2016

An analysis of treatment integrity of response interruption and redirection

Candice L. Colon-Kwedor
Western New England University

An Analysis of Treatment Integrity of Response Interruption and Redirection

Candice L. Colón-Kwedor

Western New England University

The May Institute, May Center for Autism and Developmental Disabilities

A dissertation submitted to the Department of Psychology in the School of Arts and Sciences at

Western New England University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Behavior Analysis

Supervised by:

William H. Ahearn, Ph.D., BCBA-D

Abstract

Response Interruption and Redirection (RIRD) has been shown to effectively decrease stereotypy but its application outside an experimental setting has not been well studied. In Experiment 1, decreases in automatically-maintained vocal stereotypy were obtained following RIRD treatment in a controlled setting for three participants diagnosed with an autism spectrum disorder (ASD). Descriptive data on the consistency and accuracy of treatment implementation were then collected in the classroom setting. Results showed that the consistency of treatment implementation varied across participants (i.e., Participant 1, M=60.0%; Participant 2, M=89.7%; Participant 3, M= 41.1%) and across staff members (range, 0-100%). Failure to implement the treatment was the most common error. However, when RIRD was implemented, the components were generally carried out as prescribed. In Experiment 2, three participants were exposed to a parametric analysis in a controlled setting to test the impact of consistency errors on RIRD. The results indicated that RIRD was generally effective at 50% treatment implementation or higher. Furthermore, when 25% implementation was interspersed with booster sessions at 100% treatment implementation effects were also maintained. Application of the RIRD procedure in the clinical settings is discussed.

Keywords: automatic reinforcement, parametric analysis, response interruption and redirection, treatment integrity

An Analysis of Treatment Integrity of Response Interruption and Redirection

Automatic reinforcement contingencies present a challenge for treatment because the reinforcing consequences are not accessible to caregivers (Rapp & Vollmer, 2005; Vollmer, 1994). Response blocking is one treatment used to decrease automatically-maintained behavior by preventing the behavior from occurring (e.g., Hagopian & Adelinis, 2001; Lerman & Iwata, 1996; Smith, Russo, & Le, 1999). However, Hagopian and Adelinis (2001) found that blocking alone can produce side effects such as aggression. They also reported that continuous non-contingent access to an alternative activity (identified via a preference assessment) did not suppress automatically-maintained pica. Conversely, response blocking paired with redirection decreased pica, reduced the probability of aggression, and provided exposure to an opportunity for alternative behavior.

Response interruption and redirection (RIRD) is a variant of response blocking that has been used to treat automatically-maintained stereotypic behavior. In general, the RIRD procedure entails interrupting each instance of the target behavior and redirecting to an appropriate response (Ahearn, Clark, MacDonald, & Chung, 2007). Martinez and Betz (2013) reviewed eight RIRD studies published in the *Journal of Applied Behavior Analysis* from 2007-2012. Despite procedural variations, the RIRD procedure was consistently found to produce a decrease in automatically-maintained stereotypic behavior (Ahearn et al.; Ahrens, Lerman, Kodak, Worsdell, & Keegan, 2011; Athens, Vollmer, Sloman, St. Peter Pipkin, 2008; Colón, Ahearn, Clark & Masalsky, 2012; Duffy-Cassella, Sidener, Sidener, & Progar, 2013; Love, Miguel, Fernand & LaBrie, 2012; Miguel, Clark, Tereshko, & Ahearn, 2009; Schumacher & Rapp, 2011).

Likewise, a meta-analysis conducted by Vanderkerken, Heyvaert, Maes, and Onghena (2013) indicated that single-case studies for treating vocal challenging behavior (including several RIRD studies published between 2007 and 2011) produced clinically and statistically significant effects when the interventions included both skill training (e.g., teaching appropriate replacement behavior) and decelerative procedures. Two studies not included in this meta-analysis (due to their publication date) have investigated the necessity of RIRD in relation to procedures based solely on skill training. Colón et al. (2012) directly introduced verbal operant training (VOT) in an effort to increase appropriate vocalizations and decrease vocal stereotypy. However, for 2 of the 3 participants reductions in vocal stereotypy following VOT were not clinically significant. Therefore RIRD was introduced and subsequently decreased vocal stereotypy further while appropriate vocalizations persisted. Similarly, Love et al. (2013) offered noncontingent access to matched stimulation (toys that produced noise) as a means of providing an alternative behavior and decreasing the motivation for engagement in vocal stereotypy. Matched stimulation and RIRD were then compared as a package and alone. Suppression of vocal stereotypy was obtained across all treatment conditions; however, appropriate vocalizations were highest when RIRD was a component of the treatment. Therefore both studies demonstrated that RIRD is often necessary to obtain optimal results.

Although RIRD has been shown to be an effective procedure, Duffy-Cassella et al. (2013), Kliebert, Tiger, and Toussaint (2011) and Miguel et al. (2009) suggested that RIRD may also require effortful treatment application if an individual initially engages in high rates of stereotypy. For example, if a student engages in 30 instances of stereotypy per hour, the teacher must be immediately available to implement RIRD 30 times per hour. Such high rates of treatment implementation may put the procedure at risk for treatment integrity failures. In

addition, past RIRD studies have been implemented in a controlled setting; therefore the extent to which the procedure is effective and practical in non-experimental settings has not been thoroughly examined (Martinez & Betz, 2013). Treatment Integrity (TI) has been defined as the extent to which an independent variable (IV) is implemented as intended (Gresham, 1989; Gresham, Gansle, & Noell, 1993). Clinicians often choose evidence-based procedures that are conducted in controlled settings with near perfect integrity; however, the generality of such treatment effects in the natural setting under conditions of less than optimal TI are often unstudied. Analyzing TI requires consideration of the accuracy and/or consistency with which a treatment is implemented (McIntyre, Gresham, DiGennaro & Reed, 2007; Peterson, Homer, & Wonderlich, 1982). The accuracy of treatment implementation refers to the extent to which the IV definition matches the implementation of the treatment in practice (Peterson et al.). Consistency refers to the appropriate application of the IV in time or the extent to which the IV is implemented according to a set schedule (Homer & Peterson, 1980; Peterson et al.).

A brief review of TI and intervention effectiveness completed by Fryling, Wallace and Yassine (2012) emphasized the need for descriptive research depicting treatment integrity in the natural setting in order to determine naturally occurring integrity levels in need of further empirical evaluation. Empirical questions regarding treatment outcome under degraded levels of integrity can be answered by manipulating the consistency of treatment implementation (McIntyre et al., 2007; Peterson et al., 1982). For instance, intermittent schedules of treatment delivery reflect a pattern that can be representative of naturally-occurring lapses in TI. Parametric analyses have been used to investigate such phenomenon. Parametric analyses manipulate a specific parameter of an IV to determine the extent to which it influences responding (Ahrens et al., 2011; Kliebert, Tiger, & Toussaint, 2011; Lerman & Iwata, 1996;

Smith et al., 1999; St. Peter Pipkin, Vollmer & Sloman, 2010; Vollmer, Roane, Ringdahl, & Marcus, 1999). Therefore, parametric analyses can be used to determine the effects produced by a range of TI values on treatment outcome.

Lerman and Iwata (1996) and Smith et al. (1999) used parametric analyses to investigate the behavioral mechanism responsible for the effects of response blocking which, as mentioned above, is a procedure similar to RIRD. Although it was not their purpose, these studies illustrate the possible effects of degraded TI on treatment outcomes. In each of these studies, a single participant was exposed to a parametric analysis of various treatment implementation values. Lerman and Iwata (1996) manipulated the percentage of automatically-maintained hand mouthing responses that were followed by response blocking. Following baseline (BL) measures and exposure to 100% response blocking, the proportion of responses blocked was manipulated using the values 75%, 67%, 50%, and 25%. The results indicated that response blocking decreased hand mouthing across all treatment implementation values. Smith et al. (1999) conducted a systematic replication of the procedures administered by Lerman and Iwata. In this study, continuous response blocking was implemented following BL. However, 100% of responses could not be blocked due to the rapid rate of eye poking emitted by the participant. Nonetheless, eye poking decreased under these conditions. Next, response blocking was implemented during 50%, 67%, and 80% of opportunities. The results indicated that eye poking decreased during implementation at 80% but remained near BL levels during 50% and 67% integrity. These results were similar to those found by Lerman and Iwata as both studies indicated that treatment effects could be obtained at degraded levels of TI but differed in the values with which treatment effects were maintained. In a manner similar to the aforementioned studies, Ahrens et al. (2011) implemented a parametric analysis of the RIRD procedure using the

values 50%, 25%, and 10% with one participant following an RIRD treatment analysis. The results indicated that RIRD was effective during the 50% integrity condition but treatment effects were not maintained at lower integrity values.

The purpose of the current study was to evaluate the extent that TI impacts RIRD treatment outcome.

Experiment 1

Method

Participants. All participants attended a private school program consisting of approximately 160 day and residential students. The program specialized in the education and treatment of children diagnosed with an ASD. A request for participants was sent to the programs clinical and educational team leaders. Participants were nominated by their clinical and educational team due to engagement in vocal stereotypy that interfered with daily activities (e.g., interference with attention to task, social interactions and vocational jobs in the community).

The participants selected were 1 female and 2 males, ages 15-21, all diagnosed with an autism spectrum disorder (ASD) by a clinician unassociated with the clinical or research teams. For all participants, IEP goal summaries and teacher report were used to assess their verbal repertoire. Kent was a 21-year-old male, whose communication was primarily nonvocal, and supported with a voice output device. He did not vocally participate in conversational volleys or emit spontaneous tacts or mands. He did, however, engage in echoics when prompted. Cora was a 16-year-old female who emitted spontaneous mands in the form of one-word utterances such as “computer” and “book.” When given a prompt Cora tacted items and would sometimes spontaneously tact the actions of others (e.g., “Girl is crying”). Noah was a 15-year-old male

who emitted spontaneous mands throughout the day in the form of full sentences, participated in 2-3 conversational volleys, and labeled items or attributes when prompted (e.g., "what is it?"). Each participant was assigned to a different classroom within the private school setting and taught in a 1:2 staff to student ratio.

The teachers who were observed for TI purposes were bachelor and master-level teaching professionals. They were all accustomed to frequent TI observations due to routine workplace practices and were all trained to implement the RIRD procedure prior to observation sessions. A detailed description of these training procedures is outlined in phase 2 of the methods section.

Setting. For Participant 1 experimental sessions were conducted in an experimental room (1.5 m by 3 m) equipped with a video camera, a table and chairs. For Participants 2 and 3 experimental sessions were conducted behind partitions in the classroom equipped with a video camera, a table and chairs. During Phase 2, observations were conducted in each participant's classroom as they participated in regularly scheduled activities.

Response Measurement. The definition of all target behavior remained the same during analyses and observations. *Vocal stereotypy* was defined as any instance of noncontextual or nonfunctional vocalizations, including repetitive babbling, grunts, squeals, and phrases unrelated to the present situation emitted at conversation level or above. Examples included vocalizations such as "bebebebe" and "ahhhhh" outside the context of prompted echoics and phrases such as "Flip the pancakes" outside of the appropriate context of talking about pancakes. Non-examples included functional speech to request an item or activity (e.g., "I want to go outside"), labeling items in the environment and/or making sounds in compliance with echoics prompted by the experimenter. During the treatment analysis a second by second continuous duration data collection procedure was used to measure vocal stereotypy which is reported as the proportion of

the session with vocal stereotypy (i.e., total duration with vocal stereotypy/total session duration).

The frequency and duration of each RIRD treatment implementation was recorded. RIRD treatment implementation was defined as the therapists immediate (within 2 s of the target behavior) presentation of the RIRD procedure (see RIRD section below for a description).

During the TI observation phase, data were collected on the frequency of opportunities to implement the procedure which were when episodes of vocal stereotypy occurred. Vocal stereotypy that occurred during treatment implementation was excluded from data collection. Opportunities to implement the procedure were separated by the absence of vocal stereotypy for 2 seconds or more. To measure treatment consistency the occurrence or nonoccurrence of treatment implementation (Peterson et al., 1982; Vollmer, Sloman, & Pipkin, 2008) was recorded as the percentage of occurrences out of total opportunities to implement the procedure. Furthermore, each time the RIRD procedure was implemented; accuracy data were collected on the occurrence or nonoccurrence of each RIRD component (depicted in Table 2) and reported as the percentage of components implemented out of the total number of possible components.

Interobserver Agreement. Interobserver agreement was assessed by a second observer either in vivo or via videotaped sessions. Interobserver agreement was calculated by dividing the number of intervals with agreements by the total number of intervals with agreements plus disagreements and multiplying by 100%. A second observer scored a minimum of 30% of sessions for each participant. During the treatment analysis, the mean agreement for vocal stereotypy was 99% for Kent, 93% for Cora, and 89% for Noah. The mean agreement for the frequency of treatment implementation was 100% for all participants. The mean agreement for the duration of treatment implementation was 98% for Kent, 99% for Cora, 97% for Noah.

During TI observations, the mean agreement for accuracy was 98% for Kent, 97% for Cora, 95% for Noah. The mean agreement for consistency was 94% for Kent, 98% for Cora, 91% for Noah.

Functional analysis. A brief functional analysis of vocal stereotypy was conducted using a subset of the conditions described in Iwata, Dorsey, Slifer, Bauman and Richman (1982/1994) alternated in a multi-element design. Sessions were 5 min in duration and the conditions were alternated as in Roscoe, Carreau, MacDonald, and Pence (2008) with the assumption that stereotypy was suspected to be automatically maintained. Therefore, the play condition was omitted and a series of no interaction conditions were conducted with periodic probes of the escape and attention conditions to determine whether the maintaining variable(s) of vocal stereotypy for each participant were likely to be socially mediated. The social conditions (escape and attention) served as the comparison to the no interaction condition which did not include social consequences for the target behavior. Higher responding in the no interaction condition relative to the attention and escape conditions indicates presumed maintenance by automatic reinforcement. As noted in Ahearn et al. (2007), RIRD would be contraindicated in the case of attention-maintained behavior as the treatment necessitates that attention in the form of vocal redirection be provided contingent on the target behavior. For each participant, the results of the functional analysis supported the hypothesis that vocal stereotypy occurred independent of social consequences (data available from authors).

Preference assessment. Paired-stimulus preference assessments were conducted with each participant to identify preferred leisure items for use during the treatment analysis. Eight stimuli were presented during the paired-stimulus assessment (similar to Fisher et al., 1992). The items assessed were selected by classroom staff that worked with the individual regularly. The

items were isolated for use in the preference assessment sessions for the duration of the assessment.

Phase 1

Treatment analysis. A withdrawal design was used to analyze the effects of RIRD on vocal stereotypy.

Baseline (BL). The participant and the experimenter sat at a table together. The participant had access to a preferred leisure item throughout the session. Kent engaged with small manipulatives such as carabiners, Cora looked through various movie cases and Noah built structures with Lincoln Logs®. There were no programmed consequences for vocal stereotypy. All sessions were 5 min in duration.

Response Interruption and Redirection (RIRD). The participant and the experimenter sat at a table together. The participant had access to the same high preference activity as in BL. Contingent on each instance of vocal stereotypy the experimenter paused the session timer, removed the preferred item and provided vocal prompts. Each participant was prompted to engage in vocal verbal behavior that was appropriate for their verbal repertoire. Participants were required to engage in intraverbals (e.g., "Where do you live?", "Where do you go to school?") to tact an object (e.g., "The block is blue" in reference to the current activity) or engage in echoic responses (e.g., "hi", "bye", "book"). Participant 1 was prompted to engage in echoics of familiar words. Participants 2 and 3 were prompted to emit a tact or intraverbal. If the participant did not respond within 2 seconds of the prompt the experimenter moved on to another prompt for appropriate language. The procedure was implemented until the participant responded to 3 consecutive prompts in the absence of vocal stereotypy. At the end of the procedure, the experimenter delivered social praise for using appropriate language (e.g., "Good job talking to

me!") and returned the preferred leisure item. The session timer was then resumed to ensure a 5-min sample for comparison to BL.

Results

Figure 1 shows the percentage of the session duration with vocal stereotypy during the treatment analysis. For Kent (top panel) and Cora (middle panel), an ascending trend was observed for vocal stereotypy during BL (M=24.7% and 17.8%, respectively) followed by an immediate decrease in vocal stereotypy to near-zero levels when RIRD was introduced (M=0.7% and 2.5% respectively). There was then an increase in vocal stereotypy upon reintroduction of BL conditions for both participants (M=34.2% and M=22.8%, respectively). Furthermore, Kent's vocal stereotypy was highly variable (range, 0-80%). Levels of vocal stereotypy obtained in the initial RIRD condition were replicated in the last phase as responding again decreased to low levels for both participants (M=2.7% and 1.8% respectively). For Noah (bottom panel), a moderate level of vocal stereotypy was observed throughout BL (M=43.5%). Once RIRD was implemented, there was an immediate decrease to low levels of vocal stereotypy (M=11.7%) Upon the removal of RIRD, there was an immediate increase in vocal stereotypy (M=31.7%). Levels of responding obtained in the previous RIRD condition were replicated in the last RIRD condition (M=10.5%).

Table 1 shows the frequency of RIRD treatment implementation, the duration of treatment implementation per session and the average duration per implementation for both RIRD conditions. For Kent and Cora the frequency of treatment delivery was low and variable throughout the treatment analysis (range, 0-5 and range, 3-11, respectively). For Kent, the treatment duration per session was also low and variable (range, 0 s – 54 s). However the average duration per treatment implementation was relatively consistent across RIRD conditions

(M=12.0 s and 14.5 s). For Cora, the duration of treatment implementation per session was also variable (range, 40 s -56 s); however, a decreasing trend was evident during RIRD 1 and again during RIRD 2. Lastly, the average duration per treatment implementation decreased from RIRD 1 (M=22.4 s) to RIRD 2 (M=14.8 s). For Noah, the frequency of treatment was high and variable (range, 16-38) throughout both RIRD series. Similarly, the duration of treatment implementation per session was long and variable in both RIRD conditions (range, 119 s-428 s) and was consistently longer in comparison to Kent and Cora. Lastly, the average duration per treatment implementation remained consistent across RIRD conditions (M=8.5 s and 8.7 s).

Phase 2

Staff training. Classroom staff members were provided with written guidelines outlining the RIRD procedure inclusive of the target behavior definition and an outline of treatment times (e.g., work sessions, group activities) versus non-treatment times (e.g., toileting, meals and earned leisure breaks). This document was also readily available to staff members throughout the day via each participant's behavioral treatment plan which was kept on a clipboard and accompanied the student throughout the school day.

Following a review of the procedures, staff members were given the opportunity to ask the trainer questions. Next they viewed a video model of the lead experimenter implementing the procedure with the participant. They were then required to participate in a minimum of three role plays with feedback. In order to complete the training, each staff member was required to demonstrate 100% accuracy for two consecutive role plays.

Treatment integrity observations. Following staff training, observations were conducted to provide information regarding the accuracy and consistency of treatment delivery in the classroom setting. TI observations were 15 min in duration and were administered over a

period of approximately 2 months for each participant. Data were collected at least twice per week across various days, times of day, activities, and staff members (Vollmer, Sloman, & Pipkin, 2008).

Results

The top panel of Figures 2-4 depict the percentage of treatment implementation (consistency) and vocal stereotypy episodes (opportunities to implement treatment) during classroom observations. The bottom panel of these figures shows the percentage of treatment implementation (consistency) across staff members. For Kent (Figure 2), the opportunities to implement the procedure varied (M =14.3, range, 1-82) and decreased over the first 9 sessions and then remained at low levels for the remainder of the observations (M=3.2; range, 1-10). There was a slight increasing trend in consistency over the first 9 observations, then treatment consistency was highly variable in subsequent observations (M=60.0%, range, 0-100%). Staff A implemented the procedure at a low and variable level (M=20.8% range, 0-50%). There was an increasing trend for Staff B and (M=66.0%, range, 17-91%) a decreasing trend for Staff C (M=75.7%, range, 33-100%).

For Cora (Figure 3), opportunities to implement the procedure were low throughout all observations (M=14.0, range, 1-12). Overall treatment consistency was highly variable (M=89.7%, range, 50-100%). All staff (A, B and C) implemented the procedure at a high level with some variability (M=83.3%, range, 50-100%; M =96.88%, range, 83-100%; and M =87.7%, range, 67-100%, respectively).

For Noah, (Figure 4) there were frequent opportunities to implement the procedure (M=29.1, range, 2-49). Overall, treatment consistency was low and highly variable (M=41.1%, range, 0-100%). Staff A, B and D's treatment implementation emulated this overall trend

(M=48.6%, range, 0-100%, M=43.4%, range, 22-67%, and M=27.0%, range, 11-64% respectively). However for Staff B, there was an increasing trend in treatment implementation from observation 16 throughout the remainder of the observation period (M=48.9%, range, 17-91%).

Table 2 shows the average percentage of implementation for each component of the RIRD procedure across participants during observations. For Kent and Cora, all components were implemented with a high level of integrity (range, 86%-100% and range, 97%-100% respectively). For Noah component 5 (repeating the procedure until the participant complies with 3 responses in the absence of the target response) was implemented below 80% integrity and all other components were implemented with a high level of integrity (range, 79%-100%). Component 1 was not applicable for Noah, as his preferred activity was building with blocks and he independently paused engagement with the activity each time the procedure was initiated.

Discussion

Experiment 1 suggested that the consistency of RIRD implementation was variable in the natural environment. However, the component assessment indicated that when the procedure was implemented, it was implemented with a high degree of accuracy (with the exception of Component 5 for Noah). Noah engaged in significantly higher rates of vocal stereotypy (M=29.1) producing an average of twice as many opportunities to implement the RIRD procedure when compared to the rates produced by Kent and Cora (M=14.3, M=14.0, respectively). Furthermore, treatment analysis data indicate that the frequency and the duration of the RIRD treatment procedure were significantly higher for Noah than the other two participants. These findings in conjunction with the low rates of treatment consistency observed during Noah's classroom observations highlight the concerns set forth by past researchers

regarding effortful treatment application presenting potential RIRD TI concerns. These results also support the assertion that the amount of time required to implement a treatment may affect how susceptible it is to lapses in TI (Gresham, MacMillan, Beebe-Frankenberger & Bocian, 2000).

Further investigation in a controlled setting is warranted to determine the extent to which the RIRD procedure can withstand TI errors such as those observed during Experiment 1. Therefore, in Experiment 2, consistency of RIRD implementation was directly manipulated and the subsequent effect on vocal stereotypy was examined.

Experiment 2

Method

Participants. The participants were 1 female and 2 males ages 14-16 years of age, all diagnosed with an ASD by a clinician unassociated with both the clinical and research teams. Cora, from Experiment 1, also participated in this experiment. Participant 2, Morris, was a 14-year-old male who emitted spontaneous mands in the form of 1-2 word utterances such as "computer" and "green token." Morris also emitted tacts and labeled items or attributes when prompted (e.g., "what is it?"). Participant 3, Chad, was a 15-year-old male who engaged in mands via one word approximations paired with picture exchange communication and engaged in echoics. He did not participate in conversational volleys or emit spontaneous tacts. Again, all participants were selected for participation in the same manner and attended the same private school program, were assigned to different classrooms, and taught in a 1:2 staff to student ratio.

Setting. For Participant 1, all sessions were conducted behind partitions in the classroom equipped with a video camera, a table and chairs. For participants 2 and 3, all sessions were

conducted in an experimental room (1.5 m by 3 m) equipped with a video camera, a table and chairs.

Response measurement and interobserver agreement. The definition of all target behavior and response measurement was the same as Experiment 1. However, during the treatment analysis, the average latency to the first response and the average inter-treatment time (ITT) were also calculated as a measure of treatment potency. Latency was defined as the duration from the start of the session to the first instance of vocal stereotypy. ITT was defined as the duration from the end of each treatment implementation to the next instance of vocal stereotypy.

Interobserver agreement was conducted in the same manner as Experiment 1. During the treatment analysis, the mean agreement for vocal stereotypy was 97% for Chad, and 98% for Morris. The mean agreement for the frequency of treatment implementation was 100% for both participants. The mean agreement for the duration of treatment implementation was 100% for Chad and 97% for Morris. During the parametric analysis, the mean agreement for vocal stereotypy was 97% for Cora, 95 % for Chad and 96% for Morris. The mean agreement for the frequency of treatment implementation was 100% for Cora and Chad, and 99 % for Morris. The mean agreement for the duration of treatment implementation was 99% for Cora, 97% for Chad, and 97% for Morris. The mean agreement for the average latency and average ITT was 98% for Cora, 97% for Chad, and 97% for Morris.

Pre-treatment. Functional analyses and preference assessments were conducted in the same manner as Experiment 1. For each participant, the results of the functional analysis supported the hypothesis that vocal stereotypy occurred independent of social consequences.

Experimental design. A withdrawal design was used to analyze the effects of RIRD on vocal stereotypy and to further evaluate the effects of TI manipulations via parametric analysis conditions.

Treatment analysis. BL and RIRD conditions were conducted in the same manner as Experiment 1.

Parametric analysis. Three TI conditions (75%, 50% and 25%) were alternated, sequentially, with BL. The RIRD procedure was implemented for 3 of every 4 responses during the 75% condition, 2 of every 4 responses during the 50% condition and 1 of every 4 responses during the 25% condition. Following the 25% condition, the RIRD treatment was reintroduced with 100% integrity to determine if treatment effects could be recovered. Finally, in the last condition, a series of two 25% sessions was alternated with one booster session (100% TI) to determine whether booster sessions could salvage the effects of poor treatment integrity. During each treatment session, the experimenter had access to cue cards that signaled the schedule of responses to interrupt and redirect as well as those responses to withhold treatment for (e.g., during the 25% treatment implementation condition, every 4th card indicated use of the RIRD procedure).

Results and Discussion

Figure 5 shows the percentage of the total duration of session time with vocal stereotypy during the treatment analysis. For Chad (top panel), an increasing trend was observed for vocal stereotypy throughout BL (M=14.4%). Once RIRD was implemented, there was an immediate decrease in vocal stereotypy to low levels (M=3.5%) Upon the removal of RIRD, there was an immediate increase in vocal stereotypy (M= 35.6%). Once RIRD was reintroduced vocal stereotypy decreased to near zero levels (M=1.0%). For Morris (bottom panel), vocal stereotypy

was emitted at a high level (M=68.2%) during the initial BL phase. There was a significant decrease in vocal stereotypy when RIRD was introduced (M=6.6%). Returning to the BL condition resulted in an increase in vocal stereotypy (M=65.8%). Upon reintroduction of the RIRD procedure vocal stereotypy returned to the level observed in the initial RIRD condition (M=9.7%).

Figure 6 shows the percent duration of session time with vocal stereotypy during the parametric analysis. For both Cora (top panel) and Chad (middle panel), vocal stereotypy was emitted at a low level during both the 75% condition (M=6.8% and M=4.1%, respectively) and the 50% condition (M=2.6% and M=5.3%, respectively) in comparison to the first two BL conditions (M= 29.2%, M= 42.3%; and M=11.8%, M=14.2%, respectively). There was an immediate increase in vocal stereotypy when BL was introduced for the third time (M=22.3% and M=30.3 %, respectively). During the 25% condition vocal stereotypy remained at BL levels (M=21.3% and M=18.3%, respectively). RIRD was then reintroduced at 100% integrity and low levels of vocal stereotypy were recovered (M=1.7% and M=3.4%, respectively). During the 25% plus booster condition, vocal stereotypy increased in comparison to the 100% condition and responding resembled the results obtained during the 50% condition (M=4.2%, and M= 6.1% , respectively).

The same pattern of responding was evident for Morris' (bottom panel); however, his results were not as robust as those of Cora and Chad. In comparison to the first two BL conditions (M=40.2 % and M=38.4 %, respectively), vocal stereotypy decreased to a lower level during the 75% condition (M=13.8%) and the 50% condition (M=15.4 %). There was an immediate increase in vocal stereotypy when BL was introduced for the third time (M=39.2%). During the 25% condition, vocal stereotypy remained at BL levels (M=44.5%). Low levels of

vocal stereotypy were recovered (M=5.8%) when RIRD was reintroduced at 100%. During the 25% plus booster condition, overall responding resembled the results obtained during the 50% condition (M=15.5%).

Table 3 shows the frequency of RIRD treatment implementation, the duration of treatment implementation in seconds per session and the average duration per implementation for both RIRD conditions. For Chad and Morris, the frequency and duration of treatment delivery per session were variable throughout the treatment analysis (range, 0-5, range, 41 s-302 s and range, 3-17, range, 113 s - 344 s, respectively). Furthermore, for Chad the average duration per treatment implementation was consistent across RIRD 1 (M=20.0 s) and RIRD 2 (M=20.6 s). For Morris, the average duration per treatment implementation increased slightly from RIRD 1 (M=11.5 s) to RIRD 2 (M= 13.9 s).

The parametric analysis indicated that the RIRD procedure suppressed vocal stereotypy when implemented at least 50% of the time. When booster sessions (100%) were interspersed during 25% integrity, the level of vocal stereotypy was similar to that observed during the 50% condition. However, treatment effects for Morris were modest in comparison to the results of the other two participants. Furthermore, despite previous exposure to TI failures for all participants, optimal treatment effects were recoverable once 100% RIRD was reintroduced.

Table 4 shows the average latency to the first response as well as the average ITT for each condition of the treatment analysis. For Cora, the average latency increased across conditions as the treatment analysis progressed. Although the average latency increased in duration overtime, a comparison of BL (M=12.0 s, M=36.7 s) and treatment (M=14.7 s, M=49.0 s) latency did not show a great difference. In addition the average ITT per condition increased from RIRD 1 (M=48.4 s) to RIRD 2 (M=62.0 s). For Chad and Morris, the difference between

BL and treatment latency was also not significant. For Chad, the average latency increased from RIRD 1 (M=14.3 s) to RIRD 2 (M=72.1 s) and similar to Cora, the average ITT per condition increased from RIRD 1 (M=29.9 s) to RIRD 2 (M=35.3 s). For Morris, however the average latency decreased from RIRD 1 (M=22.8 s) to RIRD 2 (M=1.5 s) and the average ITT per condition decreased from RIRD 1 (M=25.5 s) to RIRD 2 (M=13.8 s).

Table 5 shows the average frequency per condition and average duration (s) per treatment implementation across all parametric analysis conditions. For both Cora and Chad the average frequency of treatment implementation was low and variable across conditions (range, 2.6-5.2 & 2.7-7.8, respectively). In comparison to the other participants, Morris' average frequency of treatment implementation was high but also variable across conditions (range, 5.3-17.0). For all participants the average frequency was lowest during the 25%, 50% and 25% plus booster conditions. For Cora there was a decreasing trend in the average duration per treatment implementation as the parametric analysis progressed (range, 7.5 s-12.1 s). For Chad and Morris the average duration per treatment implementation did not show a trend and was variable (range, 7.2 s-10.1 s & 12.3 s -28.7 s respectively) as the parametric analysis progressed. For Morris the average duration per treatment implementation was consistently longer in comparison to Cora and Chad's results.

With regards to latency and ITT, a potent treatment effect would be indicated if an increasing or stable trend was observed. On the other hand, a mild treatment effect may be indicated if a decrease in latency and/or ITT was observed. Given the decrease in latency and ITT observed for Morris, it is possible that the RIRD procedure was not potent enough to compete with a continuous schedule of automatic reinforcement and habituation to the intervention may have occurred. Evidence indicating that mild stimuli may be less effective

contingent on challenging behavior during intermittent schedules may explain why a continuous schedule of RIRD (treatment analysis) was effective for Morris but an intermittent schedule (parametric analysis) produced less significant response suppression in comparison to the other participants. For example, Cipani, Brendinger, McDowell and Usher (1991) compared the contingent application of lemon juice versus manual guidance overcorrection without the use of extinction. They found that lemon juice produced near zero levels of stereotypical and self-injurious behavior on both a continuous schedule (FR1) and a variable ratio 4 (VR4) schedule. Overcorrection decreased the target response when implemented continuously however it did not produce significant suppression of the target response on a VR4 schedule. In addition, lemon juice produced greater overall response suppression than overcorrection. Further investigation comparing the magnitude of common stimuli used contingently to decrease challenging behavior in clinical settings is warranted.

Furthermore, as Noah's results from Experiment 1 demonstrated, high rate target behavior and long treatment durations translate to possible TI challenges in the classroom setting. Morris' results also resembled this pattern highlighting the importance of generating solutions to address such treatment challenges. An approach to decrease the duration of the RIRD procedure without diminishing treatment effects was offered by Saini, Gregory, Uran and Fantetti (2015), as they reduced the response requirement from three demands to one demand. Future research should determine whether this procedural variation can bolster TI in non-experimental settings. Moreover, clinicians should acknowledge that a treatment package may be a necessary solution for clients who engage in high rates of stereotypy. For example, Love et al. (2012) offered another potential solution by using matched stimulation to decrease the motivating operation for stereotypy. In addition, the use of direct training for appropriate

replacement behavior (e.g. Colón et al., 2012) may assist with suppression of the target behavior by making appropriate behavior more probable.

General Discussion

The current study indicates that despite evidence-based training, treatment implementation may be highly variable in the natural setting. Staff members in most school settings are not likely to be trained in experimental methods (McIntyre et al., 2007) and are exposed to various setting demands (e.g., attending to other students, data collection). Therefore, resource-intensive procedures characterized by long treatment durations or high rates of target behavior may result in inconsistent treatment delivery. Nonetheless, obtaining a successful treatment outcome may be dependent on the proportion of responses followed by treatment. In order to understand the conditions under which treatment inconsistency affects treatment outcome, treatment strength should be considered. Contingency strength, or the varying probabilities of responses and consequences in the context of environmental events (Vollmer, Borrero, Wright, Van Camp & Lalli, 2001) can be conceptualized along a continuum from very strong (FR1) to neutral and negative contingencies (Hammond, 1980). Likewise, overall treatment strength is the probability that target responses are reliably followed by programmed consequences.

In the case of automatically-maintained behavior, the response immediately produces reinforcement that the clinician does not have access to and cannot readily control. When extinction is not feasible, as is the case with automatically-maintained vocal stereotypy, reinforcement remains available contingent on each response. Given this concurrent schedule arrangement, maintaining a strong treatment contingency would likely make the procedure more effective. The current findings suggest that a mean schedule at or above 50% presumably

constitutes a strong treatment contingency that competes with continuous automatic reinforcement.

The parametric analysis conducted by Ahrens et al. (2011), as well as a study conducted by Shawler and Miguel (2015) in which motor demands and vocal demands both had similar suppressive effects on vocal stereotypy, indicate that one of the behavioral mechanisms responsible for RIRD's effectiveness may be positive punishment. Very few studies have examined the effects of intermittent punishment without the use of extinction. Nonetheless, basic research conducted by Azrin, Holz, and Hake (1963) with non-humans showed that intermittent punishment without the use of extinction did not produce response suppression as rapid or significant as continuous punishment.

Further knowledge regarding the above mentioned parameters of punishment (e.g., intermittent punishment without extinction, availability of alternative reinforcement, magnitude of currently relevant stimuli) is imperative to ensure that clinicians are fully equipped with the technology to provide behavioral treatment in an effective and efficient manner. Although punishment procedures are generally used sparingly within our clinical practice and recommended only when paired with reinforcement procedures, they are sometimes clinically necessary. Thus these procedures deserve further attention in both the basic and applied literature. A review conducted by Lerman and Vorndran (2002) indicated various gaps in our knowledge of punishment-based procedures, highlighting a decline in basic research and a lack of applied research. Unfortunately, these assertions have caused little change regarding our knowledge of decelerative procedures over the past decade.

In addition, there is some evidence that recent history can alter the effects of intermittent schedules (St. Peter Pipkin & Vollmer, 2009). Unfortunately, recent history effects have not been

analyzed thoroughly with regard to punishment procedures. Nonetheless, the basic literature suggests that exposing the responses of non-human subjects to intermittent punishment prior to continuous punishment does not lead to sufficient response suppression during a continuous schedule (Banks, 1966; Shemer & Feldon, 1984). However, applied studies have shown intermittent punishment to be effective following recent exposure to continuous punishment (Rollings & Baumeister, 1981; Romancyzk, 1977). As previously mentioned, the participant in the Smith et al. (1999) study was not exposed to a recent history of 100% TI due to an inability to block every response the participant emitted during the continuous blocking condition. Therefore recent history effects may explain the difference between their results and those of Lerman and Iwata (1996), Ahrens et al. (2011) and the current study. Furthermore, the results of a parametric analysis conducted by Kliebert, Tiger, and Toussaint (2011) also indicated that recent history effects may have been responsible, at least in part, for differences between their participant's findings. In this study, the authors manipulated the delay (3 s, 15 s and 30 s) to implementation of response interruption (only) in relation to treatment effects for automatically-maintained behavior (skin picking and hair twirling) with two participants. They found that delayed response interruption was effective for one of two participants. The participant with whom delayed response interruption was effective was first exposed to immediate response interruption initially. Furthermore, for this participant delayed response interruption was only effective during conditions in which the staff implementing the procedure was present (versus conditions when the staff member was absent from the room at the beginning of the session) and it is important to note that in the context of the staff member being absent the participant had not been exposed to immediate response interruption prior to experiencing the delay. By contrast, the

second participant (for whom the delayed response interruption was ineffective) was first exposed to all three delay conditions prior to the immediate response interruption condition. Finally, it is also possible that mixed results between studies is associated with the behavioral mechanism responsible for treatment effects. For instance, Ahrens et al., and Lerman and Iwata attributed treatment effects for their participants to punishment whereas Smith, Russo and Le indicated that the behavioral mechanism responsible for the results of their participant was extinction**.**

Nevertheless, it appears that the extent to which inconsistency impacts the effects of RIRD is minimal if the treatment is conducted in at least half of the opportunities and follows previous exposure to full RIRD implementation. Further research is required to determine whether treatment effects would be stronger if treatment failure conditions are directly preceded by 100% integrity conditions rather than BL (St. Peter Pipkin, Vollmer & Sloman, 2010; Vollmer, Roane, Ringdahl, & Marcus, 1999). Although it remains unknown how recent or how long the exposure to 100% TI must be to afford a successful treatment outcome during TI failures the existing research generally indicates that clinicians should provide exposure to 100% TI prior to the onset of possible TI failures in an effort to increase the likelihood of treatment success in the natural setting. In addition, the results of the 25% plus booster condition seem to indicate that using 100% booster sessions is a promising strategy that may ameliorate poor treatment integrity in the natural setting. In the current study using booster sessions to bolster TI to a mean of 50% or greater was effective, yet, further investigation of what booster session parameters reliably salvage previous treatment effects is justified. Lastly, the impact of RIRD TI failures on appropriate behavior was not investigated. Therefore, it would be imperative for

future studies to determine the effects of RIRD TI failures on appropriate behavior (e.g., appropriate vocalizations).

References

Ahearn, W. H., Clark, K. M., MacDonald, R. P. F., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 40,* 263-275. doi:10.1901/jaba.2007.30-06

Ahrens, E. N., Lerman, D. C., Kodak, T., Worsdell, A. S., & Keegan, C. (2011). Further evaluation of response interruption and redirection as treatment for stereotypy. *Journal of Applied Behavior Analysis, 44*, 95-108. doi:10.1901/jaba.2011.44-95

Athens, E. S., Vollmer, T. R., Sloman, K. N., & Pipkin, C. S. P. (2008). An analysis of vocal stereotypy and therapist fading. *Journal of Applied Behavior Analysis, 41*(2), 291-297. doi.org/10.1901/jaba.2008.41-291

Azrin, N. H., Holz, W. C., & Hake, D. F. (1963). Fixed-ratio punishment. *Journal of the Experimental Analysis of Behavior, 6*(2), 141-148. doi:10.1901/jeab.1963.6-141

Banks, R. K. (1966). Persistence to continuous punishment following intermittent punishment training. *Journal of Experimental Psychology, 71*(3), 373-377. doi:10.1037/h0023016

Cipani, E., Brendlinger, J., McDowell, L., & Usher, S. (1991). Continuous vs. intermittent punishment: A case study. *Journal of Developmental and Physical Disabilities, 3*(2), 147-156. doi:10.1007/BF01045930

Colón, C. L., Ahearn, W. H., Clark, K. M., & Masalsky, J. (2012). The effects of verbal operant training and response interruption and redirection on appropriate and inappropriate

vocalizations. *Journal of Applied Behavior Analysis, 45*(1), 107-120. doi: 10.1901/jaba.2012.45-107

Duffy-Cassella, M., Sidener, T. M., Sidener, D. W., & Progar, P. R. (2011). Response interruption and redirection for vocal stereotypy in children with autism: A systematic replication. *Journal of Applied Behavior Analysis, 44*, 169-173. doi:10.1901/jaba.2011.44-169.

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25*, 491-498. doi:10.1901/jaba.1992.25-491

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

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257-263. doi:10.1901/jaba.1993.26-257

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented?. *Learning Disabilities Research & Practice, 15*(4), 198-205. doi: http://dx.doi.org/10.1207/SLDRP1504_4

Hagopian L.P., & Adelinis, J. D. (2001). Response blocking with and without redirection for the treatment of PICA. *Journal of Applied Behavior Analysis, 34***,** 527–530. doi: 10.1901/jaba.2001.34-527

Hammond, L. J. (1980). The effect of contingency upon the appetitive conditioning of free-operant behavior. *Journal of the Experimental Analysis of Behavior, 34,* 297-304. doi:10.1901/jeab.1980.34-297

Homer, A. L., & Peterson, L. (1980). Differential reinforcement of other behavior: A preferred response elimination procedure. *Behavior Therapy, 11*, 449-471. doi:10.1016/S0005-7894(80)80063-3

Iwata, B.A., Dorsey, M. F., Slifer, K .J. Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27,* 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2,* 3-20, 1982). doi: 10.1002/bin.297

Kliebert, M. L., Tiger, J. H., & Toussaint, K. A. (2011). An approach to identifying the conditions under which response interruption will reduce automatically reinforced problem behavior. *Behavior analysis in practice*, *4*(1), 17-26.

Lerman, D. C., & Iwata, B. A. (1996). A methodology for distinguishing between extinction and punishment effects associated with response blocking. *Journal of Applied Behavior Analysis, 29,* 231-233. doi:10.1901/jaba.1996.29-231

Lerman, D. C. and Vorndran, C. M. (2002), On the Status of Knowledge for Using Punishment: Implications for Treating Behavior Disorders. *Journal of Applied Behavior Analysis,* 35: 431–464. doi: 10.1901/jaba.2002.35-431

Liu-Gitz, L., & Banda, D. R. (2010). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. *Behavioral Interventions, 25*, 77-87.

Love, J. J., Miguel, C. F., Fernand, J. K., & LaBrie, J. K. (2012). The effects of matched stimulation and response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 45*, 549-564. doi:10.1901/jaba.2012.45-549.

Martinez, C. K., & Betz, A. M. (2013). Response interruption and redirection: Current research trends and clinical application. *Journal of Applied Behavior Analysis, 46*, 549-554. doi: 10.1002/jaba.38

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991-2005. *Journal of Applied Behavior Analysis, 40*, 659-672. doi: 10.1901/jaba.2007.659-672

Miguel, C. F., Clark, K., Tereshko, L., & Ahearn, W. H. (2009). The effects of response interruption and redirection and sertraline on vocal stereotypy. *Journal of Applied Behavior Analysis, 42*, 883-888. doi: 10.1901/jaba.2009.42-883

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492. doi:10.1901/jaba.1982.15-477

Pipkin, C. S. P. and Vollmer, T. R. (2009), Applied implications of reinforcement history effects. *Journal of Applied Behavior Analysis*, 42: 83–103. doi: 10.1901/jaba.2009.42-83

Pipkin, C. S. P., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis*, *43*(1), 47-70.

Rapp, J.T. & Vollmer, T. R., (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities,* 26, 527-547. doi: 10.1016/j.ridd.2004.11.005

Rollings, J. P.; Baumeister, A. A. (1981). Stimulus control of stereotypic responding: effects on target and collateral behavior. *American Journal of Mental Deficiency*, 86(1), 67-77.

Romanczyk, R. G. (1977). Intermittent punishment of self-stimulation: Effectiveness during application and extinction. *Journal of Consulting and Clinical Psychology,* 45(1), 53-60. doi: 10.1037/0022-006X.45.1.53

Roscoe, E. M., Carreau, A., MacDonald, J., & Pence, S. T. (2008). Further evaluation of leisure items in the attention condition of functional analyses. *Journal of Applied Behavior Analysis*, *41*, 351-364. doi:10.1901/jaba.2008.41-351

Saini, V., Gregory, M. K., Uran, K. J., & Fantetti, M. A. (2015). Parametric analysis of response interruption and redirection as treatment for stereotypy. *Journal of applied behavior analysis*, *48*(1), 96-106. doi: 10.1002/jaba.186

Schumacher, B. I., & Rapp, J. T. (2011). Evaluation of the immediate and subsequent effects of response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 44*, 681-685. doi: 10.1901/jaba.2011.44-681

Shawler, L. A., & Miguel, C. F. (2015). The Effects of Motor and Vocal Response Interruption and Redirection on Vocal Stereotypy and Appropriate Vocalizations. *Behavioral Interventions, 30*(2), 112-134. doi: 10.1002/bin.1407

Shemer, A., & Feldon, J. (1984). Long-term partial reinforcement extinction effect and long-term partial punishment effect in a one-trial-a-day paradigm. *Bulletin of the Psychonomic Society, 22*(3), 221-224.

Smith, R. G., Russo, L., & Le, D. D. (1999). Distinguishing between extinction and punishment effects of response blocking: A replication. *Journal of Applied Behavior Analysis, 32*, 367-370. doi:10.1901/jaba.1999.32-367

Vanderkerken, L., Heyvaert, M., Maes, B., and Onghena, P. (2013). Psychosocial interventions for reducing vocal challenging behavior in persons with autistic disorder: A multilevel meta-analysis of single-case experiments. *Research in Developmental Disabilities, 34*, 4515-4533. doi:10.1016/j.ridd.2013.09.030

Vollmer, T. R. (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities, 15*, 187-207.

Vollmer, T. R., Borrero, J. C., Wright, C. S., Van Camp, C., & Lalli, J. S. (2001). Identifying possible contingencies during descriptive analyses of severe behavior disorders. *Journal of Applied Behavior Analysis,* 34, 269-287.

Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32,* 9-23. doi:10.1901/jaba.1999.32-9

Vollmer, T. R., Sloman, K. N., & Pipkin, C. S. P. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice, 1*, 4-11.

Table 1

RIRD Treatment Implementation during Treatment Analysis for Experiment 1

Participant	Condition	Session	Frequency	Total Duration (s)	Average Duration (s) Per Implementation
Kent					
	RIRD 1	1	0	0	
		2	1	17	
		3	2	19	12.0
	RIRD 2	1	0	0	
		2	1	17	
		3	5	54	
		4	1	7	
		5	1	38	14.5
Cora					
	RIRD 1	1	6	156	
		2	5	131	
		3	5	71	22.4
	RIRD 2	1	11	153	
		2	3	59	
		3	3	40	14.8
Noah					
	RIRD 1	1	19	180	
		2	31	236	
		3	16	119	
		4	32	293	8.5
	RIRD 2	1	25	181	
		2	23	428	
		3	35	167	
		4	30	293	
		5	38	370	
		6	15	130	
		7	35	293	8.7

Table 2

Percentage of Accuracy across all components of RIRD during Observations for Experiment 1

Component	Description	Kent	Cora	Noah
1	Remove preferred item	92	97	---
2	Prompt/gain attention of participant (if necessary)	100	100	86
3	Provide prompts for appropriate vocalizations	99	100	96
4	If participant does not respond within 2-3 seconds provide the next prompt	90	100	100
5	Repeat components 3 and 4 until the participant complies with 3 responses in the absence of the target response	86	98	79
6	Deliver praise for appropriate vocalizations	97	100	85

Table 3

Frequency and Duration of RIRD during Treatment Analysis for Experiment 2

Participant	Condition	Session	Frequency	Total Duration (s)	Average Duration (s) per Implementation
Chad					
	RIRD 1	1	11	123	
		2	5	139	
		3	7	199	20.0
	RIRD 2	1	5	87	
		2	8	147	
		3	6	202	
		4	17	302	
		5	6	192	
		6	8	140	
		7	3	41	
		8	5	67	
		9	4	113	
		10	4	88	20.6
Morris					
	RIRD 1	1	12	180	
		2	15	143	
		3	8	113	
		4	17	208	
		5	14	137	11.5
	RIRD 2	1	25	344	
		2	16	228	
		3	15	212	
		4	13	259	13.9

Table 4

Average Latency and ITT during Treatment Analysis for Experiment 2

Participant	Condition	Average Latency	Average ITT
Cora			
	BL	12.0	
	RIRD 1	14.7	48.4
	BL	36.7	
	RIRD 2	49.0	62.0
Chad			
	BL	33.0	
	RIRD 1	14.3	29.9
	BL	10.0	
	RIRD 2	72.1	35.3
Morris			
	BL	8.7	
	RIRD 1	22.8	25.5
	BL	0.3	
	RIRD 2	1.5	13.8

Table 5

RIRD Treatment Implementation during Parametric Analysis for Experiment 2

Participant	Condition	Average Frequency	Average Duration (s) per implementation
Cora			
	75%	5.2	12.1
	50%	2.6	10.0
	25%	3.6	9.8
	100%	4.0	9.1
	25% + Booster	2.8	7.5
Chad			
	75%	4.6	10.0
	50%	3.9	8.5
	25%	2.7	7.2
	100%	7.8	10.1
	25% + Booster	3.3	8.2
Morris			
	75%	17.0	23.1
	50%	9.9	14.4
	25%	5.3	25.0
	100%	11.2	12.3
	25% + Booster	6.5	28.7

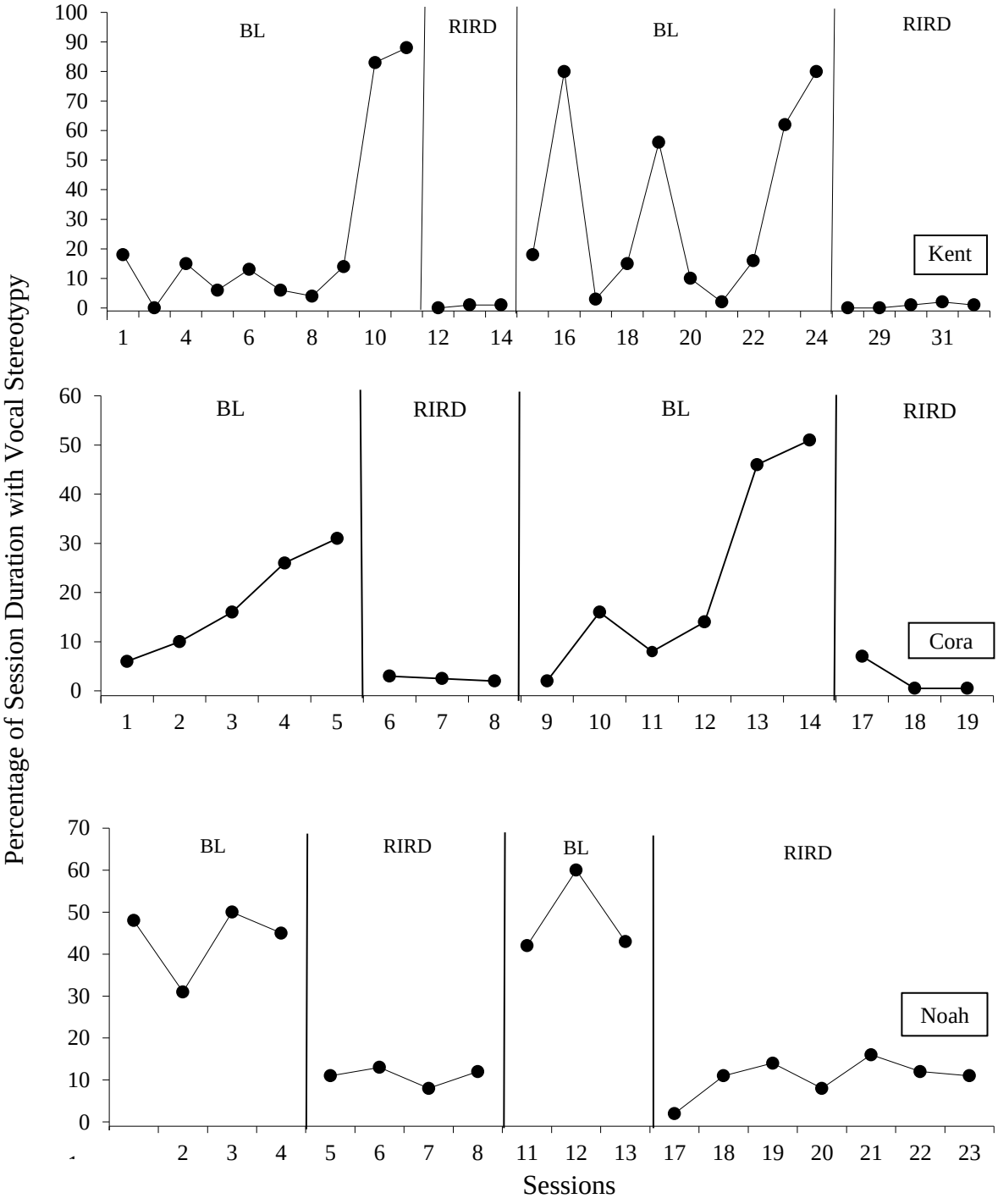

Figure 1. Percent duration of session time with vocal stereotypy during RIRD treatment analysis for Exp. 1 participants.

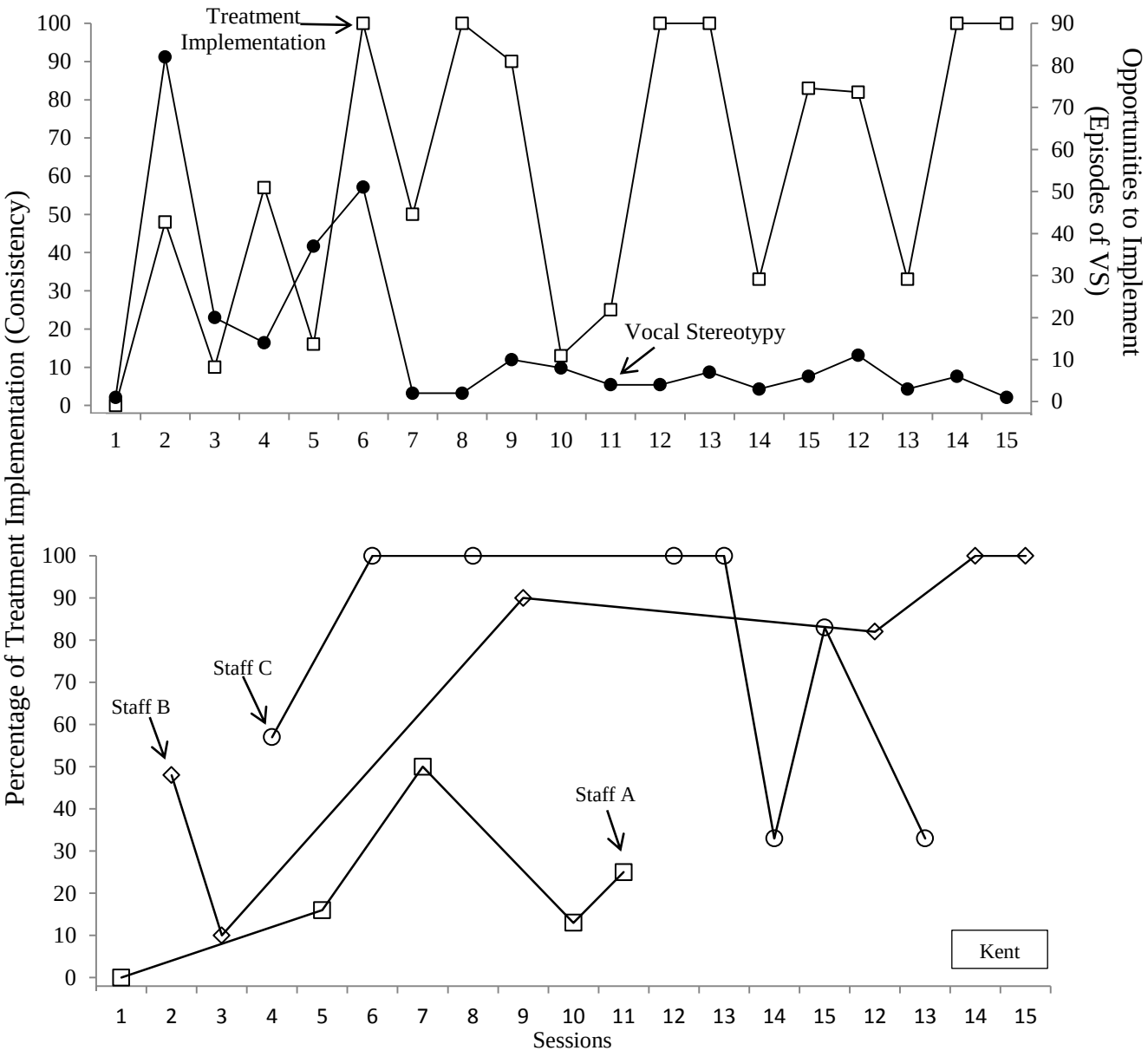

Figure 2. Percentage of treatment consistency during treatment integrity observations is depicted on the primary *y* axis. In the Top Panel, opportunities to implement the RIRD procedure are depicted on the secondary *y* axis. The Bottom panel depicts a comparison of treatment consistency across staff members. Each letter represents a different staff member implementing the procedure in the classroom setting with Kent.

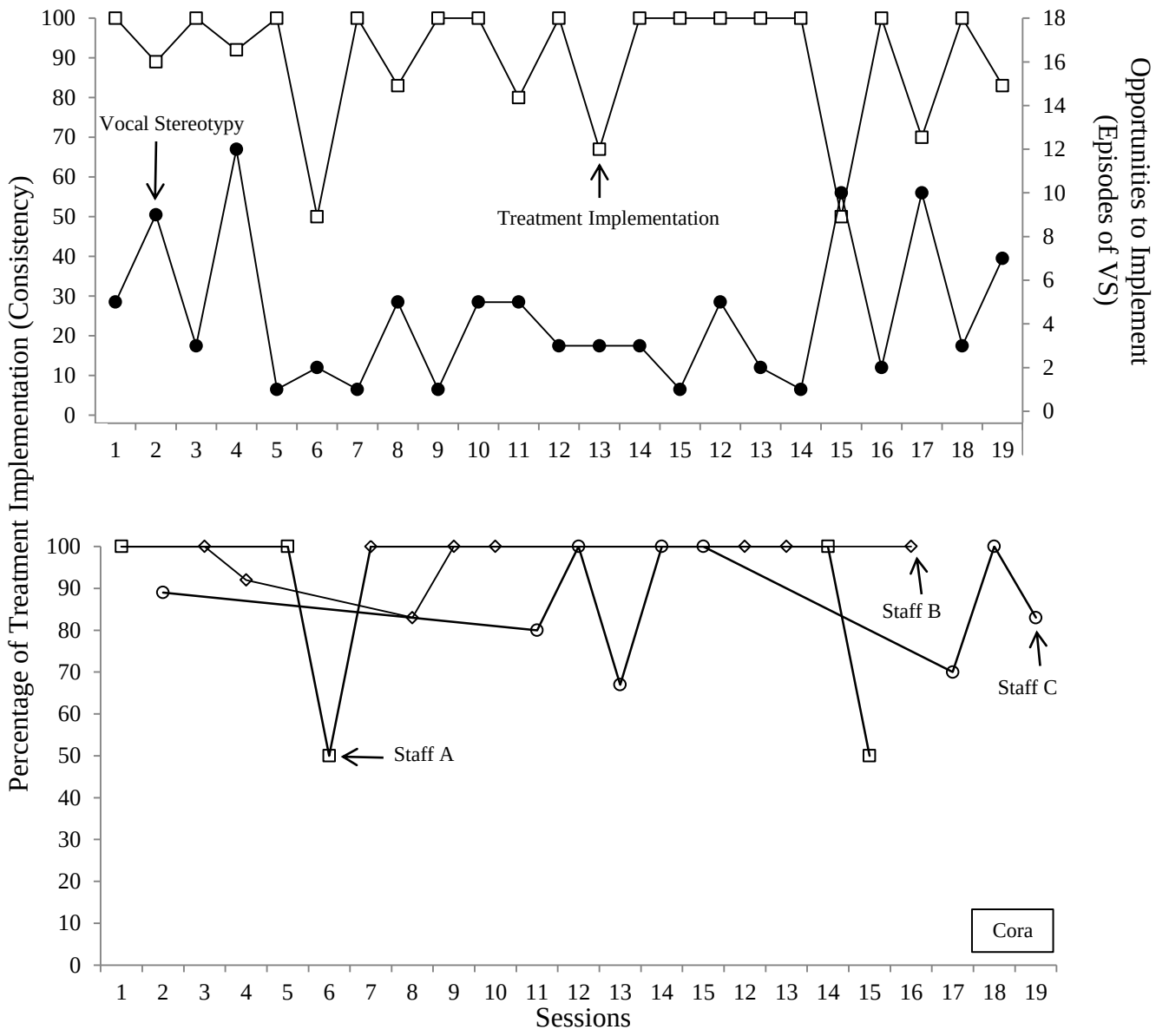

Figure 3. Percentage of treatment consistency during treatment integrity observations is depicted on the primary *y* axis. In the Top Panel, opportunities to implement the RIRD procedure are depicted on the secondary *y* axis. The Bottom panel depicts a comparison of treatment consistency across staff members. Each letter represents a different staff member implementing the procedure in the classroom setting with Cora.

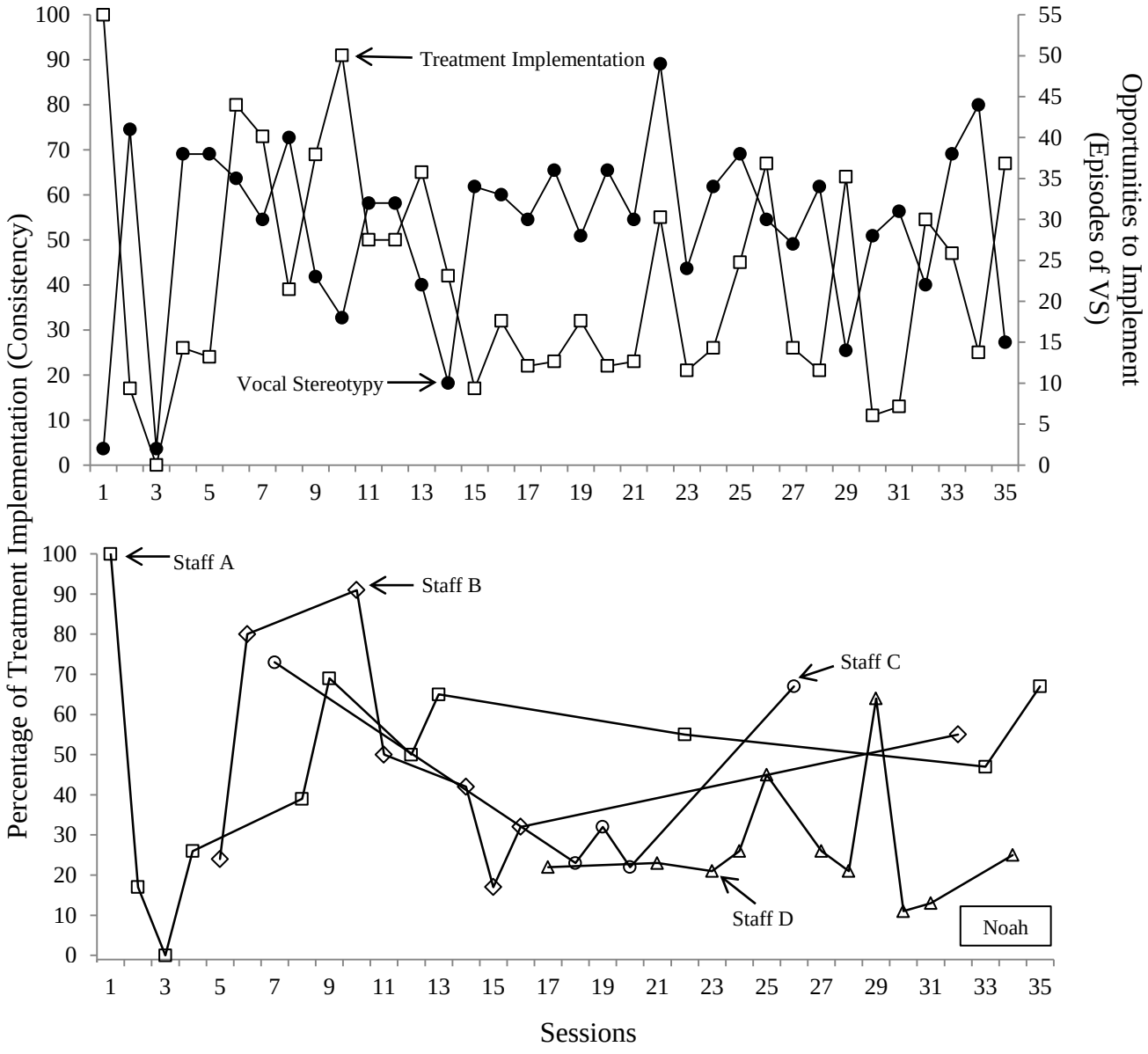

Figure 4. Percentage of treatment consistency during treatment integrity observations is depicted on the primary *y* axis. In the Top Panel, opportunities to implement the RIRD procedure are depicted on the secondary *y* axis. The Bottom panel depicts a comparison of treatment consistency across staff members. Each letter represents a different staff member implementing the procedure in the classroom setting with Noah.

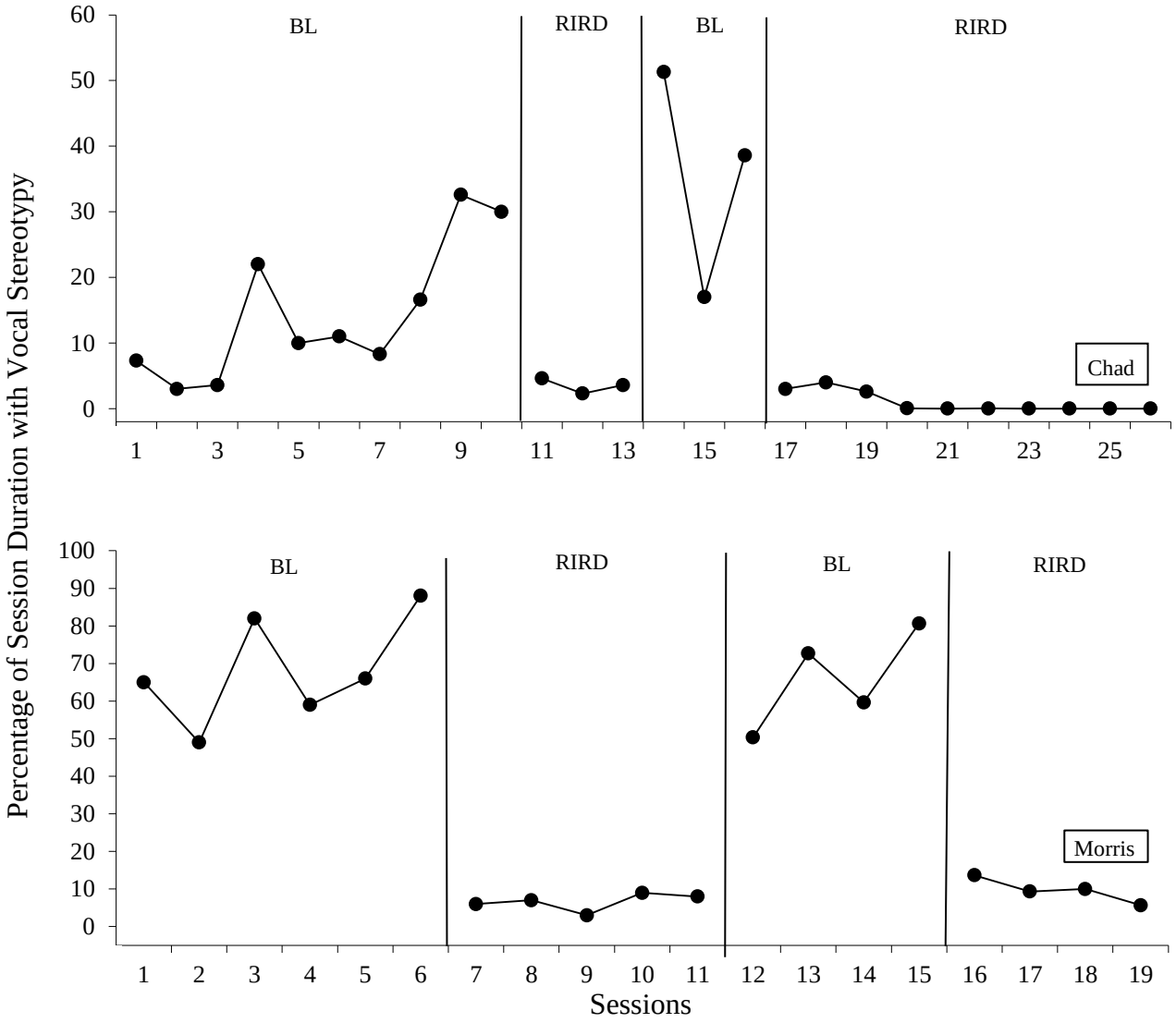

Figure 5. Percent duration of session time with vocal stereotypy during the RIRD treatment analysis for Chad and Morris (Exp.2)

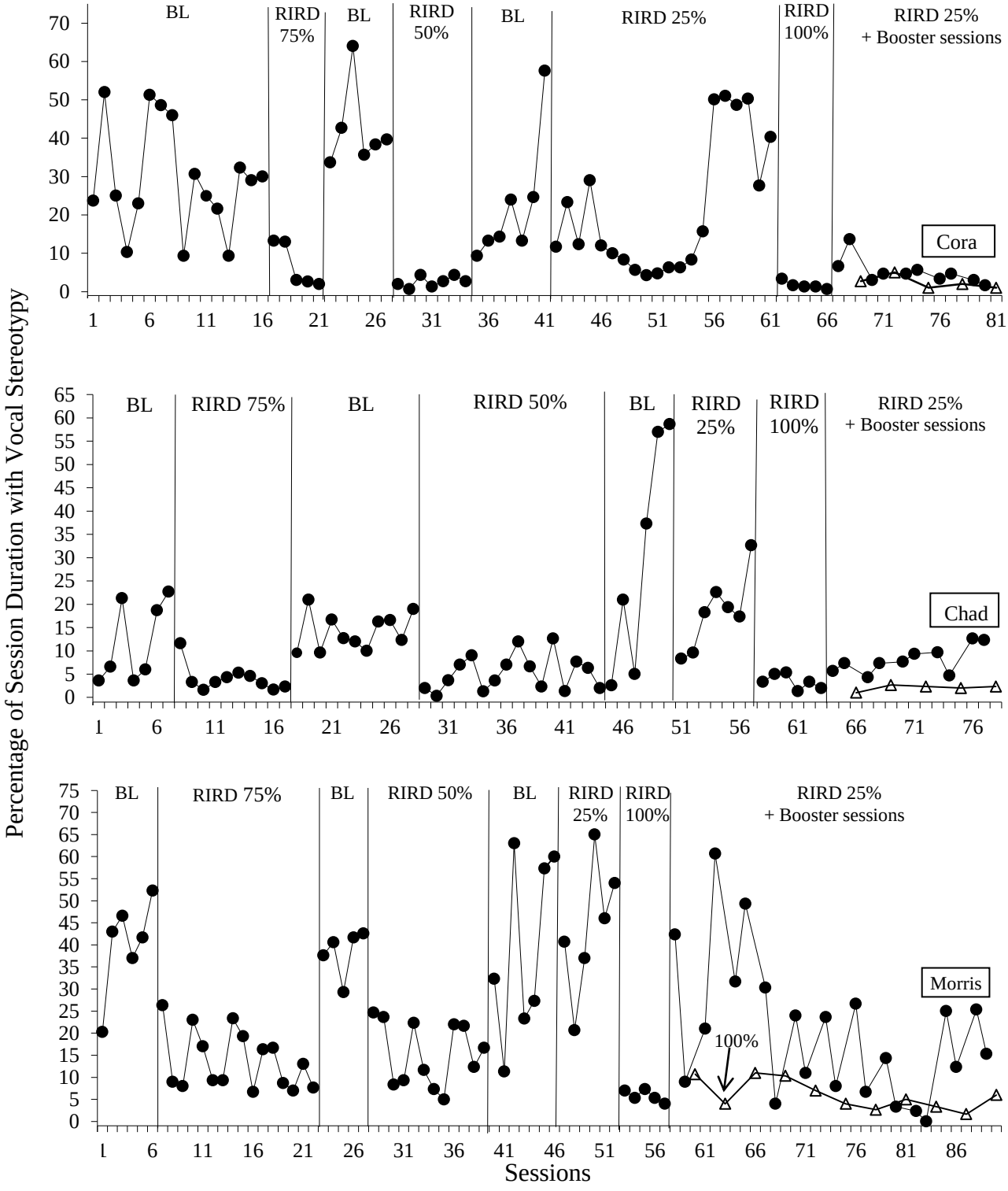

Figure 6. Percent duration of session with vocal stereotypy during RIRD parametric analysis for exp. 2 participants.

Appendix A

-Treatment Integrity Observation Data sheet-

Student (initials): _____ *Observation #*_____
Date: _____ Time: _____-_____ Observer (Circle one): Primary or IOA
Observer Initials:_____ Staff observed (initials):_____
Location: _____ Activity:_____

Target Response Definition:
Vocal stereotypy – Any instances of non-contextual speech, to include babbling, and the repetitive production (within 5 seconds) of sounds or actual words emitted at conversation level or above. Not to include speech paired with eye contact to another person, requesting items, words produced in response to teacher directives, or recognizable words produced in the context of play (ex. pushing a truck and saying "truck" or playing with a pig and saying "oink oink").

Treatment Implementation (Schedule)

Teacher action (upon each instance of Vocal Stereotypy)	**Number of opportunities**	**Number of times implemented**	**Percentage**
The teacher implements the RIRD procedure immediately (2 sec)- whether or not all steps below are included or implemented correctly			

Treatment Component Checklist (Accuracy)

Teacher action (upon each instance of vocal stereotypy)	**Number of opportunities**	**Number of times implemented**	**Percentage**
1. **Removal of Preferred item** (*if not applicable do not count/tally in # of opportunities)*			
2. **Prompts/gains attending** -*if necessary*- (if Not Applicable do not count/tally in # of opportunities)			
3. **Provides prompts for appropriate language** (randomized order of questions)			
4. **If student does not respond within 2-3 seconds,** moves on by providing the next prompt for appropriate language			
5. *Repeats components 4 & 5 until the student complies with 3 consecutive responses in the absence of the target response.*			
6. **Delivers praise** (e.g., "super job talking to me!")			
TOTAL			