Summer 8-13-2021

Translational Evaluation of Treatment Integrity Following Training with Varying Destructive Behavior Rates

Alexandra Hardee
University of Nebraska Medical Center

TRANSLATIONAL EVALUATION OF TREATMENT INTEGRITY FOLLOWING TRAINING WITH CARYING DESTRUCTIVE BEHAVIOR RATES

by

Alexandra Hardee

A DISSERTATION

Presented to the Faculty of

the University of Nebraska Graduate College

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

Medical Sciences Interdepartmental Area Graduate Program

(Applied Behavior Analysis)

Under the Supervision of Professor Amanda N. Zangrillo

University of Nebraska Medical Center

Omaha, Nebraska

July, 2021

Supervisory Committee:

Amanda N. Zangrillo, Psy.D.

Keith D. Allen, Ph.D.

Bethany A. Hansen, Ph.D.

Amy S. Nordness, Ph.D.

ACKNOWLEDGEMENTS

Thank you to everyone who has made this possible for me. To Riley Ruzicka, Samantha Bryan, Heather Anderson, Sarah Martinez, Marley Brown, Jordan Rogers, Amanda Morris, Jessie Weber, Jordan DeBrine, Brinea Osborne, Rachel Norton, Cati Millier, Seth Walker, and Jason Ramsey for helping run participants and take data. Thank you to Dr. Amanda Zangrillo for guidance and support throughout my doctoral studies and project. Thank you to Dr. Billie Retzlaff for guidance throughout my years in graduate school and helping throughout this project. Thank you to my family who have supported me from multiple states across all of my years in school and life. Without your support I would not be the person I am today.

TRANSLATIONAL EVALUATION OF TREATMENT INTEGRITY FOLLOWING TRAINING WITH CARYING DESTRUCTIVE BEHAVIOR RATES

Alexandra M. Hardee, Ph.D.

University of Nebraska, 2021

Supervisor: Amanda N. Zangrillo

Behavior analysts often use behavioral skills training (BST) to teach caregivers to implement treatment for their child's destructive behavior with high levels of integrity. Even when trained to high levels of integrity, caregivers may revert to undesirable behaviors, and treatment integrity may decrease if high rates of destructive behavior occur (i.e., relapse). In the present study, we implemented a translational evaluation with adult participants to determine the impact of training under low- or high-rate destructive behavior on procedural integrity during subsequent simulated treatment challenges involving exposure to high rates of destructive behavior only. Participants trained to implement the treatment package under conditions of high-rate destructive behavior maintained higher levels of procedural integrity during treatment challenges compared to those exposed to low-rate destructive behavior during training. We discuss potential implications of these results for clinicians to consider when training caregivers in applied settings as well as areas for future research.

Keywords: behavioral skills training, destructive behavior, relapse, translational, treatment integrity

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

FCT	Functional communication training
DBx	Destructive behavior
FCR	Functional communication response
BST	Behavioral skills training
SR+	Reinforcement interval
IOA	Interobserver agreement

INTRODUCTION

Functional communication training (FCT) is an efficacious treatment for destructive behavior (e.g., Carr & Durand, 1985; Falligant & Hagopian, 2020; Gerow et al., 2018; Ghaemmaghami et al., 2021; Kurtz et al., 2011; Tiger et al., 2008). The application of FCT typically involves teaching an appropriate communicative response (i.e., functional communication response [FCR]) and delivering the reinforcer maintaining destructive behavior contingent on the alternative response while simultaneously placing destructive behavior on extinction. After establishing an FCR, behavior analysts gradually thin the schedule of reinforcement to reach terminal schedules that are feasible for caregivers to implement in the natural environment (e.g., using multiple or chained schedules; Fisher & Bouxsein, 2011; Fisher et al., 1993; Fisher et al., 2015; Greer et al., 2016; Hagopian et al., 2011; Tiger et al., 2008). Once the terminal treatment package is achieved, analysts must train caregivers to implement their child's treatment protocols with high levels of treatment integrity in both analog and naturalistic settings.

Behavioral skills training (BST) has emerged as a reliable and efficacious means for training caregivers to implement a wide range of skills with high integrity (e.g., Conklin & Wallace, 2019; Drifke et al., 2017; Miles & Wilder, 2009; Mitteer et al., 2018; Spiegel et al., 2016; Tarbox et al., 2007; Unholz-Bowden et al., 2020). For example, BST has been used to train caregivers to implement three-step prompting (e.g., Drifke et al., 2017; Miles & Wilder, 2009; Tarbox et al., 2007), differential reinforcement (e.g., Conklin & Wallace), and FCT (e.g., Mitteer et al., 2018; Saini et al., 2018) to high levels of treatment integrity. The BST procedure typically involves introducing written materials and instructions, modeling, in-vivo role-play or rehearsal, and feedback. However, an important question remains as to when and how in the treatment progression clinicians should introduce caregivers to BST and in-vivo implementation of the treatment package with their child.

There may be several advantages to training caregivers in the clinic and subsequently in the natural context (e.g., Greer et al., 2019) only after achieving an efficacious treatment package and clinically significant reductions in destructive behavior. First, because clinically significant reductions in destructive behavior are achieved (e.g., 80%-90% reduction from baseline rates), we optimize caregiver, child, and clinician safety during training. In addition, we only train caregivers on the critical components of the treatment package responsible for reductions in destructive behavior, which requires a comprehensive treatment evaluation prior to training. Last, caregivers implement the treatment package under more realistic schedules of reinforcement (i.e., terminal schedules) that are more directly generalizable to their natural setting (i.e., home or community).

However, training at this point in the treatment progression may not ultimately prepare caregivers to sustain implementation at high levels of integrity during situations in which destructive behavior may return. For example, destructive behavior may reemerge following the introduction of caregivers or transitions to new settings (e.g., stimulus or context change; Falligant et al., 2020; Ibañez et al., 2019; Kelley et al., 2018; Muething et al., 2020; Saini et al., 2018), a phenomenon termed renewal (e.g., Bouton et al., 2011; Kelley et al., 2015; Kelley et al., 2018). Much of the research on renewal comes from the basic research literature. However, a few notable translational and applied studies examined the phenomenon of renewal and its prevalence. Specifically, Saini and Mitteer (2020) identified renewal in 83% (77 of 93) of cases reviewed in the extant literature. Similarly, Muething et al. (2020) conducted a consecutive case-series analysis of clinical records for individuals who received assessment and treatment of their destructive behavior. After reviewing cases in which context changes occurred, the authors found that renewal of destructive behavior occurred in 67.2% of cases.

One notable translational study conducted by Saini et al. (2018) evaluated the renewal of destructive behavior during caregiver implementation of FCT in the home setting. Once baseline rates of destructive behavior were established in the home with caregivers, therapists implemented FCT treatment in the clinic setting until reductions of behavior occurred. Finally, caregivers implemented FCT treatment in the home setting with 100% treatment integrity across all sessions. Renewal occurred for three of the four participants during the last phase of treatment in the home setting. These results provide additional empirical support that caregivers are likely to experience renewal of destructive behavior in the home setting, even after being trained to competency in the procedures. Thus, preparing caregivers for this possibility during the training process is a critical consideration for clinicians.

When destructive behavior returns, caregivers may revert to undesirable behaviors (e.g., reinforcing destructive behavior, omitting treatment components), thus leading to corresponding decreases in treatment integrity (e.g., Mitteer et al., 2018; Mitteer et al., 2021). Maintaining high levels of treatment integrity is critically important to the ongoing suppression of destructive behavior and durability of the treatment package (e.g., Fryling & Wallace, 2012). Therefore, clinicians must understand when and how to train caregivers to high levels of integrity that persist under both ideal (i.e., once clinically significant reductions are obtained) and renewal-like (i.e., when baseline rates occur) treatment conditions.

One variable that has yet to be considered is how varying rates of destructive behavior experienced during training may impact maintenance of treatment integrity when exposed to renewal-like conditions. Therefore, the purpose of this translational study was to evaluate how varying rates (i.e., low and high) of destructive behavior

during training impacts treatment integrity during exposure to programmed elevations in destructive behavior.

CHAPTER 1: METHOD

Participants

We recruited 18 adults to participate in this study. Eight females and 10 males ranging in age between 22 and 55 years old (M = 30) participated in the study. We recruited participants from the community, local colleges, and incoming staff in an autism treatment center. Participants were included based on the following criteria: (1) spoke English as their first language, (2) obtained at least a high school diploma, (3) did not identify as a caregiver of a child with destructive behavior, and (4) self-reported work history not including management of destructive behavior or implementation of chained schedules. Table 1 displays participant demographics.

Participant Number	Group	Gender	Ethnicity	Highest Education
1	High	Female	White	Bachelor's Degree
2	Low	Male	White	Bachelor's Degree
3	High	Male	White	Bachelor's Degree
4	Low	Male	White	Bachelor's Degree
5	High	Male	White	Post-Graduate Degree
6	Low	Male	White	Bachelor's Degree
7	High	Female	White	High School Diploma
8	Low	Female	White	Bachelor's Degree
9	Low	Male	White	Some College, No Degree
10	High	Female	White	Bachelor's Degree
11	Low	Female	White	Bachelor's Degree
12	High	Male	White	Master's Degree
13	High	Male	White	Some College, No Degree
14	Low	Female	White	Bachelor's Degree
15	High	Female	White	Bachelor's Degree
16	Low	Male	White	Bachelor's Degree
17	Low	Female	White	Associate Degree
18	High	Male	White	Bachelor's Degree

Table 1: Participant Demographic Information.

Confederates

A board-certified behavior analyst (BCBA®) or registered behavior technician (RBT®) served as a child confederate during BST, role-play with feedback, and role-play without feedback phases. Across role-play with feedback and role-play without feedback phases, the confederate engaged in scripted responding prompted by a researcher that mimicked a child's severe destructive behavior, FCRs, and compliance.

Setting and Materials

All sessions were conducted in 3 m by 3 m padded therapy rooms equipped with a table and chairs adjacent to a room with a one-way observation mirror and two-way speaker. Materials used in BST, role-play with feedback, and role-play without feedback included a notecard-sized laminated FCR card, discriminative stimuli (i.e., notecard-sized red and green paper), demand materials (i.e., matching folder tasks), timer, paper datasheets, a web camera, and pens. During sessions in both role-play with feedback and role-play without feedback, confederates wore an earpiece connected to a walkie-talkie. A researcher from the booth read a script prompting the confederate to engage in destructive behavior, FCRs, and compliance. Scripts included destructive behavior at either a low (1 response per min) or high (12 responses per min) rate. The authors determined these values by approximating rates of renewal reported by Saini et al. (2018). Specifically, the authors extracted data from the results of the renewal challenge with Zack, who displayed moderate levels of renewal (i.e., six instances of destructive behavior per minute). The authors multiplied this value (i.e., six per minute) by two, resulting in an average high rate value of 12 responses per minute, with a range of 10 to 13 responses per minute. For the low behavior value, the authors chose an average of one response per minute, with a range of zero to three responses per minute. We chose this value as this is the lowest value that would still present opportunities for the

participant to engage in each protocol component, which required participants to ignore destructive behavior.

Response Measurement

Trained data collectors collected data on participant and confederate responding in the session room during BST and behind the one-way observation window during role-play with feedback and role-play without feedback phases using paper and pencil data.

Participant Responses. Our primary dependent variable was the average percentage of correct implementation across components of the chained-schedule procedure. Table 2 provides operational definitions of correct and incorrect responding for all 11 protocol components. The data collector scored correct or incorrect for each implementation of the 11 protocol components given the opportunity for occurrence. Each script programmed equal opportunities for each of the 11 protocol components (e.g., there were five opportunities in each script to implement component 1), with the exception of component 8 (e.g., ignores destructive behavior), as this differed according to the programmed destructive behavior rates. In each session, data collectors calculated the percentage of correct implementation per component by dividing the total number of correct responses for that component by the total number of correct and incorrect responses for the component and multiplying by 100. We then generated the percentage of correct implementation for each session by summing the percentage of correct implementation for each of the 11 protocol components and dividing by 11 (number of components). We used this aggregated measure such that the differential opportunity for implementation of component eight did not inflate scores for the high-behavior group.

#	Correct Implementation	Incorrect Implementation
1	Signals work interval (red) at the appropriate time (i.e., end of SR+ interval)	Includes: a) uses incorrect signal (e.g., does not switch color or state the switch), b) signals work interval (red) at the incorrect time, or c) no signal occurred at the appropriate time
2	Delivers a verbal prompt instruction (e.g., does not phrase as a question) at the appropriate time	Includes: a) delivers verbal prompt instruction incorrectly (e.g., phrases instruction as a question), b) delivers instruction at the incorrect time, or c) no instruction given at the appropriate time
3	Delivers task-specific praise (e.g., “good job matching”) within 10 s of correct confederate compliance with the verbal prompt (i.e., compliance without destructive behavior)	Includes: a) delivers a neutral statement (e.g., “that’s matching”, “good”, “thank you”) instead of task-specific praise for correct confederate compliance with the verbal prompt, b) provides task-specific praise for incorrect confederate compliance (e.g., compliance with destructive behavior), or c) no praise given following correct confederate compliance
4	Delivers correct model prompt (i.e., repeats the instruction while modeling task completion) at the appropriate time	Includes: a) delivers the model prompt incorrectly, b) delivering the model prompt at the incorrect time, or c) no model prompt given at the appropriate time
5	Delivers task-specific praise within 10 s of correct confederate compliance with the model prompt without destructive behavior	Includes: a) delivers a neutral statement (e.g., “that’s matching”, “thank you”) instead of task-specific praise for correct confederate compliance with the model prompt, b) provides task-specific praise for incorrect confederate compliance, or c) no praise given following correct confederate compliance with the model prompt
6	Delivers correct physical guidance prompt (i.e., physically guides confederate through the task while repeating the instruction) at the appropriate time	Includes: a) delivers incorrect physical guidance prompt, b) delivers correct physical guidance prompt at the incorrect time, or c) no correct physical guidance prompt given at the appropriate time
7	Delivers neutral statement (e.g., “that’s matching”, “that’s how you match red”) in a neutral tone following a physically guided response	Includes: a) delivers task-specific praise instead of a neutral statement following a physically guided response, or b) delivers a neutral statement at the incorrect time, or c) no neutral statement given following a physically guided response
8	Ignores destructive behavior (i.e., does not provide any form of attention following confederate destructive behavior)	Includes: a) delivers any form of attention following confederate destructive behavior (e.g., comments on the behaviors, reprimands, laughs audibly, etc.)

#	Correct Implementation	Incorrect Implementation
9	Correctly signals SR+ (green) interval (i.e., switches to green and states the switch) at the appropriate time (i.e., following work interval)	Includes: a) incorrectly signals SR+ (green) interval, b) signals SR+ (green) at the incorrect time, or c) no reinforcement (green) signal occurred at the appropriate time
10	Reinforces correct FCR (i.e., FCR without destructive behavior) with task-specific praise and delivery of the reinforcer (i.e., Legos) at the appropriate time	Includes: a) delivers incorrect reinforcement following a correct FCR (e.g., doesn't deliver praise, doesn't deliver Legos) at the appropriate time, b) delivers correct reinforcement at the incorrect time, orc) no reinforcement occurred at the appropriate time
11	Ignores incorrect FCRs (i.e., does not provide any form of attention or access to Legos following confederate destructive behavior)	Includes: a) delivers reinforcement (i.e., praise and Legos) following an incorrect FCR, or b) delivers the form of attention (e.g., commenting on behavior, delivers reprimand, delivers rule statement) following an incorrect FCR

Table 2: Chained-Schedule Protocol Components. Definitions of correct and incorrect implementation of the 11 chained-schedule protocol components. SR+ = reinforcement; FCR = functional communication response.

Confederate Responses. Data collectors also measured treatment integrity of script-prompted confederate responding across 100% of sessions in BST, role-play with feedback, and role-play without feedback. A correct script-prompted response was defined as the confederate engaging in the correct response within 5 s of the programmed time (i.e., within 5 s of the script being read to the confederate). An incorrect script-prompted response was defined as the confederate engaging in (1) the incorrect number of the programmed response, (2) a different response than was programmed, or (3) no response within 5 s of the programmed response. Script-prompted responses included destructive behavior (i.e., property destruction, to minimize the risk of harm to participants and confederates), FCRs (i.e., card exchange), and compliance as target responses. We defined property destruction as hitting or kicking furniture or surfaces and throwing, swiping, or tearing materials. We programmed both correct and incorrect FCRs to ensure opportunities for participants to demonstrate

the skills detailed in components 10 and 11 (See Table 2 below). A correct FCR was defined as exchanging a card for a programmed reinforcer (i.e., laminated card that says, “Legos, please” with a picture of Legos exchanged for tangible reinforcers) without destructive behavior at the correct time (i.e., following the participant signaling the reinforcement interval). An incorrect FCR was defined as exchanging the FCR card within 5 s of the occurrence of destructive behavior(s) or exchanging the FCR card at the incorrect time (i.e., when on the red side, during the 20 s reinforcement interval with continuous access to Legos). We defined correct compliance as the confederate correctly completing the participant’s instruction within 5 s of the initial verbal or model prompt without engaging in destructive behavior. The mean correct confederate treatment integrity across all participants was 94.2% (range, 88.2% to 100%) across all phases.

Interobserver Agreement

A second independent observer collected data simultaneously with the primary data collector or asynchronously via video recordings across all phases. We obtained interobserver agreement measures across a mean of 37.3% (range, 34.6% to 42.9%), 51.7% (range, 33.3% to 100%), and 34.3% (range, 33.3% to 50.0%) of sessions for BST, role-play with feedback, and role-play without feedback phases, respectively, for correct and incorrect participant implementation across components. Due to technical issues with the video recording device, Participants 6 and 8 did not have video recordings to score interobserver agreement data for BST and role-play with feedback. Table 3 shows IOA values for all participants. We assessed total agreement per protocol component for correct and incorrect participant responses by comparing the recorded totals from the two data collectors. Data collectors calculated interobserver agreement by calculating the number of correct responses per component scored by each observer,

Participant Number	Group	BST		Role-Play with Feedback		Role-Play without Feedback	
		% Trials	IOA Coefficients	% Sessions	IOA Coefficients	% Sessions	IOA Coefficients
		M	*M* (range)	*M*	*M* (range)	*M*	*M* (range)
1	High	36.4%	80.1% (79.5%-81.8%)	80.0%	90.3% (87.1%-93.5%)	33.3%	84.6% (80.5%-88.7%)
2	Low	42.9%	79.5% (77.3%-81.8%)	50.0%	100% (-)	33.3%	87.1% (86.4%-87.7%)
3	High	38.5%	91.8% (79.5%-100%)	40.0%	91.6% (90.8%-92.5%)	33.3%	92.4% (90.8%-94.0%)
4	Low	0%*	-	0%*	-	33.3%	96.0% (94.6%-97.4%)
5	High	38.5%	89.1% (86.4%-100%)	66.7%	95.7% (94.1%-97.2%)	33.3%	83.8 (81.5%-86.1%)
6	Low	0%*	-	50.0%	94.7% (90.0%-99.4%)	33.3%	84.3% (84.1%-84.5%)
7	High	34.6%	91.8% (81.8%-100%)	42.9%	93.8% (87.8%-97.7%)	33.3%	96.3% (93.3%-99.4%)
8	Low	42.9%	97.3% (93.2%-100%)	100%	95.8% (95.0%-96.5%)	33.3%	92.8% (90.2%-95.3%)
9	Low	35.5%	95.7% (86.4%-100%)	50.0%	95.5% (-)	33.3%	93.6% (89.9%-95.5%)
10	High	36.4%	87.3% (81.8%-93.2%)	42.9%	88.0% (83.1%-92.7%)	33.3%	85.8% (85.6%-86.0%)
11	Low	36.4%	85.8% (81.8%-95.5%)	50.0%	87.4% (-)	50.0%	88.1% (80.3%-92.8%)
12	High	37.5%	97.0% (95.5%-100%)	40.0%	91.1% (87.7%-94.5%)	33.3%	95.3% (90.6%-99.9%)
13	High	37.5%	97.7% (93.2%-100%)	50.0%	89.2% (88.6%-89.8%)	33.3%	91.2% (91.0%-91.4%)
14	Low	35.7%	92.8% (84.8%-100%)	50.0%	89.5% (-)	33.3%	91.2% (89.3%-93.1%)
15	High	35.7%	96.4% (86.4%-100%)	33.3%	90.8% (-)	33.3%	90.3% (88.2%-92.3%)
16	Low	36.8%	95.5% (86.4%-100%)	50.0%	94.5% (-)	33.3%	83.6% (80.1%-87.1%)
17	Low	35.0%	88.2% (79.5%-100%)	50.0%	87.6% (87.1%-88.2%)	33.3%	87.4% (86.4%-88.4%)
18	High	36.8%	97.7% (93.2%-100%)	33.3%	92.9% (-)	33.3%	91.2% (90.5%-91.9%)

Table 3: Participant Interobserver-Agreement Values. IOA = interobserver agreement

dividing the smaller number by the larger number, and converted the resulting quotient into a percentage. Data collectors then summed the number of incorrect responses per

component scored by each observer, divided the smaller number by the larger number, and converted the resulting quotient into a percentage. The percentage of agreement for the correct and incorrect responses per component was then averaged across components to calculate the percentage of interobserver agreement across components for each session and phase.

Social Validity

We assessed the social validity of the procedures by having the participants complete a questionnaire in which they rated their perception of how prepared they felt to implement the procedures with an actual child. A copy of the social validity questionnaire is available upon request. Researchers asked participants to provide a rating for three questions using a 1 to 6 Likert scale, 1 representing strongly disagree and 6 representing strongly agree. Questions included: 1) the training I received prepared me to implement this protocol with high integrity, 2) I feel prepared to implement this protocol with a child, and 3) I feel prepared to implement this protocol with a child while they are engaging in problem behavior. Researchers collected social validity ratings at the end of each study phase (i.e., BST, role-play with feedback, and role-play without feedback phases).

Experimental Design and Data Analysis

We employed a between-groups design to evaluate the impact of high versus low rates of destructive behavior during the role-play with feedback phase on subsequent performance in the role-play without feedback phase. An online randomization generator was used to facilitate randomization of participants across the two experimental groups (i.e., low- and high-behavior groups). We calculated the average percentage of correct implementation across components of the protocol per session across BST, role-play with feedback, and role-play without feedback phases. We averaged the last two

sessions in the role-play with feedback phase (referred to as the pretest), to mean scores obtained across the six role-play without feedback phase sessions (referred to as the posttest). We used a two-tailed Mann Whitney U test with an alpha of .05 using XLSTAT (Addinsoft, 2020) software to compare the difference in percent change from mean scores obtained in the pretest and posttest across groups.

Behavioral Skills Training (BST)

We implemented BST with each participant until the participant demonstrated mastery of each condition. The mastery criterion for each condition is defined under each condition section below. Training conditions included: 1) written material, 2) modeled implementation, and 3) role-play.

BST: Written Material. During this condition of BST, the trainer provided a written description of the protocol at an 8th-grade reading level. The trainer reviewed the entire protocol with the participant by reading the protocol aloud and providing opportunities for the participant to ask questions throughout the review of the written materials. After the trainer reviewed the written materials fully, the trainer ensured that the participant did not have any questions or answered all questions until they had no further questions. The trainer asked the participant to complete a multiple-choice quiz with questions related to the treatment components reviewed. A copy of the quiz is available upon request. Once the participant completed the quiz, the trainer checked the quiz answers for accuracy, reviewed any incorrect item(s), and again provided opportunities for the participant to ask questions. All 18 participants completed the quiz with 100% accuracy within one or two attempts following the written materials condition of BST. No participant scored lower than 80% correct responding on any quiz attempt following the written materials component of BST. Data on individual participant quiz

scores are available upon request. The written materials condition was considered complete when the participant answered all questions of the quiz correctly.

BST: Modeled Implementation. During the modeled implementation condition of BST, the trainer modeled one alternation between a work interval and reinforcement interval, which was defined as one trial. During modeled implementation, the trainer modeled 10 trials of the protocol with a confederate using a matching task for instructions. During each trial, the confederate displayed scripted destructive behavior, FCRs, and compliance. Six trials included destructive behavior, compliance, and FCRs. Four trials included compliance and FCRs without destructive behavior, and the distribution of destructive behavior present and absent trials were randomized across the 10 trials. The trainer provided opportunities for participants to ask questions or model the trial again at the end of each trial. After the 10 trials were complete, the trainer ensured that the participant did not have any questions or answered all questions until they did not have any further questions. The participant then completed another post-training quiz identical to the quiz in the written materials condition described above. If the participant answered any of the quiz questions incorrectly, the trainer reviewed the incorrect item(s) and again provided opportunities for the participant to ask questions. All 18 participants completed the quiz with 100% accuracy on the first attempt following the modeled implementation condition of BST. The modeled implementation condition was considered complete when the participant answered all quiz questions correctly.

BST: Role-Play. During the role-play condition of BST, the participant implemented a minimum of 10 trials with a confederate. During each trial, the confederate displayed scripted destructive behavior and FCRs similar to those programmed in the modeled implementation condition. The trainer collected data on the integrity of participant implementation of each of the 11 protocol components during role-

play trials. That is, the trainer scored either a correct or an incorrect response based on participant responding. If a participant engaged in an incorrect response for any protocol component during role-play, the trainer provided corrective feedback at the end of the trial and repeated the trial until the participant completed the trial without errors before conducting additional trials. The role-play condition of BST was considered complete when the participant completed a minimum of 10 trials with the last three trials at 100% accuracy (e.g., for example, trials 8, 9, and 10). If a participant did not meet these criteria, the trainer continued presenting trials until the participant completed three consecutive trials with 100% accuracy.

Role-Play with Feedback Phase

In this phase, participants in each of the two experimental groups were differentially exposed to varying rates of destructive behavior during role-play training scenarios. Specifically, we evaluated the extent to which exposure to varying rates of destructive behavior during role-play training impacted performance during later exposures to treatment challenges (i.e., high rates). Therefore, following BST, the participant implemented the protocol with a confederate in a separate room from the trainer. The confederate engaged in destructive behavior, compliance, and FCRs during the role-play with feedback sessions according to the relevant low- or high-behavior scripts.

Sessions were 5 min in duration and contained the same 11 protocol components that participants had mastered during BST. Researchers asked participants to implement the protocol using compliance with a matching task as the programmed response requirement for the chained-schedule protocol and to continue to switch between the work interval (red side) and the reinforcement interval (green side). Participants repeated the protocol components as described in the chained-schedule

protocol until the session ended. The trainer entered the session room and provided feedback based on participants responding in each of the 11 protocol components immediately following each session. This phase continued until the participant implemented procedures with an 80% or higher average percentage of correct implementation across each of the 11 protocol components for two consecutive sessions. If participants were unable to achieve mastery within 10 sessions of the role-play with feedback phase, we would have conducted a booster session and then conducted up to 10 additional sessions. All 18 participants completed the role-play with feedback phase within 10 sessions, and therefore, no participants required this additional training.

Role-Play without Feedback Phase

Participants came in for the role-play without feedback phase within one week following the end of the role-play with feedback phase. Before initiating sessions in this phase, the researcher allowed the participant to review the written protocol document, ask any questions regarding the protocol and components within the protocol, and practice any protocol components for up to 5 minutes before initiating sessions. Researchers informed participants that sessions would “look similar” to sessions conducted on the previous research day, except that the trainer would provide no feedback on performance in between sessions. During this phase, sessions were identical to sessions conducted with the high-behavior group in the pretest (i.e., the role-play with feedback phase), with the exception that the researchers did not provide feedback on participant performance following each session. Regardless of which group and corresponding rate of destructive behavior (i.e., low- or high-behavior) the participant experienced during the role-play with feedback phase, the confederate engaged in high rates of destructive behavior (i.e., an average of 12 responses per

minute). The researchers exposed the participant to six consecutive sessions and terminated the phase following the completion of the sixth session, regardless of participant performance.

CHAPTER 2: RESULTS

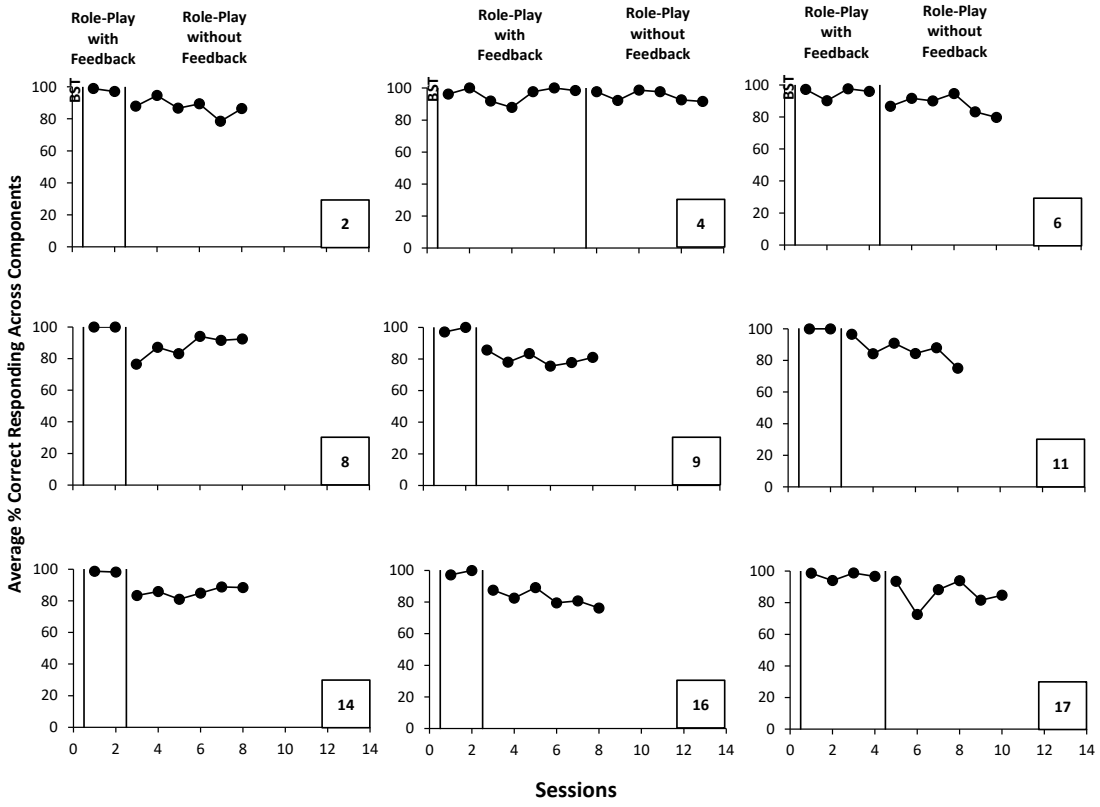

Figure 1: Individual Participant Results for Low-Behavior Group. Results from the role-play with feedback (i.e., 1 destructive behavior per minute programmed) and role-play without feedback (i.e., 12 destructive behaviors per minute programmed) phases for all participants in the low-behavior group. Participant numbers are labeled in the bottom right corner of each panel. The closed circle data path depicts average percentage correct responding across components per 5 min session.

All participants achieved mastery during BST in 40 trials or less for both the low-

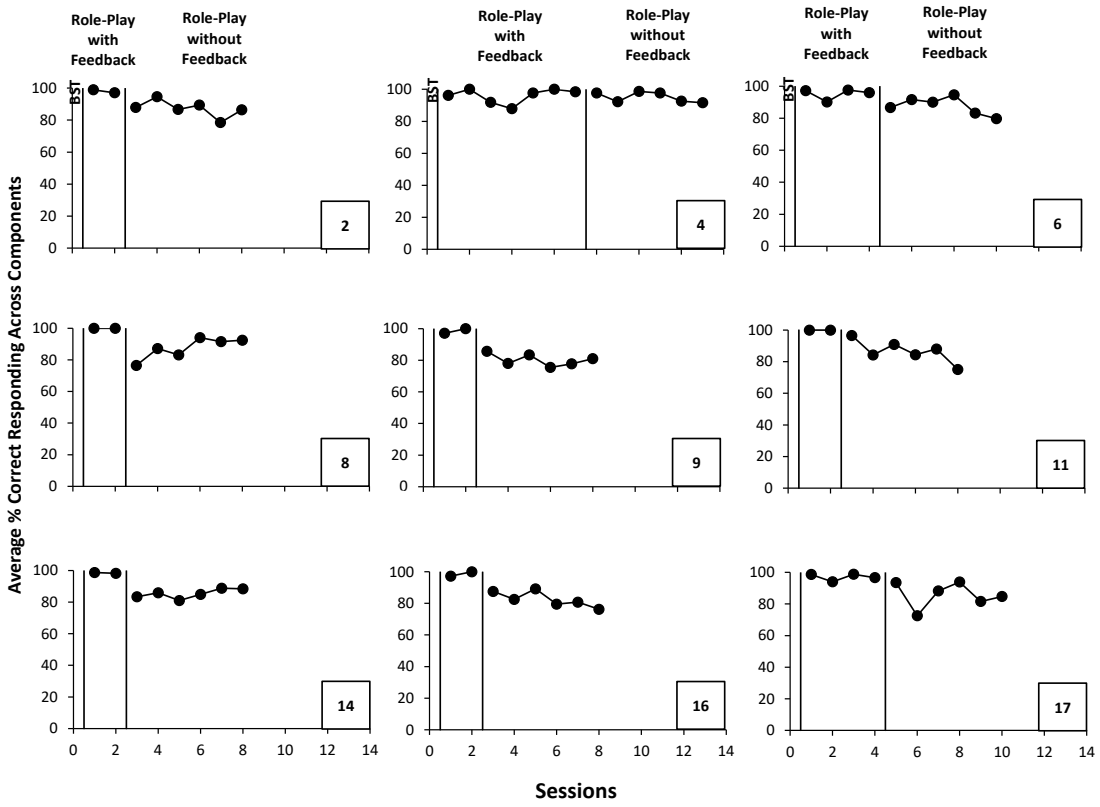

Figure 2: Individual Participant Results for High-Behavior Group. Results from the role-play with feedback (i.e., 12 destructive behaviors per minute programmed) and role-play without feedback (i.e., 12 destructive behaviors per minute programmed) phases for all participants in the high-behavior group. Participant numbers are labeled in the bottom right corner of each panel. The closed circle data path depicts average percentage correct responding across components per 5 min session.

behavior (*M* = 21.8; range, 14 to 35) and high-behavior groups (*M* = 17.4; range, 11 to 23). These data are available upon request from the corresponding author. Participants required an average of 3.0 sessions (range, 2 to 7) to reach mastery in the role-play with feedback phase in the low-behavior group, and an average of 4.7 sessions (range, 3 to 7) for participants in the high-behavior group. The average number of days between completion of the role-play with feedback phase, and initiation of the role-play without feedback phase across all 18 participants was 2.7 days (range, 1 to 7 days). The average number of days between completion of the role-play with feedback phase, and initiation of the role-play without feedback phase for participants in the low-behavior

group was 2.8 days (range, 1 to 7 days), and was 2.4 days (range, 1 to 7 days) for participants in the high-behavior group.

Figures 1 and 2 display individual participant data across role-play with feedback and role-play without feedback phases for all participants in the low- and high-behavior groups. Each panel in each figure represents a different individual participant's data. Participants across both the low- and high-behavior groups required seven sessions or less to reach mastery in the role-play with feedback phase averaging 3.8 sessions (range, 2 to 7). Participants in the low-behavior group required an average of 3.0 sessions (range, 2 to 7) to reach mastery in the role-play with feedback phase. Participants in the high-behavior group required an average of 4.7 (range, 3 to 7) sessions to reach mastery.

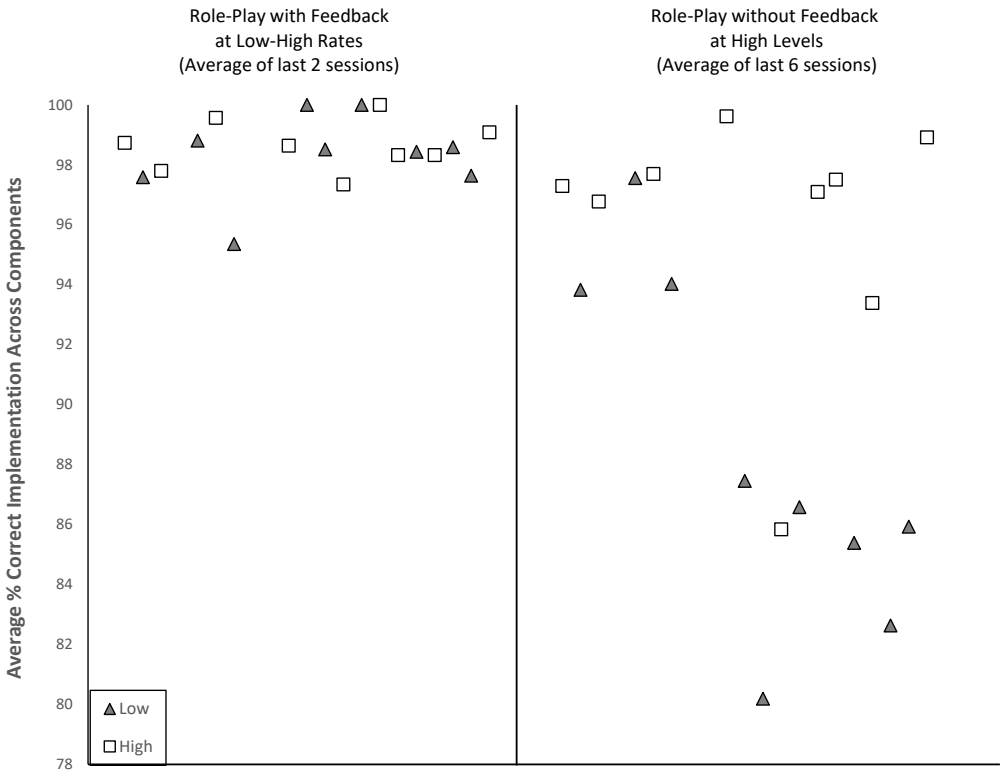

Figure 3: Average Individual Correct Responding Across Groups at Pre- and Post-Test. Average percentage correct implementation across components for pre- (i.e., an average of last 2 sessions in role-play with feedback phase) and posttest measures (i.e., an average of all 6 sessions in role-play without feedback phase) for all participants across the low- and high-behavior groups. DBx = programmed rate of confederate destructive behavior.

To compare participant responding across groups in the pretest and posttest, we averaged the percentage of correct implementation across components in the last two sessions of the pretest (i.e., the sessions that met mastery) and all six sessions of the posttest for each individual. Figure 3 displays these data. In the role-play with feedback phase, participants showed similar levels of responding in the high- and low-behavior groups, which was expected given we aggregated data that conformed to mastery criteria for this phase (i.e., 80% or higher correct responding for two consecutive sessions). In the role-play without feedback phase, participants in the high-behavior group generally displayed higher percentages of correct implementation across components than the low-behavior group. However, all participants in both the low- and high-behavior groups showed relatively high percentages of correct implementation. When looking at posttest scores, participants in the low-behavior group averaged 88.2% (range 80.2% to 97.6%) correct implementation across components, and participants in the high-behavior group averaged 96.0% (range 85.8% to 99.6%) correct implementation across components. We observed a small degree of overlap (i.e., 4 out of 18 data points) across the two groups. Specifically, three participants (i.e., Participants 2, 4, and 6) in the low-behavior group displayed average percentages of correct implementation across components at or near levels of average percentages of correct implementation displayed by participants in the high-behavior group. Additionally, one participant (i.e., Participant 10) in the high-behavior group displayed an average percentage of correct implementation across components at or near the percentages displayed by participants in the low-behavior group during the posttest.

Table 4 shows the aggregated pretest and posttest means and the difference in the aggregated pretest and posttest means across the low- and high-behavior groups. To compare the difference in the aggregated pretest and posttest means across groups,

Group	*n*	Pretest Mean	Posttest Mean	Difference in Pre- and Post-Test Means
		Averages		
Low	9	98.33	88.18	10.15
High	9	98.65	96.02	2.63

Table 4: Aggregated Pre- and Post-Test Outcomes Across Groups. Average pretest (i.e., last two sessions in role-play with feedback phase), posttest (i.e., six sessions of role-play without feedback phase) means, and difference in pre- and posttest means across the low- and high-behavior groups.

we subtracted the average of the six sessions of the posttest from the average of the last two sessions of the pretest for each participant. The difference between aggregated pretest and posttest means for the low-behavior group averaged 10.2 percentage points (range, 1.25 to 18.33), where the difference between the aggregated pretest and posttest means for the high-behavior group averaged 2.6 percentage points (range, 2.6 to -1). In the low-behavior group, six of the nine participants (i.e., 66.7%) displayed a 10 point or more decrease in the average percentage of correct implementation across components from the pretest to the posttest. Only one of nine participants (i.e., 11.1%) displayed a 10 point or more decrease in the high-behavior group. We used a two-tailed Mann-Whitney U test with an alpha of .05 to compare the difference across pretest and posttest for the low- and high-behavior groups. The critical *U* value for our sample size (n = 18) at .05 is 17. The obtained *U* value is 12 (p = .01), suggesting the difference between pretest and posttest scores for the low- and high-behavior groups is statistically significant.

Table 5 shows average social validity ratings for the low- and high-behavior groups across questions 1, 2, and 3 after BST, role-play with feedback, and role-play without feedback phases. When assessing the differences between social validity ratings across phases, 44.4% of (n = 4) participants in the low-behavior group displayed lower ratings for the question related to preparedness to implement in the presence of

destructive behavior (i.e., question 3) following the role-play without feedback phase. In contrast, 11.1% (*n* = 1) of participants in the high-behavior group displayed lower ratings on question 3 following the role-play without feedback phase.

Group	Phase	Question 1: The training I received prepared me to implement this protocol with high integrity	Question 2: I feel prepared to implement this protocol with a child	Question 3: I feel prepared to implement this protocol with a child while they are engaging in problem behavior
Low				
	BST	5.8 (range, 5 to 6)	4.7 (range, 2 to 6)	4.6 (range, 1 to 6)
	Role-Play with Feedback	5.9 (range, 5 to 6)	4.9 (range, 1 to 6)	4.9 (range, 3 to 6)
	Role-Play without Feedback	5.8 (range, 5 to 6)	5.0 (range 4 to 6)	4.6 (range, 2 to 6)
High				
	BST	5.8 (range, 5 to 6)	5.2 (range, 4 to 6)	4.9 (range, 3 to 6)
	Role-Play with Feedback	5.9 (range, 5 to 6)	5.4 (range, 4 to 6)	5.2 (range, 3 to 6)
	Role-Play without Feedback	5.9 (range, 5 to 6)	5.6 (range, 4 to 6)	5.3 (range, 4 to 6)

Table 5: Group Social Validity Ratings. Average social validity ratings for the low- and high-behavior groups across BST, role-play with feedback, and role-play without feedback phases for questions 1, 2, and 3 of the social validity questionnaires.

Discussion

Child participation in clinical services often results in caregiver training on the key components of the treatment package. The intention is to train caregivers to implement the treatment package at high levels of treatment integrity to facilitate continued implementation at high integrity following discharge from services and return to the natural context. This is critical because maintaining high levels of treatment integrity is important to the ongoing durability of treatment effects and reductions in destructive behavior (e.g., Fryling & Wallace, 2012). In this translational evaluation, we trained 18 adult participants to clinically acceptable levels of treatment integrity in BST (e.g., written instructions, modeling, role-play). We then successfully trained caregivers to implement

the same protocol in a role-play with feedback scenario while experiencing varying rates of destructive behavior. This outcome is promising as it echoes the previous literature on the efficacy of BST procedures for teaching caregivers a variety of skills to high levels of integrity. Of note, training under high versus low rates of destructive behavior produced statistically significant differences in levels of treatment integrity when both groups were exposed to high rates of destructive behavior post-training. Specifically, participants in the high-behavior group displayed higher average percentages of correct implementation across components in the role-play without feedback phase (i.e., posttest) relative to the low-behavior group. In addition, participants in the high-behavior group reported higher perceived preparedness to implement the protocol with a child engaging in destructive behavior on social validity measures relative to participants in the low-behavior group.

The current evaluation is the first study of its kind to specifically evaluate the impact of varying rates of destructive behavior on treatment integrity during caregiver training. We think research questions of this kind are important and encourage continued research in this area to understand the broader impact of the training context on continued treatment integrity in the natural environment. Destructive behavior is likely to reemerge at some point during the generalization of treatment to caregivers or new settings (e.g., Bouton et al., 2011; Falligant et al., 2020; Ibañez et al., 2019; Kelley et al., 2015; Kelley et al., 2018; Muething et al., 2020; Saini et al., 2018). When destructive behavior returns, levels of caregiver treatment integrity are likely to decrease (e.g., Mitteer et al., 2018; Mitteer et al., 2021). Maintenance of high levels of treatment integrity is essential to the durability of treatment effects (e.g., Fryling & Wallace, 2012). Therefore, clinicians need to understand how to train caregivers to high levels of treatment integrity that persist even when faced renewal-like conditions, which is

reported to occur in nearly half of context changes experienced (Muething et al., 2020; Saini & Mitteer, 2020). The results of the current study provide evidence for prioritizing programmed rates of destructive behavior during caregiver training. In addition, clinicians may further inform how they program rates of destructive behavior during caregiver training by considering rates that may be experienced if renewal were to occur (e.g., during reemergence of destructive behavior).

Although our results are promising, they are preliminary, and therefore, we must caution readers given the limitations of the translational preparation used. Our participant population included adults not identifying as caregivers of children with destructive behavior. Therefore, it is unclear how these results would generalize to more applied populations, such as actual caregivers of children who engage in severe destructive behavior.

Second, we purposely constructed this evaluation to maximize safety to the participants (e.g., only disruptive behavior targeted) and control for the impact of extraneous variables (e.g., unpredictable child responding, distractors in the home). The current study's analog setting does not represent how implementation would occur in the natural environment (e.g., home or community setting with multiple distractions). For example, typically, caregivers would be managing protocol implementation with the target child (e.g., giving instructions to complete the morning routine), attending to multiple disruptive/destructive responses at one time, and potentially monitoring other children in the home. We encourage that future research considers systematic evaluations of other variables approximating the natural environment more closely, thus enhancing ecological validity. Specifically, researchers may consider programming additional responses into the training context (our evaluation only targeted disruptive

behavior in isolation) and more naturalistic conditions and settings (e.g., home context, competing activities, distractions).

Additionally, confederates were used rather than actual children for all phases of the study. The use of confederates may have influenced participant responding. Caregivers of children with destructive behavior often have a long history of reinforcing destructive behavior for many reasons (e.g., safety, lack of knowledge or training in behavioral procedures, competing responsibilities). Future research should replicate these procedures primarily with caregivers of children who engage in severe destructive behavior to examine more specific characteristics of this population and their patterns of responding. However, the use of confederates was a critical component for ensuring that participants experienced the programmed rates of severe destructive behavior corresponding to each experimental group.

Additionally, all participants across both the low- and high-behavior groups in this study maintained relatively high levels of integrity during the posttest challenge. When looking at posttest scores, participants in both the low- and high-behavior groups showed clinically acceptable percentages of correct implementation across components (i.e., all participants averaged above 80% correct implementation). Therefore, all participant posttest levels of correct implementation were above the percentage required in the mastery criteria defined in this study (i.e., 80% or higher), regardless of the rate experienced during training.

We did not control for the number of trials experienced in BST, number of sessions experienced in the role-play with feedback phase, or the time between the role-play with feedback and role-play without feedback phases. Therefore, there may have been differences in the amount of exposure to training across groups. To evaluate this, we completed a post-hoc two-tailed Mann-Whitney U test with an alpha of .05 to

compare the number of trials to mastery in BST, number of sessions to mastery in the role-play with feedback phase, and the number of days between the role-play with feedback and role-play without feedback phases across participants in the low- and high-behavior groups. For the number of trials to mastery in BST, the critical *U* value for our sample size (*n* = 18) at .05 is 17. The obtained *U* value is 26 (*p* = .21), suggesting the difference between trials to mastery in BST for the low- and high-behavior groups is not statistically significant. For the number of days between the role-play with feedback and role-play without feedback phase, the critical *U* value for our sample size (*n* = 18) at .05 is 17. The obtained *U* value is 31 (*p* = .43), suggesting the difference between time between phases for the low- and high-behavior groups is not statistically significant. For the number of sessions to mastery in the role-play with feedback phase, the critical *U* value for our sample size (*n* = 18) at .05 is 17. The obtained *U* value is 15 (*p* = .03), suggesting the difference between sessions to mastery in the role-play with feedback phase for the low- and high-behavior groups are statistically significant. Therefore, some participants in the high-behavior group experienced more sessions in the role-play with feedback phase than the low-behavior group on average, which may have influenced responding in the last phase. However, participants required an average of 3.0 sessions (range, 2 to 7) to reach mastery in the role-play with feedback phase in the low-behavior group, and an average of 4.7 sessions (range, 3 to 7) for participants in the high-behavior group. Therefore, the difference between the average number of sessions to mastery between groups was only 1.7 sessions that were 5 min in length. Nonetheless, future research should consider matching the number of training sessions experienced to control for this potential difference.

We find these preliminary results very promising because it shifts the focus to important aspects of the training environment when training caregivers for acquisition

and success in the natural environment. Given the novelty of this research question, there are many areas for future research, as discussed above. Therefore, we hope that this study inspires extensions, such as the inclusion of more applied populations, other topographies of destructive behavior (in isolation or combination), and inclusion of children. These extensions may have implications for clinicians when determining when and how to introduce caregivers into the treatment of their child's destructive behavior. Additionally, research in this area may help clinicians determine how to train caregivers to high levels of treatment integrity that maintain even when faced with treatment challenges (e.g., renewal).

BIBLIOGRAPHY

Addinsoft (2020). XLSTAT Mann-Whitney Test in Excel Tutorial (Version 2021.3.1) [Computer software]. https://help.xlstat.com/s/article/mann-whitney-test-in-excel-tutorial?language=en_US

Bouton, M. E., Todd, T. P., Vurbic, D., & Winterbauer, N. E. (2011). Renewal after the extinction of free operant behavior. *Learning & Behavior*, *39*(1), 57–67. https://doi.org/10.3758/s13420-011-0018-6

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis*, *18*(2), 111–126. https://doi.org/10.1901/jaba.1985.18-111

Conklin, S. M., & Wallace, M. D. (2019). Pyramidal parent training using behavioral skills training: Training caregivers in the use of a differential reinforcement procedure. *Behavioral Interventions, 34(3), 377-387*. https://doi.org/10.1002/bin.1668

Drifke, M. A., Tiger, J. H., & Wierzba, B. C. (2017). Using behavioral skills training to teach parents to implement three-step prompting: A component analysis and generalization assessment. *Learning and Motivation*, *57*, 1–14. https://doi.org/10.1016/j.lmot.2016.12.001

Falligant, J. M., & Hagopian, L. P. (2020). Further extensions of precision medicine to behavior analysis: A demonstration using functional communication training. *Journal of Applied Behavior Analysis, 53(4), 1961-1981*. https://doi.org/10.1002/jaba.739

Falligant, J. M., Kranak, M. P., McNulty, M. K., Schmidt, J. D., Hausman, N. L., & Rooker, G. W. (2020). Prevalence of renewal of problem behavior: Replication

and extension to an inpatient setting. *Journal of Applied Behavior Analysis, 54(1), 367-373*. https://doi.org/10.1002/jaba.740

Fisher, W. W., & Bouxsein, K. (2011). Developing function-based reinforcement procedures for problem behavior. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis.* (pp. 335–347). The Guilford Press.

Fisher, W. W., Greer, B. D., Fuhrman, A. M., & Querim, A. C. (2015). Using multiple schedules during functional communication training to promote rapid transfer of treatment effects. *Journal of Applied Behavior Analysis*, *48*(4), 713–733. https://doi.org/10.1002/jaba.254

Fisher, W. W., Piazza, C. C., Cataldo, M. F., Harrell, R., Jefferson, G., & Conner, R. (1993). Functional communication training with and without extinction and punishment. *Journal of Applied Behavior Analysis*, *26*(1), 23–36. https://doi.org/10.1901/jaba.1993.26-23

Fisher, W. W., Saini, V., Greer, B. D., Sullivan, W. E., Roane, H. S., Fuhrman, A. M., Craig, A. R., & Kimball, R. T. (2018). Baseline reinforcement rate and resurgence of destructive behavior. *Journal of the Experimental Analysis of Behavior*, *111*(1), 75–93. https://doi.org/10.1002/jeab.488

Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, *45*(2), 449–453. https://doi.org/10.1901/jaba.2012.45-449

Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2020). Functional communication training: From efficacy to effectiveness. *Journal of Applied Behavior Analysis*, *54*(1), 122-143. https://doi.org/10.1002/jaba.762

Gerow, S., Davis, T., Radhakrishnan, S., Gregori, E., & Rivera, G. (2018). Functional communication training: The strength of evidence across disabilities. *Exceptional Children*, *85*(1), 86–103. https://doi.org/10.1177/0014402918793399

Greer, B. D., Fisher, W. W., Saini, V., Owen, T. M., & Jones, J. K. (2016). Functional communication training during reinforcement schedule thinning: An analysis of 25 applications. *Journal of Applied Behavior Analysis*, *49*(1), 105–121. https://doi.org/10.1002/jaba.265

Greer, B. D., Fisher, W. W., Briggs, A. M., Lichtblau, K. R., Phillips, L. A., & Mitteer, D. R. (2019). Using schedule-correlated stimuli during functional communication training to promote the rapid transfer of treatment effects. *Behavioral Development*, *24*(2), 100–119. https://doi.org/10.1037/bdb0000085

Hagopian, L. P., Boelter, E. W., & Jarmolowicz, D. P. (2011). Reinforcement schedule thinning following functional communication training: Review and recommendations. *Behavior Analysis in Practice*, *4*(1), 4–16. https://doi.org/10.1007/bf03391770

Ibañez, V. F., Piazza, C. C., & Peterson, K. M. (2019). A translational evaluation of renewal of inappropriate mealtime behavior. *Journal of Applied Behavior Analysis*, *52*(4), 1005–1020. https://doi.org/10.1002/jaba.647

Kelley, M. E., Jimenez-Gomez, C., Podlesnik, C. A., & Morgan, A. (2018). Evaluation of renewal mitigation of negatively reinforced socially significant operant behavior. *Learning and Motivation*, *63*, 133–141. https://doi.org/10.1016/j.lmot.2018.05.003

Kelley, M. E., Liddon, C. J., Ribeiro, A., Greif, A. E., & Podlesnik, C. A. (2015). Basic and translational evaluation of renewal of operant responding. *Journal of Applied Behavior Analysis*, *48*(2), 390–401. https://doi.org/10.1002/jaba.209

Kurtz, P. F., Boelter, E. W., Jarmolowicz, D. P., Chin, M. D., & Hagopian, L. P. (2011). An analysis of functional communication training as an empirically supported treatment for problem behavior displayed by individuals with intellectual disabilities. *Research in Developmental Disabilities*, *32*(6), 2935–2942. https://doi.org/10.1016/j.ridd.2011.05.009

Miles, N. I., & Wilder, D. A. (2009). The effects of behavioral skills training on caregiver implementation of guided compliance. *Journal of Applied Behavior Analysis*, *42*(2), 405–410. https://doi.org/10.1901/jaba.2009.42-405

Mitteer, D. R., Greer, B. D., Fisher, W. W., Briggs, A. M., & Wacker, D. P. (2018). A laboratory model for evaluating relapse of undesirable caregiver behavior. *Journal of the Experimental Analysis of Behavior*, *110*(2), 252–266. https://doi.org/10.1002/jeab.462

Mitteer, D. R., Greer, B. D., Randall, K. R., Kimball, R. T., & Smith, S. W. (2021). Empirically deriving omission and commission errors for relapse tests: A demonstration of reverse translation. *Behavior Analysis: Research and Practice*. https://doi.org/10.1037/bar0000218

Muething, C., Call, N., Pavlov, A., Ringdahl, J., Gillespie, S., Clark, S., & Mevers, J. L. (2020). Prevalence of renewal of problem behavior during context changes. *Journal of Applied Behavior Analysis*, *53*(3), 1485–1493. https://doi.org/10.1002/jaba.672

Saini, V., & Mitteer, D. R. (2020). A review of investigations of operant renewal with human participants: Implications for theory and practice. *Journal of the Experimental Analysis of Behavior*, *113*(1), 105–123. https://doi.org/10.1002/jeab.562

Saini, V., Sullivan, W. E., Baxter, E. L., DeRosa, N. M., & Roane, H. S. (2018). Renewal during functional communication training. *Journal of Applied Behavior Analysis*, *51*(3), 603–619. https://doi.org/10.1002/jaba.471

Spiegel, H. J., Kisamore, A. N., Vladescu, J. C., & Karsten, A. M. (2016). The effects of video modeling with voiceover instruction and on-screen text on parent implementation of guided compliance. *Child & Family Behavior Therapy*, *38*(4), 299–317. https://doi.org/10.1080/07317107.2016.1238690

Tarbox, R. S. F., Wallace, M. D., Penrod, B., & Tarbox, J. (2007). Effects of three-step prompting on compliance with caregiver requests. *Journal of Applied Behavior Analysis*, *40*(4), 703–706. https://doi.org/10.1901/jaba.2007.703-706

Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice*, *1*(1), 16–23. https://doi.org/10.1007/bf03391716

Unholz-Bowden, E., McComas, J. J., McMaster, K. L., Girtler, S. N., Kolb, R. L., & Shipchandler, A. (2020). Caregiver Training Via Telehealth on Behavioral Procedures: A Systematic Review. *Journal of Behavioral Education*, *29*(2), 246–281. https://doi.org/10.1007/s10864-020-09381-7