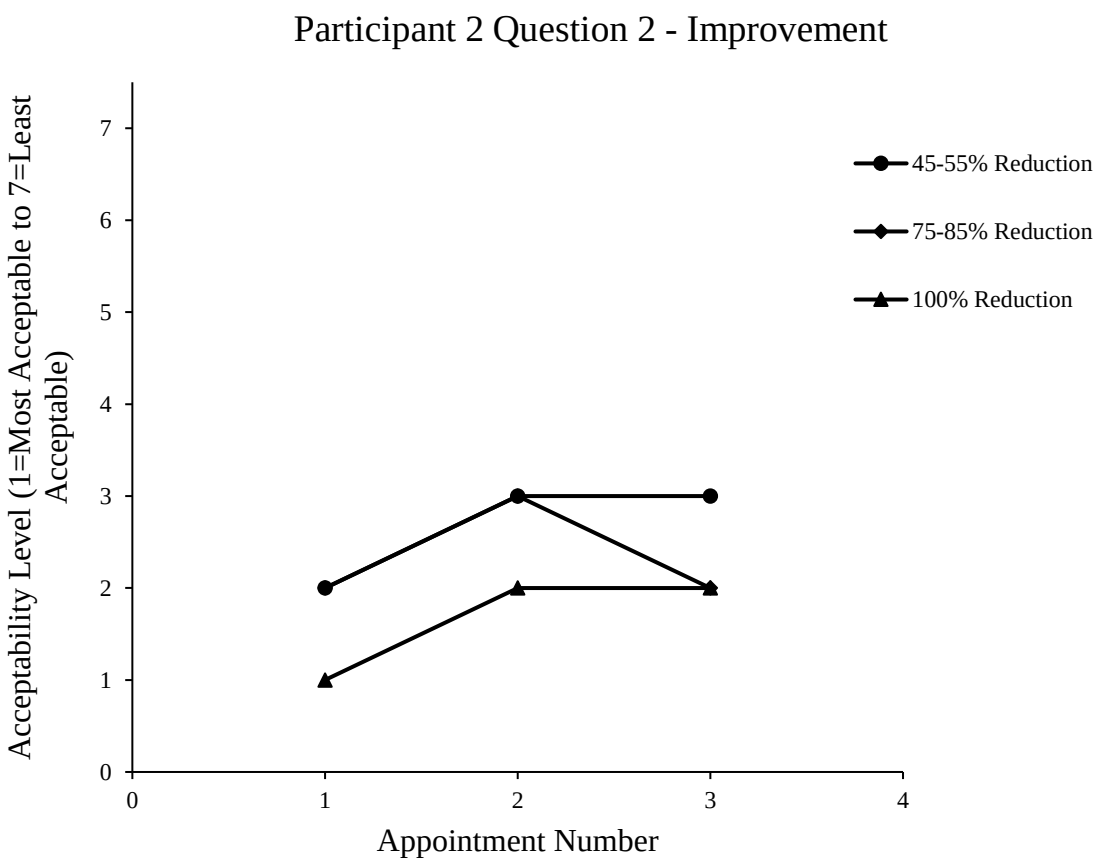

Figure 3.5

Median Improvement Ratings for Each Video and Each Appointment for Participant Two

Figure 3.5 shows participant two's median improvement scores based on reduction level and appointment.

Participant Three

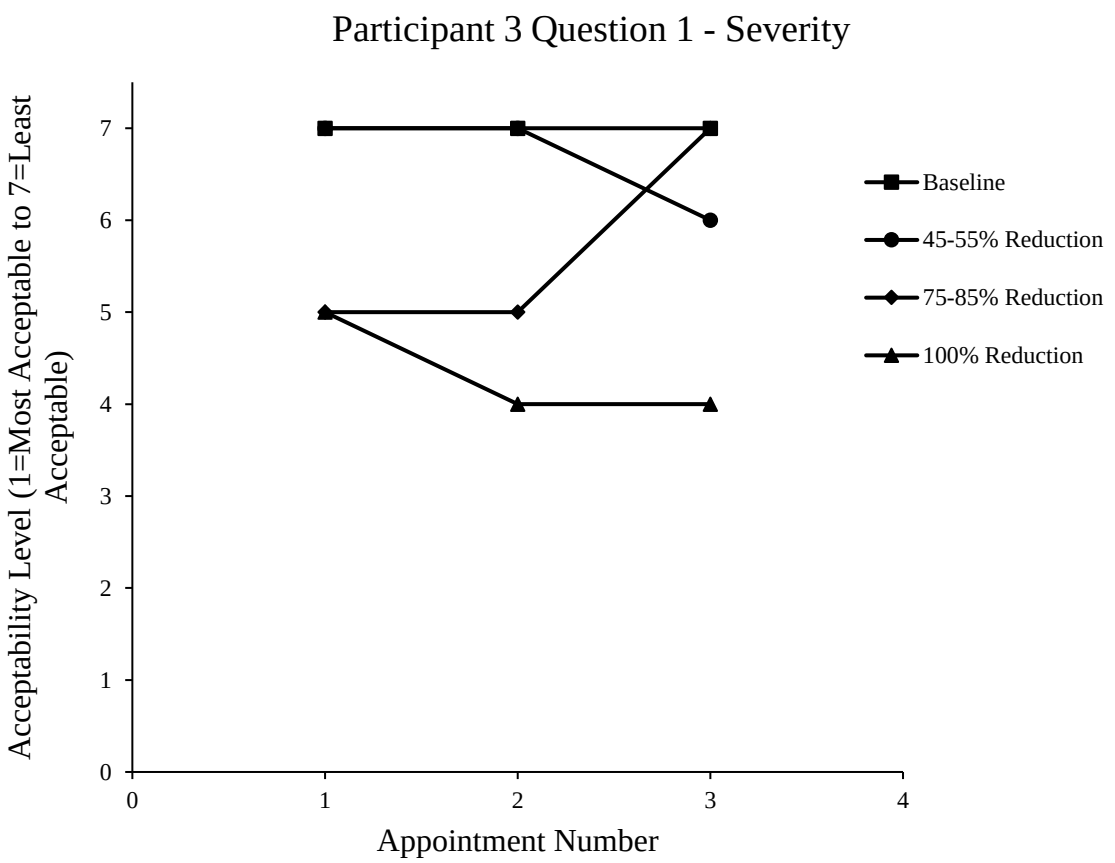

Figure 3.6

Median Severity Ratings for Each Video and Each Appointment for Participant Three

Figure 3.6 shows participant three's median severity scores based on reduction level and appointment.

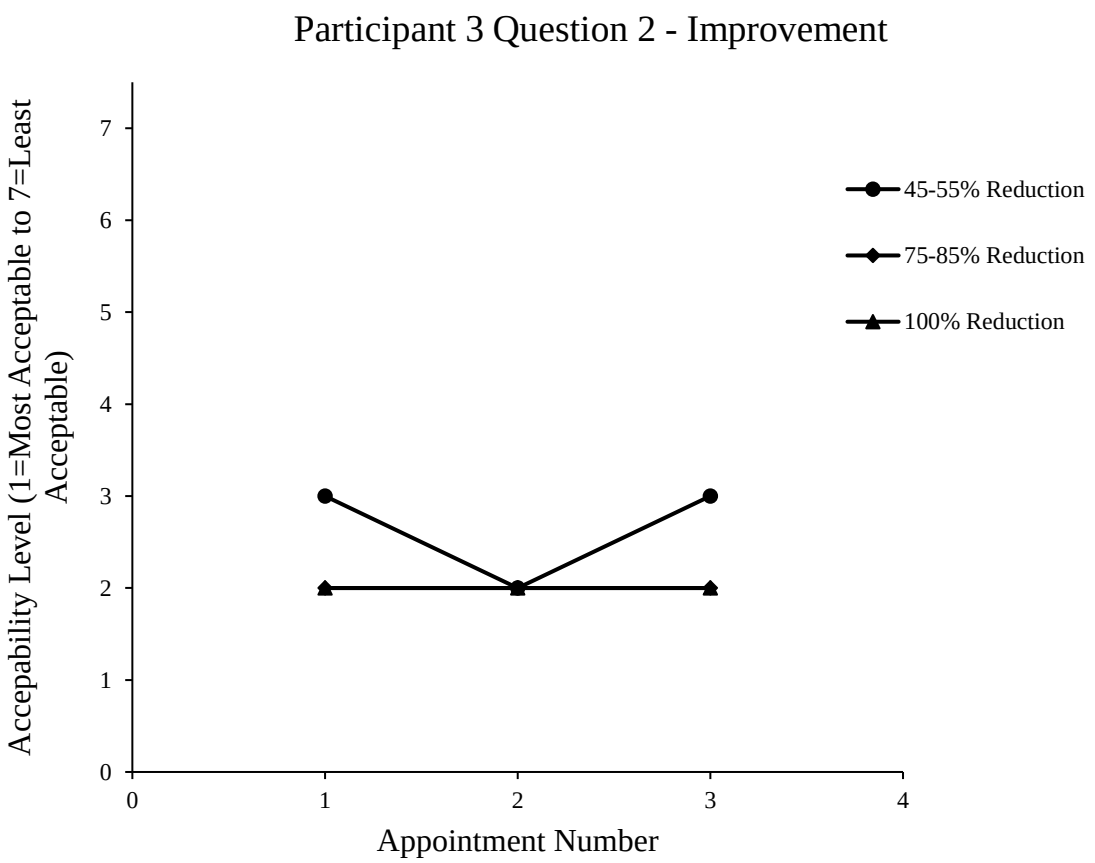

Figure 3.7

Median Improvement Ratings for Each Video and Each Appointment for Participant Three

Figure 3.7 shows participant three's median improvement scores based on reduction level and appointment.

CHAPTER 4

AN ANALYSIS OF THE SOCIAL VALIDITY OF THE 80% REDUCTION LINE

DISCUSSION

This study used a social validity rating scale to collect caregivers' assessments of the severity, impact, and level of improvement for different levels of reduction in aggressive behavior. Given the prevalence of utilizing an 80% reduction goal for behavioral interventions, the experimenter sought to specifically investigate how an 80% reduction compares to other levels of behavior reduction.

Based on visual inspection of Figure 3.1, a hierarchy of levels of behavior reduction was formed. This hierarchy outlines least acceptable ratings to most acceptable ratings with the order of baseline, 45-55% reduction, 75-85% reduction, and 100% reduction as the most acceptable. While predictable, this information is important in demonstrating caregivers can distinguish between varying levels of behavior reduction.

When examining how the median acceptability scores compare, some interesting inferences can be made. Ideally, if a 100% reduction in aggressive behavior represents perfection, it would be best if the 75-85% reduction acceptability scores, for each question, were close to the 100% reduction acceptability scores. This would help support the use of an 80% reduction in aggression as a goal for treatment outcomes since the score would be close to perfection.

When comparing each participant's graphs from question one, it is clear that baseline is rated consistently as the least acceptable and the 100% reduction is rated as the most acceptable.

However, for participant two and three, the 100% reduction is not rated as perfect. The graphs for question one also show that the 45-55% reduction videos and 75-85% reduction videos are not consistently rated on the same level of acceptability, meaning each participant's views of the acceptability of the severity of the aggressive behavior in these two video categories differ.

The second question about the improvement in aggressive behavior also had some interesting results as well. The hierarchy of videos is much harder to see in these graphs and, for participant two and three, there is not much of a difference in the median scores for the 45-55% reduction videos and the 75-85% reduction videos. Most of the data for all of the participants for question two is compiled all at the same level at one, two, and three.

Considering the ordinal data produced by the Likert-type of data collected, an area that could be further explored is the manner with which statistical analyses were conducted. MANN Whitney U Test could be an avenue to compare the scores more accurately across participants since the data collected in the study are not interval data (Sullivan & Artino, 2013). Since it is not interval data, it cannot be assumed that the distance between each response is the same (ex., the distance between 3-Mild to 4-Moderate is the same as the difference between 4-Moderate and 5-Marked) (Sullivan & Artino, 2013). It also cannot be assumed that the data are normally distributed (Sullivan & Artino, 2013). This is also why the best measure of central tendency to use in the study was median instead of mean (Sullivan & Artino, 2013).

Although the data from this study rendered mixed results, there are limitations which may have affected these results. Firstly, the study was conducted in an autism center specializing in behavior intervention for children with complex and intense aggressive behavior. Therefore, results may not have generality for adult populations, populations of individuals with lower intensity aggressive behavior, and/or populations of children who engage in lower frequency

aggressive behavior. Another limitation is all the clients in the reduction exemplar videos were males. So, there is no way to know how viewing a female engaging in aggressive behavior would have affected the caregiver's scores of the reduction exemplar videos. The third limitation is, we did not collect any demographic information on the caregiver participants who were rating the reduction exemplar videos. This information could have given some interesting insight as to how various characteristics (e.g., gender, age, etc.) and family dynamics (e.g., number of children in the household, number of caregivers, etc.) could affect the caregiver participants ratings.

Other than limitations, this study and its results leave a number of future directions for research in this area. Specifically, research could be conducted to further analyze how the type of intervention being used in the reduction exemplar videos could affect the caregiver's ratings of the acceptability of the severity, impact, and treatment outcomes. Similarly, given the singular focus on aggressive behavior in the reduction exemplar videos, an area that can be further explored, is the relation between topography and level of acceptability.

Another future direction or advancement that could be made, is creating a statistically validated measure of the acceptability of treatment outcomes. The current literature concerning the social acceptability of treatment outcomes is extremely scarce. Carter and Wheeler (2019) identify around 15 measures of treatment acceptability in The Social Validity Manual. However, no measures of the acceptability of treatment outcomes were identified (Carter & Wheeler, 2019). In ABA, a statistically validated measure for the acceptability of treatment outcomes would greatly impact behavior analysts' abilities to ensure the effect of treatment is significant enough to positively impact the lives of our clients and their families.

Overall, the results of this study highlight an incredibly important question for behavior analysts striving for behavior reduction. Is an 80% reduction enough? Or better yet, what level of

behavior change would make the most impactful and socially valid difference for the client and their family?

References

Baer, D.M., Wolf, M.M. and Risley, T.R. (1968), Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1: 91-97. https://doi.org/10.1901/jaba.1968.1-91

Bearss, K., Johnson, C., Smith, T., Lecavalier, L., Swiezy, N., Aman, M., ... & Scahill, L. (2015). Effect of parent training vs parent education on behavioral problems in children with autism spectrum disorder: A randomized clinical trial. *JAMA*, 313(15), 1524-1533. doi:10.1001/jama.2015.3150

Briggs, A.M., Dozier, C.L., Lessor, A.N., Kamana, B.U. and Jess, R.L. (2019), Further investigation of differential reinforcement of alternative behavior without extinction for escape-maintained destructive behavior. *Journal of Applied Behavior Analysis*, 52: 956-973. https://doi.org/10.1002/jaba.648

Carter, S. L., & Wheeler, J. J. (2019). *The social validity manual: Subjective evaluation of interventions* (2nd ed.). Academic Press.

Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology*, *67*(3), 308–319. https://doi.org/10.1037/0022-006X.67.3.308

Frank-Crawford, M.A., Borrero, J.C., Newcomb, E.T., Doan, T., Fisher, A. and Rooker, G.W. (2021), Accumulated and distributed response–reinforcer arrangements during the treatment of escape-maintained problem behavior. *Journal of Applied Behavior Analysis*, 54: 1566-1585. https://doi.org/10.1002/jaba.870

Greer, B.D., Shahan, T.A., Fisher, W.W., Mitteer, D.R. and Fuhrman, A.M. (2023), Further evaluation of treatment duration on the resurgence of destructive behavior. *Journal of Applied Behavior Analysis*, 56: 166-180. https://doi.org/10.1002/jaba.956

Guy, W. (1976). US Dept Health, Education, and Welfare publication; ADM No. 76-338. Rockville, MD: National Institute of Mental Health; 1976. ECDEU Assessment Manual for Psychopharmacology; pp. 218–222.

Lomas Mevers, J., et al. (2020). A pilot randomized clinical trial of a multidisciplinary intervention for encopresis in children with autism spectrum disorder. *Journal of autism and developmental disorders*, 50(3): 757-765.

Rose, J.C. and Beaulieu, L. (2019), Assessing the generality and durability of interview-informed functional analyses and treatment. *Journal of Applied Behavior Analysis*, 52: 271-285. https://doi.org/10.1002/jaba.504

Schwartz, I.S. and Baer, D.M. (1991), Social validity assessments: is current practice state of the art?. *Journal of Applied Behavior Analysis, 24*, 189-204. https://doi.org/10.1901/jaba.1991.24-189

Sheridan, E., Gillespie, S., Johnson, C. R., Lecavalier, L., Smith, T., Swiezy, N., Turner, K., Pritchett, J., Mruzek, D. W., Evans, A. N., Bearss, K., & Scahill, L. (2021). Using parent target problem narratives to evaluate outcomes in children with autism spectrum disorder. *Research on Child and Adolescent Psychopathology*, 49(11), 1527–1535. https://doi.org/10.1007/s10802-021-00843-8

Snodgrass, M. R., Chung, M. Y., Meadan, H., Halle, J. W. (2018). Social validity in single case research: a systematic literature review of prevalence and application. *Research in Developmental Disabilities. 74*, 160-173. https://doi.org/10.1016/j.ridd.2018.01.007

Stuesser, H.A. and Roscoe, E.M. (2020), An evaluation of differential reinforcement with stimulus fading as an intervention for medical compliance. *Journal of Applied Behavior Analysis*, 53: 1606-1621. https://doi.org/10.1002/jaba.685

Sullivan, G. M., Artino, A. R. (2013), Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5 (4): 541–542. https://doi.org/10.4300/JGME-5-4-18

Toolan, C., Holbrook, A., Schlink, A., Shire, S., Brady, N., & Kasari, C. (2022). Using the Clinical Global Impression scale to assess social communication change in minimally verbal children with autism spectrum disorder. *Autism Research*, 15(2), 284–295. https://doi.org/10.1002/aur.2638

Tsami, L. and Lerman, D.C. (2020), Transfer of treatment effects from combined to isolated conditions during functional communication training for multiply controlled problem behavior. *Journal of Applied Behavior Analysis*, 53: 649-664. https://doi.org/10.1002/jaba.629

Verriden, A.L. and Roscoe, E.M. (2019), An evaluation of a punisher assessment for decreasing automatically reinforced problem behavior. *Journal of Applied Behavior Analysis*, 52: 205-226. https://doi.org/10.1002/jaba.509

Wolf, M.M. (1978), Social Validity: for the case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11: 203 214. https://doi.org/10.1901/jaba.1978.11-203

Appendix

Question 1: How impactful is the child's level of behavioral difficulties? By impactful, consider safety, resources expended by family, stress, etc. Rating is between 1, as if the individual had no unique needs beyond typical needs of a child, to 7, impossible to perform regular activities of daily life.

1= Normal (i.e. as if the individual had no unique needs beyond typical needs of a child)

2= Borderline

3= Mild

4= Moderate

5= Marked

6= Severe

7= Extreme (i.e. impossible to perform regular activities of daily life)

Appendix A

Rating Scale Given After Baseline

This rating scale was developed with a seven choice Likert scale for each question. Each scale codes from one to seven, one being most acceptable and five being the least acceptable.

Question 1: How impactful is the child's level of behavioral difficulties? By impactful, consider safety, resources expended by family, stress, etc. Rating is between 1 as if the individual had no unique needs beyond typical needs of a child, to 7, impossible to perform regular activities of daily life.

1= Normal (i.e. as if the individual had no unique needs beyond typical needs of a child)

2= Borderline

3= Mild

4= Moderate

5= Marked

6= Severe

7= Extreme (i.e. impossible to perform regular activities of daily life)

Question 2: (Treatment Only): Based on the original video, how acceptable is this amount of change in challenging behavior? By acceptable, consider safety, resources expended by family, stress, etc. Rating is between 1, as if the individual had no unique needs beyond typical needs of a child/ treatment is not needed, to 7, very much worse.

1= Very Much Improved (i.e. as if the individual had no unique needs beyond typical needs of a child/treatment is not needed)

2= Much Improved

3= Minimally Improved

4= No Change (i.e. the child's behavior is exactly the same as the original video shown/treatment is necessary)

5= Minimally Worse

6= Much Worse

7= Very Much Worse

Appendix B

Rating Scale Given After Reduction Exemplar Videos

This 2-question rating scale was developed with a seven choice Likert scale for each question. Each scale codes from one to seven, one being most acceptable and five being the least acceptable.

You will watch a series of four 5-minute videos of a client where they are engaging in aggression. The first video shows their aggression before treatment, and the other 3 videos that we will show you are after treatment. After each video, we will provide you with one or two questions to answer with your opinion.

As these videos may be hard to watch at times, please let me know at any time during the appointment if you would like a break or need to end the appointment for the day. Remember, participation is voluntary and you can remove yourself from the study at any time.

Once you begin watching the videos we ask that you focus and refrain from engaging in conversation with staff members. As a reminder, you are to be focusing on the aggressive behavior in these videos.

Before the first video: This video represents the child's behavior before treatment starts.

After video one: Here is the first question for you to answer. Let me know when you are finished.

Before video two, three, and four: For this video, keep in mind the change in behavior from video one.

After video two, three, and four: Here are your questions to answer. Let me know when you are finished.

Appendix C

Appointment Script

This is the script the experimenter read at each meeting with caregiver participants.