

2021

Effects of Blended Integrity Failures on Reinforcement-Based Interventions

Stephanie Hope Jones
shj0001@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>



Part of the [Applied Behavior Analysis Commons](#), and the [Experimental Analysis of Behavior Commons](#)

Recommended Citation

Jones, Stephanie Hope, "Effects of Blended Integrity Failures on Reinforcement-Based Interventions" (2021). *Graduate Theses, Dissertations, and Problem Reports*. 8126.
<https://researchrepository.wvu.edu/etd/8126>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

2021

Effects of Blended Integrity Failures on Reinforcement-Based Interventions

Stephanie Hope Jones

**Effects of Blended Integrity Failures on Responding during Reinforcement-Based
Interventions**

Stephanie Hope Jones

**Dissertation submitted to the Eberly College of Arts and Sciences at West Virginia
University in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Psychology**

Claire St. Peter, Ph.D., Chair

Christina Duncan, Ph.D.

Kimberly Floyd, Ph.D.

Kathryn Kestner, Ph.D.

Michael Perone, Ph.D.

Department of Psychology

Morgantown, West Virginia

2021

Keywords: Differential Reinforcement of Alternative Behavior; Fixed-Time Schedules; Human
Operant; Treatment Integrity

Copyright 2021 Stephanie Jones

ABSTRACT

Effects of Blended Integrity Failures on Responding during Reinforcement-Based Interventions

Stephanie Hope Jones

Reinforcement-based interventions reduce problematic behavior when implemented as designed. However, the effectiveness of these interventions may decrease when deviations from treatment protocols (i.e., treatment-integrity errors) occur. Treatment-integrity errors differentially impact reinforcement-based interventions based on multiple factors, including how frequently errors occur and the intervention type. Even nominally acceptable integrity values (e.g., 80%) may be detrimental depending on the intervention. To evaluate this possibility and directly compare the effectiveness of multiple reinforcement-based interventions, we conducted two within-subject evaluations using laboratory arrangements. For both experiments, we recruited four undergraduate students to participate in a computer task that involved clicking on moving circles to earn points. During Experiment 1, we compared the effectiveness of Fixed-Time Schedules (FT) and Extinction (i.e., Noncontingent Reinforcement) and Differential Reinforcement of Alternative Behavior (DRA) when implemented at 80% integrity. During Experiment 2, we compared the effectiveness of ratio-based DRA (DRA with a ratio schedule maintaining alternative behavior) and interval-based DRA (DRA with an interval schedule maintaining alternative behavior). Results were idiosyncratic across participants. However, DRA with a ratio schedule was the only consistently effective intervention when implemented at 80% integrity; neither FT nor interval-based DRA were consistently effective when implemented at 80% integrity. Implications for research on effects of treatment-integrity errors and applied practice are discussed.

Table of Contents

Introduction.....	1
General Method.....	8
Participants and Setting.....	8
Procedure.....	9
Integrity Errors.....	10
Experiment 1 Method.....	11
Experiment 1 Results and Discussion.....	13
Experiment 2 Method.....	17
Experiment 2 Results and Discussion.....	19
General Discussion.....	22
References.....	29
Appendix A.....	34
Appendix B.....	35
Tables & Figures.....	36

Effects of Blended Integrity Failures on Responding during Reinforcement-Based Interventions

An estimated 27% of children with Autism Spectrum Disorder (ASD) engage in at least one form of problem behavior (Soke et al., 2016), and between 5% and 10% of individuals with intellectual disabilities engage in life-threatening problem behavior (Holden & Gitlesen, 2006; Lowe et al., 2007). Aside from potential risk of harm for the individual with problem behavior and their caregivers, occurrence of problematic behavior is also associated with increased academic challenges, isolation, and family stress (Hagopian et al., 2013). Fortunately, reinforcement-based interventions reduce problematic behavior when implemented as designed and have convincing empirical support for their effectiveness for multiple populations (e.g., adults and children with disabilities), settings (e.g., home, clinic, and school), and topographies of problematic behavior (e.g., aggression, self-injury, inappropriate vocalizations; Hagopian et al., 2013). Broadly, reinforcement-based interventions reduce problematic behavior by shifting reinforcer access to periods other than those following problematic behavior (e.g., following alternative behavior).

Although reinforcement-based interventions decrease problematic behavior when implemented as designed, numerous descriptive evaluations suggest that deviations from treatment protocols (termed *treatment-integrity errors*; Vollmer et al., 2008) are likely (Arkoosh et al., 2007; Carroll et al., 2013; Donnelly & Karsten, 2017; Wood et al., 2007). Impacts of integrity errors on treatment outcomes are affected by at least three variables: error type, error frequency, and intervention type. Unfortunately, multiple types of integrity errors may occur more frequently than do correct implementer responses during implementation of reinforcement-based interventions (Arkoosh et al., 2007; Wood et al., 2007). This finding is especially troubling because certain types and frequencies of errors are likely to reduce intervention efficacy (Brand et al., 2019).

Although multiple types of errors can occur during reinforcement-based interventions, two commonly evaluated errors include failure to implement a treatment component (termed an

omission error) and implementation of a procedure not specified by the protocol (termed a *commission error*). Omission and commission errors can occur in isolation or combination, with the latter often termed a *blended error*. Commission and blended errors are more likely to be detrimental to reinforcement-based interventions than are omission errors (St. Peter Pipkin et al., 2010). Unfortunately, blended errors are common in clinical practice (Arkoosh et al., 2007; Carroll et al., 2013). Thus, understanding how blended errors affect common interventions is important for maximizing positive client outcomes.

Blended errors may be particularly problematic when they occur frequently. Previous studies have attempted to establish a minimum necessary integrity value (i.e., maximum allowable frequency of errors) for multiple reinforcement-based interventions. These studies have expressed integrity as a percentage based on the ratio of correct implementer responses to total opportunities to implement (the sum of correct and incorrect implementer responses). Of the limited number of reinforcement-based interventions evaluated at reduced-integrity, most are no longer efficacious when integrity falls below 50% with commission or blended errors (e.g., Brand et al., 2019; St. Peter Pipkin et al., 2010).

Although most interventions are negatively affected when integrity drops below 50% with blended errors, the extent to which intervention effects are degraded appears to differ across reinforcement-based interventions. Therefore, the following sections describe effects of integrity errors during three reinforcement-based interventions: ratio-based differential reinforcement of alternative behavior (DRA), interval-based DRA, and fixed-time schedules (FT). Each section will follow a similar pattern in which the interventions, effects of integrity errors during the interventions, and research on response allocation during simultaneously available reinforcement schedules (i.e., concurrent schedules) is described.

Differential Reinforcement of Alternative Behavior

Effects of integrity failures have been most frequently evaluated in the context of an intervention called differential reinforcement of alternative behavior (DRA; Leon et al., 2014;

St. Peter Pipkin et al., 2010; Vollmer et al., 1999). Initial application of DRA often involves reinforcing each instance of appropriate alternative behavior (i.e., a fixed-ratio [FR] 1 schedule) and withholding reinforcers following problematic behavior (i.e., extinction; Hagopian et al., 2013; Vollmer & Iwata, 1992). For example, a teacher may use DRA with a student by providing attention each time the student raises his hand and withholding attention when the student talks out. DRA is a well-established, effective intervention when implemented with high levels of integrity. However, naturalistic and empirical evaluations suggest that DRA is less efficacious when implemented with frequent errors (Leon et al., 2014; St. Peter Pipkin et al., 2010; Vollmer et al., 1999; Wood et al., 2007).

In a series of experiments, St. Peter Pipkin et al. (2010) evaluated impacts of omission errors in isolation, commission errors in isolation, and blended errors at five levels of integrity (100%, 80%, 60%, 40%, and 20%) during ratio-based DRA. During the initial experiments, college students served as participants, and treatment effects were evaluated for arbitrary responses (clicking circles on a computer screen) that served as analogs to problematic and alternative behavior. The researchers classified omission errors as failing to deliver points following alternative responding, commission errors as delivering a point following the response targeted for decrease (i.e., target responding), and blended errors as a combination of commission and omission errors. These errors occurred probabilistically (according to random-ratio [RR] schedules). Each participant experienced baseline, full-integrity DRA, and several levels of reduced-integrity DRA in a within-subject reversal design. During baseline, participants received a point for each target response (i.e., an FR 1 schedule) and no points following alternative responding (i.e., extinction), which resulted in frequent target responding. During the full-integrity (100%) DRA condition, a point followed each instance of alternative behavior; points were no longer delivered for target behavior, which resulted in suppressed target responding. Participants also experienced a series of reduced-integrity conditions: 80%, 60%, 40%, and 20% integrity. The probability of integrity errors differed at each integrity level

and varied partially due to participant response allocation. For example, during 80% integrity, one in five target responses on average resulted in point delivery (.2 probability of reinforcement).

St. Peter Pipkin et al. (2010) found that the error type and error frequency differentially impacted DRA. Specifically, they found that DRA remained efficacious when omission errors occurred, regardless of error frequency. This finding contrasted with effects of other error types; DRA was no longer effective when integrity fell below 60% with either commission or blended errors. In sum, omission errors in isolation were unlikely to decrease the effectiveness of DRA, but commission or blended errors were likely to decrease the effectiveness of DRA when they occurred frequently. After completing the experiments with college students, St. Peter Pipkin et al. (Experiments 2 and 3) replicated the results with two individuals with disabilities who engaged in problematic behavior in schools.

The negative impacts of commission and blended errors during DRA are logical in the context of research on response allocation when two or more reinforcement schedules are simultaneously in operation (i.e., concurrent schedules). In general, research on concurrent schedules demonstrates that responding will be allocated toward the reinforcement schedule(s) that result in the highest reinforcement rate (Herrnstein, 1961; Borrero et al., 2010; St. Peter Pipkin et al., 2010). Consider again the study by St. Peter Pipkin et al., in which omission errors were evaluated in isolation. This arrangement of concurrent schedules resulted in continued extinction for target responding (no reinforcers) and varying reinforcement rates for appropriate behavior. Thus, participants could only access reinforcers by engaging in alternative behavior, regardless of the programmed integrity level. In contrast, the concurrent schedules during commission and blended error phases resulted in reinforcers being available for both target and alternative responding. When the reinforcement schedule for the target response resulted in higher reinforcement rates than the reinforcement schedule for the alternative response (which occurred when integrity fell below 60%), responding was allocated toward the target response.

Thus, the extent to which DRA remains efficacious when integrity errors occur may partially depend on the programmed reinforcement rate and frequency of errors.

To date, all empirical evaluations of effects of integrity errors during DRA have maintained alternative behavior on an FR-1 schedule (Vollmer et al., 1999; St. Peter Pipkin et al., 2010; Leon et al., 2014). Although common during DRA, an FR 1 is by no means the only schedule that is used to reinforce alternative behavior, particularly in clinical treatment. Indeed, reinforcing alternative behavior each time it occurs may not be feasible for caregivers. Thus, analyzing impacts of integrity errors with schedules that do not include reinforcer delivery following each alternative response (i.e., intermittent schedules) is extremely important.

One alternative schedule of particular interest is the fixed-interval (FI) schedule. Fixed-interval schedules are intermittent reinforcement schedules that involve providing a reinforcer following the first response that occurs after some specified interval elapses. Fixed-interval schedules are likely to occur in real-world settings (Critchfield et al., 2003; St. Peter Pipkin & Vollmer, 2009) and are sometimes used when reinforcing academic responding or on-task behavior (Henderson et al., 1986). For example, a teacher may use DRA with a student by providing attention following the first hand raise after 5-min elapses and withholding attention when the student talks out. Although FI schedules can be used to reinforce alternative behavior during DRA (see Hanley et al., 2001 or Henderson et al., 1986 for examples), there has yet to be an evaluation of effects of treatment-integrity errors during DRA when fixed-interval schedules, rather than fixed-ratio schedules, are used to maintain appropriate behavior.

As with fixed-ratio schedules, however, research on concurrent schedules may provide insight into how integrity errors would impact DRA when alternative behavior is maintained by an interval schedule. For example, Rider (1981) evaluated response allocation of five rats when concurrent FI and variable-ratio (VR) schedules were in place. This arrangement is analogous to DRA with an FI schedule for alternative behavior and frequent probabilistic commission errors for problem behavior. The rats consistently responded more on the lever associated with the

ratio schedule, even when the experimenters programmed the reinforcement rate to be higher on the interval schedule. These results suggest that ratio-based commission errors or blended errors may be particularly detrimental during DRA when alternative responding is maintained on an interval schedule. If commission errors are highly detrimental to interval-based DRA, intervention effects may be lost even when the overall integrity level is nominally acceptable (e.g., 80% integrity, an integrity level that has historically been considered adequate implementation; Fiske et al., 2008; Hagermoser Sanetti & Kratochwill, 2008).

Broadly, 80% has been considered an adequate integrity level in behavior-analytic research. In a review and commentary of treatment integrity in published research, Hagermoser Sanetti and Kratochwill (2008) described “high levels of treatment integrity” as 80% or greater and posited that articles may be unlikely to be published if integrity was lower than 80%. Additionally, staff-training studies have historically considered 80% integrity as mastery-level implementation (e.g., Dart et al., 2017; Fiske et al., 2008). Although contemporary behavior-analytic literature has suggested that 80% may be adequate, it is possible that implementation of certain treatments, such as interval-based DRA, at 80% integrity would result in degraded treatment outcomes.

Noncontingent Reinforcement (Fixed-Time) Interventions

Nominally acceptable integrity values may also be particularly detrimental to interventions based on noncontingent reinforcement. Most often, noncontingent reinforcement involves delivering reinforcers independently of responding after a period of time (according to a fixed-time [FT] schedule) and withholding reinforcers following problematic behavior (Hagopian et al., 2013). For example, a teacher may deliver a reinforcer every 5 minutes, regardless of hand raising or talking out, and provide no reinforcers contingent on talking out. Interventions that use FT schedules and extinction (hereafter referred to as FT) are thought to be easy to implement because they do not require the monitoring of behavior (Vollmer et al., 1993), but this assertion has not been verified empirically. Additionally, no published evaluation has analyzed naturalistic

integrity during FT or effects of reduced-integrity implementation on problematic behavior, despite the common use of FT as an intervention for problematic behavior (Brand et al., 2019).

In a series of pilot studies, our laboratory has evaluated impacts of probabilistic commission, omission, and blended errors during FT schedules in applied and laboratory arrangements. Although results were idiosyncratic across participants, FT implemented with 80% integrity with commission errors sometimes resulted in increased rates of target responding. Thus, FT may need to be implemented with higher than 80% integrity to be consistently effective.

Although there are no published evaluations of integrity failures during FT, several studies have evaluated concurrent FT FR schedules (akin to commission errors) with thinned FT schedules (akin to omission errors). Generally, these studies suggest that concurrent FT FR schedules result in increased target responding relative to FT schedules in isolation. For example, Wallace et al. (2012) simultaneously introduced response-dependent reinforcers for problematic behavior and decreased the frequency of reinforcer delivery during FT. In this experiment, the FT schedule was thinned from continuous access to an FT 15-s schedule (during which response-independent reinforcers were delivered every 15 s regardless of target responding), and response-dependent reinforcers were delivered following each instance of problem behavior. This arrangement resulted in reemergence of frequent problematic behavior, suggesting that frequent blended integrity errors would be detrimental to therapeutic outcomes during FT.

Statement of Purpose

In summary, the existing body of research on integrity errors during reinforcement-based interventions has many notable gaps. First, impacts of integrity errors on some reinforcement-based interventions (e.g., interval-based DRA and FT) have been insufficiently evaluated. Evaluating these impacts may inform the boundary conditions for efficacious use of reinforcement-based interventions. Second, there is a lack of direct comparisons of multiple reinforcement-based interventions implemented with reduced integrity. It is possible that participant variables (e.g., response rate during full-integrity treatment) play a role in the later efficacy of reduced-integrity

interventions. Thus, direct comparisons of multiple interventions within a single experiment with each participant as their own control will help us explore this area and inform treatment recommendations. Third, the nominally acceptable integrity level of 80% should be carefully evaluated in the context of multiple reinforcement-based interventions. The present studies aimed to address some of the aforementioned gaps in the treatment integrity literature. Experiment 1 directly compared impacts of 80% blended errors on ratio-based DRA and FT. Experiment 2 directly compared and ratio- and interval-based DRA.

General Method

Participants

We recruited ten college students between the ages of 19 and 20 through a university website (SONA) that lists research opportunities for extra credit in Psychology courses (see Table 1 for additional demographic information). Two participants' data were excluded from analysis because they withdrew before completing the experiment. Participants received a fixed amount of extra credit for each hour of participation, regardless of their responding or point earnings. Participants completed a phone or email screening for COVID-19 symptoms within 48 hours of their appointment and wore a mask for the duration of the experiment.

Apparatus and Setting

Participants sat alone in a laboratory room at a computer with a monitor and mouse. The experimenter arranged contingencies using a custom Visual Basic Program prior to the participant's arrival. The computer screen displayed a message box that said, "Press Ok to Start." and an "Ok" button. After the participant clicked the "Ok" button to start the session, the computer screen continuously displayed a black circle and a white circle (each 38.1 mm diameter) that moved across the computer screen at 20 mm/s. The response targeted for decrease (analogous to problem behavior) was clicking on the black circle (hereafter referred to as the *target response*). The alternative response (analogous to appropriate behavior) was clicking the

white circle (hereafter referred to as the *alternative response*). Due to a programming error, both circles were black during the second replication of all conditions for P7. The circles traveled in a line across the screen and changed direction when they reached the screen's edge. A point counter in the bottom left of the screen displayed cumulative points in the session and briefly flashed orange and incremented by one each time a point was earned. The background color of the screen varied across conditions (red or blue during Experiment 1 and green or purple during Experiment 2). The program recorded every mouse click and output the data into a file that included a location (e.g., on one of the circles or the background) and timestamp in milliseconds for each click. The program also recorded a timestamp and triggering event for each point delivery.

The experimenter monitored the participant through a one-way mirror during the experiment. If the participant left the work area before the session was over or put their head down, the experimenter entered the room and said, "Please attend to the experiment." This occurred during Experiment 2 when P8 put their head down at minute 52 (denoted by an asterisk on the bottom graph of Figure 1).

Procedure

Each participant completed one appointment comprised of two 60-min sessions with a 15-min break between sessions. Each session included 60 alternating 1-min components associated with one of the experimental conditions.

Participants completed an informed-consent process before the experiment. During the consenting process, the experimenter reviewed the consent document and experiment using a script. The script described the purpose of the study as follows: "The purpose of this study is to learn more about how rewards affect behavior." The script also included only a vague description of procedures: "You will use only the mouse to earn as many points as possible during the experiment."

After participants consented, they were asked to silence phones and all other electronic

devices (e.g., watches) and place them out of reach from the computer station. The experimenter then oriented the participant to the computer station and said:

This is where you will be working. You will work for two rounds of 60 minutes, with a 15-minute break between each round. Please be sure to keep your mask on for the duration of the experiment. Remember to use only the mouse to earn as many points as possible. The instructions you see on the screen are the only ones that you will get; it is up to you to figure out how to earn points. What you need to do may change during the experiment. When a thank you message appears, please come and get me to let me know it is your break time. Good luck!!

After 1 hr elapsed, the screen turned white and a thank you message appeared. The experimenter then permitted the participant to take a break outside of the laboratory with any of their personal belongings. The experimenter programmed the computer for the next session while the participant was out of the laboratory and asked the participant to return once the computer was programmed. When the participant returned, the experimenter said, “Thanks for coming back! As before, I need you to silence or turn off all electronic devices including watches, cell phones, and tablets, and leave them on the table. This session is also an hour. Please keep your mask on for the duration of the experiment. Once the thank you message appears, please knock on the door to let me know that you are done.”

After the session ended, the experimenter asked participants to complete a brief demographics form and debriefed participants. The experimenter debriefed differently depending on the study in which the participant was enrolled (see Appendix A for the script for Experiment 1 and Appendix B for the script for Experiment 2).

Integrity Errors

Across both experiments, blended errors were programmed probabilistically. Each time a triggering event occurred (i.e., target response, criterion for alternative reinforcer delivery met) during reduced-integrity conditions, the computer program generated a

random number between 0 and 1 and compared the number to the programmed probability. If the number was less than the programmed probability for errors (.2) when an alternative reinforcer should have been delivered, the alternative reinforcer was withheld (i.e., an omission error occurred). Similarly, if the number was less than the programmed probability (.2) when a target response occurred, a reinforcer was delivered (i.e., a commission error occurred).

Because errors were programmed probabilistically, obtained integrity was likely to vary from programmed integrity. Therefore, we calculated obtained commission, omission, and blended integrity per replication of each reduced-integrity condition across experiments. To calculate commission integrity, the number of target responses that did not result in point delivery (i.e., correct withholding of the reinforcer) was divided by the total number of target responses, and the quotient multiplied by 100. Commission integrity was only calculated if a target response occurred at least one time during the condition. Omission integrity was calculated by dividing the number of alternative responses that met criteria for reinforcement (ratio- and interval-based DRA) or seconds (FT) that resulted in point delivery by the number of alternative responses that met criteria for reinforcement (ratio- and interval-based DRA) or 59 (FT) and the quotient multiplied by 100. To calculate blended integrity, the sum of all appropriately withheld reinforcers and appropriately delivered reinforcers was divided by the sum of all appropriately withheld reinforcers, commission errors, appropriately delivered reinforcers, and omission errors.

Experiment 1

Experimental Design. The purpose of Experiment 1 was to compare effects of 80% integrity with blended errors during DRA and FT. Effects of 80% integrity were evaluated during DRA and FT using a reversal (ABAC-ABAC) design with an embedded multielement design. In this experiment, A refers to the baseline, B refers to the full-integrity treatment phase,

and C refers to the reduced-integrity treatment phase. Within each phase, two distinct background colors signaled which condition (DRA, signaled by a blue screen, or FT, signaled by a red screen) was in effect (see Table 2 for descriptions of screen color and reinforcement schedules per condition). Background colors and associated conditions alternated randomly without replacement each minute.

Baseline. The purposes of baseline were to establish a target response that could later be treated, and to determine if responding occurred at differential rates in the presence of different background colors. The only difference between conditions during baseline was screen color. In both conditions, each target response resulted in a point delivery (an FR-1 schedule). No points were awarded for the alternative response or clicking on the background. Participants experienced a total of 40 min of baseline across four, 10-min baseline phases. Each baseline phase consisted of 5 min of each condition.

Full-Integrity Treatment. The purpose of the full-integrity phase was to evaluate the effectiveness of DRA and FT when implemented as designed. During the full-integrity treatment phase, full-integrity DRA and FT alternated. During the full-integrity DRA condition, each alternative response resulted in point delivery (FR-1 schedule). No points were awarded for target responses or clicking on the background. During the full-integrity FT condition, a point was delivered every second, regardless of participant responding (an FT 1-s schedule). No points were awarded for responses, alternative responses, or clicks on the background during the full-integrity FT condition. Participants experienced a total of 40 min of full-integrity treatment across two, 20-min full-integrity phases. Each full-integrity treatment phase consisted of 10 min of each condition.

Reduced-Integrity Treatment. Treatment-integrity errors occurred probabilistically during the reduced-integrity treatment phase. During the reduced-integrity DRA condition, approximately 80% of alternative responses resulted in point delivery according to a random-ratio (RR) 1.25 schedule (i.e., 80% omission integrity). Additionally, approximately 80% of

target responses had no programmed consequence (i.e., 80% commission integrity); the other 20% of target responses on average resulted in point delivery, according to an RR 5 (similar to St. Peter Pipkin et al., 2010).

During the reduced-integrity FT condition, points were delivered regardless of participant responding approximately every 1.25 s on average (according to a random-time 1.25-s schedule). The contingencies for target responding were identical to the reduced-integrity DRA condition; approximately 20% of the clicks on the black circle produced a point (according to a RR 5 schedule). Thus, the reinforcement schedules in this condition omitted 20% of FT reinforcers (80% omission integrity) and provided a point following 20% of instances of target behavior (80% commission integrity). Participants experienced a total of 40 min of reduced-integrity treatment across two, 20-min reduced-integrity treatment phases. Each reduced-integrity treatment phase consisted of 10 min of each condition.

Experiment 1 Results and Discussion

Figure 1 shows target responses per min for participants in Experiment 1. For all graphs in Figure 1, target responses per min is on the y-axis, component is on the x-axis, the black circles correspond with DRA conditions, and the white circles correspond with FT conditions. Each solid phase-change line indicates phase changes. Phase labels of “BL” denote baseline, 100% denote full-integrity, and 80% denote reduced-integrity.

For all participants, baseline contingencies resulted in frequent target responding regardless of color changes, and DRA was effective at suppressing target responding across participants and integrity levels. Target responding during FT conditions was more variable across participants. Three response patterns occurred with respect to responding during FT: inconsistent suppression across both 100% and FT 80% conditions (P1), inconsistent suppression during FT 80% (P2), and consistent suppression across 100% and 80% conditions (P3 and P4; see Table 3 for condition means per participant).

For P1 (top graph of Figure 1), DRA was consistently effective, but FT was not

consistently effective regardless of integrity level. Target responding occurred 29 times more frequently during FT 100% than DRA 100%. Similarly, target responding occurred 16 times more frequently during FT 80% than DRA 80%. Target responding occurred at similar rates during FT 100% and FT 80%. Thus, in contrast to DRA (which suppressed target behavior at both 100% and 80% integrity), FT did not consistently suppress target responding during either 100% or 80% conditions. Thus, treatment type had a larger impact on rates of target responding than did integrity level for P1.

For P2 (top-middle graph of Figure 1), DRA was consistently effective across integrity levels but FT was not consistently effective during 80% integrity. Target responding was nearly eliminated during 100% integrity DRA and FT conditions. During DRA 80%, target responding continued to occur infrequently. When FT was implemented with 80% integrity, high rates of target responding persisted for at least one component presentation before decreasing to near-zero rates. Overall, target responding occurred 17 times more frequently during FT 80% than FT 100%. Thus, reducing the integrity level seemed to slow the transition to effective treatment for FT, but not for DRA.

For P3 and P4 (bottom graphs in Figure 1), DRA and FT were both consistently effective regardless of integrity level. For both participants, target responding was nearly eliminated during 100% and 80% integrity DRA and FT conditions. There was no differentiation in rates of target responding across the two conditions in any phase of the experiment.

Tau U effect sizes and P values were also calculated to compare DRA 100% vs FT 100%, DRA 100% vs DRA 80%, FT 100% vs FT 80%, and DRA 80% vs FT 80% for each participant. These values were calculated using the online Tau U calculator (Vannest et al., 2016) and are displayed in Table 4. Tau U calculations suggested that there were significant differences and medium effect sizes between FT 100% and DRA 100% and FT 80% and DRA 80% for P1 and a significant difference but small effect size between DRA 100% and DRA 80% for P2. All other Tau U comparisons yielded small effect sizes and insignificant P values.

Tau U calculations suggested there was a significant difference between DRA 100% and DRA 80% for P2. There was variability in target responding for P2 during the latter DRA 80% condition (range = 0-44). However, the effect size was small and suppression relative to baseline was very similar for DRA 100% and DRA 80% (97% suppression and 96% suppression, respectively).

Results of Experiment 1 were consistent within-subject across replications, but showed considerable intersubject variability. Our results suggest that it is difficult to predict response patterns during FT. The lack of intersubject replication in FT conditions replicates a recent study conducted by our research team that evaluated impacts of commission errors on FT for children with problem behavior in a school setting (Jones & St. Peter, 2020). Additionally, the lack of intersubject replication in FT conditions replicates a study conducted by Borrero et al. (2011) that evaluated impacts of response-dependent reinforcers (i.e., commission errors) during FT on an arbitrary task with individuals with psychiatric disorders in a residential facility (Borrero et al., 2011). Collectively, these studies suggest that further research exploring possible sources of the failures to replicate across participants (e.g., historical variables, patterns of experienced errors) are needed.

In light of intersubject failures to replicate, Borrero et al. (2011) hypothesized that inconsistent suppression during full-integrity treatments could be indicative of decreased treatment efficacy during reduced-integrity treatments. This was the case for P1 who engaged in variable rates of target responding during both full- and reduced-integrity FT. However, response rates during full-integrity FT were very similar for P2, P3, and P4 although P2 engaged in target responding during reduced-integrity FT. Thus, response rate during full-integrity may be a lackluster predictive variable. Baseline response rates were also not predictive of degraded outcomes during reduced-integrity FT; P4 had the highest mean response rate during baseline and target responding was consistently suppressed during reduced-integrity FT. Thus, future

researchers should identify other participant variables that may serve as predictors (e.g., sensitivity to reinforcers).

Figure 2 shows alternative responses per min for all participants of Experiment 1. Similar to Figure 1, all graphs in Figure 2 have alternative responses per min on the y-axis, component is on the x-axis, the black circles correspond with DRA conditions, and the white circles correspond with FT conditions. Each solid phase-change line indicates phase changes. Phase labels of “BL” represent baseline, 100% represents full-integrity, and 80% represents reduced-integrity. For all participants, alternative responding only consistently occurred during DRA conditions during full- and reduced-integrity phases.

Although no participants engaged in consistent alternative responding during FT conditions, P1 engaged in variable rates of alternative responding ($M = 74.35$, range = 0-230) during FT 100% conditions. This responding occurred despite a lack of response-dependent reinforcers during the FT condition. It is possible that the rapid alternation between DRA and FT accounts for some of this variability. It is also possible that inadvertent contiguity between alternative responding and response-independent reinforcers occurred.

Recall that responding may be influenced by obtained reinforcement rates. Therefore, average points per min from each source was calculated per participant for the reduced-integrity conditions. Results are shown in Table 5. As a notable function of the contingencies, overall reinforcement rate was much higher in the DRA condition relative to the FT condition. Thus, it remains possible that the high reinforcement rate, rather than the DRA contingency itself, resulted in more consistent response suppression during reduced-integrity phases. Future studies could yoke FT reinforcer delivery to obtained reinforcer delivery during DRA to control this variable. However, it is also worth noting that overall reinforcement rates were not predictive of intersubject variability during reduced-integrity FT; P3 and P4 obtained the fewest and most reinforcers from alternative behavior (respectively), but outcomes were consistent across the two participants.

Figure 3 displays obtained integrity per condition and participant. The left column shows obtained integrity values during DRA conditions; the right column shows obtained integrity values during FT conditions. The top two graphs show percent blended integrity per replication of DRA and FT conditions (y-axis) for each participant (x-axis) for the first (black bars) and second (grey bars) replication of each condition. The bottom two graphs show percent integrity separated by error type (y-axis) for each participant (x-axis) for commission integrity (bars with black data points) and omission integrity (bars with white data points). Each data point represents percent integrity per error type for each replication of reduced-integrity phases. Asterisks show that commission integrity could not be calculated due to lack of target responding. Obtained blended integrity values hovered near 80% for each condition and participant ($M = 80.91\%$; range = 79.11%-82.24%). Similarly, omission integrity hovered near 80% (FT $M = 81.52$, range = 78.90-82.28; DRA $M = 80.17$, range = 79.11-81.61). Commission integrity was more variable (FT $M = 86.125$, range = 80.97-100; DRA $M = 85.42$, range = 77.78-96.07).

In summary, the current investigation suggests that ratio-based DRA is more consistently effective than FT when implemented with 80% and possibly 100% integrity. Perhaps it is the timed-based nature of FT, rather than the lack of dependency, that contributes to variable outcomes. This question could be addressed by evaluating errors on a schedule that retains the dependency of DRA but includes a time-based aspect of FT. Interval schedules may be able to disentangle these two variables.

Experiment 2

Experimental Design. The purpose of Experiment 2 was to compare effects of 80% blended errors during DRA with an interval or ratio schedule for appropriate behavior. Effects of 80% integrity with blended errors during two variants of DRA (either interval or ratio schedule for appropriate behavior) were evaluated in an ABAC-ABAC design with an embedded multielement design. As with Experiment 1, A refers to baseline, B refers to full-integrity

treatment, and C refers to reduced-integrity treatment. Within each phase, two distinct background colors signaled which condition. DRA with a ratio schedule (hereafter referred to as FR) was signaled by a green screen. DRA with an interval schedule (hereafter referred to as FI) was signaled by a purple screen (see Table 6 for descriptions of screen color and reinforcement schedules per condition). Each component presentation was 1-min and alternated randomly without replacement.

Baseline. Similar to Experiment 1, the purpose of baseline was to establish a response to treat and to determine if differential responding occurred in the presence of different background colors. The color of the screen alternated between green and purple in 1-min components. In both conditions, each target response resulted in a point (an FR-1 schedule). No points were awarded for alternative responding or clicking on the background. Participants experienced a total of 40 min of baseline across four 10-min baseline phases. Each phase consisted of 5 min of each condition.

Full-Integrity Treatment. Similar to Experiment 1, the purpose of full-integrity treatment conditions was to assess the effectiveness of both treatments when implemented as designed. During the full-integrity treatment phase, the reinforcement schedule for alternative responding alternated between an FR 1 and a fixed-interval (FI) 1-s schedule across components. When the screen was green, each alternative response resulted in a point (an FR-1 schedule; identical to DRA in Experiment 1). When the screen was purple, the first alternative response after 1s elapsed resulted in a point (a fixed-interval [FI] 1-s schedule). Target responses or clicks on the screen background did not produce points during the full-integrity conditions. Participants experienced a total of 40 min of full-integrity treatment across two 20-min phases. Each phase consisted of 10 min of each condition.

Reduced-Integrity Treatment. During the reduced-integrity treatment phases, the points were delivered probabilistically. The FR 80% condition was identical to Experiment 1; approximately 80% of alternative responses resulted in point delivery according to an RR 1.25

schedule. Additionally, approximately 80% of target responses had no programmed consequence; the other 20% of target responses on average resulted in point delivery, according to an RR 5.

During the FI 80% condition, approximately 80% of alternative responses that would have met reinforcement criteria (a click after 1 s elapsed) resulted in point delivery according to a random-interval (RI) 1.25 s schedule. As in the FR 80% condition, about 20% of target responses resulted in point delivery (according to a RR 5 schedule). Participants experienced a total of 40 min of reduced-integrity treatment across two 20-min phases. Each phase consisted of 10 min of each condition.

Experiment 2 Results and Discussion

Figure 4 shows target responses per min for participants in Experiment 2. For all graphs in Figure 4, target responses per min is on the y-axis, component is on the x-axis, the black circles correspond with FR conditions, and the white circles correspond with FI conditions. Each solid phase-change line indicates phase changes. Phase labels of “BL” denote baseline, 100% denote full-integrity, and 80% denote reduced-integrity.

For all participants, baseline contingencies resulted in frequent target responding regardless of color changes, and full-integrity ratio-based and interval-based DRA nearly eliminated rates of target responding. Although differences in rates were often small, all participants engaged in more frequent mean rates of target responding during FI 80% conditions than FI 100% and FR 80% conditions (see Table 7 for condition means and ranges).

The top two graphs of Figure 4 show P5 and P6’s target responses per min. For both participants, FI 80% conditions resulted in more frequent target responding than FR 80% conditions. FI 80% conditions resulted in 10 times and 5 times more target responding than FR 80% conditions, respectively. However, target responding was transient for both participants and therapeutic effects, similar to those observed during FR 80%, were recovered by the end of the second FI 80% replication.

The bottom-middle graph shows target responses per min for P7. Similar to P5 and P6, FI 80% conditions resulted in 6 times more target responding than FR 80% conditions. Unlike P5 and P6, target responding during FI 80% was elevated above target responding during FR 80% throughout most components of both reduced-integrity replications. Additionally, response patterns during the second replication of all conditions were similar to the first replication although both response options were the same color.

The bottom graph shows target responses per min for P8. During the second baseline phase, responding was on a decreasing trend before the 80% condition began. During the 80% condition, the experimenter observed the participant put their head down and asked the participant to attend to the experiment at min 52 (denoted by an asterisk on the graph). Before and after this reminder, the participant engaged in more frequent target responding during FI 80% than FR 80%. The second FI 80% condition also resulted in more frequent target responding than the FR 80% condition. Overall, target responding occurred 4 times more frequently during FI 80% than FR 80%. Similar to P7, target response rates during reduced-integrity FI were more consistently elevated above target response rates during reduced-integrity FR.

Tau U effect sizes and P values were also calculated to compare FI 100% vs FR 100%, FI 100% vs FI 80%, FR 100% vs FR 80%, and FR 80% vs FI 80% per participant using the online Tau U calculator (Vannest et al., 2016) and are displayed in Table 8. Tau U calculations suggested that there were significant differences and medium effect sizes between FR 80% and FI 80% for all participants. There was also a significant difference and medium or large effect size between FI 100% and FI 80% for P6, P7, and P8. There was also a medium effect size and significant difference between FR 100% and FR 80% for P8. All other Tau U comparisons yielded small effect sizes and insignificant P values.

In summary, all participants engaged in more frequent target responding during FI 80% conditions relative to FR 80% conditions, although these differences were sometimes small (P5

and P6). Recall that during Experiment 1, the two treatments that were compared differed on multiple dimensions including dependency and the time-based delivery of reinforcers. When controlling for dependency during Experiment 2, results were less variable across participants. Thus, it is possible that the lack of dependency during FT influenced the variability in Experiment 1 rather than the time-based nature of the reinforcement schedules. Additionally, response rate during full-integrity conditions and baseline conditions was not indicative of degraded outcomes during reduced-integrity conditions. All participants engaged in near 0-levels during full-integrity treatments and P8 and P7 had the lowest and highest response rate during baseline (respectively) and engaged in consistent target responding during interval-based DRA.

Figure 5 shows alternative responses per min for all participants of Experiment 2. Similar to Figure 4, all graphs in Figure 5 have alternative responses per min on the y-axis, component on the x-axis, the black circles correspond with DRA conditions, and the white circles correspond with FT conditions. Each solid phase-change line indicates phase changes. Phase labels of “BL” denote baseline, 100% denote full-integrity, and 80% denote reduced-integrity.

All participants engaged in less frequent alternative responding during interval-based DRA than ratio-based DRA. This is logical given that participants could maximize reinforcement rate by engaging in high rates of clicking during ratio-based DRA, but could only earn a point a second for clicking, regardless of how frequently they clicked, during interval-based DRA. The differentiation between alternative response rates during interval- and ratio-based DRA may suggest that participants were sensitive to the schedule differences between the alternative reinforcement schedules.

Reinforcement rates (shown in Table 9) were higher during ratio-based DRA than interval-based DRA. These results resemble those of Experiment 1, in which ratio-based DRA resulted in more frequent reinforcer delivery than FT. Thus, it remains possible that the high reinforcement rate, rather than inclusion of a ratio contingency, resulted in more consistent

response suppression during reduced-integrity phases. Future studies could yoke interval-based DRA reinforcer delivery to obtained reinforcer delivery during ratio-based DRA to control for reinforcement rate.

Figure 6 displays obtained integrity per condition and participant and is formatted identically to Figure 3. Obtained blended integrity values hovered near 80% for each condition and participant ($M = 80.19\%$; range = 78.02%-81.72%). Commission and omission integrity varied partially depending on participant response rate (Commission; FI $M = 81.46$, range = 78.61-90.38; FR $M = 80.34$, range = 72.90-89.47; Omission; FI $M = 79.88$, range = 74.94-82.64; FR $M = 79.33$, range = 72.29-81.24).

General Discussion

The current experiments compared the efficacy of common reinforcement-based treatments (DRA and FT) when implemented with 80% blended integrity. In both experiments, we systematically replicated effects obtained by St. Peter Pipkin et al. (2010) by demonstrating that ratio-based DRA was consistently effective when implemented with 80% integrity. We extended existing research by demonstrating that even fairly infrequent blended errors (80% integrity) sometimes negatively affected FT and interval-based DRA. However, these negative effects were inconsistent across participants, and did not seem to be clearly related to response rates in baseline or full-integrity treatment. Additionally, the obtained overall reinforcement rates did not seem to be predictive of for which participants integrity failures would be detrimental. A crucial next step in this line of study will be to identify factors responsible for the intersubject variability observed in both experiments.

Although our findings were not consistent across participants, our results add to the treatment-integrity literature by providing direct comparisons of multiple reduced-integrity reinforcement-based interventions. Overall, ratio-based DRA was more robust in the face of integrity errors than was interval-based DRA or FT. However, the variables leading to these

differences are unclear because schedule type and reinforcement rate were confounded across interventions. We did not control for reinforcement rate because these interventions are likely to result in different reinforcement rates when used to reduce problem behavior. However, future research should parse apart causal variables by controlling for reinforcement rate while manipulating intervention type.

The present experiments suggest that programmed reinforcement schedules associated with treatments may be a more important factor than intervention type in the analysis of integrity errors during reinforcement-based interventions. Previous researchers have suggested that DRA is a generally robust treatment in the face of integrity errors (St. Peter Pipkin et al., 2010; Vollmer et al., 1999). However, our results suggest that DRA may be robust specifically when an FR-1 schedule is used to maintain alternative behavior. Even though interval-based DRA resulted in frequent reinforcer delivery (every 1.25 s on average), it was inconsistently effective. In practice, schedules that include such frequent reinforcer delivery are often thinned to more manageable values. It is possible that reducing reinforcement rates may exacerbate negative impacts of blended errors. Future research should evaluate effects of integrity errors throughout the schedule-thinning process (i.e., when alternative reinforcement rate systematically varies to enhance the feasibility of treatments).

The present experiments underscore that 80% integrity, a value that has been accepted as nominally sufficient (Fiske et al., 2008; Hagermoser Sanetti & Kratochwill, 2008), may be inappropriate for some interventions. For example, 80% integrity appears sufficient for DRA with an FR-1 schedule maintaining alternative behavior but insufficient for interval-based DRA or FT. Rather than setting arbitrary levels of adequate integrity, it may be more beneficial for practitioners and researchers to consider adequate levels of integrity in the context specific interventions. Future research should identify minimal necessary integrity levels per intervention.

Although relative reinforcement rate is an important variable, our results suggest that

consideration of reinforcement rate across response options is potentially insufficient in predicting suppression of target responding during reduced-integrity interventions. In fact, research on responding during concurrent schedules suggests that responding will be allocated proportionally across the schedules based on reinforcement rate (Herrnstein, 1961; Borrero et al., 2010). In some cases, responding on multiple response options may increase overall reinforcement rate. However, this did not appear to be the case during Experiment 2. Recall that P5 and P6 rarely responded on both response options, yet reinforcement rates were similar to P7, who frequently responded on both response options. Additionally, P8 who also frequently responded on both response options, had the lowest overall reinforcement rate during reduced-integrity interval-based DRA conditions.

Recall that errors occurred probabilistically in the current evaluation. The probabilistic nature of errors has at least three implications. First, the use of probabilistic errors differs from previous research examining effects of concurrent ratio and FT schedules on problem behavior (e.g., Wallace et al., 2012). Recall that Wallace et al., implemented FT with a continuous reinforcement schedule for target behavior (i.e., an FR-1 schedule), which resulted in increased target responding for all participants. In contrast, we implemented FT with an intermittent, probabilistic schedule and our participants did not uniformly engage in target responding. Of particular note are P3 and P4, who rarely engaged in target responding during reduced-integrity FT. It is possible that our participants engaged in less frequent target responding than did those in previous research because our study included probabilistic errors instead of a continuous reinforcement schedule (e.g., Wallace et al., 2012).

Second, the use of probabilistic errors may have increased the variability of target responding within and across participants. We chose to use probabilistic errors to remain consistent with previous research on integrity errors during reinforcement-based interventions (e.g., St. Peter Pipkin et al., 2010). However, because errors were random, there were frequently runs of target responses in which local commission integrity deviated from the programmed

probability. At the level of the individual component, commission integrity ranged between 33%-100%; reinforcers were more likely to be delivered following target behavior during some components than other components. This variation may have contributed to the variability of target responding during reduced-integrity conditions.

Third, the use of probabilistic errors may be a limitation to the generality of our study. Errors occurring during the treatment of problem behavior may not occur randomly. Rather, errors may be more likely when the implementer is balancing multiple tasks (e.g., teaching math facts to a class and implementing a behavior intervention plan) or after a burst of problem behavior. Future research should work to identify how errors are happening naturalistically and evaluate errors that mimic those that occur naturalistically (e.g., Carroll et al., 2013; Foreman et al., 2021).

In addition to considering how integrity errors are programmed, researchers should consider how integrity errors are calculated and reported. Because of an interaction between probabilistic errors and response rate, participants experienced different frequencies of omission and commission errors, although all participants experienced global integrity close to 80% blended integrity. In the evaluation conducted by St. Peter Pipkin et al. (2010), commission errors alone and blended errors were equally likely to result in increases in target responding. This finding suggests that commission errors are especially detrimental during reinforcement-based interventions implemented with blended errors. Thus, it is possible that reduced-integrity conditions were effective for some participants because there was insufficient exposure to commission errors. For example, P3 experienced only omission errors during FT phases that nominally included blended errors. To permit interpretation of findings in the context of obtained integrity failures, researchers should report integrity values for all components of their intervention rather than reporting integrity solely as an aggregate value. Reporting only an aggregate would conceal the differences in error types and possible sources of behavioral control

(Cook et al., 2015).

The likelihood of participants contacting commission errors and subsequent degraded therapeutic outcomes during reduced-integrity conditions may be impacted by condition sequence (St. Peter Pipkin et al., 2010; Colón & Ahearn, 2019). Treatments implemented with reduced integrity are more likely to result in degraded outcomes when following baseline than full-integrity phases (St. Peter Pipkin et al., 2010). In the current experiments, baseline preceded reduced-integrity phases and the component order in each phase was randomized. It is possible that the sequence of components impacted results. For example, response rates were differentiated between DRA and FT in four of the eight reduced-integrity phases in Exp 1; the first component was FT in all four of those phases, but in only one of the four phases without differentiation. However, because component presentations were randomized, we cannot make conclusive statements regarding sequence effects. Future studies should deliberately manipulate sequence to identify impacts of various sequences on responding when treatment integrity is degraded.

The use of rapid alternation of conditions may have inadvertently enhanced the effectiveness of reduced-integrity FT and interval-based DRA. Alternating between low-integrity, historically ineffective treatments and full-integrity, historically effective treatments, increases the effectiveness of the former. For example, Colón and Ahearn (2019) compared a treatment at 25% integrity and 100% integrity and found that the treatment reduced behavior at 100% integrity but not at 25% integrity. They then rapidly alternated between 25% integrity and 100% integrity conditions and found that therapeutic effects maintained during both 25% integrity and 100% integrity conditions. Thus, it is possible that if ratio-based DRA was not frequently presented during reduced-integrity phases more consistent target responding would have occurred during reduced-integrity FT or interval-based DRA. However, no research has identified increased therapeutic outcomes when alternating two different treatment types with

reduced integrity. Future researchers could evaluate this possibility by initially exposing participants to various reduced-integrity interventions in isolation using a reversal design and then exposing participants to those same interventions in a multielement design (similar to Colón & Ahearn, 2019).

The rapid alternation of conditions also may have impacted our results by capturing transition states. For example, target responding was on a slight increasing trend during last three components of reduced-integrity ratio-based DRA for P8. The second-by-second data show that all target responses for the first two of these minutes occurred within 5 s of the transition from the interval component. This may suggest that target responses occurred because of the transition and may not have occurred if rapid alternation did not occur. Future research utilizing multielement designs should carefully assess responding during transitions between components. Future researchers could also consider procedural variations that may mitigate impacts of rapid alternation of conditions (e.g., increasing the time between components [Barlow & Hayes, 1979; McGonigle et al., 1987] or the use of a reversal design).

Recall that the screen color varied depending on treatment (e.g., DRA or FT) but remained the same for baseline, full-integrity, and reduced-integrity within each treatment. In clinical practice, reduced-integrity conditions may be associated with salient stimulus changes (e.g., a regular teacher may signal full-integrity implementation and a substitute teacher may signal reduced-integrity implementation). In fact, stimulus changes alone can be evocative of challenging behavior (Podlesnik et al., 2017). Thus, a stimulus change paired with decreased treatment integrity may be even more likely to result in increased, and possibly sustained, rates of problem behavior than are suggested by the findings of the current studies. Thus, future researchers should assess impacts of integrity errors with accompanied stimulus changes from full-integrity treatments.

Laboratory investigations appear to be a promising avenue for future researchers in pursuit

of understanding effects of reduced-integrity interventions because they allow for enhanced experimental control to isolate impacts of integrity errors. As variables become uncontrolled in naturalistic implementation, these variables may exert stronger effects than integrity errors or they may interact with and exacerbate effects of integrity errors. For example, if a child inconsistently receives medication before experimental sessions, variability induced by inconsistent medication complicates evaluation of impacts of integrity errors. Additionally, laboratory investigations allow for identification of important variables that can later be manipulated with individuals with problem behavior (see St. Peter Pipkin et al., 2010 for an example). Most importantly, there is evidence that findings from laboratory investigations replicate when conducted with clinical populations who engage in problem behavior. Our study replicates the laboratory and applied findings of St. Peter Pipkin et al. (2010) and of our pilot study with children with problem behavior in a school setting (Jones & St. Peter, 2020).

In summary, the present experiments highlight the need for further analysis of reduced-integrity reinforcement-based interventions. Because descriptive evaluations suggest that integrity errors are commonplace in practice (e.g., Arkoosh et al., 2007), it is essential to understand the effects of errors on reinforcement-based interventions and to identify predictive variables for degraded therapeutic outcomes during reduced-integrity interventions. While additional research is being conducted, behavior-analytic professionals should consider treatments in the context of concurrent schedules and avoid arbitrary standards for acceptable levels of integrity. If it is not possible to analyze concurrent schedules, practitioners should consider ensuring frequent reinforcement delivery outside of the context of problem behavior to bolster the effectiveness of their treatments.

References

- Arkoosh, M. K., Derby, K. M., Wacker, D. P., Berg, W., McLaughlin, T. F., & Barretto, A. (2007). A descriptive evaluation of long-term treatment integrity. *Behavior Modification, 31*(6), 880-895. <https://doi.org/10.1177/0145445507302254>
- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*(2), 199-210. <https://doi.org/10.1901/jaba.1979.12-199>
- Borrero, J. C., Bartels-Meints, J. A., Sy, J. R., & Francisco, M. T. (2011). Fixed-time schedule effects in combination with response-dependent schedules. *Journal of Applied Behavior Analysis, 44*(1), 163-167. <https://doi.org/10.1901/jaba.2011.44-163>
- Borrero, C. S., Vollmer, T. R., Borrero, J. C., Bourret, J. C., Sloman, K. N., Samaha, A. L., & Dallery, J. (2010). Concurrent reinforcement schedules for problem behavior and appropriate behavior: Experimental applications of the matching law. *Journal of the Experimental Analysis of Behavior, 93*(3), 455-469. <https://doi.org/10.1901/jeab.2010.93-455>
- Brand, D., Henley, A. J., DiGennaro Reed, F. D., Gray, E., & Crabbs, B. (2019). A review of published studies involving parametric manipulations of treatment integrity. *Journal of Behavioral Education, 28*(1), 1-26. <https://doi.org/10.1007/s10864-018-09311-8>
- Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis, 46*(2), 379-394. <https://doi.org/10.1002/jaba.49>
- Colón, C. L., & Ahearn, W. H. (2019). An analysis of treatment integrity of response interruption and redirection. *Journal of Applied Behavior Analysis, 52*(2), 337-354. <https://doi.org/10.1002/jaba.537>
- Cook, J. E., Subramaniam, S., Brunson, L.Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-

- trial training. *Behavior Analysis in Practice*, 8(1), 37-47.
<https://doi.org/10.1007/s40617-014-0039-7>
- Critchfield, T. S., Haley, S. R., Sabo, B., Colbert, J., & Macropoulis, G. (2003). A half century of scalloping in the work habits of the United States Congress. *Journal of Applied Behavior Analysis*, 36(4), 465-486. <https://doi.org/10.1901/jaba.2003.36-465>
- Dart, E. H., Radley, K. C., Furlow, C. M., & Murphy, A. N. (2017). Using behavioral skills training to teach high school students to implement discrete trial training. *Behavior Analysis Research and Practice*, 17(3), 237-249. <http://dx.doi.org/10.1037/bar0000075>
- Donnelly, M. G., & Karsten, A. M. (2017). Effects of programmed teaching errors on acquisition and durability of self-care skills. *Journal of Applied Behavior Analysis*, 50(3), 511-528.
<https://doi.org/10.1002/jaba.390>
- Fiske, K. E. (2008). Treatment integrity of school-based behavior analytic interventions: a review of the research. *Behavior Analysis in Practice*, 1(2), 19-25.
<https://doi.org/10.1007/BF03391724>
- Foreman, A. P., St. Peter, C. C., Mesches, G. A., Robinson, N., & Romano, L. M. (2020). Treatment integrity failures during timeout from play. *Behavior Modification*, 145445520935392. Advance online publication.
<https://doi.org/10.1177/0145445520935392>
- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy*, 4(1), 95-114. <https://doi.org/10.1037/h0100835>
- Hagopian, L. P., Dozier, C. L., Rooker, G. W., & Jones, B. A. (2013). Assessment and treatment of severe problem behavior. *APA Handbook of Behavior Analysis: Vol 2. Translating Principles into Practice*, Ed. G. J. Madden. <https://doi.10.1037/13938-014>
- Hanley, G. P., Iwata, B. A., & Thompson, R. H. (2001). Reinforcement schedule thinning

- following treatment with functional communication training. *Journal of Applied Behavior Analysis*, 34(1), 17-38. <https://doi.org/10.1901/jaba.2001-34-17>
- Henderson, H. S., Jenson, W. R., & Erken, N. F. (1986). Using variable interval schedules to improve on-task behavior in the classroom. *Education and Treatment of Children*, 9(3), 250-263. <http://www.jstor.org/stable/42898977>
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4(3), 267-272. <https://doi.org/10.1901/jeab.1961.4-267>
- Holden, B., & Gitlesen, J. P. (2006). A total population study of challenging behavior in the county of Hedmark, Norway: Prevalence, and risk markers. *Research in Developmental Disabilities*, 27(X), 456-465. <https://doi.org/10.1016/j.ridd.2005.06.001>
- Jones, S. H., & St. Peter, C. C. (2020). *Effects of commission and omission errors on the effectiveness of noncontingent reinforcement* [Manuscript in preparation].
- Leon, Y., Wilder, D. A., Majdalany, L., Myers, K., & Saini, V. (2014). Errors of omission and commission during alternative reinforcement of compliance: The effects of varying levels of treatment integrity. *Journal of Behavioral Education*, 23(1), 19-33. <https://doi.org/10.1007/s10864-013-9181-5>
- Lowe, K., Allen, D., Jones, E., Brophy, S., Moore, K., & James, W. (2007). Challenging behaviours: Prevalence and topographies. *Journal of Intellectual Disability Research*, 51(X), 625-636. <https://doi.org/10.1111/j.1365-2788.2006.00948.x>
- McGonigle, J. J., Rojahn, J., Dixon, J., & Strain, P. S. (1987). Multiple treatment interference in the alternating treatments design as a function of the intercomponent interval length. *Journal of Applied Behavior Analysis*, 20(2), 171-178. <https://doi.org/10.1901/jaba.1987.20-171>
- Podlesnik, C. A., Kelley, M. E., Jimenez-Gomez, C., & Bouton, M. E. (2017). Renewed behavior produced by context change and its implications for treatment maintenance:

- A review. *Journal of Applied Behavior Analysis*, 50(3), 675-697.
<https://doi.org/10.1002/jaba.400>
- Rider, D. P. (1981). Concurrent fixed-interval variable-ratio schedules and the matching relation. *Journal of the Experimental Analysis of Behavior*, 36(3), 317-328.
<https://doi.org/10.1901/jeab.1981.36-317>
- Shuler, N. J., & St. Peter, C. C. (2018). *Effects of integrity failures during FT* [Manuscript in preparation].
- Soke, G. N., Rosenberg, S. A., Hamman, R. F., Fingerlin, T., Robinson, C., Carpenter, L., Giarelli, E., Lee, Li-Ching, Wiggins, L. D., Durkin, M. S., & DiGuseppi, C. (2016). Brief report: Prevalence of self-injurious behaviors among children with autism spectrum disorder – A population-based study. *Journal of Autism and Developmental Disorders*, 46(11), 3607-3614. <https://doi.org/10.1007/s10803-016-2879-1>
- St. Peter Pipkin, C., & Vollmer, T. R. (2009). Applied implications of reinforcement history effects. *Journal of Applied Behavior Analysis*, 42(1), 83-103.
<https://doi.org/10.1901/jaba.2009.42-83>
- St. Peter Pipkin, C. 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. <https://doi.org/10.1901/jaba.2010.43-47>
- Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single Case Research: web based calculators for SCR analysis. (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University. Retrieved Sunday 11th October 2020. Available from singlecaseresearch.org
- Vollmer, T. R., & Iwata, B. A. (1992) Differential reinforcement as treatment for behavior disorders: Procedural and functional variations. *Research in Developmental Disabilities*, 13(X), 393-417. <https://doi.org/10.1016/0891->

4222(92)90013-V

- Vollmer, T. R., Iwata, B. A., Zarcone, J.R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. *Journal of Applied Behavior Analysis, 26*(1), 9-21. <https://doi.org/10.1901/jaba.1993.26-9>
- 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*(1), 9-23. <https://doi.org/10.1901/jaba.1999.32-9>
- Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice, 1*(2), 4-11. <https://doi.org/10.1007/BF03391722>
- Wallace, M. D., Iwata, B. A., Hanley, G. P., Thompson, R. H., & Roscoe, E. M. (2012). Noncontingent reinforcement: A further examination of schedule effects during treatment. *Journal of Applied Behavior Analysis, 45*(4), 709-719. <https://doi.org/10.1901/jaba.2012.45-709>
- Wood, B. K., Umbreit, j., Liaupsin, C. J., & Gresham, F. M. (2007). A treatment integrity analysis of function-based intervention. *Education and Treatment of Children, 30*(4), 105-120. <https://doi.org/10.1353/etc.2007.0035>

Appendix A
Script for Debriefing Participants following Experiment 1

We use experiments like this one to understand how reward-based treatments work when people don't always implement them perfectly. Initially, we provided rewards for a "problem behavior" (clicking on the black circle), and no rewards for an "alternative behavior" (clicking on the white circle). We then "treated" your problem behavior when you no longer earned points for clicking on the black circle. We tried this treatment in two different ways. One, we gave you a point each time you clicked on the white circle, and the other we gave you a point after a second passed even if you didn't click anywhere. Both of these treatments are commonly used to reduce problem behavior. We then "goofed up" the treatments; sometimes we "forgot" to give you a point when you should have gotten one, and sometimes we gave you a point for engaging in "problem behavior" (that is, when you clicked on the black circles). We want to see which of the two treatments is most resistant to these kinds of "goof ups". We will use this information to make better treatment recommendations for people who are seeking behavioral treatments.

Appendix B
Script for Debriefing Participants following Experiment 2

We use experiments like this one to understand how reward-based treatments work when people don't always implement them perfectly. Initially, we provided rewards for a "problem behavior" (clicking on the black circle), and no rewards for an "alternative behavior" (clicking on the white circle). We then "treated" your problem behavior when you no longer earned points for clicking on the black circle. We tried this treatment in two different ways. One, we gave you a point each time you clicked on the white circle, and the other we gave you a point after a second passed and you clicked on the circle. We then "goofed up" the treatments; sometimes we "forgot" to give you a point when you should have gotten one, and sometimes we gave you a point for engaging in "problem behavior" (that is, when you clicked on the black circle). We want to see which of the two treatments is most resistant to these kinds of "goof ups". We will use this information to make better treatment recommendations for people who are seeking behavioral treatments.

Table 1*Demographic Information per Participant in Experiments 1 and 2*

Participant	Exp. #	Age	Sex	Race/Ethnicity	Color Deficiency?
P1	1	19	Female	White	No
P2	1	19	Female	Latine	No
P3	1	20	Female	Black	No
P4	1	20	Male	White	No
P5	2	19	Female	Black	No
P6	2	19	Female	White	No
P7	2	20	Male	White	No
P8	2	19	Female	White	No

Table 2*Condition Components during Experiment 1*

	DRA Baseline	FT Baseline	DRA 100%	FT 100%	DRA 80%	FT 80%
Screen Color	Blue	Red	Blue	Red	Blue	Red
Target Response Schedule	FR 1	FR 1	EXT	EXT	RR 5	RR 5
Alternative Response Schedule	EXT	EXT	FR 1	EXT	RR 1.25	EXT
Response-Independent Schedule	N/A	N/A	N/A	FT 1''	N/A	RT 1.25''

Table 3*Average Response Rate and Range per Participant, Condition, and Response Option during Experiment 1*

	P1		P2		P3		P4	
	Target	Alternative	Target	Alternative	Target	Alternative	Target	Alternative
DRA Baseline	239.40 (185-280)	4.30 (0-53)	244.25 (211-270)	1.10 (0-5)	177.45 (114-230)	2.00 (0-9)	276.55 (222-320)	2.00 (0-6)
FT Baseline	239.60 (202-281)	1.85 (0-9)	249.25 (214-275)	.95 (0-4)	173.15 (76-220)	1.70 (0-4)	271.15 (228-324)	1.70 (0-18)
DRA 100%	1.95 (0-14)	239.45 (187-271)	6.65 (0-100)	238.90 (117-281)	1.20 (0-9)	190.15 (114-231)	1.60 (0-15)	304.45 (255-333)
FT 100%	56.20 (0-179)	74.35 (0-230)	5.00 (0-29)	6.10 (0-14)	.35 (0-3)	2.90 (0-13)	3.20 (0-31)	5.55 (0-28)
DRA 80%	3.65 (0-27)	212.00 (132- 263)	10.50 (0-44)	236.05 (175-345)	1.00 (0-12)	189.40 (119-249)	1.95 (0-18)	286 (206-316)
FT 80%	57.60 (0-173)	21.40 (1-104)	82.55 (0-295)	4.85 (0-16)	0.00	2.00 (0-7)	0.95 (0-12)	4.80 (0-14)

Table 4*Tau U per participant during Experiment 1*

	P1		P2		P3		P4	
	Tau U	P Value	Tau U	P Value	Tau U	P Value	Tau U	P Value
DRA 100 vs FT 100	.8400	.0000*	.2925	.1136	-.2025	.2733	.0125	.9461
DRA 100 vs DRA 80	.1275	.4903	.3850	.0373*	-.1475	.4249	-.1725	.3507
FT 100 vs FT 80	.0225	.9031	.1650	.3720	-.1500	.4171	-.2000	.2793
DRA 80 vs FT 80	.7300	.0001*	.1050	.5700	-.2000	.2793	-.0525	.7764

Table 5

Average Points per Minute delivered per Reduced-Integrity Condition, Response Option, and Participant during Experiment 1

	P1		P2		P3		P4	
	DRA 80	FT 80	DRA 80	FT 80	DRA 80	FT 80	DRA 80	FT 80
Target Response	0.50	10.85	1.53	14.67	0.15	0	0.30	0.10
Alternative Response	168.45	0	189.00	0	152.75	0	223.80	0
Response Independent	0	47.65	0	45.95	0	45.90	0	48.35

Table 6*Condition Components during Experiment 2*

	Ratio Baseline	Interval Baseline	Ratio 100%	Interval 100%	Ratio 80%	Interval 80%
Screen Color	Green	Purple	Green	Purple	Green	Purple
Target Response Schedule	FR 1	FR 1	EXT	EXT	RR 5	RR 5
Alternative Response Schedule	EXT	EXT	FR 1	FI 1''	RR 1.25	RI 1.25''

Table 7*Average Response Rate per Participant, Response Option, and Condition Experiment 2*

	P5		P6		P7		P8	
	Target	Alternative	Target	Alternative	Target	Alternative	Target	Alternative
FI Baseline	259.95 (165-302)	1.55 (0-11)	278.65 (152-337)	4.10 (0-23)	338.10 (218-428)	3.30 (0-19)	178.45 (36-219)	4.80 (0-53)
FR Baseline	276.30 (245-317)	0.60 (0-5)	284.55 (200-363)	4.25 (0-36)	361.85 (236-461)	3.50 (0-33)	175.30 (66-238)	1.40 (0-11)
FI 100%	0.85 (0-5)	75.60 (57-127)	1.25 (0-5)	149.75 (63-227)	3.50 (0-13)	172.60 (43-244)	2.65 (0-15)	155.25 (46-225)
FR 100%	5.20 (0-97)	294.45 (171-327)	3.80 (0-31)	317.15 (267-351)	2.70 (0-33)	352.10 (237-460)	2.75 (0-33)	194.15 (128-242)
FI 80%	9.35 (0-43)	68.15 (43-90)	19.25 (0-158)	116.95 (34-207)	55.00 (5-130)	122.55 (54-189)	53.80 (0-178)	85.05 (0-178)
FR 80%	0.95 (0-19)	264.70 (225-310)	3.95 (0-45)	291.65 (212-364)	9.50 (0-92)	361.05 (54-189)	12.00 (0-66)	159.75 (24-214)

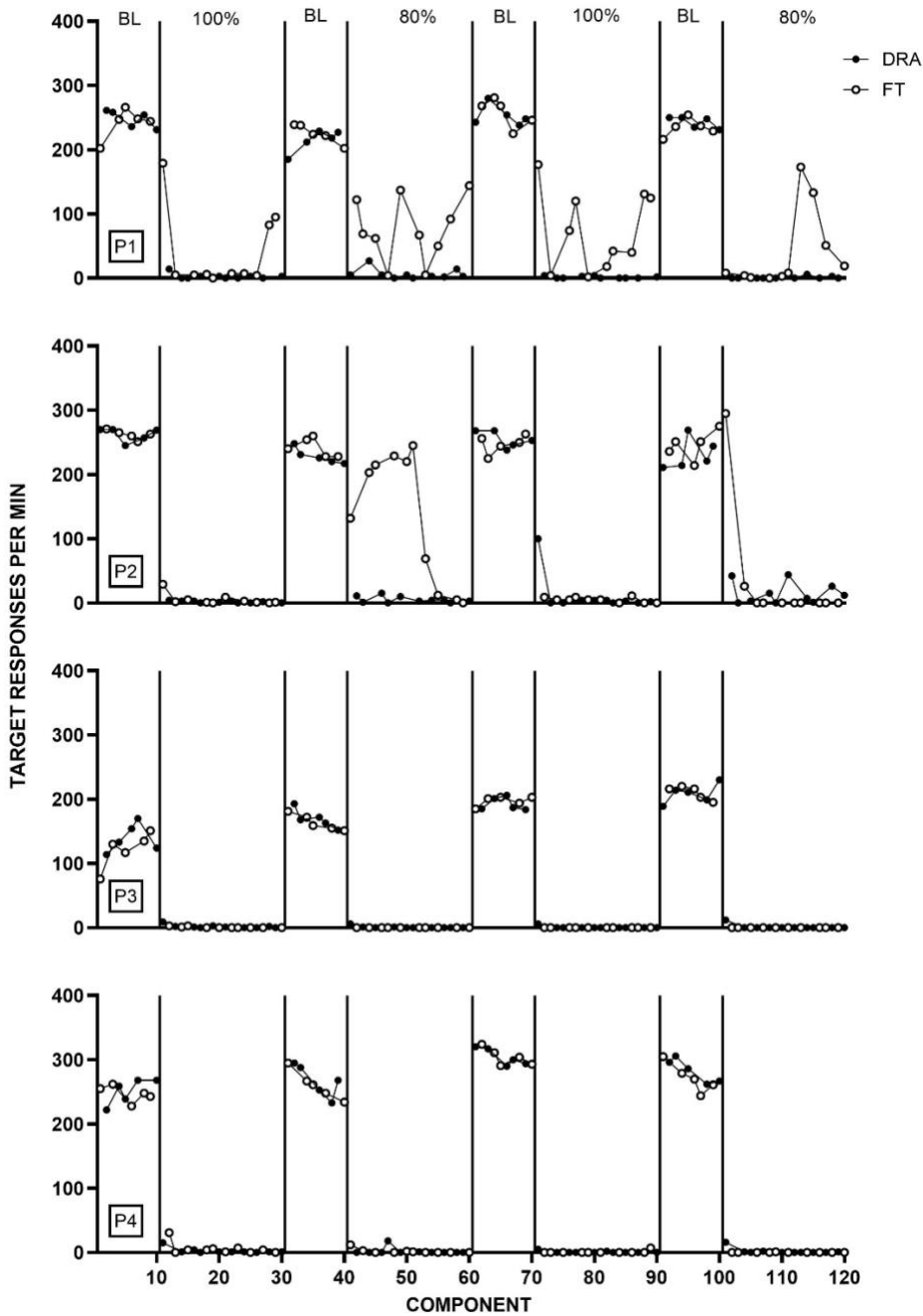
Table 8*Tau U per participant during Experiment 2*

	P5		P6		P7		P8	
	Tau U	P Value	Tau U	P Value	Tau U	P Value	Tau U	P Value
FI 100 vs FR 100	.1250	.4989	.0625	.7353	.3375	.0679	.3425	.0639
FI 100 vs FI 80	.2974	.1124	.3975	.0315*	.9300	.0000*	.8500	.0000*
FR 100 vs FR 80	-.0500	.7868	-0.215	.2448	.2050	.2674	.5950	.0013*
FR 80 vs FI 80	.4211	.0246*	.4750	.0102*	.8175	.0000*	.6675	.0003*

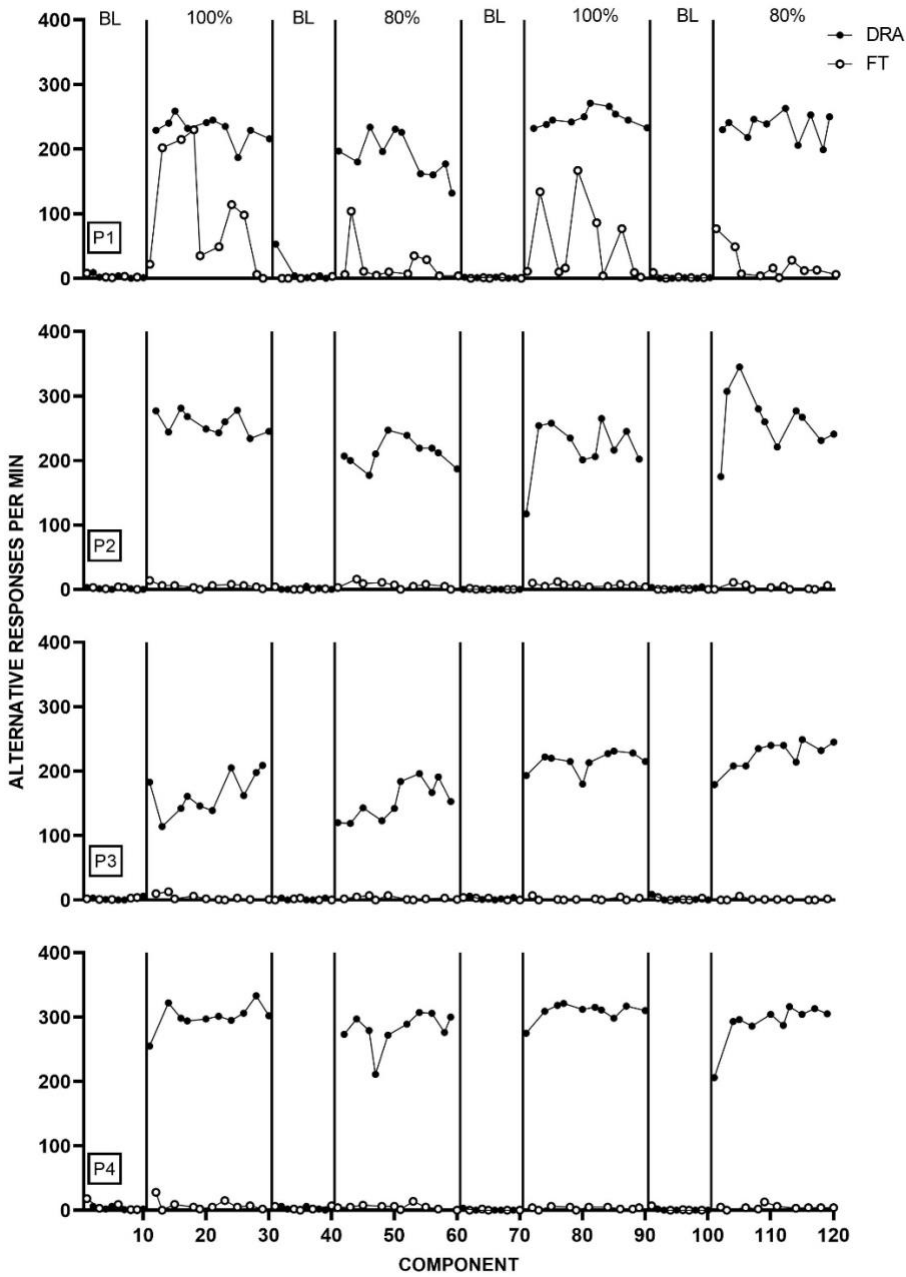
Table 9

Average Points per Minute delivered per Reduced-Integrity Condition, Response Option, and Participant during Experiment 2

	P5		P6		P7		P8	
	FI 80	FR 80	FI 80	FR 80	FI 80	FR 80	FI 80	FR 80
Target Response	1.60	0.10	3.75	0.80	10.45	1.90	11.40	2.85
Alternative Response	41.55	213.00	43.60	233.15	32.85	292.00	20.85	127.05

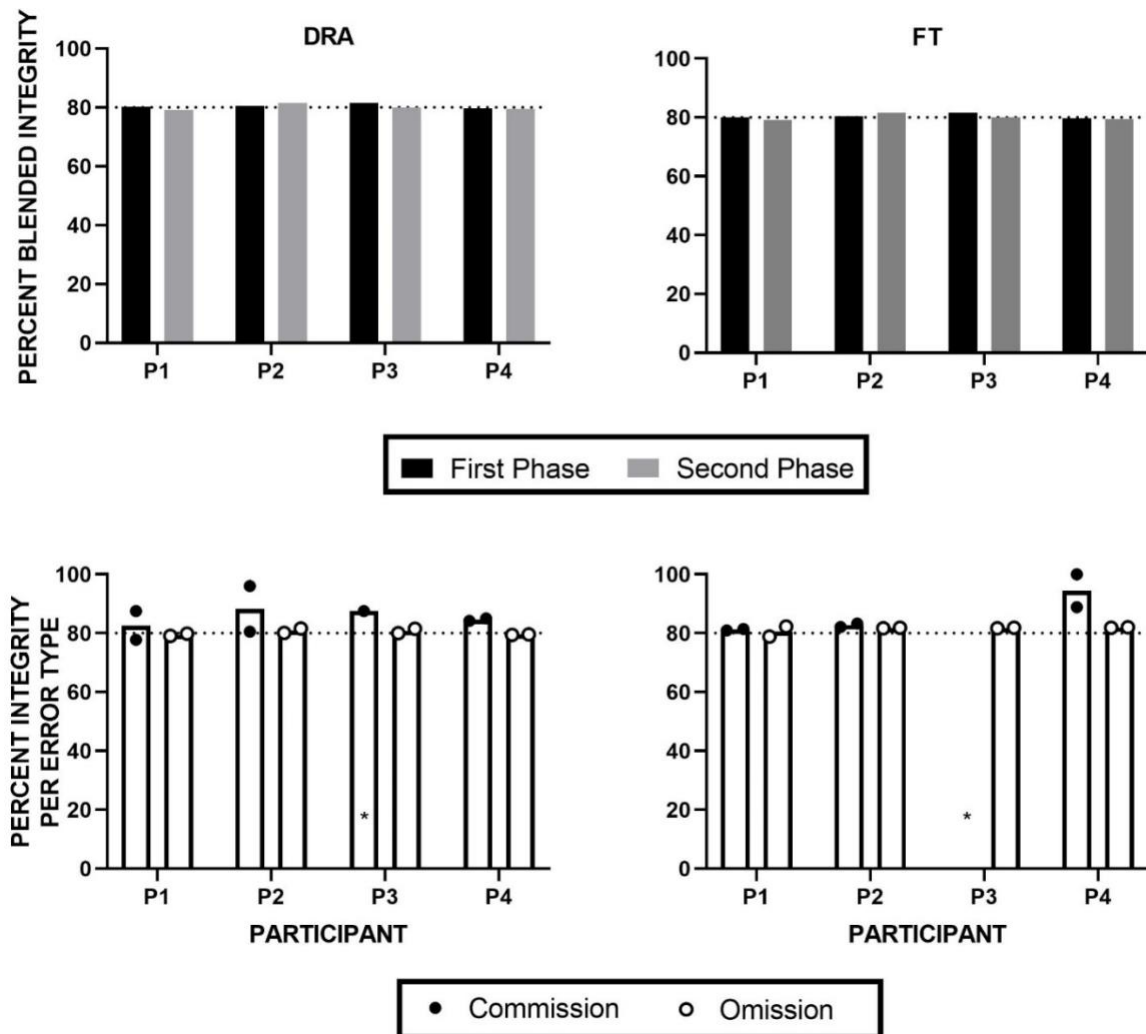
Figure 1*Target Responses per Participant during Experiment 1*

Note. Closed circles indicate DRA conditions; open circles indicate FT conditions. BL indicates baseline phases, 100% indicates full-integrity treatment phases, and 80% indicated reduced-integrity treatment phases.

Figure 2*Alternative Responses per Participant during Experiment 1*

Note. Closed circles indicate DRA conditions; open circles indicate FT conditions. BL indicates baseline phases, 100% indicates full-integrity treatment phases, and 80% indicated reduced-integrity treatment phases.

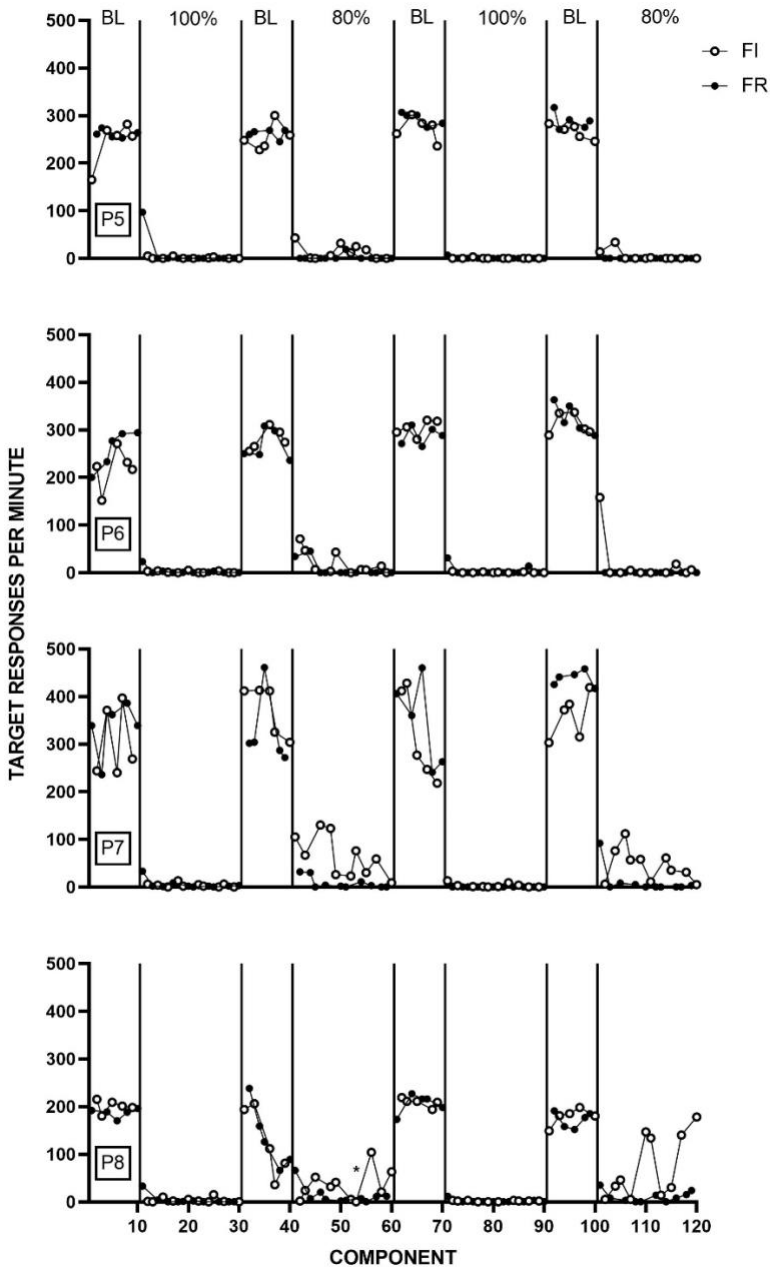
Figure 3
Obtained Integrity Values during Experiment 1



Note. The top graph shows percent blended integrity per DRA replication; black bars correspond with the first DRA phase and the grey bars correspond with the second DRA phase. The horizontal dotted line on both graphs shows the programmed integrity level (80%). The bottom graph shows percent integrity for commission and omission integrity. The black data points correspond with commission integrity; the white data points correspond with omission integrity. The presence of an asterisk in a bar shows that during one condition commission integrity could not be calculated.

Figure 4

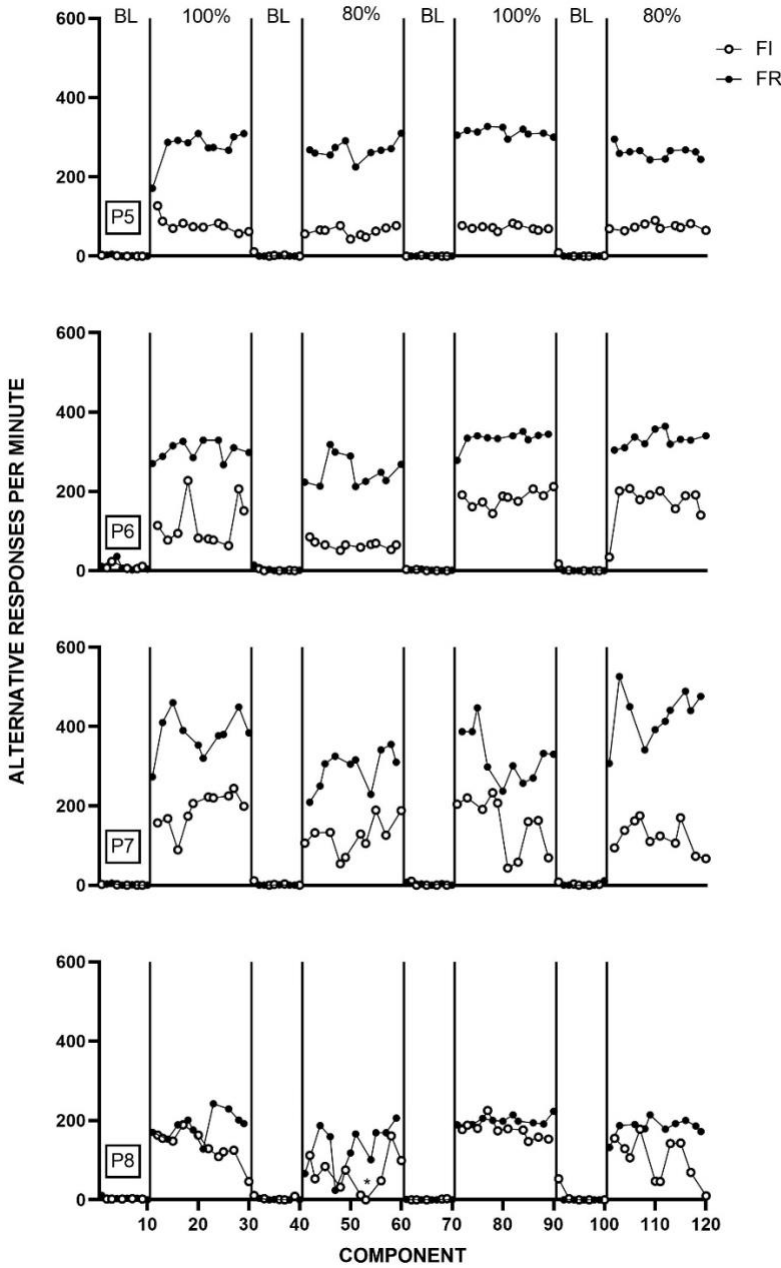
Target Responses per Participant during Experiment 2



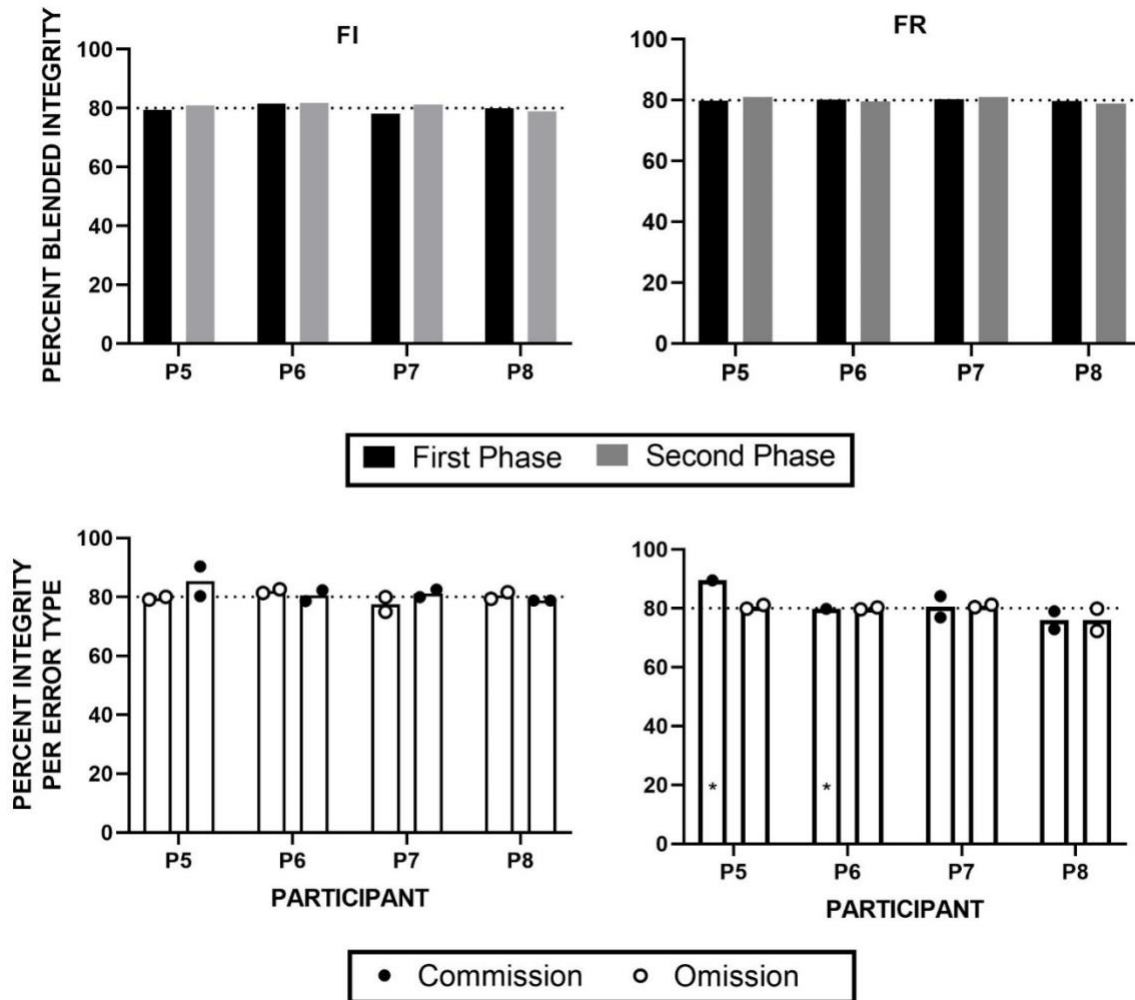
Note. Closed circles indicate DRA conditions; open circles indicate FT conditions. BL indicates baseline phases, 100% indicates full-integrity treatment phases, and 80% indicated reduced-integrity treatment phases.

Figure 5

Alternative Responses per Participant during Experiment 2



Note. Closed circles indicate DRA conditions; open circles indicate FT conditions. BL indicates baseline phases, 100% indicates full-integrity treatment phases, and 80% indicated reduced-integrity treatment phases. Note the y-axis difference relative to Figure 4.

Figure 6*Obtained Integrity during Experiment 2*

Note. The top two graphs show percent blended integrity per reduced-integrity replication; black bars correspond with the first phase and the grey bars correspond with the second phase. The horizontal dotted line on all graphs shows the programmed integrity level (80%). The bottom graphs show percent integrity for commission and omission integrity. The black data points correspond with commission integrity; the white data points correspond with omission integrity. The presence of an asterisk indicates commission integrity could not be calculated during a condition.

