8-2022

A Translational Comparative Investigation of Two Token Reinforcement Systems for ASD Programming

Kate Brockevelt

A Translational Comparative Investigation of Two Token Reinforcement Systems for ASD Programming

by

Kate Brockevelt

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Applied Behavior Analysis

August, 2022

Thesis Paper Committee:
Benjamin Witts, Chairperson
Odessa Luna
Michele Traub

Abstract

Despite the popularity of token reinforcement in applied settings, little research has been conducted on strategies for optimizing its effectiveness in autism service delivery; however, this work is well-suited for a translational exploration. Employing an alternating treatments design, this paper serves as a pilot study, demonstrating how a random ratio (RR) token-exchange schedule can be arranged and compared to a standard fixed ratio (FR) schedule to examine differences in preference, in trial duration, rate, and accuracy, and in challenging behaviors. Two children diagnosed with autism spectrum disorder (ASD) participated during regularly scheduled therapy sessions. During Phase 1, we arranged the random ratio token board (RRTB) to allow participants to draw from a cup of red and blue tokens. Tokens could be exchanged for the backup reinforcer contingent on drawing 10 red tokens or one blue token. This was compared to a static token board (STB), during which the client always had to draw 10 white tokens before exchanging. During this phase, one participant showed a strong fixation on the blue tokens, consistent with a phenomenon known as sign tracking. This led researchers to revise the arrangement of the RRTB so that staff used a random number list instead of different colored tokens to determine when participants could exchange. Following Phase 2, differences in trial rate and challenging behaviors revealed that one participant performed better with the RRTB, despite preferring the STB, while the other participant performed better with the STB, despite preferring the RRTB. Limitations and future directions are discussed.

Keywords: token board, delay reduction theory, DRT, research to practice, exchange production, generalized conditioned reinforcer

Table of Contents

Page

Chapter Page

Appendices.....Page

List of Tables

List of Figures

Chapter 1: Introduction and Literature Review

The first applied study in behavior analysis is often cited as Ayllon and Michael's 1959 paper "The Psychiatric Nurse as a Behavioral Engineer," which appeared in the *Journal of the Experimental Analysis of Behavior*. A keen interest is taken in the choice of the wording *behavioral engineer* and what that could mean for current practitioners in ABA.

An "engineer" is defined by Merriam-Webster (n.d.) as "a person who has scientific training and who designs and builds complicated products, machines, systems, or structures." Behavioral engineering, then, can be defined as using scientific principles to design the environment "to optimize human behavior" (Lattal, 2015, para. 4). According to Homme et al. (1968), behavioral engineering combines the use of contingency management and stimulus control. That is, behavioral engineers manipulate antecedents to evoke or abate a particular behavior and alter consequences to increase or decrease the future probability of that behavior. To truly *engineer* an environment, however, behavior analysts must draw from the basic literature of our field, which has sought to discover the defining principles of behavior. When Ayllon and Michael (1959) published the first applied study of behavioral engineering, they set a standard for the integration of basic and applied work in behavior analysis.

The work of the behavioral engineer can be compared to that of the translational researcher. As Mace and Critchfield (2010) argued, translational research is innovative, requiring "synthesis of basic and applied questions" (p. 293). According to the authors, "pure basic" research in behavior analysis can become too detached from real-world problems, while "pure applied" research is prone to neglecting the fundamental behavioral principles that explain the real-world problems targeted for improvement. Mace and Critchfield encourage a translational

approach to behavior analytic research, while recognizing that most behavior analysts will fall either on the more applied or more basic side. The authors' advice for applied researchers is to incorporate the basic literature and frame everyday problems using the fundamental guiding principles of behavior.

In this sense, then, practitioners as behavioral engineers can seek inspiration for improving and refining their work in the basic literature and then reworking those principles and procedures into existing frameworks. This paper serves as an exemplar of behavioral engineering, in which translational efforts are developed and tested for the purposes of improving practical work by focusing on a widely-accepted mechanism of stimulus function change (i.e., delay reduction theory [DRT]) and its role in a ubiquitous reinforcement delivery system (i.e., the token board). Before discussing how these two issues relate, it is worth first exploring each topic so the reader can better understand the engineering process.

Token Reinforcement System

A token reinforcement system consists of six components: target responses, tokens, back-up reinforcement, a token-production schedule, an exchange-production schedule, and a token-exchange schedule (Hackenberg, 2009; Ivy et al., 2017; Kazdin & Bootzin, 1972). By engaging in select responses, clients earn tokens that can be exchanged for a back-up reinforcer. Three schedules of reinforcement are required to operate a token reinforcement system. The schedule of token production refers to when tokens will be delivered, the schedule of exchange production details when clients will be permitted to exchange tokens for a back-up reinforcer, and the schedule of token exchange refers to the number of tokens required to "purchase" a back-up reinforcer. A standard 10-token token board might consist of an FR 1 schedule of token

production for targets in acquisition, an FR 10 for exchange production, and an FR 10 for token exchange. In this type of token reinforcement system, a client must "fill their board" by earning 10 tokens before being permitted to exchange their tokens for a back-up reinforcer. Additionally, all back-up reinforcers in this token reinforcement system would require the same token-exchange schedule, as opposed to other token economies that use a "reinforcer menu" consisting of back-up reinforcers of various values.

There are multiple benefits of using token economies in applied work. Token economies can be used to maintain responding over extended periods while bridging the delay to back-up reinforcement (Kazdin & Bootzin, 1972). Additionally, when used correctly, tokens function as generalized conditioned reinforcers, limiting the effect of momentary changes in motivating operations (Ivy et al., 2017; Kazdin & Bootzin, 1972). Token economies are also convenient and portable. Tokens can be delivered more discretely than most back-up reinforcers, and token delivery prevents interruption of the target response. Lastly, token economies are highly customizable. Each of the three reinforcement schedules can be individualized for each client and target response (Ivy et al., 2017). Tokens and back-up reinforcers can also be tailored to clients' interests (Kazdin & Bootzin, 1972).

Despite the popularity of token economies in applied settings, Hackenberg (2018) argued that little is known about the basic behavioral mechanisms responsible for the success of token economies due to little integration between applied and basic research on the topic. Although, token economies are often a component of treatment packages employed in applied research, recently, token economies have not been the subject of the research per se. Hackenberg suggests that the success of token economies in real-world settings might actually be the reason for the

decline in applied research on token reinforcement. That is, the wide-spread effectiveness of token economies might signal that additional research is not needed. However, token economies can still be optimized for better performance where responding tends to be weaker. For example, research has shown that across species, response rates increase with temporal proximity to exchange-production opportunities but drop off again shortly after an exchange, similar to the break-and-run pattern observed in simple FR schedules (see Hackenberg, 2018 for a review).

In response to findings like these, basic researchers have been studying strategies for optimizing the conditioned reinforcement effect of tokens in pigeons and rats for decades by altering exchange-production schedules (e.g., Foster et al., 2001; Waddell et al., 1972; Webbe & Malagodi, 1978). For example, Foster et al. (2001) discovered higher rates of responding in pigeons when VR exchange-production schedules were used compared to when FR exchange-production schedules were used. However, when recent translational work has posed these basic questions in human studies, results can be less conclusive (Argueta et al., 2019; Holmes et al., 2022). Argueta and colleagues (2019) systematically replicated Foster et al. (2001), but Argueta and colleagues found no significant differences in rates of responding in a child diagnosed with autism spectrum disorder (ASD); although, they did find preference for the VR schedule. Discrepancies like these that can appear when replicating basic research arrangements with human participants, highlight the importance of conducting translational research to discover effective methods of applying basic principles to our work with our clients.

Holmes et al. (2022) compared the efficacy of FR and time-based exchange-production schedules on four preschool-aged participants' rate of responding to a visual motor coordination task while holding token production schedules constant. Similar to Argueta et al. (2019), Holmes

and colleagues did not find any significant differences in rate of responding under the two schedule arrangements for any of the participants; however, they did see clear, albeit different, preferences for one of the two schedules with two of the four participants. Holmes et al. concluded that decisions regarding exchange-production schedule arrangement should therefore be based on convenience or client preference; however, they also suggested future research in this area should be conducted in applied settings to account for factors such as session durations and response effort requirements more typical of natural clinic arrangements.

Delay Reduction Theory

It is generally accepted that tokens maintain responding because they are conditioned reinforcers. A traditional view of how stimuli develop conditioned reinforcement functions is through the pairing hypothesis, which states that these functions develop when a stimulus precedes a primary or other conditioned reinforcer. Over time then, with enough pairings, the stimuli come to take on reinforcing properties of their own (see Fantino, 2008, for a brief review). However, this view has largely fallen out of favor, being challenged by an opposing view, which claims that it is not simple pairing that leads to conditioned reinforcement effects, but rather it is any pairing that signals a reduction in delay to reinforcement that would have the best chance at developing a reinforcement function. Initially termed the delay reduction hypothesis, after several decades of work the concept graduated to a theory (cf. Fantino, 1969; Fantino et al., 1993).

To help shed light on the difference between the pairing hypothesis and DRT, consider Schuster's (1969, as cited in Fantino, 2008) unpublished dissertation. In Schuster's dissertation, pigeons were required to peck on either of two concurrent chain VI 60 s VI 30 s schedules, in

which completion of one VI 60 s schedule provided access to the terminal VI 30 s schedule after which reinforcement was provided. However, in one of the two chains, a superimposed FR 11 was included that produced brief stimuli during the terminal link of the chain. Fantino described how each of the prominent models of conditioned reinforcement would have predicted the results of Schuster's study. First, if the brief stimulus presentations functioned as conditioned reinforcers, responding to the terminal link of the chain with the superimposed FR 11 should have been higher than responding in the terminal link of the chain without the FR 11. In fact, this is what Schuster found. However, the pairing hypothesis also suggests that preference would be observed for the chain that included the paired brief stimulus presentations if they indeed acquired conditioned reinforcement functions. DRT, though, would lead one to predict that preference would not be observed for the chain with the FR 11 because the paired stimuli more often signaled nonreinforcement, and therefore, did not indicate a reduction in time to reinforcement. Schuster's second finding supported DRT. All pigeons showed a preference for the chain schedule without the superimposed brief stimuli. Considering ratio schedules tend to produce high rates of responding, Fantino (2008) suggested that the cost of responding to the chain schedule with the FR 11 could have caused that chain to become more aversive, leading to preference for the other chain.

Squires (1972, as cited in Fantino 2008) extended the work of Schuster (1969, as cited in Fantino 2008) by replacing the superimposed FR 11 in the terminal link of one chain with a VI 15 s schedule to avoid the high rates of responding in the terminal link. In Experiment A, Squires set up two concurrent chains similar to Schuster's; however, brief stimuli were paired with food reinforcement in the terminal link of the chain with the superimposed VI 15 s schedule.

Experiment B was set up identically to Experiment A, except that the brief stimuli were unpaired with food. Squires discovered the same results for both Experiment A and B, suggesting that whether the stimuli were paired with primary reinforcement or not, choice for the schedule was unaffected. In her next experiment, Squires directly compared responding to paired and unpaired brief stimuli presentations by arranging one of the VI 60 s VI 30 s schedules to include a VI 15 s terminal link schedule with paired stimuli and the other with unpaired stimuli. Squires discovered a small increase in responding during the terminal link of the schedule with the paired brief stimulus presentations compared to the terminal link of the schedule with the unpaired brief stimuli, which reached statistical significance for half of the pigeons. This difference in response rate suggested a small conditioned reinforcement effect, consistent with Schuster's (1969, as cited in Fantino 2008) results; however, once again, all pigeons preferred the chain with the unpaired brief stimulus presentations. Fantino and Romanowich (2007) suggested that the preference data in these studies are more revealing than the response rate data, citing multiple studies that have discovered how additional factors can impact rate of responding. Ultimately, Squires concluded that pairing a stimulus with a primary reinforcer was insufficient for conditioning that stimulus as a reinforcer, as evidenced by the lack of preference for the schedule with the higher rates of paired stimuli.

The work in DRT is often surprisingly similar to the structure of common token boards used in autism service delivery. Take, for example, Gollub's (1958, as cited in Kelleher & Gollub, 1962) work on the role of conditioned reinforcers in chain schedules. Gollub compared pigeons' responding to chain schedules with various numbers of links. He found that when there were only two links in the chain and each was signaled with different colored lights, the pigeons

continued to respond 20 times per minute to the initial link; however, as the number of links increased, responding to the first few links of the chain decreased, possibly because the beginning of the chain signaled a long delay to reinforcement. In the five-link chain, the first two links maintained responding at near zero rates, while responding increased rapidly in the last three links. Gollub also compared these results to response rates on a tandem control schedule, during which, overall response requirements to obtain the primary reinforcer were the same, but each link of the schedule was unsignaled. Instead, one light was present throughout the entire tandem schedule. Gollub noticed higher response rates on the tandem schedule than on the chain schedule. When reviewing Gollub's work, Fantino (2008) suggested that the lights in the chain schedule were not effective conditioned reinforcers because the early lights were more often followed by non-reinforcement (i.e., they signaled a long delay to primary reinforcement). On the other hand, the light that was present throughout the tandem schedule was consistently paired with primary reinforcement, instead of additional lights.

Gollub's (1958, as cited in Kelleher & Gollub, 1962) chain schedules can be compared to the token boards often used in ABA clinics. In a typical token board, the client could earn, for example, 10 tokens for performing certain requirements which are then exchanged for some backup reinforcer. These 10 tokens can be conceptualized as 10 steps in a chain of behaviors. The first few tokens signal a large upcoming workload and a long delay to reinforcement, maintaining responding less effectively than the last few tokens do. While the client is working toward the first few tokens, therapists could observe a break in the chain, which might manifest as escape-maintained challenging behavior or decreased attending or response rate.

A more recent study by Bullock and Hackenberg (2015) systematically replicated Gollub (1958, as cited in Kelleher & Gollub, 1962) and found similar results. Although Fantino used Gollub's findings to suggest that the tokens (i.e., lights) at the beginning of the chain schedule were less effective conditioned reinforcers compared to those at the end of the chain schedule, Bullock and Hackenberg explained that weak early-link responding suggested that tokens were not functioning as conditioned reinforcers at all, but rather tokens were serving a discriminative function (i.e., the first token functioned as an S^{Δ} for the backup reinforcer, while the last token served as an S^{D}).

Using either theoretical explanation of how tokens function to maintain responding, the selection of the tokens used in token economies are of particular concern when we look at work like Gollub's (1958, as cited in Fantino, 2008) and Bullock and Hackenberg's (2015). Consider how a token board would work if tokens consisted of one type, alternated type consistently (e.g., sequenced through a series of colors), or alternated type randomly (e.g., a colored token was randomly selected as opposed to being sequenced). If, for example, a token board used a consistent color pattern, such as a rainbow, the first red token would always signal a long delay to reinforcement (or an S^{Δ}), while the blue and purple tokens would signal reductions in the delay to reinforcement (or S^{D}s). If tokens consisted of one type or alternated type randomly, there would not be a single token that consistently signaled a long or short delay to reinforcement, making tokens more effective at maintaining responding. However, to further increase the effectiveness of tokens, the number of tokens required to exchange for the backup reinforcer must also be randomized to prevent a correlation between the number of earned tokens and the delay to backup reinforcement (see Hackenberg, 2018).

More recent research demonstrates how DRT is likely to be extended and adapted to explain human behavior in novel experimental arrangements. An experimental study on risk-taking examined humans' preference between chain schedules with an FR 50 terminal link and those with a mixed-ratio 1/99 terminal link, in which an FR 1 and an FR 99 were equally likely outcomes (Meyer et al., 2011). The initial links for the two concurrent chains were always the same; however, the authors varied the time spent in the initial links (i.e., FI 1 s, VI 15 s, and VI 30 s) between groups. Results showed that participants tended to prefer the mixed ratio schedules, especially as the time spent in the initial link increased. The authors explained that despite equal average rates of reinforcement between the mixed ratio and equivalent FR schedules, the mixed ratio link inevitably included shorter delays to contact reinforcement compared to the equivalent FR schedule. That is, the possibility of the FR 1 in the mixed ratio link signaled a reduction in time to reinforcement compared to the FR 50 link. The possibility of decreased delay to reinforcement is even more salient (i.e., there is an even greater reduction in delay) after a longer initial link. Therefore, adding this gambling component, wherein participants have a chance to contact immediate reinforcement, might help establish conditioned reinforcement functions by decreasing the predictability of the delay to reinforcement.

To apply this finding to clinical work with individuals with ASD and maintain responding more effectively, behavior analysts can use progressive random ratio schedules of reinforcement to prevent clients from being able to predict when reinforcement will be delivered. Therefore, the purpose of this study is to pilot a method for examining the effects of a token board that decreases the predictability of the delay to back-up reinforcement on participants' rate

of responding, rate of acquisition, and rate of challenging behavior during discrete trial training (DTT) blocks.

Chapter 2: Method

Participants and Settings

Participants were two children who had been diagnosed with ASD and were receiving in-home applied behavior analysis (ABA) services from a local autism service provider.

Colin is a 10-year-old White male who communicates mostly with sign language and gestures and often engages in work avoidance and property destruction. Therapy sessions were 3.5 hours in duration, typically 6-8 times per week, and they took place in Colin's upstairs living room and basement sensory room. The living room was furnished with a table, two stools, two chairs, a couch, and a coffee table. Colin's common reinforcers included playing with landscaping equipment (e.g., leaf blowers, rakes, shovels) outside, watching TV, and playing in the ball pit. Colin mands for reinforcers using signs or pictures presented on the back of the token board. During the study, Colin's programs included answering social questions (e.g., "What is your favorite color?"), tacting familiar people, receptive identification of community helpers (e.g., firefighter, vet, police officer), copying block patterns, fine motor imitation, answering yes and no to factual info (e.g., "Do you smell with your ears?"), calmly wearing goggles for increasing durations, receptive identification of parts of objects (e.g., "Touch the door on the car"), signing attributes when presented a vocal label (e.g., "Show me 'wet'"), and signing colors when presented a color card and vocal label.

Mandy is a 4-year-old White female who speaks in full sentences but shows delays in articulation. Mandy's most common challenging behaviors include work avoidance and tantrums, which often last longer than 20 minutes and can last upwards of an hour. ABA therapy sessions ranged from 2.25 to 3.25 hours in duration, typically 7 times per week, and were

conducted either in a furnished downstairs living room or her bedroom. The living room was equipped with a table, two chairs, a couch, and multiple toy shelves. Reinforcers included stuffed animals, animal figurines, baby dolls, movies, candy, books, and social play, such as piggy-back rides, and chase. Mandy mands for reinforcers vocally or using pictures on the back of the token board. During the study, Mandy's DTT objectives included tacting objects, receptive identification of gender (i.e., boy, girl, man, woman), matching pictures by categories (e.g., clothing, people, places), short vocal imitation (e.g., consonant-vowel combinations), following instructions with verbs (e.g., "Show me clapping"), copying simple drawings (e.g., vertical line, letters) social conversation (e.g., "Thank you," "You're welcome"), tacting body parts, receptive identification of environmental objects (e.g., table, chair), identification of pronouns (e.g., "Touch my shirt"), and receptive identification of parts of objects.

Design

An alternating treatment design was used across eight weeks to compare the effects of two types of token boards on participants' average trial duration, trial rate, percentage of trials correct, rate of challenging behavior, and duration of challenging behavior. During each session, only one of the two token boards was used. The sequencing of the two token boards was randomized, ensuring each board was used exactly twice per four sessions. Randomization was accomplished by selecting the four session conditions from a bag (two for each condition). Every fifth session, the participant was provided a choice of which board to use during the session, after which the sequence repeated. Additional probes were conducted during one program each session to collect data on average trial duration (i.e., how quickly clients are responding to instructions). Following eight weeks of alternating treatments, the token board yielding better

outcomes in the majority of performance variables was extended for two additional weeks, while the less effective token board was terminated.

Response Definitions and Reinforcement Schedules

Token boards were used to reinforce correct responding during DTT blocks. Targets in acquisition were reinforced with tokens on an FR 1 schedule for independent or prompted responses that matched the definition on the data sheet. Maintenance targets were reinforced with tokens on an unspecified intermittent schedule for correct and independent responses but not for prompted responses. The agency where the study was conducted also encouraged therapists to deliver tokens intermittently (i.e., 30-50% of tokens) for attending skills (e.g., hands in lap, eye contact, feet on floor) during DTT blocks. Therefore, each DTT block varied in the number of instructions provided and responses required to earn all 10 tokens and obtain the reinforcer. However, any differences in the number of trials presented or the rate of token delivery would have been washed out throughout the 3.5-hour sessions and would not have been related to the type of token board in use. Non-probe DTT blocks during baseline and intervention included token delivery for a mix of acquisition responses, maintenance responses, and attending skills. Probes, however, only included token delivery for acquisition responses.

Tokens were not delivered when the participant produced no response to a maintenance target or when they produced any incorrect response to a maintenance or acquisition target. Instead, the therapist implemented the error correction procedure, which consisted of 7 steps. If the client did not respond within 5 seconds or responded incorrectly, the therapist 1) turned their head away for 1-3 seconds, 2) brought the client's hands back to their lap to re-establish attention, 3) provided a full vocal or full physical prompt depending on the required response, 4)

delivered neutral feedback (e.g., “That’s say ‘shoe’”), 5) re-presented the instruction, 6) provided an immediate full vocal or full physical prompt, and 7) delivered neutral feedback (e.g., “Okay) without delivering a token. Following step 7 of the error correction procedure, therapists delivered an instruction to which the client was likely to respond correctly based on previous data. Because the error correction procedure included re-presenting the trial, steps 5-7 were recorded as a second trial for the purposes of calculating average trial duration during probes.

Independent Variables

Static Token Board

In Phase 1 of the study, the *static token board* (STB_1) included a white plastic clipboard, a white cup, and 30 white tokens. For this token board, the participant was required to earn 10 white tokens to receive the reinforcer. In this type of token reinforcement system, each token was equal in value, and the exchange rate for accessing the reinforcer remained consistent. When the participant earned a token, they had the choice of drawing a token out of a cup and placing it on the token board or having the therapist do it for them. Under this condition, reinforcement was provided on a 10:1 exchange rate.

After four weeks, when the random ratio token board (see below) was revised, the static token board was also revised; the new static token board, STB_2, consisted of a green clipboard and a green cup with 15 green tokens. Otherwise, the STB_2 remained functionally equivalent to the STB_1.

Random Ratio Token Board

The *random ratio token board* ($RRTB_1$) was designed to decrease the predictability of the delay to reinforcement that was seen with the static token boards. In Phase 1, the $RRTB_1$

consisted of a blue plastic clipboard, a blue cup, 28 red tokens, and 2 blue tokens. Participants were required to earn 10 red tokens or 1 blue token to receive the reinforcer. When the participant earned a token, they had the choice of drawing a token out of a cup and placing it on the token board or having the therapist do it for them. Because this token board included 28 red tokens and 2 blue tokens, the probability of selecting a blue token on the first draw was 1 in 15, or 6.67%. If the participant selected a red token on the first draw, the probability of selecting a blue token on the second draw was 1 in 14.5. A complete list of probabilities is provided in Table 2.1.

Table 2.1

Token Selection Probability Chart

Token number	Number of red remaining before drawing	Total tokens remaining before drawing	Probability of ending the block on this drawing
1	28	30	6.67%
2	27	29	6.90%
3	26	28	7.14%
4	25	27	7.41%
5	24	26	7.69%
6	23	25	8.00%
7	22	24	8.33%
8	21	23	8.70%
9	20	22	9.09%
10	N/A	N/A	100%

Four weeks into the intervention, the $RRTB_1$ was revised in Phase 2 to prevent fixation on the blue token and avoid the possibility of the red tokens signaling a longer delay to reinforcement. An updated board, the $RRTB_2$, consisted of a yellow clipboard, a yellow cup, and 15 yellow tokens. When using the yellow $RRTB_2$, therapists referred to a pre-made randomly-generated number list to determine how many tokens to deliver for each token exchange. To create the list, the primary researcher used Microsoft Excel's random number generator feature to produce a number between 1 and 20. Any number from 1 through 9 was retained as the exchange rate for that given token exchange, and any number 10-20 would require a full 10 tokens to earn reinforcement. After a few days the randomly-generated number list was revised to better match the rate of reinforcement of the blue $RRTB_1$. The revised number list was

generated by inputting 8 repetitions of numbers 1 through 13 into an Excel spreadsheet. The order of the list was randomized, and any number above 10 was replaced with a 10. Using this arrangement, each number below 10 had a 7.7% chance of occurring, and the client had a 69% chance of receiving reinforcement early (before earning 10 tokens). See Appendix A for samples of each random number list.

Daily Probe

Each day, a probe was conducted during the same target program for each participant using the token board assigned for that day to collect data on the average duration of each trial. During probes, an FR 1 schedule of token delivery was used for correct responses to acquisition targets. To keep the rate of reinforcement consistent, probes did not include token delivery for attending skills, and maintenance responses were excluded. Additionally, when using the RRTB during probes, 10 tokens were always delivered prior to providing the back-up reinforcer to ensure the rate of reinforcement was consistent between the two token boards. This was accomplished by either 1) removing the blue tokens from the cup without the participant's awareness, so that only red tokens could be selected when using the $RRTB_1$ or 2) disregarding the randomly generated number list during probes using the $RRTB_2$. For Colin, probes were initially conducted while implementing the acquisition responses in his Answers Social Questions program, for which Colin used a Dynavox speech-generating device to respond. On session 47, Colin's probe program was changed to Labels Emotions using ASL following concerns from Colin's supervisor that staff were prompting too often and Colin was becoming prompt dependent with this program. Because staff were providing immediate prompts during most probe trials when Answers Social Questions was implemented, average trial duration was

more likely to be a reflection of staff behavior rather than client behavior. Switching the probe program to Labels Emotions was thought to provide a better reflection of the client's interresponse time when using each token board because Colin did not require immediate full physical prompts during every trial. For Mandy, probes were conducted while implementing the acquisition responses in her Follows Instructions with Verbs program. During probes, the therapist started the stopwatch immediately before delivering the first instruction (e.g., "What's your name?"/"How does he feel?" for Colin or "Show me clapping?" for Mandy) and stopped it immediately after delivering the reinforcer at the end of the token board. Throughout the probe, the therapist tallied the number of trials presented and the number of error corrections implemented. Following reinforcer delivery, the therapist recorded the total duration of the probe on the data sheet so the average trial duration could be calculated by dividing the total duration by the number of trials conducted during the probe.

Staff Training

Participants' therapists were trained using behavioral skills training on the use of the new token boards, but not client programming, as each therapist has been trained and received ongoing training on client programs by their supervising Board Certified Behavior Analyst. First, staff received written instructions detailing the steps for implementing the procedures, including the schedule of reinforcement to use, when and how to conduct probes, and how to collect additional data for the study. See Appendix B for a copy of the study instructions for Phase 1. Two revisions were made to the instructions throughout the course of the study. First, a couple weeks into the study, both participants began exhibiting new challenging behaviors related to drawing their own tokens, leading the clinical team to request the primary researcher to write a new

section of instructions on how to prevent and respond to these challenging behaviors. Second, when Phase 2 of the study began with revisions to the token boards, a new set of instructions was provided to staff. See Appendix C for a copy of the study instructions for Phase 2 with the additional challenging behavior protocol attached. Included in all versions of the written instructions was a section detailing the purpose of the study. To control for experimenter expectancies regarding which token board should yield better performance, the instructions stated that the main purpose of the study was to determine whether participants performed better when they were able to choose their own tokens and that the secondary question was whether participants preferred consistent, predictable reinforcement systems or randomized reinforcement systems. After providing written instructions, the primary researcher met with the therapists in person either one-on-one or in groups of two for training. Next, the primary researcher reviewed the procedure with the therapist and modeled a DTT block with the therapist acting as the participant. The researcher or a second therapist then acted as the participant, referring to a loose script of responses, while the first therapist rehearsed the procedure. When training was conducted in groups of two, the second therapist then rehearsed the procedure while the first therapist acted as the participant. See Appendix D for a copy of the role-play script. After the role-play, the researcher provided positive and corrective feedback to the therapist. After therapists met mastery criteria for the standard DTT block procedure, this training protocol was repeated to teach the probe procedure. Mastery criteria was set at two consecutive role-plays with 100% accuracy for each procedure; however, due to time constraints, some therapists were considered fully trained following at least one role play of each procedure at 95% accuracy or higher. See Appendix E for a copy of the staff training checklist. Starting on the first session of

intervention for both participants, the primary researcher began collecting treatment integrity data on each procedural step and provided feedback to therapists throughout the study.

Additional Staff Training for Colin's Team. Following staff training for two of Colin's therapists, Colin's family extended their vacation. Therefore, a booster training was conducted via telehealth for two of Colin's therapists. Following a brief summary of the study instructions, therapists rehearsed the standard DTT block procedure and the probe procedure in a role-play scenario. The researcher provided additional positive and corrective feedback, and training continued until both therapists reached at least 95% accuracy during one role-play for each procedure.

During Phase 3 of intervention, one of Colin's therapists resigned and was replaced with a therapist who had previously worked with Colin but had not been working with the agency during the four months prior to her return to Colin's team; therefore, the new therapist received additional training on Colin's programming from the supervisor on Colin's team during sessions 88 through 90. On session 90, the primary researcher provided an abbreviated training on the study procedures via telehealth. This training included a brief summary of Phase 3 of the study instructions, modeling and rehearsal of the implementation of the $RRTB_2$ during a role-play scenario, and positive and corrective feedback. The new therapist only began implementing DTT with Colin after reaching 100% mastery on three role-plays during training on session 90. The new therapist did not receive training on conducting probes; therefore, she never conducted probes but was always paired up with a second therapist who conducted the probe for that session.

Dependent Variables

Choice Selections

This study examined six dependent variables, some of which were more highly controlled than others. The first dependent variable was participants' choice selections during the concurrent chains procedure conducted every fifth session. See Appendix F for a sample token board selection data sheet with choice sessions highlighted.

Daily Probes

The remaining five dependent variables were measures of participant performance. Probes provided a relatively more controlled method of collecting data on how quickly therapists and participants were completing DTT trials under each token board condition. During probes, therapists used a stopwatch to measure the duration of the DTT block and kept a tally of the number of trials and error corrections conducted to calculate average trial duration. See Appendix G for a sample data sheet.

Full Session Data

The remaining four dependent variables were relatively less controlled, but they provided a measure of the effects of the token boards under real-world conditions. For the third and fourth dependent variables, the primary researcher analyzed session data to collect information on trial rate (i.e., number of trials conducted divided by session duration) and average percentage of correct responses across all of participants' DTT programs.

The last two dependent variables were rate and average duration per hour of all challenging behaviors exhibited by the participants. Duration was recorded for any behavior persisting more than 10 seconds (e.g., tantrums).

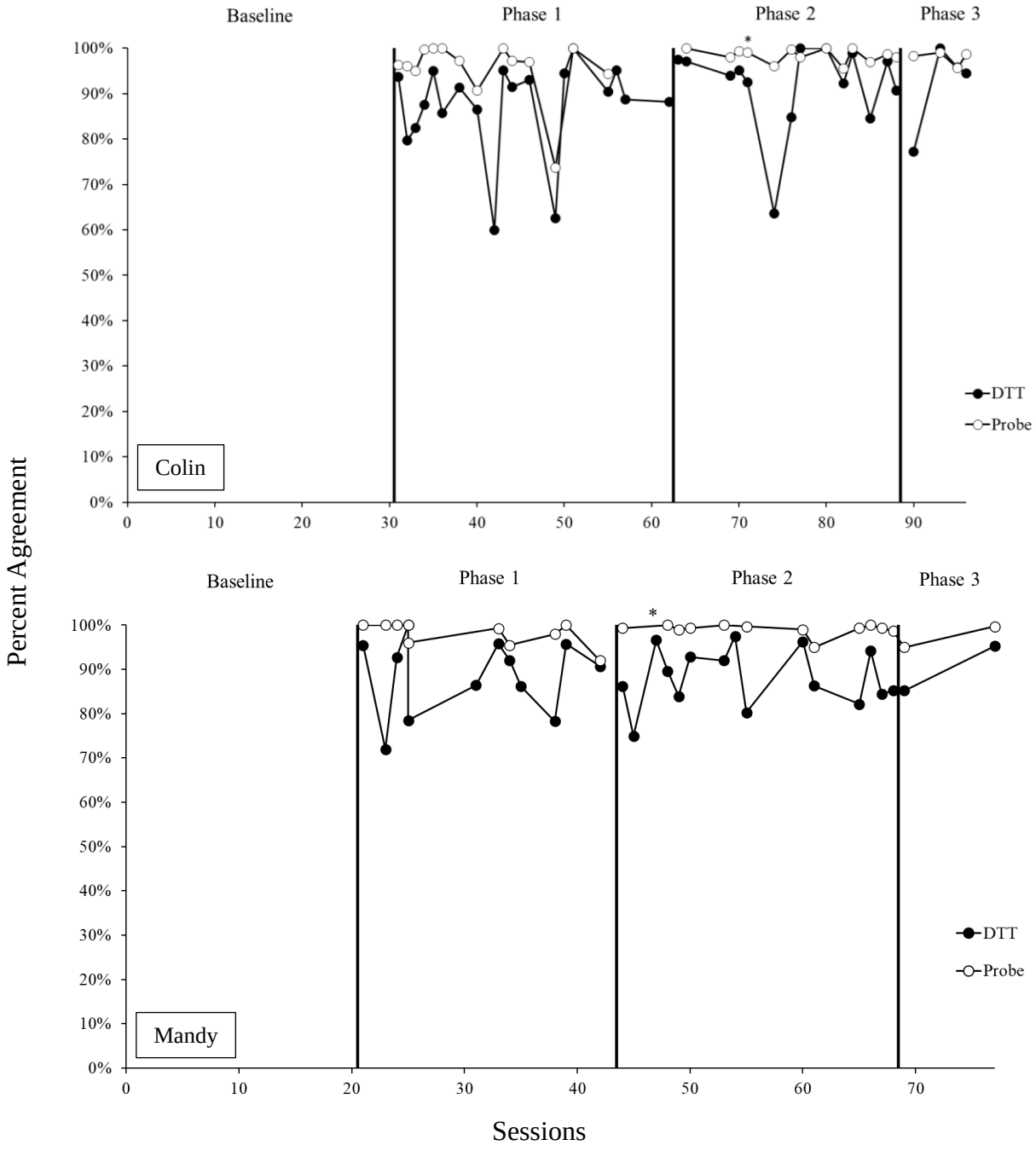
Interobserver Agreement and Treatment Integrity

The primary researcher collected treatment integrity (TI) and interobserver agreement (IOA) data on Colin's DTT data either in-person or via telehealth during 56.1% of sessions and 47.0% of probes. IOA on Colin's challenging behaviors was collected during 28.8% of sessions. TI and IOA were collected on Mandy's DTT data either in-person or via telehealth during 54.4% of sessions and 42.1% of probes. IOA on Mandy's challenging behaviors was collected during 31.6% of sessions. Due to scheduling constraints, during Phase 3 of the intervention, TI and IOA were only collected on DTT data and challenging behavior during 22.2% of sessions and probes.

IOA for participants' correct and incorrect responses during DTT was calculated using exact trial-by-trial agreement by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100%. DTT IOA averaged 89.6% for Colin (60.0%-100%) and 88.5% for Mandy (72.0%-100%). During probes, total count IOA was collected on the number of trials and the number of error corrections, while total duration IOA was calculated for the duration of the probe during each session. These three IOA scores were then averaged, resulting in one probe IOA score for each session. Probe IOA averaged 97.0% (73.7%-100%) for Colin and 98.5% (92.0%-100%) for Mandy. DTT and probe IOA is displayed in Figure 2.1 for Colin and Mandy.

Figure 2.1

DTT and Probe IOA for Colin and Mandy

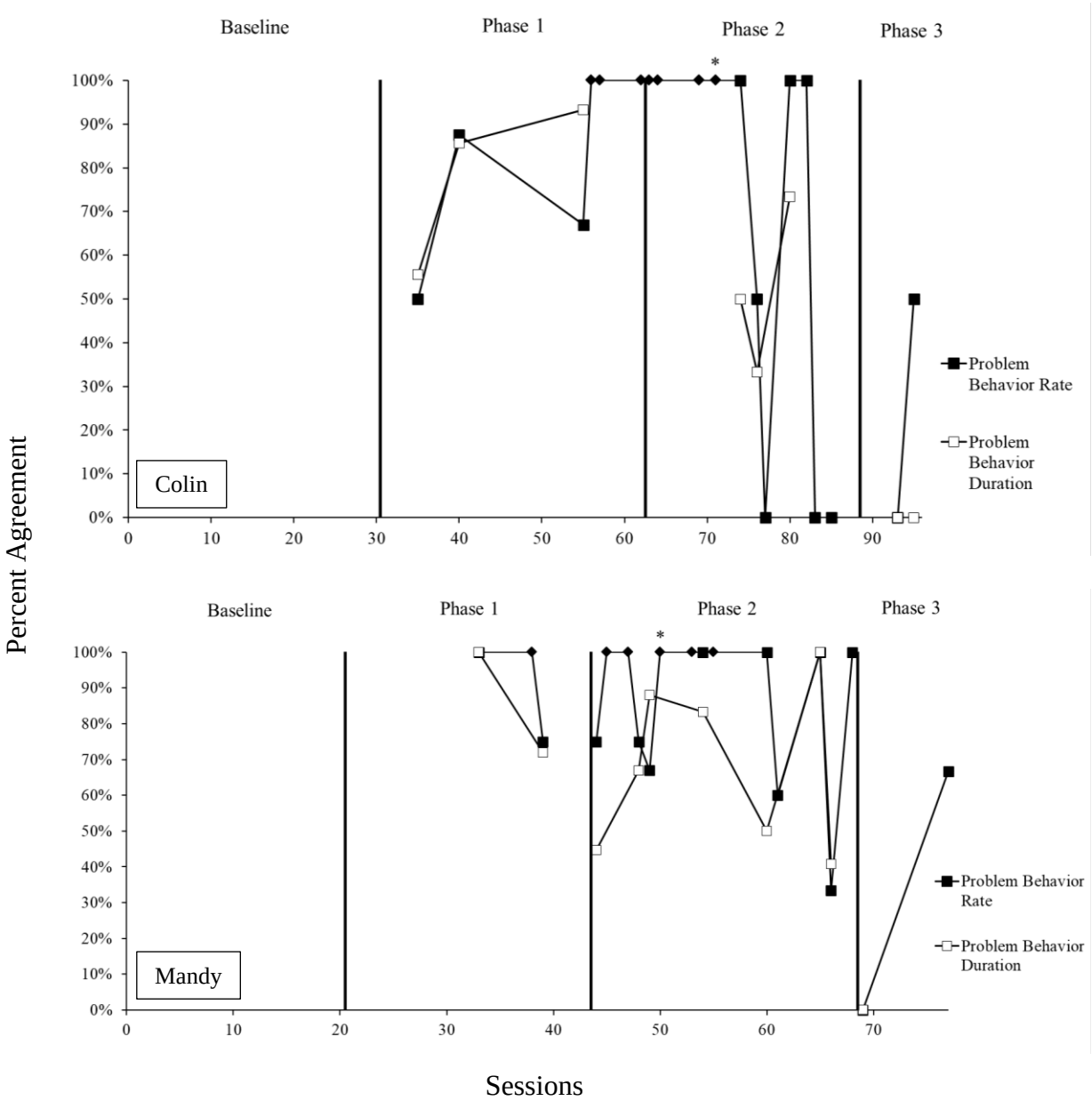


Note. * = Revised reinforcement rate for $RRTB_2$.

Challenging behavior IOA is depicted in Figure 2.2 for Colin and Mandy. IOA for frequency of challenging behaviors was calculated using total count IOA by dividing the smaller frequency by the larger frequency and multiplying by 100% at the end of each observation. Challenging behavior frequency IOA averaged 69.7% (0.0%-100%) for Colin and 81.7% (0.0%-100%) for Mandy. Duration IOA was calculated for tantrums and work avoidance for both participants using total duration IOA. At the end of each observation, the smaller total duration was divided by the larger total duration and multiplied by 100%. Total duration IOA averaged 48.0% for Colin (0.0%-93.0%) and 66.0% (0.0%-100%) for Mandy. Although total duration IOA is a less conservative measure of duration IOA, it was chosen because the same episode of tantrumming or avoidance was sometimes recorded as two separate events by one observer but as only one by the other observer if they occurred in quick succession. Using duration-per-occurrence IOA would have underestimated agreement in these situations. As it was calculated, total duration IOA was sometimes low even when the two observers' data agreed within one minute just because participants' challenging behaviors were short.

Figure 2.2

Challenging Behavior IOA for Colin and Mandy

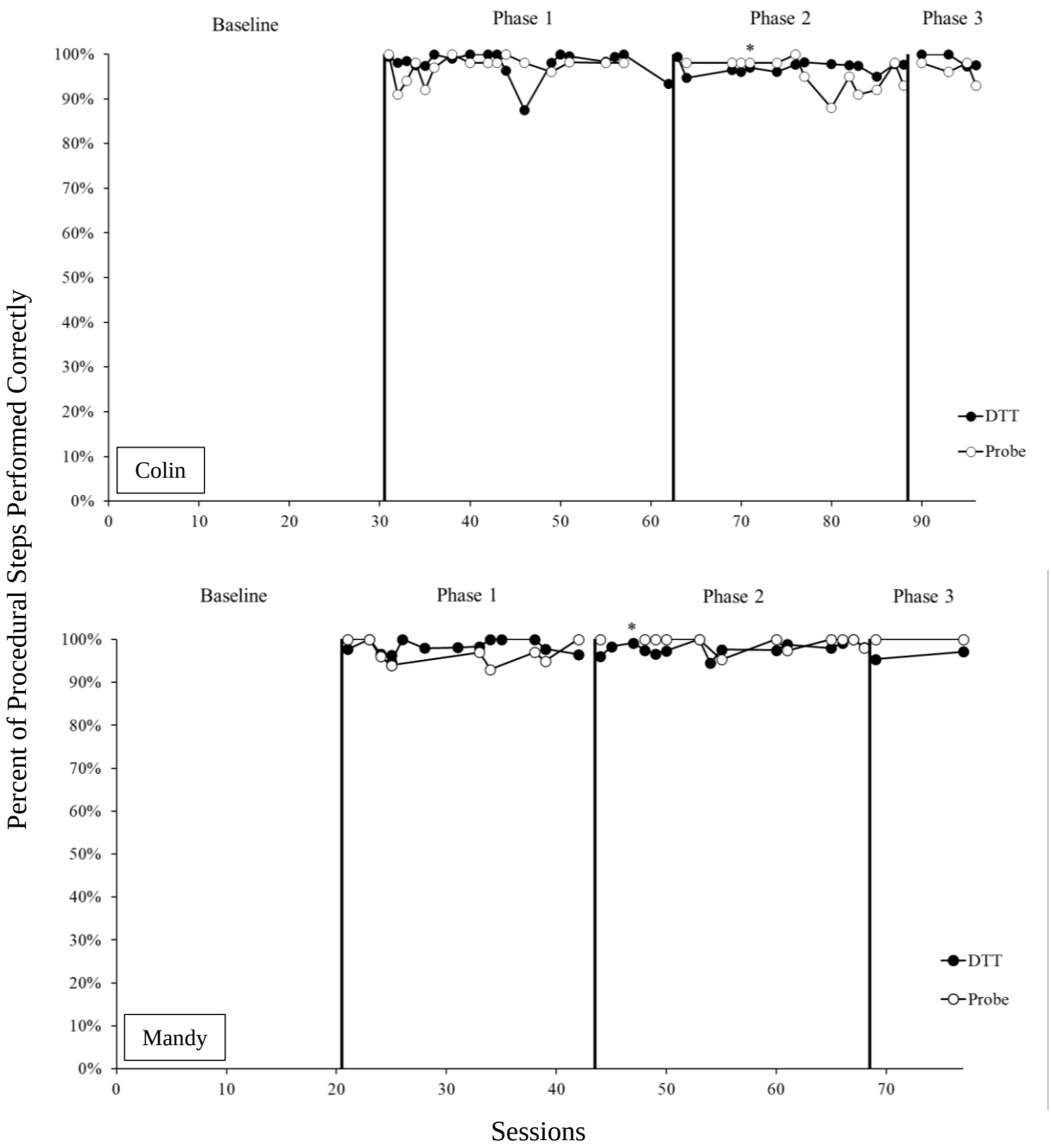


Note. * = Revised reinforcement rate for $RRTB_2$. Diamond markers = Both observers agreed there were no occurrences.

Procedural steps assessed for TI during standard DTT blocks included using the correct token board and correct bag of tokens, using an FR 1 schedule of token delivery during acquisition trials and an intermittent schedule during maintenance trials, and using a vocal prompt with a 5-second time delay to prompt the client to choose a token. See Appendix H for the DTT block TI data sheet. Some of the steps included on the treatment integrity checklist could occur multiple times during one DTT block (e.g., providing behavior-specific praise, implementing the error correction procedure following an incorrect response); therefore, those steps were scored for each opportunity, while others (e.g., using the correct token board) were only scored once per DTT block. TI was calculated for up to 6 DTT blocks per session and averaged. TI for Colin's sessions averaged 97.7% (87.5%-100%) and TI for Mandy's sessions averaged 97.9% (94.5%-100%). A unique TI data sheet was used for probes and included additional procedural steps, such as conducting one probe per day, implementing only acquisition targets from the correct program while conducting the probe, starting the stopwatch within 3 seconds of delivering the first instruction, stopping the stopwatch within 3 seconds of delivering the reinforcer, and recording the total duration of the DTT block and the number of trials and error corrections conducted during the block. See Appendix I for the probe TI data sheet. TI averaged 96.4% (88.0%-100%) for Colin's probes and 98.2% (93.0%-100%) for Mandy's probes. TI is depicted in Figure 2.3 for Colin and Mandy.

Figure 2.3

DTT and Probe TI for Colin and Mandy



Note. * = Revised reinforcement rate for $RRTB_2$.

Procedures

Baseline

Session data from the past 6 weeks of ABA therapy was collected retroactively to serve as the baseline for this study. Baseline data were collected on four of the dependent variables: trial rate, percentage of trials correct, and frequency and duration of challenging behavior. During baseline, participants used whatever type of token board was in place at that time. Colin's baseline token board consisted of 10 star stickers adhered to the board with Velcro, which the therapist moved from the top to the bottom of a clipboard following correct responses. Colin had been using this token board for ABA therapy sessions for at least 2.5 years prior to the start of the study. Mandy's baseline token board consisted of 10 gems adhered to the board with Velcro, which the therapist moved from the top to the bottom of a clipboard following correct responses. During baseline, Mandy was also receiving token board training, which consisted of systematically increasing the token exchange schedule from 1 to 10 tokens. Mandy reached 10 tokens on the 16th day of baseline and mastered her token board training objective on the last day of baseline.

Intervention

During intervention, one of the two token boards was implemented each day for all DTT blocks throughout the session. During the first four weeks of intervention (i.e., Phase 1), the STB1 and the RRTB1 were alternated. During the seconds four weeks of the intervention (i.e., Phase 2), the STB2 and the RRTB2 were alternated. In the final two weeks of the study (i.e., Phase 3), the token board yielding better outcomes in the majority of the five performance variables in Phase 2 was implemented continuously, while the other token board was terminated.

Concurrent Chains Procedure as a Preference Assessment

Every fifth session, participants were provided a choice of token board at the beginning of the session using a concurrent chains procedure. The therapist was instructed to place both the STB and the RRTB on the table with their associated cups and tokens and ask the participant, "Which one do you want to use today?" The participant could make a selection by pointing to one or vocally tacting the color of the token board. The selected token board was then implemented during the entire session, and participants were given another opportunity to choose the token board during the next choice probe. Any mands for the other token board produced outside of the concurrent chains procedure were recorded on a data sheet and were redirected (e.g., "We are going to use this token board today, but on Friday, you can choose which token board we use.")

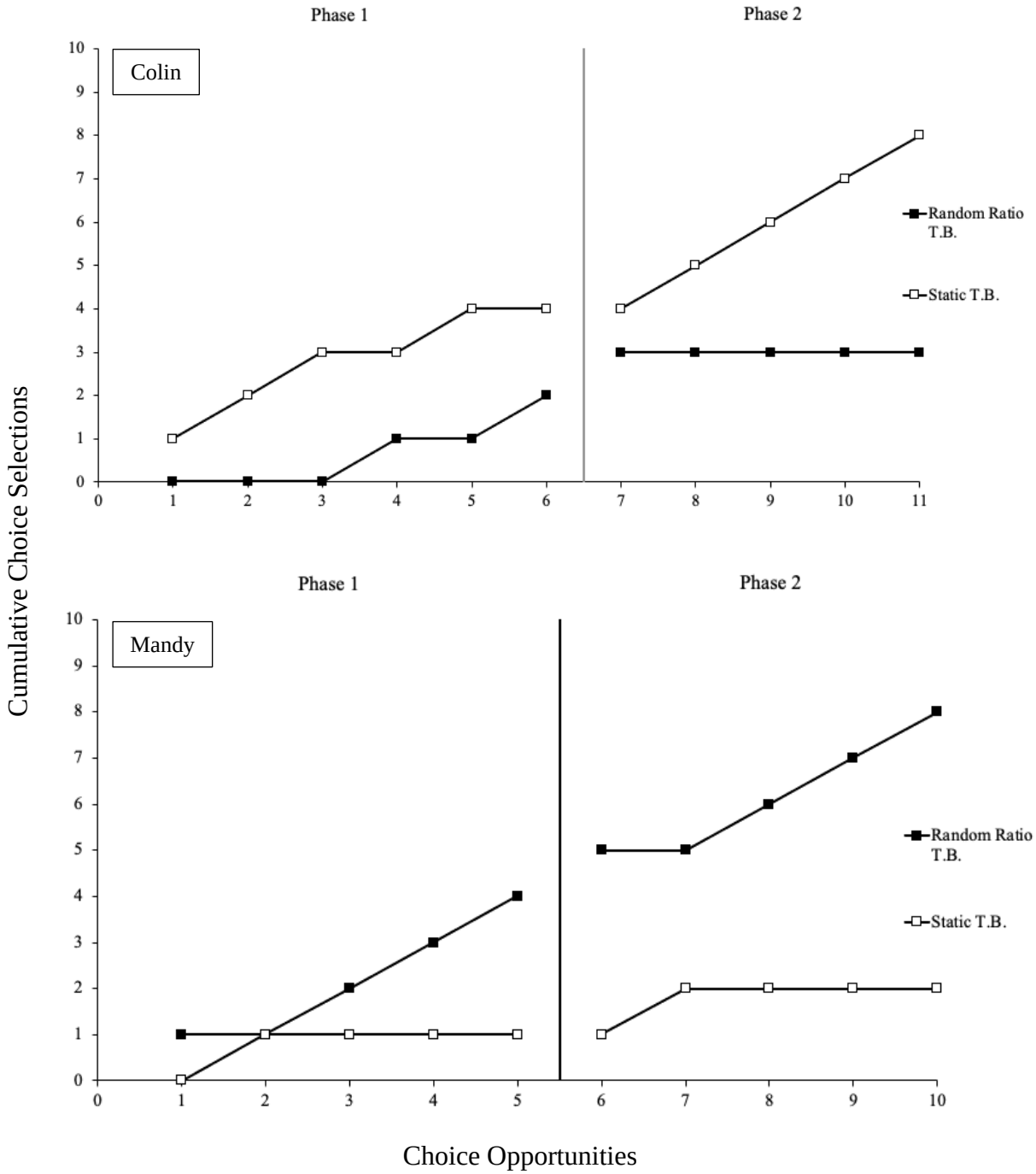
Chapter 3: Results

Choice Selections

Cumulative choice selections for Colin and Mandy are presented in Figure 3.1. During the first four weeks of the intervention, Colin had 6 opportunities from which to choose a board for the day. He chose the STB_1 four times and the $RRTB_1$ twice. Choices for the $RRTB_1$ occurred during the fourth and sixth choice probes. During the next 4 weeks, Colin had 5 opportunities to choose between STB_2 and $RRTB_2$, and he selected the STB_2 four times and the $RRTB_2$ once. Colin's choice for the $RRTB_2$ occurred during the first choice probe. Additional data suggest that Colin requested the STB_1, the STB_2, and the $RRTB_2$ each one time outside of choice selections during the course of the study. Mandy had 5 opportunities to choose a token board during the first four weeks of the intervention. Mandy chose the STB_1 once and the $RRTB_1$ four times. Mandy's choice for the STB_1 occurred during the second choice probe, though anecdotally it was reported that after selecting STB_1, Mandy immediately requested $RRTB_1$ but was denied access given the protocols. When given the choice between STB_2 and $RRTB_2$ during the next four weeks, Mandy chose the STB_2 once and the $RRTB_2$ 4 times. Once again, Mandy's choice for the STB_2 occurred during the second choice probe. Additional data suggest that Mandy requested the $RRTB_1$ once and the $RRTB_2$ 5 times outside of choice selections during the course of the study.

Figure 3.1

Cumulative Choice Selections for Colin and Mandy



Daily Probes

Average trial duration during probes is depicted in Figure 3.2 for Colin and Mandy. Neither the STB_1 nor the $RRTB_1$ produced noticeable trends in average trial duration. When the STB_1 was implemented, data remained stable, with at least 80% of the data points within ±25% of the median[1] (Ledford et al., 2017). Mean probe trial average duration for the STB_1 was 33.3 seconds, with a standard deviation of 6.6 seconds, and a range of 23.6 to 49.4 seconds. (Descriptive statistics for all of Colin's performance variables are also listed in Table 3.1.) When the $RRTB_1$ was implemented, data were initially more variable, but stabilized during the last 5 sessions. Mean probe trial average duration for the $RRTB_1$ was 39.8 seconds, with a standard deviation of 21.3 seconds, and a range of 23.3 to 102.5 seconds. Challenging behaviors, such as work avoidance, occasionally occurred during probes with both token boards, leading to an increase in average trial duration. Although, the longest probe using the $RRTB_1$ resulted in an average trial duration of 102.5 seconds, the next highest average trial duration for the $RRTB_1$ was 55.8 seconds. When Phase 2 of the intervention began, data for both token boards became more variable. On session 71, Colin's probe program was changed, and average probe trial duration decrease for both token boards. When the STB_2 was implemented during Phase 2, mean probe trial average duration was 29.5 seconds, with a standard deviation of 8.0 seconds, and a range of 19.2 to 43.4 seconds. When the $RRTB_2$ was implemented, mean probe trial average duration was 30.6 seconds, with a standard deviation of 6.3 seconds, and a range of 20.6 to 40.2 seconds. In Phase 3 of the intervention, when only the $RRTB_2$ was implemented, Colin's mean

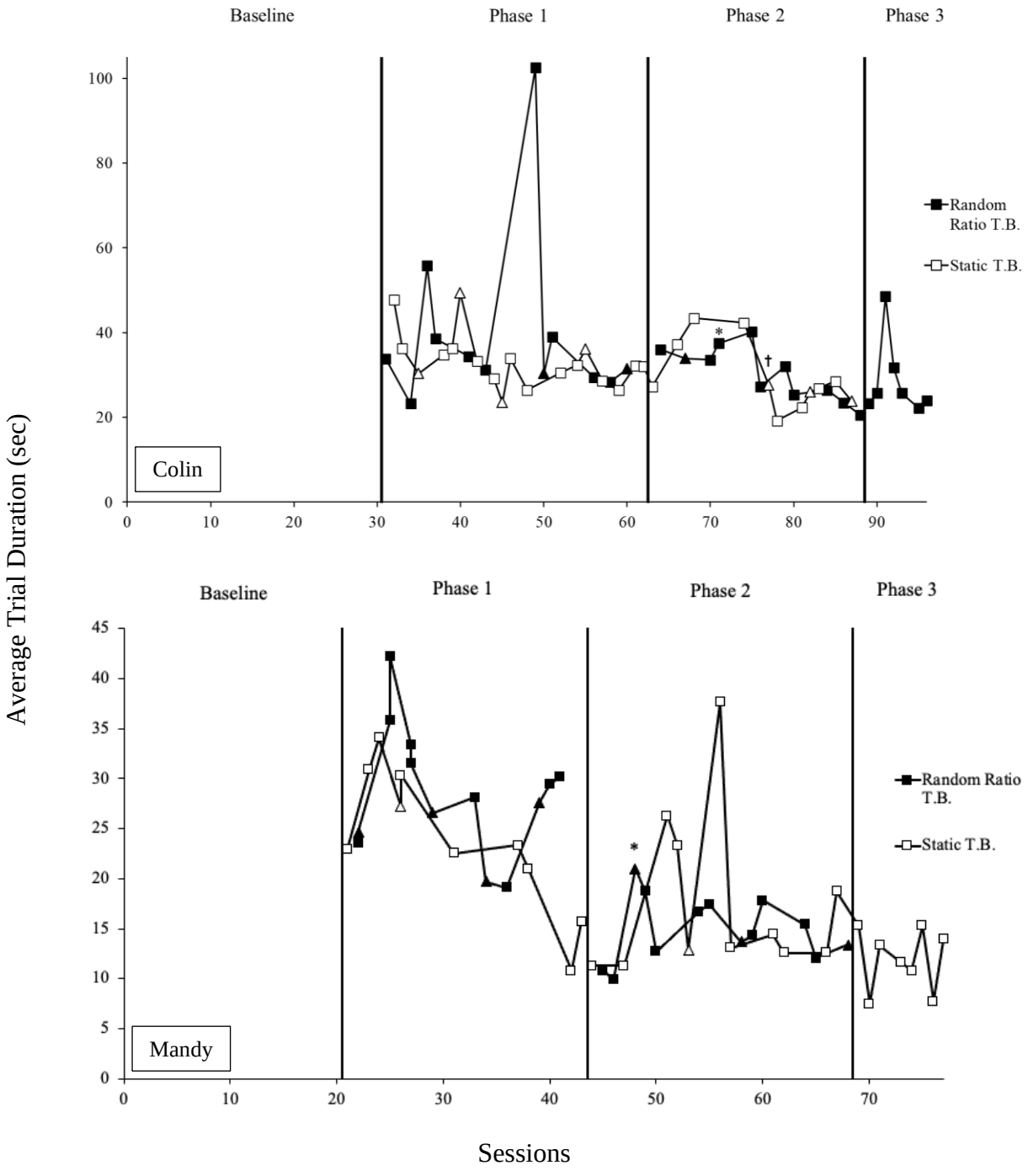
[1] Throughout this document, stability criteria is always set at 80% of data points within ±25% of the median

probe trial average duration was comparable to that in Phase 2 at 28.8 seconds, with a standard deviation of 9.3 seconds, and a range of 22.3 to 48.6 seconds.

Mandy's average trial duration for the STB_1 showed a decreasing trend. Mean probe trial average duration was 23.9 seconds, with a standard deviation of 7.1 seconds, and a range of 10.8 to 34.1 seconds. (Descriptive statistics for all of Mandy's performance variables are also listed in Table 3.2.) Probe trial average duration for the $RRTB_1$ initially increased until session 7, decreased until session 20, and then increased again until the new token boards were introduced. Mean probe trial average duration for the $RRTB_1$ was 28.6 seconds, with a standard deviation of 6.4 seconds, and a range of 19.1 to 42.2 seconds. When Phase 2 of the intervention began, average trial duration showed no discernible trends for the STB_2 or the $RRTB_2$, and neither met stability criteria. When the STB_2 was implemented, mean probe trial average duration was 17.6 seconds, with a standard deviation of 8.3 seconds, and a range of 11.3 to 37.6 seconds. Mean probe trial average duration for the $RRTB_2$ was lower than that for the STB_2 at 14.9 seconds, with a standard deviation of 3.3, and a range of 9.9 to 21.0 seconds. In the final phase of the study, when only the STB_2 was implemented, Mandy's average trial duration decreased and remained variable. Her mean probe trial average duration was 11.9 seconds, with a standard deviation of 3.1 seconds, and a range of 7.5 to 15.3 seconds.

Figure 3.2

Average Probe Trial Duration for Colin and Mandy



Note. * = Revised reinforcement rate for $RRTB_2$. † = Changed probe program.

Table 3.1

Descriptive Statistics for Colin

Dependent Variable	Baseline 1	Phase 1 STB_1	Phase 1 $RRTB_1$	Phase 2 STB_2	Phase 2 $RRTB_2$	Follow-Up $RRTB_2$
Avg. Probe Trial Duration	NA	33.3 (6.6) 23.6/49.4	39.8 (21.3) 23.2/102.5	29.5 (8.0) 19.2/43.4	30.6 (6.3) 20.6/40.2	28.8 (9.3) 22.3/48.6
Trial Rate	12.0 (2.9) 5.1/16.0	16.7 (4.8) 9.1/28.6	13.5 (4.5) 7.3/21.6	14.1 (2.6) 10.8/19.7	15.5 (6.1) 7.5/29.3	11.8 (3.3) 8.1/16.6
Percent Correct (%)	58.8 (9.8) 36.5/76.9	67.2 (8.4) 57.7/88.1	65.9 (7.5) 47.5/75.0	58.3 (8.8) 41.0/80.0	58.2 (7.1) 45.1/68.0	67.3 (11.0) 49.2/86.8
Rate of Challenging Bx	0.5 (0.5) 0.0/2.0	0.9 (1.4) 0.0/5.0	0.8 (0.8) 0.0/3.1	1.4 (1.3) 0.0/3.6	0.6 (1.0) 0.0/3.1	0.5 (0.5) 0.0/1.6
Avg. Duration of Challenging Bx/Hour	0.2 (0.6) 0.0/3.3	1.0 (2.0) 0.0/7.6	1.0 (1.6) 0.0/5.6	1.0 (1.4) 0.0/4.0	0.8 (1.5) 0.0/4.6	0.2 (0.5) 0.0/1.3

Note. The top row of data for each dependent variable represents the mean followed by the standard deviation in parentheses. The bottom row of data represents the minimum/maximum.

Table 3.2

Descriptive Statistics for Mandy

Dependent Variable	Baseline 1	Baseline 2	Phase 1 STB_1	Phase 1 $RRTB_1$	Phase 2 STB_2	Phase 2 $RRTB_2$	Follow-Up STB_2
Avg. Probe Duration	NA	NA	23.9 (7.1) 10.8/34.1	28.6 (6.4) 19.1/42.2	17.6 (8.3) 11.3/37.6	14.9 (3.3) 9.9/21.0	11.9 (3.1) 7.5/15.3
Trial Rate	20.6 (9.4) 3.5/35.4	32.7 (6.5) 23.1/39.6	32.3 (12.3) 8.8/48.4	20.4 (10.3) 4.0/37.8	33.2 (11.2) 14.9/51.3	25.2 (7.0) 15.3/43.7	34.2 (20.4) 13.2/69.0
Percent Correct (%)	83.1 (8.2) 66.7/94.6	NA	84.9 (7.7) 73.4/100.0	83.0 (9.5) 58.3/95.5	85.2 (9.4) 71.4/99.0	88.3 (7.3) 77.4/97.6	89.7 (9.2) 65.9/96.1
Rate of Challenging Bx	0.2 (0.5) 0.0/1.8	NA	0.5 (0.5) 0.0/1.3	0.8 (0.9) 0.0/3.0	0.6 (1.0) 0.0/3.2	0.4 (0.4) 0.0/1.3	0.4 (0.8) 0.0/2.3
Avg. Duration of Challenging Bx/Hour	NA	NA	5.2 (11.9) 0.0/38.4	6.1 (8.4) 0.0/27.0	1.3 (2.7) 0.0/8.5	1.0 (1.7) 0.0/6.0	0.9 (2.2) 0.0/6.5

Note. The top row of data for each dependent variable represents the mean followed by the standard deviation in parentheses. The bottom row of data represents the minimum/maximum.

Full Session Data

Trial Rate

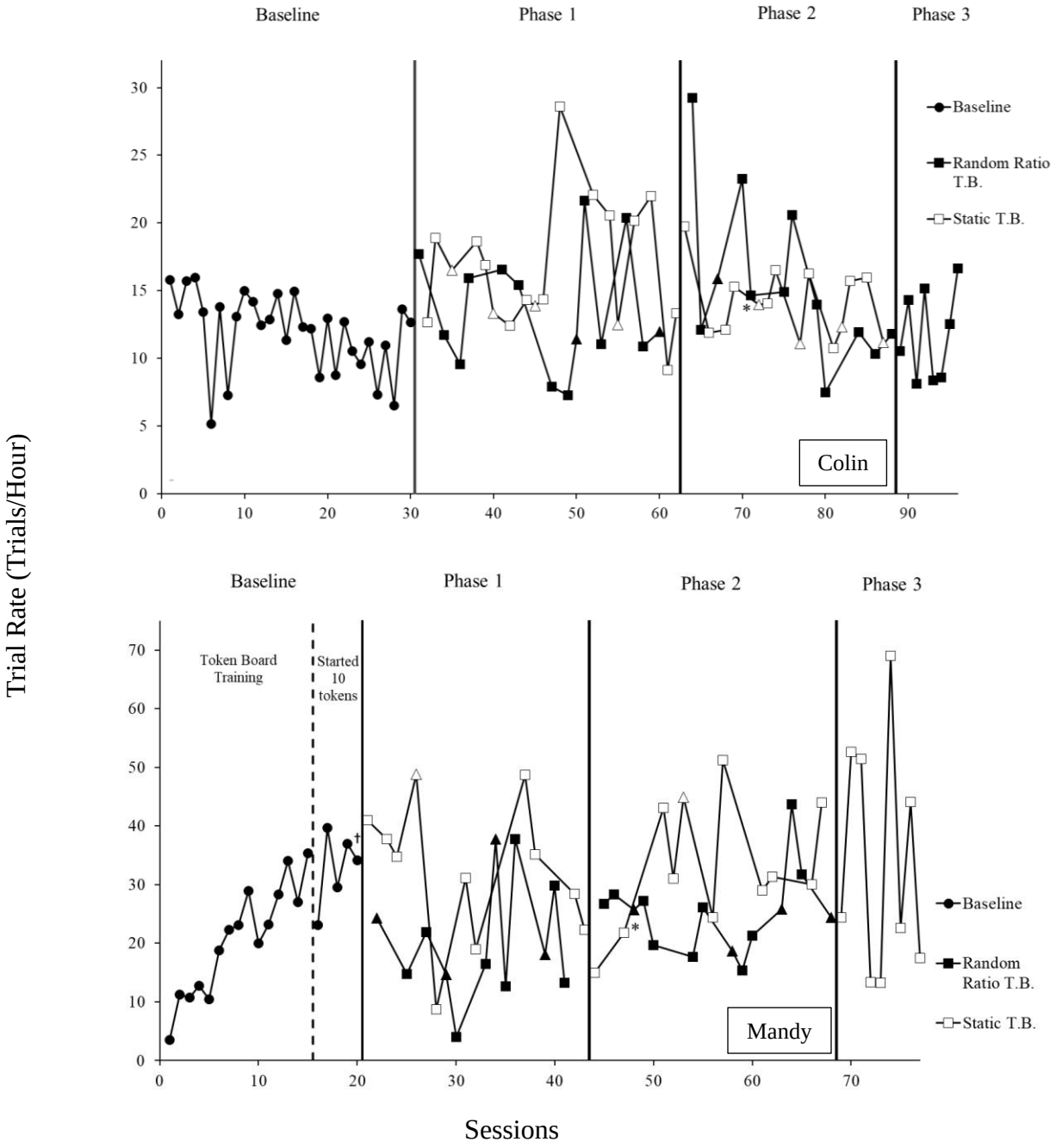
Data for Colin's and Mandy's average trial rate across all ABA therapy sessions is depicted in Figure 3.3. During baseline, Colin's mean trial rate was 12.0, with a standard deviation of 2.9, and a range of 5.1 to 16.0. A slight decreasing trend was observed during baseline trial rate. When intervention began, trial rate was variable with both token boards. Mean trial rate for the STB_1 was 16.7 trials per hour, with a standard deviation of 4.8, and a range of 9.1 to 28.6. Mean trial rate for the $RRTB_1$ was lower than that for the STB_1, but still higher than baseline at 13.5 trials per hour, with a standard deviation of 4.5, and a range of 7.3 to 21.6. When Phase 2 of the intervention began, $RRTB_2$ trial rate was initially higher than STB_2 trial rate; however, $RRTB_2$ trial rate showed a decreasing trend while STB_2 trial rate remained stable and showed no trend. Mean trial rate for the STB_2 was 14.1 trials per hour, with a standard deviation of 2.6, and a range of 10.8 to 19.7 trials per hour. Mean trial rate for the $RRTB_2$ was 15.5 trials per hour, with a standard deviation of 6.1, and a range of 7.5 to 29.3 trials per hour. Finally, when Phase 3 was implemented using only the $RRTB_2$, data remained variable and showed no trend. Mean trial rate was 11.9 trials per hour, with a standard deviation of 3.5, and a range of 8.1 to 17.4 trials per hour.

Mandy's trial rate steadily increased during baseline from session 1 to 15 while she was being trained on token board use; during this time there were systematic increases in the token exchange schedule. During token board training, mean trial rate was 20.6 trials per hour, with a standard deviation of 9.4, and a range of 3.5 to 35.4. Mandy's token exchange schedule was upgraded to the final target of 10 tokens starting in session 16 of baseline, and the program was

mastered during the final baseline session. During the last five sessions of baseline, trial rate reached stability with a mean of 32.7 trials per hour, a standard deviation of 6.5, and a range of 23.1 to 39.6. When the intervention began, mean trial rate for both token boards underperformed compared to the last five sessions of baseline and did not meet the stability criterion. The mean trial rate for the STB_1 was 32.3 trials per hour, with a standard deviation of 12.3, and a range of 8.8 to 48.8. Mean trial rate for the $RRTB_1$ was 20.4 trials per hour, with a standard deviation of 10.3, and a range of 4.0 to 37.8. When Phase 2 of the intervention was implemented, trial rate remained variable for both token boards. For the first six sessions of Phase 2, the $RRTB_2$ outperformed the STB_2, however, this trend reversed by the seventh session, shortly after the reinforcement rate was revised for the $RRTB_2$. Mean trial rate for the STB_2 was 33.2 trials per hour with a standard deviation of 11.2 and a range of 14.9 to 51.3 trials per hour. Mean trial rate for the $RRTB_2$ was 25.2 trials per hour with a standard deviation of 7.0 and a range of 15.3 to 43.7 trials per hour. When Phase 3 began, data remained variable with no discernible trend. Mean trial rate when the STB_2 was implemented was 34.2 trials per hour, with a standard deviation of 20.4, and a range of 13.2 to 69.0.

Figure 3.3

Trial Rate for Colin and Mandy



Note. † = Mastered token board training with 10 tokens. * = Revised reinforcement rate for $RRTB_2$.

Percentage Correct

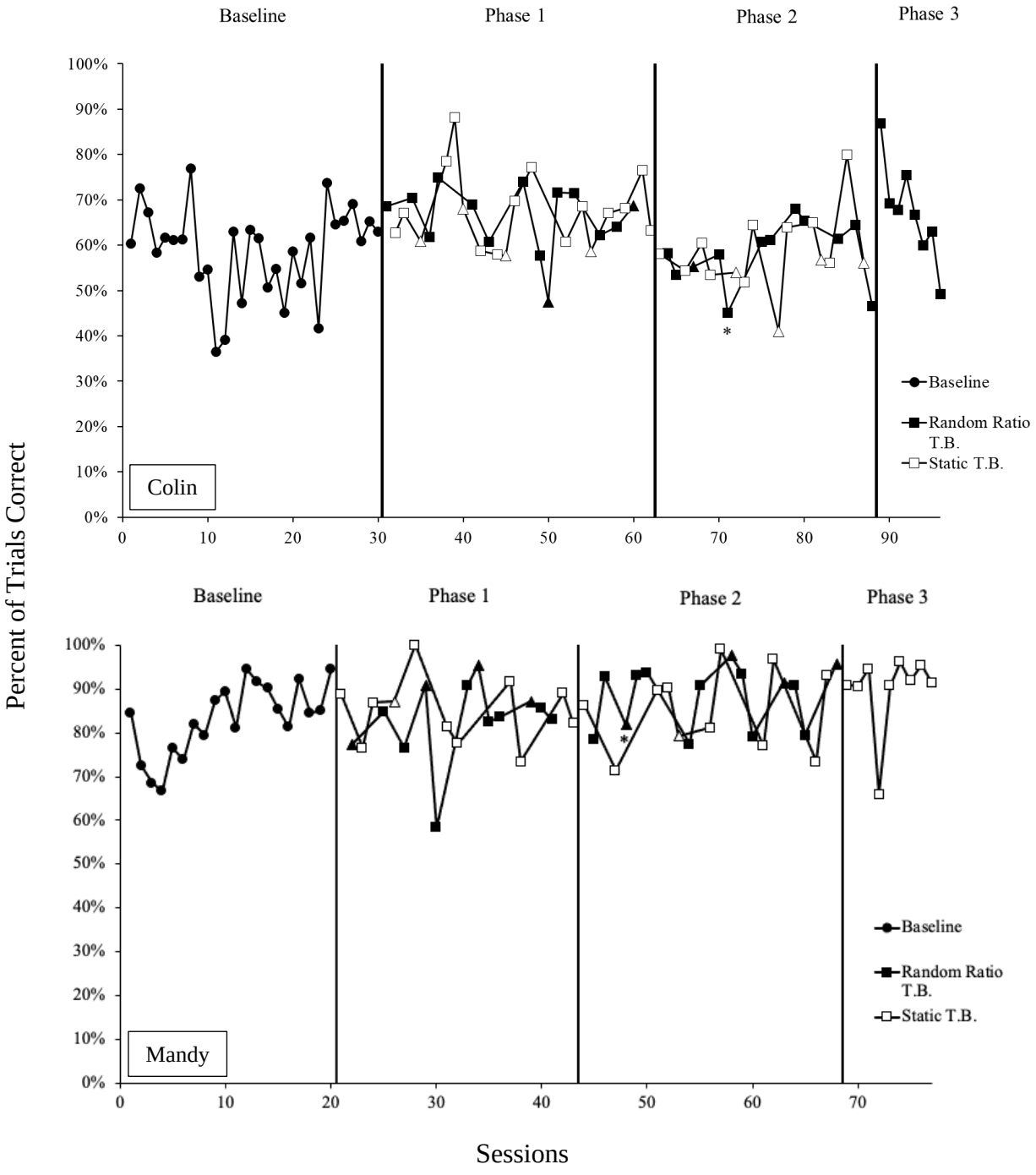
Colin's and Mandy's accuracy data across ABA therapy sessions is depicted in Figure 3.4. Colin's percentage correct data were stable during baseline and showed no discernible trend. Mean percentage of correct trials was 58.8%, with a standard deviation of 9.8%, and a range of 36.5% to 76.9%. During the intervention, data for both token boards remained stable with no noticeable trends. When the STB_1 was implemented, mean percentage of trials correct was 67.2%, with a standard deviation of 8.4%, and a range of 57.7% to 88.1%. When the $RRTB_1$ was implemented, mean percentage of trials correct was 65.9%, with a standard deviation of 7.5%, and a range of 47.5% to 75.0%. When Phase 2 of the intervention began, data remained stable with no discernible trends. Mean percentage of correct trials with the STB_2 was 58.3%, with a standard deviation of 8.8%, and a range of 41% to 80%. When the $RRTB_2$ was implemented, mean percentage of correct trials was 58.2%, with a standard deviation of 7.1%, and a range of 45.1% to 68.0%. In Phase 3, when only the $RRTB_2$ was implemented, percentage of correct trials initially increased but then continued on a decreasing trend. Mean percentage correct was 67.3%, with a standard deviation of 11.0%, and a range of 49.2% to 86.8%.

Mandy's percentage correct data showed an increasing trend during baseline but remained within the stability envelope. Mean percentage of correct trials was 83.1%, with a standard deviation of 8.2%, and a range of 66.7% to 94.6%. During the first phase of the intervention, data for both token boards became slightly more variable but again remained within the stability envelope and produced no discernible trends. When the STB_1 was implemented, mean percentage of trials correct was 84.9%, with a standard deviation of 7.7%, and a range of 73.4% to 100%. When the $RRTB_1$ was implemented, mean percentage of trials correct was

83.0%, with a standard deviation of 9.5%, and a range of 58.3% to 95.5%. When Phase 2 of the intervention was implemented, data for both token boards remained high and stable and produced no discernible trends. When the STB_2 was implemented, mean percentage of trials correct was 85.2%, with a standard deviation of 9.4%, and a range of 71.4% to 99.0%. When the $RRTB_2$ was implemented, mean percentage of trials correct was 88.3%, with a standard deviation of 7.3%, and a range of 77.4% to 97.6%. In Phase 3, when only the STB_2 was implemented, percentage of correct trials generally increased and became more stable except for one outlier at 65.9%. Overall, the mean percentage of correct trials was 89.7%, with a standard deviation of 9.2%, and a range of 65.9% to 96.1%.

Figure 3.4

Percentage of Correct DTT Trials for Colin and Mandy



Note. * = Revised reinforcement rate for $RRTB_2$.

Rate and Duration of Challenging Behaviors

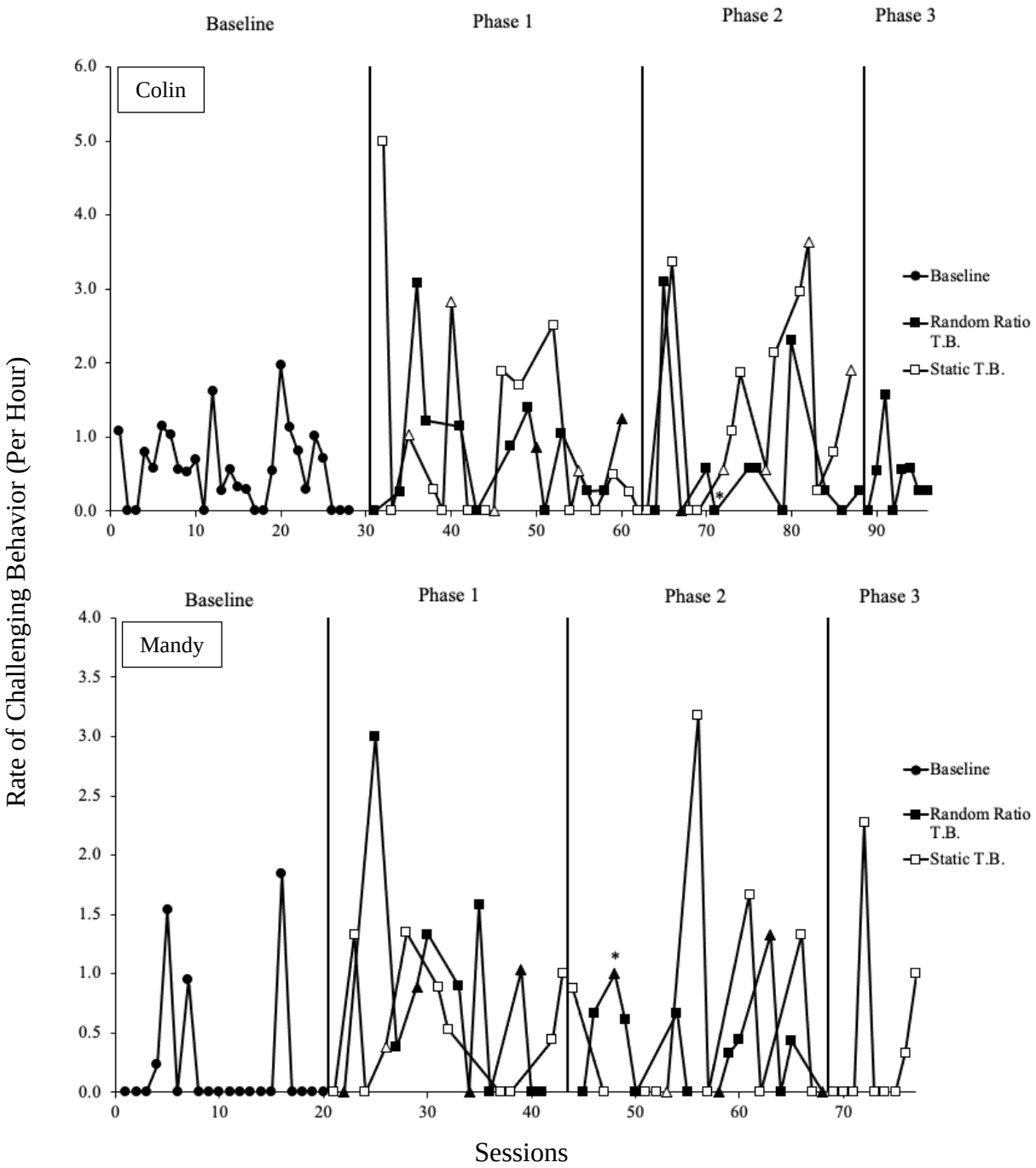
Rate of challenging behaviors is depicted in Figure 3.5 for Colin and Mandy. During baseline, Colin's data were variable and showed no discernible trend. Mean rate of challenging behaviors was 0.5 per hour, with a standard deviation of 0.5, and a range of 0.0 to 2.0. During the intervention, data for both token boards increased slightly and became more variable. When the STB_1 was implemented, Colin's mean rate of challenging behaviors was 0.8 per hour, with a standard deviation of 1.3, and a range of 0.0 to 5.0. When the $RRTB_1$ was implemented, mean rate of challenging behaviors was 0.8 per hour, with a standard deviation of 0.9, and a range of 0.0 to 3.1. When Phase 2 of the intervention began, data remained variable with no discernible trends for both token boards. When the STB_2 was implemented, mean rate of challenging behaviors was 1.4 per hour, with a standard deviation of 1.3, and a range of 0.0 to 3.6 per hour. When the $RRTB_2$ was implemented, mean rate of challenging behaviors was lower, at 0.6 per hour, with a standard deviation of 1.0, and a range of 0.0 to 3.1 per hour. When Phase 3 was implemented with only the $RRTB_2$, mean rate of challenging behaviors was comparable to that in Phase 2 at 0.5 per hour, with a standard deviation of 0.5, and a range of 0.0 to 1.6.

Mandy's rate of challenging behaviors was stable during baseline and showed no discernible trend. Mean rate of challenging behaviors was 0.2 per hour, with a standard deviation of 0.5, and a range of 0.0 to 1.8. During the first phase of the intervention, data for both token boards increased and became more variable. When the STB_1 was implemented, Mandy's mean rate of challenging behaviors was 0.5 per hour, with a standard deviation of 0.5, and a range of 0.0 to 1.3. When the $RRTB_1$ was implemented, mean rate of challenging behaviors was 0.8 per hour, with a standard deviation of 0.9, and a range of 0.0 to 3.0. When Phase 2 of the

intervention began, data remained variable with no discernible trends for either token board. When the STB_2 was implemented, the mean rate of challenging behaviors was 0.6 per hour, with a standard deviation of 1.0, and a range of 0.0 to 3.2 per hour. When the $RRTB_2$ was implemented, the mean rate of challenging behaviors was 0.4 per hour, with a standard deviation of 0.4, and a range of 0.0 to 1.3 per hour. When Phase 3 was implemented with only the STB_2, mean rate of challenging behaviors decreased to 0.4 per hour, with a standard deviation of 0.8, and a range of 0.0 to 2.3.

Figure 3.5

Rate of Challenging Behaviors for Colin and Mandy



Note. * = Revised reinforcement rate for $RRTB_2$.

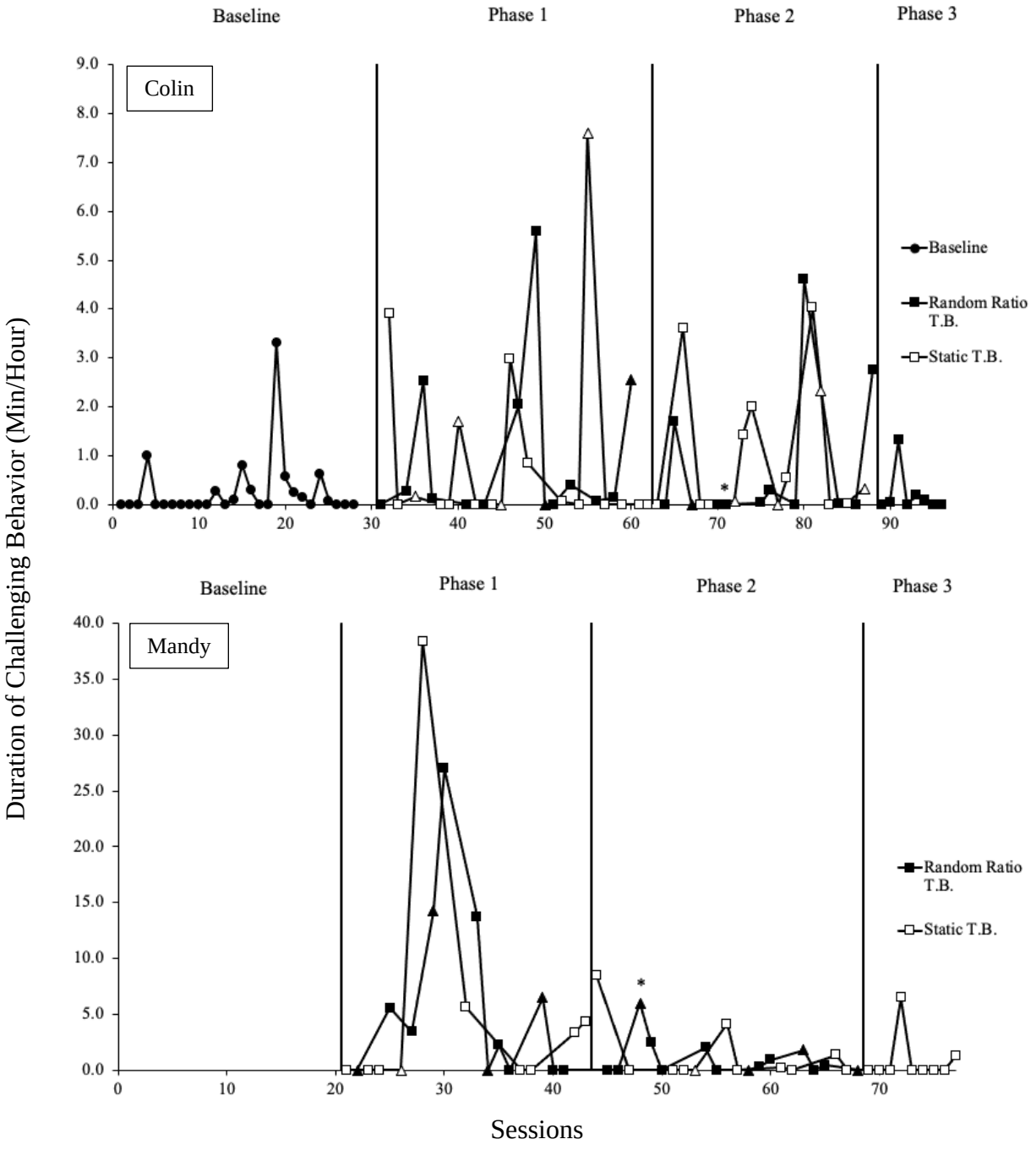
The duration of challenging behaviors per hour is depicted in Figure 3.6 for Colin and Mandy. During baseline, Colin's data were not considered stable, and toward the end of baseline, the number of non-zero data points increased, showing a slight accelerating trend. Colin's mean duration of challenging behaviors during baseline was 0.2 minutes per hour, with a standard deviation of 0.6, and a range of 0.0 to 3.3. During the intervention, data for both token boards increased and became more variable. When the STB_1 was implemented, Colin's mean duration of challenging behaviors was 1.0 minutes per hour, with a standard deviation of 2.0, and a range of 0.0 to 7.6. When the $RRTB_1$ was implemented, mean duration of challenging behaviors was 1.0 minutes per hour, with a standard deviation of 1.6, and a range of 0.0 to 5.6. When Phase 2 of the intervention began, data remained variable with no discernible trends for either token board. When the STB_2 was implemented, mean duration of challenging behaviors was 1.0 minutes per hour, with a standard deviation of 1.4, and a range of 0.0 to 4.0 minutes per hour. When the $RRTB_2$ was implemented, mean duration of challenging behaviors was 0.8 minutes per hour, with a standard deviation of 1.5, and a range of 0.0 to 4.6 minutes per hour. When Phase 3 was implemented with only the $RRTB_2$, mean duration of challenging behaviors continued to decrease at 0.2 minutes per hour, with a standard deviation of 0.5, and a range of 0.0 to 1.3 minutes per hour.

Prior to the start of intervention, therapists did not collect data on the duration of Mandy's challenging behaviors. During the intervention, data for both token boards were variable and showed an increasing trend until session 10 of intervention. Data then decreased starting in session 11. When the STB_1 was implemented, Mandy's mean duration of challenging behaviors was 5.2 minutes per hour, with a standard deviation of 11.9, and a range of 0.0 to 38.4.

When the $RRTB_1$ was implemented, mean duration of challenging behaviors was 6.1 minute per hour, with a standard deviation of 8.4, and a range of 0.0 to 27.0. When Phase 2 of the intervention was implemented, data for both token boards continued to decrease, but remained variable. When the STB_2 was used, Mandy's mean duration of challenging behaviors was 1.3 minutes per hour, with a standard deviation of 2.7, and a range of 0.0 to 8.5. When the $RRTB_2$ was implemented, mean duration of challenging behaviors was 1.0 minute per hour, with a standard deviation of 1.7, and a range of 0.0 to 6.0. When Phase 3 was implemented with only the STB_2, mean duration of challenging behaviors was 0.9 minutes per hour, with a standard deviation of 2.2, and a range of 0.0 to 6.5 minutes per hour.

Figure 3.6

Duration of Challenging Behaviors Per Hour for Colin and Mandy



Note. * = Revised reinforcement rate for $RRTB_2$.

Chapter 4: Discussion

One goal of this study was to demonstrate how ABA clinics could participate in translational practice by incorporating behavior analytic principles from the basic literature into existing treatment frameworks, such as reinforcement systems. To that end, this study largely succeeded. Specifically, this study compared the effects of two token reinforcement systems during ABA therapy sessions, one using a fixed schedule of token exchange, reflecting token reinforcement systems common in ABA clinics, and the other using a random schedule of token exchange. In general, the two boards produced different behavioral patterns, suggesting that token board arrangement does influence clinical progress.

In the first intervention phase, we aimed to arrange the $RRTB_1$ to decrease the predictability of the delay to reinforcement, but surprisingly the $RRTB_1$ did not outperform the STB_1 in any dependent variable measuring participant performance. Throughout the first phase of the intervention, we discovered a confound with the $RRTB_1$, which actually caused the token board to work counter to the theory that inspired it. The arrangement of $RRTB_1$ resulted in Mandy fixating on the blue tokens, consistent with a phenomenon known as sign tracking. In sign tracking, the organism will attend to stimuli that signal reinforcement, sometimes at the expense of the actual reinforcer (see Witts & Harri-Dennis, 2015). When the $RRTB_1$ was implemented with Mandy, many of the antecedents of her challenging behaviors were related to drawing a red token instead of a blue token or the therapist denying the opportunity to peek inside the cup when she was drawing her token. The interference from the blue token fixation was so troublesome that during one overlap by the primary researcher Mandy was observed tantrumming from receiving a red token as her 10^{th} and final token on that board. In other words,

despite having just earned her reinforcement, Mandy was so fixated on the blue tokens that even the opportunity for reinforcement derived from red tokens was enough to induce challenging behavior. Because our research question aimed to determine the effects of a token board whose tokens did not signal differing delays to reinforcement, we changed the arrangement of the RRTB halfway through the study.

Revising the RRTB in Phase 2 of the study had the following effects on the data. First, Colin's trial rate initially increased when implementing the $RRTB_2$ compared to the $RRTB_1$, the STB_1, and the STB_2; however, once we revised the reinforcement rate in session 70 to better match that of the $RRTB_1$, we observed a decreasing trend in trial rate when using the $RRTB_2$. Second, we saw a decrease in level and a change in trend for Mandy's probe trial average duration with the $RRTB_2$ compared to the $RRTB_1$, suggesting Mandy was responding more quickly when the $RRTB_2$ was implemented. Anecdotally, the primary researcher observed many requests for the blue token during Phase 1, but similar requests (e.g., to be done early) did not occur during Phase 2, which could have contributed to the decrease in average probe trial duration. Lastly, the duration of Mandy's challenging behaviors continued to decrease for both boards once Phase 2 began. Additional data collected on the antecedents of Mandy's challenging behaviors also suggests that fewer challenging behaviors were preceded by RRTB-related antecedents (e.g., choosing a red token instead of a blue one, therapist denying access to peeking inside the cup, therapist shaking cup of tokens). In Phase 1, 10 RRTB-related antecedents were recorded, while in Phase 2, there was only one recorded (i.e., denial of the $RRTB_2$ on a day that the STB_2 had been selected). These data suggest that changing the arrangement of the RRTB from relying on different colored tokens to signal early reinforcement to using a randomly-

generated number list, resulted in a decrease in challenging behaviors, as well as a decrease in average trial duration during controlled probes for one of the participants.

Phase 3 consisted of a two-week extension of the best performing token board for each participant in an attempt to wash out contrast effects, a limitation of the alternating treatments design. However, two weeks might not have been sufficient for this purpose. For Colin, we continued the $RRTB_2$ and terminated the STB_2 because although Colin showed a preference for the STB_2, Colin was exhibiting more frequent and slightly longer challenging behaviors when the STB_2 was in place. During Phase 2, average probe trial duration, percentage of trials correct, and trial rate remained comparable between the STB_2 and the $RRTB_2$. When Phase 3 was implemented with the $RRTB_2$, trial rate did not continue on a decreasing trend. Instead, trial rate increased compared to the last four data points for the $RRTB_2$ in Phase 2. Interestingly, percentage of correct trials immediately increased at the beginning of Phase 3 but continued on a decreasing trend for the remainder of the two weeks. It is unclear if this would have continued if Phase 3 had been extended longer than two weeks, but future researchers might consider a longer wash out period.

For Mandy, we continued the STB_2 and terminated the $RRTB_2$ because even though Mandy showed preference for the $RRTB_2$, trial rate was far lower with the $RRTB_2$ than with the STB_2. Despite seeing a slightly lower average probe trial duration with the $RRTB_2$ compared to the STB_2, this trial speed was not translating to a quicker trial rate with the $RRTB_2$ when we examined full session data. Considering percentage of correct trials and frequency and duration of challenging behaviors was comparable between the STB_2 and the $RRTB_2$, we decided it was clinically necessary to extend the STB_2 with Mandy and terminate the $RRTB_2$. However, when

we extended the STB_2, we noticed even more variability in trial rate, suggesting contrast effects may have played a part in the data we obtained in Phase 2.

One limitation of this study is that we saw misalignment between participants' preferences and their performance with the two token boards. Although participant preference was considered when selecting the token board to extend in Phase 3, in the end, researchers decided to implement the token board that either supported less challenging behavior or supported a higher rate of trial completion for the last two weeks. The misalignment between preference and performance suggests that more work can be done to optimize these token boards and ensure participant preference is consistent with the more effective treatment, as the ultimate goal is for therapy to be both effective and enjoyable for our clients. Future research should continue to assess participant choice and explore why participants like Colin might show preference for the treatment that evokes more challenging behavior.

Another limitation of this study is the lack of control that accompanies research conducted in a non-university-based clinic. For example, due to staffing shortages, therapists were occasionally required to complete additional clinic tasks during direct sessions while participants were on longer breaks, which surely contributed to the variability in trial rate. Additionally, during the study, variables such as the number of trials conducted for each target skill and the schedule of reinforcement for maintenance targets and attending skills were not controlled. When the $RRTB_2$ was implemented, we were able to collect data on the number of tokens delivered during each session. We then calculated the average number of trials implemented per token delivered (range = 0.58 trials/token to 1.47 trials/token). From these data, it was clear that during some sessions, staff provided high rates of non-trial tokens for behaviors

such as appropriate attending skills, while on other days, staff likely implemented more maintenance trials, leading to fewer tokens delivered per trial. The current study was designed to test the effects of two types of token boards under real-world conditions; therefore, the fact that we saw unique behavioral patterns with each of the token boards even with limited control could be considered a strength. However, future research should be conducted in a setting that can better control the number and types of trials implemented and the rate of reinforcement for each response.

Applied research is also vulnerable to extraneous variables, such as staffing changes, illness, and vacations, all three of which occurred between the end of baseline and the start of intervention for both participants in this study. For example, Mandy's new therapist started on the fifth session of intervention (session 25), which had the highest rate of challenging behaviors during Phase 1. In addition, the five sessions with the highest rates of challenging behaviors throughout the intervention all occurred while the new staff was working with Mandy. Because this staff did not work with Mandy during baseline, staffing changes were likely a contributor to the higher rates of challenging behaviors throughout the intervention.

One limitation affecting Colin's data is that Colin's treatment plan included multiple objectives that used teaching techniques other than DTT (e.g., chaining), but still involved token reinforcement systems. Because having Colin select his own tokens would interrupt the behavior chains therapists were trying to teach, Colin's previous token board remained in place for programs such as independent play and simple conversation so that therapists could deliver tokens less intrusively throughout the chain. After baseline, staff tended to reduce the time spent implementing chaining programs, which therefore increased DTT trial rate in intervention. This

difference in time spent in chaining programs surely accounts for some of the differences in trial rate for Colin between baseline and intervention. As a confound, because more time was being devoted to DTT programs, it would make sense that those DTT programs would enjoy higher accuracies as Colin was practicing those skills more often.

A limitation of an alternating treatments design is the possibility of a contrast effect between the two treatments. When one of the two treatments is implemented on its own, it may be less effective than when compared directly with an alternative. A contrast effect could manifest either within the client's behavior or within the therapist's behavior. For example, by using this design, therapists might experience a contrast effect in how they manage their time during therapy sessions. Knowing there will be multiple opportunities to use each token board throughout the week, the therapist may conduct sessions at a slower pace with one board compared to the other (e.g., spending more time on reinforcer breaks, delivering more tokens for attending skills); however, if this study was conducted using an ABABC design, in which each phase was 3 months, contrast effects could be avoided, allowing researchers more control over day-to-day therapist behavior.

Future research should also explore the effects of RR and FR token-exchange schedules while varying the average token-exchange schedule used for the RRTB. One possibility as to why the RRTB did not outperform the STB across more variables is that the average rate of reinforcement on the RRTB was too dense in this study. This could have resulted in early reinforcement becoming an expectation instead of an unlikely surprise, producing frustration when the participant did not receive early reinforcement under the RRTB. If the schedule of reinforcement was leaner, we might not have seen Mandy's sign tracking or her increased rate

and duration of challenging behaviors when the RRTB was implemented. In addition, with a leaner reinforcement schedule for the RRTB, less time would have been spent in reinforcement, and therefore, full-session trial rate might have been more comparable between the STB and RRTB, while other variables may have continued to favor the RRTB. Future researchers could even use a RRTB with a higher maximum number of tokens (e.g., 15) than that of the STB (e.g., 10) and set the average token-exchange schedule for the RRTB to one that would result in equal reinforcement rates for both token boards. In this arrangement, performance data and participants' choice selections would not be influenced by the RRTB having a higher rate of reinforcement.

By changing the method by which therapists determined when early reinforcement would occur from the $RRTB_1$ arrangement with the different colored tokens to the $RRTB_2$ with the randomly-generated number list, we avoided creating an additional stimulus that signaled immediate or delayed reinforcement. Future research should continue to use a method similar to the randomly-generated number list, in which only the staff knows when early reinforcement is coming. One drawback to the random number list though, was that because staff knew ahead of time how many tokens they were going to deliver, sometimes they chose what trials to implement based on that number, despite being advised against doing so by the primary researcher. For example, when a low number of tokens was prescribed, sometimes therapists would implement more difficult or longer programs, perhaps to get through the difficult ones quickly. On the other hand, sometimes staff would deliver all tokens for attending skills only, possibly deciding it was not worth setting up a program for just one or two trials. Both of these staff behaviors could affect the trial rate or the percentage of correct trials when the RRTB is

implemented; therefore, a possible solution to this problem would be to use a data collection software that could tell staff in the moment when to end the token board. For example, after the therapist inputs data for the 6th trial, a message could appear telling the staff that the 6th trial was the last trial of that token board.

The results of this study are consistent with the results of other translational efforts to study exchange production schedules in token economies. Consider Holmes et al. (2022) and Argueta et al. (2019), both of which replicated basic research designs in applied settings, and found inconclusive results, despite previous research showing clear results under tightly controlled conditions. Hackenberg (2018) argued that token economy research is well suited for a bi-directional translational approach, where token economies can be revised in applied settings to account for recent advances in the basic literature, and in return, new applied questions can inspire additional laboratory and efficacy research for further optimization. Therefore, future research should focus on the efficacy of FR and RR token exchange schedules using more controlled arrangements until we come to a better understanding of how RR token-exchnage schedules affect response rates in human participants, and what, if any, adverse side effects (e.g., sign tracking) might appear. Although, token reinforcement systems have been used in applied settings since the inception of our field (Ayllon & Azrin, 1965), we have only just begun to scratch the surface on the variety of manipulations that can be made to token systems to improve performance in applied settings.

References

Argueta, T., Leon, Y., & Brewer, A. (2019). Exchange schedules in token economies: A preliminary investigation of second-order schedule effects. *Behavioral Interventions, 34*(2), 280–292. https://doi.org/10.1002/bin.1661

Ayllon, T., & Azrin, N. H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior, 8*(6), 357–383. https://doi.org/10.1901/jeab.1965.8-357

Ayllon, T., & Michael, J. (1959). The psychiatric nurse as a behavioral engineer. *Journal of the Experimental Analysis of Behavior, 2*(4), 323-334. https://doi.org/10.1901/jeab.1959.2-323

Bullock, C. E., and Hackenberg, T. D. (2015). The several roles of stimuli in token reinforcement. *Journal of the Experimental Analysis of Behavior, 103*(2), 269-287. https://doi.org/10.1037/e625362007-001

Fantino, E. (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12*(5), 723-730. https://doi.org/10.1901/jeab.1969.12-723.

Fantino, E. (2008). Choice, conditioned reinforcement, and the Prius effect. *The Behavior Analyst, 31*(2), 95-111. https://doi.org/10.1007/BF03392164.

Fantino, E., Preston, R. A., & Dunn, R. (1993). Delay reduction: Current status. *Journal of the Experimental Analysis of Behavior, 60*(1), 159-169. https://doi.org/10.1901/jeab.1993.60-159

Fantino, E., & Romanowich, P. (2007). The effects of conditioned reinforcement rate on choice: A review. *Journal of the Experimental Analysis of Behavior, 87*(3), 409–421. https://doi.org/10.1901/jeab.2007.44-06

Foster, T. A., Hackenberg, T. D., & Vaidya, M. (2001). Second-order schedules of token reinforcement with pigeons: Effects of fixed- and variable-ratio exchange schedules. *Journal of the Experimental Analysis of Behavior, 76*(2), 159–178. https://doi.org/10.1901/jeab .2001.76-159

Hackenberg, T. D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior, 91*(2), 257-286. https://doi.org/10.1901/jeab.2009.91-257

Hackenberg, T. D. (2018). Token reinforcement: Translational research and application. *Journal of Applied Behavior Analysis, 51*(2), 393-435. https://doi.org/10.1002/jaba.439

Holmes, S. C., Lozy, E. D., Ruby, A. M., & Donaldson, J. M. (2022). Effects of and preference for fixed-ratio and time-based exchange-production schedules in a token system. *Behavior Analysis: Research and Practice, 22*(1), 1-13. https://doi.org/10.1037/bar0000220

Homme, L., C'de Baca, P., Cottingham, L., Homme, A. (1968). What behavioral engineering is. *The Psychological Record, 18*(3), 425–434. https://doi.org/10.1007/BF03393790

Ivy, J. W., Meindl, J. N., Overley, E., & Robson, K. M. (2017). Token economy: A systematic review of procedural descriptions. *Behavior Modification, 41*(5), 708-737. https://doi.org/10.1177/0145445517699559

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis, 5*(3), 343-372. https://doi.org/10.1901/jaba.1972.5-343

Kelleher, R. T., & Gollub, L. R. (1962). A review of positive conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 5*, 543-597. https://doi.org/10.1901/jeab.1962.5-s543

Lattal, A. (2015, September 2). Bad behavioral engineering. *Aubrey Daniels International.* https://www.aubreydaniels.com/blog/bad-behavioral-engineering

Ledford, J. R., Lane, J. D., & Severini, K. E. (2017). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment, 19*(1), 4-17. https://doi.org/10.1017/BrImp.2017.16

Mace, F. C., & Critchfield, T. S. (2010). Translational research in behavior analysis: Historical traditions and imperative for the future. *Journal of the Experimental Analysis of Behavior, 93*(3), 293–312. https://doi.org/10.1901/jeab.2010.93-293

Merriam-Webster. (n.d.). Engineer. In *Merriam-Webster.com dictionary*. Retrieved November 5, 2021, from https://www.merriam-webster.com/

Meyer, S. F., Schley, D. R., & Fantino, E. (2011). The role of context in risky choice. *Behavioural Processes, 87*, 100-105. https://doi.org/10.1016/j.beproc.2011.01.010.

Waddell, T. R., Leander, J. D., Webbe, F. M., & Malagodi, E. F. (1972). Schedule interactions in second-order fixed-interval (fixed-ratio) schedules of token reinforcement. *Learning and Motivation, 3*(1), 91–100. https://doi.org/10.1016/0023-9690(72) 90050-1

Webbe, F. M., & Malagodi, E. F. (1978). Second-order schedules of token reinforcement: Comparisons of performance under fixed-ratio and variable ratio exchange schedules.

Journal of the Experimental Analysis of Behavior, 30(2), 219–224. https://doi.org/10.1901/jeab.1978.30-219

Witts, B. N., & Harri-Dennis, E. (2015). Free-operant research in the experimental analysis of human slot machine gambling. *Analysis of Gambling Behavior, 9*(2), 59-70. https://repository.stcloudstate.edu/agb/vol9/iss2/2

Appendix A

$RRTB_2$ Random Number Lists

Figure A1

Original $RRTB_2$ Random Number List

Date, AM/PM, Initials	Number of tokens	Check	Date, AM/PM, Initials	Number of tokens	Check	Date, AM/PM, Initials	Number of tokens	Check
	8			10			8	
	3			3			1	
	10			4			10	
	1			10			10	
	10			10			10	
	8			10			10	
	7			8			1	
	2			7			6	
	8			4			10	
	10			9			10	
	8			10			6	
	9			10			10	
	10			8			10	
	10			10			10	
	6			2			1	
	10			10			8	
	10			9			5	
	10			4			6	
	1			2			10	
	8			10			2	
	10			10			10	
	4			9			6	
	3			10			10	
	3			8			4	
	1			9			10	
	1			5			10	
	8			1			7	
	10			3			2	
	10			5			10	
	10			5			1	
	10			10			10	
	8			10			4	
	10			10			10	
	4			8			10	

Figure A2

Revised $RRTB_2$ Random Number List

Date, AM/PM, Initials	Number of tokens	Check	Date, AM/PM, Initials	Number of tokens	Check	Date, AM/PM, Initials	Number of tokens	Check
	10			8			4	
	10			7			3	
	9			5			1	
	6			2			10	
	10			9			5	
	5			1			10	
	1			10			6	
	7			3			10	
	2			10			7	
	8			4			10	
	10			9			5	
	7			3			10	
	2			10			7	
	5			1			10	
	10			6			2	
	1			10			6	
	4			10			9	
	3			10			8	
	9			5			8	
	3			10			10	
	8			4			10	
	6			2			10	
	5			1			7	
	2			10			2	
	10			6			10	
	7			3			9	
	4			10			4	
	10			8			1	
	9			5			8	
	3			10			3	
	10			7			10	
	6			2			10	
	8			4			6	
	1			10			9	
	4			10				

Appendix B

Example of Phase 1 Token Board Research Study Instructions for Mandy

Purpose

The purpose of this study is to determine whether providing clients more autonomy during token reinforcement by allowing clients to draw their own tokens will affect clients' rate of responding, rate of acquisition, or rate of challenging behavior. A secondary purpose of this study is to determine whether clients prefer consistent, predictable reinforcement systems or randomized reinforcement systems.

Two Token Boards

Each day you will use one of the two token boards for the entire session.

The white token board will include a cup of 30 white tokens. Each time the client earns a token, the client will be able to select a token from the cup and put it on the token board. The white token board requires 10 tokens for the client to earn the reinforcer.

The blue token board will include a cup of 28 red tokens and 2 blue tokens. Each time the client earns a token, the client will be able to select a token from the cup and put it on the token board. The blue token board requires *either* 10 red tokens *or* 1 blue token for the client to earn the reinforcer. If the client draws a blue token at any point, the DTT session is over and the client gets the reinforcer.

Instructions

Intervention

1. At the beginning of each shift, you will randomly select one piece of paper from a bag of four, which says "Draw." The paper will tell you which token board to use during that shift.
    a. Record the selection on the data sheet.
    b. Put the piece of paper in the bag marked "Discard."
2. Use the selected token board for all of the client's designated DTT programs (see Program List).
3. Use the correct cup of tokens that matches the board.
    a. If the white board was selected, use the white cup, and ensure it has 30 white tokens.
    b. If the blue board was selected, use the blue cup, and ensure it has 28 red tokens and two blue tokens.
4. Before each DTT session, establish a reinforcer.

5. During each DTT session, present S^Ds, prompt, correct, and reinforce according to program write-ups.
    a. Acquisition responses are reinforced on a continuous token schedule for prompted or independent responses.
    b. Maintenance responses are reinforced on an intermittent token schedule for independent responses only.
    c. Provide intermittent tokens for attending skills (e.g., hands down, feet on floor, eye contact)
    d. Provide behavior specific praise after each correct response.
    e. Any incorrect response (or no response to a maintenance target within 3-5 seconds) results in the error correction procedure and no token.
6. When the client earns a token, tell them they can choose a token, and hold the cup of tokens up high enough that the client can't see inside.
7. Allow 5 seconds for the client to choose a token.
8. After 5 seconds, if the client has not chosen a token, provide a vocal prompt (e.g., "Time to choose one.")
9. If the client still does not choose a token, randomly select a token from the cup for them, and place it on the token board.
10. If the client chooses more than one token, vocally prompt them to drop the tokens back in the cup and try again. You can remind them to "Choose just one."
11. Repeat steps 4-10 until the client receives the correct number of tokens.
12. Deliver the back-up reinforcer after the client has 10 red or white tokens or after the client draws 1 blue token, whichever comes first.
13. During the session, record skill acquisition data as normal.
14. Record frequency of challenging behavior on the histogram, and highlight the box if the challenging behavior occurs during the DTT session (i.e., between delivery of the first instruction and delivery of the reinforcer).
15. When upsets or work avoidance occurs, record the duration on the histogram, highlighting the box if the upset occurred during the DTT session.

Probes

Probes will be conducted once per shift to measure the duration of a DTT session and the number of trials and error corrections presented. Probes will be implemented using only acquisition responses from the client's __Follows Verb Instructions ___ program. Do not provide tokens for attending skills. The same token board that was selected at the beginning of the session should be used for the probe that day. If using the blue token board that day, ensure the token cup only has red tokens during the probe.

1. Conduct one probe per day.
2. Use the same token board and token cup that was selected at the beginning of the shift.
3. If the blue token board was selected for the day, **when the client is not looking, remove the two blue tokens from the cup and keep them out of sight**. (The client should not know that the blue tokens have been removed.)

4. Before each DTT session, establish a reinforcer.
5. Immediately before providing the first S^D, **start the stopwatch** to record the duration of the probe session.
6. Present only acquisition trials from the client's __Follows Verb Instructions___ program.

7. During each DTT session, present S^Ds, prompt, correct, and reinforce according to program write-ups.
    a. Acquisition responses are reinforced on a continuous token schedule for prompted or independent responses.
    b. Provide behavior specific praise after each correct response.
    c. Any incorrect response results in the error correction procedure and no token.
8. **Avoid providing tokens for attending skills or bonus tokens during probes.**
9. When the client earns a token, tell them they can choose a token, and hold the cup of tokens up high enough that the client can't see inside.
10. Allow 5 seconds for the client to choose a token.
11. After 5 seconds, if the client has not chosen a token, provide a vocal prompt (e.g., "Time to choose one.")
12. If the client still does not choose a token, randomly select a token from the bag for them, and place it on the token board.
13. If the client chooses more than one token, vocally prompt them to drop the tokens back in the cup and try again. You can remind them to "Choose just one."
14. Repeat steps 6-13 until the client receives 10 red or 10 white tokens. Then deliver the back-up reinforcer.
15. During the session, keep a tally of the number of trials presented.
    a. **Each error correction procedure in which the S^D is re-presented will count as two trials.**
    b. **If the error correction procedure is implemented incorrectly, still record the number of trials according to how many S^Ds were presented.**
        i. E.g., If the therapist forgets to re-present the S^D following the first half of the error correction procedure, count as one trial.
    c. **If you need to repeat the S^D after attending has been lost, that will count as a new trial.**
16. During the session, keep a tally of the number of error correction procedures implemented.
17. During the session, record skill acquisition data as normal.
18. Record frequency of challenging behavior on the histogram, and highlight the box if the challenging behavior occurs during the DTT session (i.e., between delivery of the first instruction and delivery of the reinforcer).
19. When upsets or work avoidance occurs, record the duration on the histogram, highlighting the box if the upset occurred during the DTT session.
20. Immediately after delivering the reinforcer, **stop the stopwatch** and record the duration of the probe on the data sheet.

Appendix C

Example of Phase 2 Token Board Research Study Instructions for Mandy

Purpose

The purpose of this study is to determine whether providing clients more autonomy during token reinforcement by allowing clients to draw their own tokens will affect clients' rate of responding, rate of acquisition, or rate of challenging behavior. A secondary purpose of this study is to determine whether clients prefer consistent, predictable reinforcement systems or randomized reinforcement systems.

Two Token Boards

Each day you will use one of the two token boards for the entire session.

The green token board will include a green cup of green tokens. Each time the client earns a token, the client will be able to select a token from the cup and put it on the token board. The green token board requires 10 tokens for the client to earn the reinforcer.

The yellow token board will include a yellow cup of yellow tokens. The yellow token board functions the same as the green token board except that the therapist will use a list of randomly generated numbers to determine how many tokens to deliver during each DTT session. The yellow token board will still always start with 0 tokens on the board, but the client won't always be required to fill up the board completely before earning their reinforcer.

Instructions

Intervention

16. At the beginning of each shift, you will randomly select one piece of paper from a bag of four, which says "Draw." The paper will tell you which token board to use during that shift.
    a. Record the selection on the data sheet.
    b. Put the piece of paper in the bag marked "Discard."
17. Use the selected token board for all of the client's designated DTT programs (see Program List).
18. Use the correct cup of tokens that matches the board.
    a. If the green board was selected, use the green cup, and ensure it has only green tokens.
    b. If the yellow board was selected, use the yellow cup, and ensure it has only yellow tokens.
19. Before each DTT session, establish a reinforcer.

20. If using the yellow token board, refer to the randomly generated number list to determine how many tokens you will deliver that session. Do not base the program you run off of how many tokens should be delivered during the session (e.g., do not choose a more difficult program after seeing that the next session will only require 3 tokens). Try to choose the program before referring to the randomly generated number list.
21. During each DTT session, present S^Ds, prompt, correct, and reinforce according to program write-ups.
    a. Acquisition responses are reinforced on a continuous token schedule for prompted or independent responses.
    b. Maintenance responses are reinforced on an intermittent token schedule for independent responses only.
    c. Provide intermittent tokens for attending skills (e.g., hands down, feet on floor, eye contact)
    d. Provide behavior specific praise after each correct response.
    e. Any incorrect response (or no response to a maintenance target within 3-5 seconds) results in the error correction procedure and no token.
22. When the client earns a token, tell them they can choose a token from the cup.
23. Allow 5 seconds for the client to choose a token.
24. After 5 seconds, if the client has not chosen a token, provide a vocal prompt (e.g., "Time to choose one.")
25. If the client still does not choose a token, randomly select a token from the cup for them, and place it on the token board.
26. If the client chooses more than one token, vocally prompt them to drop the extra tokens back in the cup and try again. You can remind them to "Choose just one."
27. Repeat steps 4-10 until the client receives the correct number of tokens.
28. Deliver the back-up reinforcer after the client has 10 green or yellow tokens or after you've delivered the designated number of tokens listed on the randomly generated number list (for the yellow board only).
    a. If the randomly generated number list calls for fewer than 10 tokens, you can end the session by telling the participant "You've done such a good job! You can get your (reinforcer) early!" or a variation of this statement.
29. During the session, record skill acquisition data as normal.
30. Record frequency of challenging behavior on the histogram, and highlight the box if the challenging behavior occurs during the DTT session (i.e., between delivery of the first instruction and delivery of the reinforcer).
31. When upsets or work avoidance occurs, record the duration on the histogram, highlighting the box if the upset occurred during the DTT session.

Probes

Probes will be conducted once per day to measure the duration of a DTT session and the number of trials and error corrections presented. Probes will be implemented using only acquisition responses from the client's __Follows Verb Instructions ___ program. Do not provide tokens for attending skills. The same token board that was selected at the beginning of the session should be

used for the probe that day. If using the yellow token board that day, continue running trials until all 10 tokens have been delivered (do not refer to the randomly generated number list for the probe).

21. Conduct one probe per day.
22. Use the same token board and token cup that was selected at the beginning of the shift.
23. Before each DTT session, establish a reinforcer.
24. Immediately before providing the first S^D, **start the stopwatch** to record the duration of the probe session.
25. Present only acquisition trials from the client's __Follows Verb Instructions__ program.
26. During each DTT session, present S^Ds, prompt, correct, and reinforce according to program write-ups.
    a. Acquisition responses are reinforced on a continuous token schedule for prompted or independent responses.
    b. Provide behavior specific praise after each correct response.
    c. Any incorrect response results in the error correction procedure and no token.
27. **Avoid providing tokens for attending skills or bonus tokens during probes.**
28. When the client earns a token, tell them they can choose a token from the cup.
29. Allow 5 seconds for the client to choose a token.
30. After 5 seconds, if the client has not chosen a token, provide a vocal prompt (e.g., "Time to choose one.")
31. If the client still does not choose a token, randomly select a token from the cup for them, and place it on the token board.
32. If the client chooses more than one token, vocally prompt them to drop the extra tokens back in the cup and try again. You can remind them to "Choose just one."
33. Repeat steps 6-13 until the client receives 10 green or 10 yellow tokens. Then deliver the back-up reinforcer.
34. During the session, keep a tally of the number of trials presented.
    a. Each error correction procedure in which the S^D is re-presented will count as two trials.
    b. If the error correction procedure is implemented incorrectly, still record the number of trials according to how many S^Ds were presented.
        i. E.g., If the therapist forgets to re-present the S^D following the first half of the error correction procedure, count as one trial.
    c. If you need to repeat the S^D after attending has been lost, that will count as a new trial.
35. During the session, keep a tally of the number of error correction procedures implemented.
36. During the session, record skill acquisition data as normal.
37. Record frequency of challenging behavior on the histogram, and highlight the box if the challenging behavior occurs during the DTT session (i.e., between delivery of the first instruction and delivery of the reinforcer).
38. When upsets or work avoidance occurs, record the duration on the histogram, highlighting the box if the upset occurred during the DTT session.

39. Immediately after delivering the reinforcer, **stop the stopwatch** and record the duration of the probe on the data sheet.

Reminders

- Record DTT data for shift 2 in a new column on both the acquisition and maintenance pages. Mark AM or PM with the date and your initials in each column.

- There's an AM and a PM histogram now.

- Highlight histogram box if behavior occurred during work session with blue or white token board (between instruction to transition and delivery of reinforcer). Do not highlight if behavior occurred during break, snack, reinforcer, or program that uses a different reinforcement system (e.g., tangibles only)

Troubleshooting checklist for increased challenging behaviors

Proactive Strategies

- Make sure you have a strong reinforcer.

- Ask the client if she wants to select the tokens herself or if she wants the therapist to choose them. You can ask this as often as needed, and you can switch during the token board. Honor her request at any time unless she continues to want to select them herself, but this continuously results in challenging behaviors (See below).

- If the client is choosing tokens herself, remind her to select only one token at a time (you can give this reminder as often as needed). You can also remind the client that she can earn extra or earn treats for selecting only one at a time.

- Make sure you're providing bonus tokens, treats, and enthusiastic praise for calm behavior/selecting one token during non-probe token boards

    - Aim for 2-3 bonus tokens or edibles during the first token board of the day and at least one bonus token or edible during all other non-probe token boards.

    - Continue to praise these behaviors often (>3 times per token board).

- If the client is asking about getting done early/fewer tokens/etc., say, "Usually we fill up our whole board, but sometimes we get done early. It's a surprise!"

Reactive Strategies

1. If the client ever protests, turns away, etc. ask the client if she wants to pick the tokens or if the therapist should pick them.

2. If the client is continuing to request that she chooses the tokens herself, but this is continuously leading to challenging behaviors (3+ in a shift), the therapist should deliver the tokens for the rest of the shift by randomly selecting tokens from the cup and putting them on her token board.

3. If during a probe, the client has an avoidance episode or upset longer than 10 minutes, you may choose to end the probe and finish the token board by reinforcing a mix of attending skills and maintenance responses. Just mark an X in the “Invalid?” box so we can exclude that data.

Appendix D

Behavioral Skills Training Role-Play Script

Role play option 1

- 2 errors:
  - 2 incorrect responses
- 2 discrete challenging behaviors
  - 1 vocal protest
  - 1 aggressive behavior (e.g., hit, kick)
- 1 trial in which the "client" takes more than 5 seconds to choose a token from the bag
  - "Client" will respond to a vocal prompt to select a token

Role play option 2

- 3 errors:
  - 2 trials of no response
  - 1 incorrect response
- 1 persistent challenging behavior
  - 1 60-second tantrum
- 2 trials in which the "client" takes more than 5 seconds to choose a token from the bag
  - "Client" will not respond to a vocal prompt to select a token

Role play option 3

- 1 error
  - 1 incorrect response
- 1 discrete challenging behavior
  - 1 vocal protest
- 1 persistent challenging behavior
  - 1 30-second tantrum
- 2 trials in which the "client" takes more than 5 seconds to choose a token from the bag
  - 1 trial "client" will not respond to a vocal prompt to select a token
  - 1 trial "client" will respond to a vocal prompt to select a token

Appendix E

Sample of Staff Training Checklist

Intervention: DTT Block	**Role-play 1**	**Role-play 2**	**Role-play 3**	**Percentage Correct**
Uses correct token board				
Uses correct cup of tokens				
Establishes a reinforcer prior to DTT session				
Uses FR1 schedule for acquisition trials (independent or prompted)				
Uses intermittent schedule for maintenance trials (independent only)				
Provides behavior specific praise after each correct trial				
Implements error correction procedure following incorrect responses or no response				
Holds cup high enough that participant doesn't see inside				
Allows participant 5 seconds to select a token from the cup				
If client has not selected a token, provides a vocal prompt and allow another 5 seconds				
If the client has not selected a token, therapist chooses a token a puts it on the board				
Records frequency of problem behavior (and highlights it) within 10 seconds of occurrence				
Records duration of problem behavior (and highlights it) within 10 seconds of the end of the behavior				
Percentage Correct				

Probe	**Role-play 1**	**Role-play 2**	**Role-play 3**	**Percentage Correct**
Implements correct program for probe				
Uses correct token board				
Uses cup with all red tokens				
Ensures client isn't looking when removing blue tokens				
Uses FR1 schedule of token delivery				
Avoids maintenance trials				
Avoids token delivery for attending skills				
Provides behavior specific praise after each trial				
Holds cup high enough that participant doesn't see inside				
Allows participant 5 seconds to select a token from the cup				
If client has not selected a token, provides a vocal prompt and allow another 5 seconds				
If the client has not selected a token, therapist chooses a token a puts it on the board				
Records frequency of problem behavior (and highlights it) within 10 seconds of occurrence				
Starts stopwatch within 3 seconds of delivering first instruction				
Stops stopwatch within 3 seconds of delivering back-up reinforcer				
Records total duration of probe session				
Keeps tally of number of trials presented				
Keeps tally of number of error correction procedures implemented				
Percentage Correct				

Appendix F

Sample Token Board Selection Data Sheet for Phase 1

Instructions: Draw a piece of paper from the bag labeled "Draw" at the beginning of the shift. Record the date (always include "AM" or "PM") and the token board selection. Put the piece of paper in the bag labeled "Discard." At the beginning of every 5th session (highlighted), put the two token boards and their corresponding materials in front of the client and allow them to choose which token board they want to use that day (e.g., "Which one should we use today?"). Record the client's selection in the highlighted boxes. At the beginning of the 5th session, also switch all the papers from the "Discard" bag to the "Draw" bag.

Date	Token Board Selected (White/Blue)	Date	Token Board Selected (White/Blue)

Appendix G

Sample Probe Data Sheet

Date/(AM/PM)	Duration of Probe	# of Trials	# of Error Corrections	Challenging Bx Y/N?	Invalid?
4/25/22 AM	4:07	10		N	
4/28/22 PM	5:03	14	2	N	
4/29/22 PM	3:53	10		Y	
4/30/22 AM	4:36	12	1	N	X
4/30/22 PM	4:10	10		N	

Appendix H

Example of DTT TI Data Sheet for Colin

Intervention: DTT Block	**Block 1**	**Block 2**	**Block 3**	**Block 4**	**Block 5**
Uses correct token board					
Uses correct cup of tokens					
Delivers correct number of tokens (Phase 2 only)					
Establishs a reinforcer prior to DTT session					
Uses FR1 schedule for acquisition trials (independent or prompted)					
Uses intermittent schedule for maintenance trials (independent only)					
Provides at least one bonus token/edible and praise for choosing one token if all 10 are delivered					
Provides prompt and reinforcement following no response to acquisition target					
Implements error correction procedure following incorrect responses, aberrant behavior, or no response to maintenance target					
Provides behavior specific praise after each correct trial					
Holds cup high enough that participant can't see inside (Phase 1 only)					
Allows participant 5 seconds to select a token from cup					
If the client has not selected a token, therapist provides vocal or physical prompt					
If p.d., therapist delivers 1-5 simple instructions to regain compliance					
If p.d. or 3+ instances of grabbing a handful of tokens, therapist delivers the rest of the tokens, choosing randomly					
Records frequency of problem behavior within 30 seconds of the end of the token board					
Records duration of problem behavior within 30 seconds of the end of the token board					
Highlights histogram box if bx occurred					
Percentage Correct					

Appendix I

Example of Probe TI Data Sheet for Colin

Probe	**Y/N**
Conducts one probe per day	
Implements correct program for probe	
Uses correct token board	
Uses correct cup of tokens (all red or all white tokens for Phase 1)	
Delivers all 10 tokens before providing reinforcer	
Ensures client isn't looking when removing blue tokens (Phase 1 only)	
Establishes reinforcer prior to starting DTT session	
Uses FR1 schedule of token delivery	
Avoids maintenance trials	
Avoids token delivery for attending skills	
Provides prompt and reinforcement following no response to acquisition target	
Implements error correction procedure following incorrect responses or aberrant behavior	
Provides behavior specific praise after each correct response	
Holds cup high enough that participant can't see inside (Phase 1 only)	
Allows participant 5 seconds to select a token from the cup	
If client has not selected a token or chooses more than one, provides a vocal prompt to choose one or drop all tokens back into cup and allows another 5 seconds	
If the client has not selected a token, therapist provides physical prompt or chooses a token and puts it on the board	
Records frequency of problem behavior within 30 seconds of the end of the token board	
Records duration of problem behavior within 30 seconds of the end of the token board	
Highlights box on histogram if behavior occurred	
Starts stopwatch within 3 seconds of delivering first instruction	
Stops stopwatch within 3 seconds of delivering back-up reinforcer	
Records total duration of probe session	
Keeps tally of number of trials presented	
Keeps tally of number of error correction procedures implemented	
Percentage Correct	

Appendix J

Institutional Review Board Approval Letter

ST. CLOUD STATE UNIVERSITY
EDUCATION FOR LIFE.

Institutional Review Board (IRB)

720 4th Avenue South AS 210, St. Cloud, MN 56301-4498

Name: Katherine Brockevelt

Email: Katherine.brockevelt@go.stcloudstate.edu

IRB PROTOCOL DETERMINATION:

Expedited Review-1

Project Title Comparing Two Types of Token Boards in an Autism Clinic

The Institutional Review Board has reviewed your protocol to conduct research involving human subjects. Your project has been: **APPROVED**

Please note the following important information concerning IRB projects:

- The principal investigator assumes the responsibilities for the protection of participants in this project. Any adverse events must be reported to the IRB as soon as possible (ex. research related injuries, harmful outcomes, significant withdrawal of subject population, etc.).

- For expedited or full board review, the principal investigator must submit a Continuing Review/Final Report form in advance of the expiration date indicated on this letter to report conclusion of the research or request an extension.

-Exempt review only requires the submission of a Continuing Review/Final Report form in advance of the expiration date indicated in this letter if an extension of time is needed.

- Approved consent forms display the official IRB stamp which documents approval and expiration dates. If a renewal is requested and approved, new consent forms will be officially stamped and reflect the new approval and expiration dates.

- The principal investigator must seek approval for any changes to the study (ex. research design, consent process, survey/interview instruments, funding source, etc.). The IRB reserves the right to review the research at any time.

If we can be of further assistance, feel free to contact the IRB at 320-308-4932 or email ResearchNow@stcloudstate.edu and please reference the SCSU IRB number when corresponding.

IRB Chair:

Dr. Mili Mathew
Chair and Graduate Director
Assistant Professor
Communication Sciences and Disorders

IRB Institutional Official:

María-Claudia Tomany

Dr. Claudia Tomany
Associate Provost for Research
Dean of Graduate Studies

OFFICE USE ONLY

SCSU IRB#: 2075 - 2704	**Type:** Expedited Review-1	**Today's Date:** 2/1/2022
1st Year Approval Date: 1/31/2022	**2nd Year Approval Date:**	**3rd Year Approval Date:**
1st Year Expiration Date: 1/30/2023	**2nd Year Expiration Date:**	**3rd Year Expiration Date:**