An evaluation of the effectiveness of and preference for differential reinforcement and response cost within token economies

by

Erica S. Jowett

Submitted to the graduate degree program in Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson: Claudia L. Dozier, Ph.D.

Pamela L. Neidert, Ph.D.

Florence D. DiGennaro Reed, Ph.D.

Derek D. Reed, Ph.D.

Eva M. Horn, Ph.D.

Date Defended: 07-16-14

The Dissertation Committee for Erica S. Jowett

certifies that this is the approved version of the following dissertation:

An evaluation of the effectiveness of and preference for differential reinforcement and response cost within token economies

Chairperson: Claudia L. Dozier, Ph.D

Date approved: 07-16-14

Abstract

Several researchers have directly compared differential reinforcement and response cost within the context of token economies, and the results have varied. That is, some studies have shown equal effectiveness across procedures, whereas other studies have shown one procedure to be more effective than the other (e.g., Brent & Routh, 1978; Iwata & Bailey, 1974; Tanol et al., 2012). Therefore, it is possible that certain variables (e.g., experimental design, back-up reinforcers, or opportunities for net tokens) may influence the efficacy of the two procedures. In addition, only two studies have empirically evaluated preference for differential reinforcement and response cost within the context of token economies (e.g., Donaldson et al., 2014; Iwata & Bailey, 1974), and the authors found that preference varied among individuals. The purposes of the current study were to (a) replicate research comparing differential reinforcement and response cost within token economies, (b) evaluate preference for these procedures, and (c) evaluate whether varying the opportunity for net tokens influences the effectiveness of or preference for these procedures. Results showed that when the opportunity for net tokens was equal in Study 1 and 2, DRA and RC were similarly effective for increasing responding for the majority of participants. However, preference for these procedures was idiosyncratic. Results from Study 2 showed that when the opportunity for net tokens was unequal across DRA and RC, all participants engaged in similar increases in the level of responding. However, when DRA resulted in more opportunity for net tokens, all participants preferred DRA; whereas, when RC resulted in the opportunity for net tokens, preferences were idiosyncratic. Results are discussed with respect to implications and areas for future research.

Acknowledgements

The current project has been a work in progress since the summer of 2008, and several people have contributed to its development. First, I'd like to express thanks to my advisor, Dr. Claudia Dozier, who provided guidance and support throughout the development and continuation of this project. In addition, her commitment to my education has furthered the development of my understanding of behavior analysis and fostered personal and professional growth. Second, I'd like to express gratitude to my colleagues and undergraduate research assistants who assisted with conducting research sessions, collecting data, and calculating overall results. Third, I'd like to acknowledge my family and friends who have provided substantial emotional support and encouragement throughout the years. Finally, I'd like to recognize my fiancé, Jason Hirst, for his unconditional love and constant support throughout this process.

Table of Contents

Abstract…………………………………………………………………………….…………iii

Acknowledgements……………………………………………………………….………...iv

Table of Contents…………………………………………………………….……….…..v-vi

Introduction to Token Economies…………………………………….……………….….…1

Study 1 Method: DRA versus RC…………………………………………………………..16

Participants and Setting……………………………………………………………..16

Materials…………………………………………………………………………...…17

Response Measurement and Interobserver Agreement...18

Procedures...20

Baseline...23

DRA...24

RC..24

Experimental Design..25

Preference Evaluation..25

Token Analysis...27

Results and Discussion...28

Study 2 Method: Token Evaluation...40

Purpose..40

Participants, Setting, and Materials...41

Response Measurement and Interobserver Agreement..42

Treatment Integrity..43

DRA...44

RC..44
Procedures..45
Baseline..45
DRA (10 tokens)..46
DRA (5 tokens)..46
RC (10 tokens)..46
RC (5 tokens)..46
Preference Evaluation..47
Experimental Design..47
Token Analysis..48
Results and Discussion..48
General Discussion..63
References..76
Tables..87
Figures..93
Appendices..103

An evaluation of the effectiveness of and preference for differential reinforcement and response cost within token economies

The token economy is one of the earliest applications of operant conditioning for changing socially important behavior and has been an important part of the history of behavior analysis (Ayllon & Azrin, 1968a; Doll, McLaughlin, & Barretto, 2013; Hackenberg, 2009; Kazdin, 1977; Kazdin & Bootzin, 1972; Matson & Boisjoli, 2009). Token economies involve earning, losing, or both earning and losing a conditioned reinforcer (e.g., tokens, points, buttons, and fake money) contingent upon the occurrence of a particular behavior or the absence of a particular behavior. In addition, at the end of a set period of time, the earned or remaining conditioned reinforcers are exchanged for back-up reinforcers (e.g., prizes, treats, and leisure activities; Ayllon & Azrin, 1968b; Boerk & Reitman, 2011; Dickerson, Tenhula, & Green-Paden, 2005; Kazdin, 1977; O'Leary & Drabman, 1971).

Numerous studies have shown that the token economy is a powerful intervention for changing socially significant behavior (Doll et al., 2013; Kazdin, 1977). That is, token economies have been used to increase various appropriate behaviors such as on-task behavior, pre-academic skills, social skills, self-care, task completion, and rule following (e.g., Lahey & Drabman, 1974; O'Leary & Becker, 1967; Rowbury, Baer, & Baer, 1976); and decrease inappropriate behaviors such as off-task behavior, rule violations, tardiness, and aggression (e.g., Burchard & Barrera, 1972; Haring & Kennedy, 1990; Iwata & Bailey, 1974; O'Leary, Becker, Eans, & Saudaras, 1969) in various populations (e.g., children, college students, psychiatric patients, and individuals with intellectual and development disabilities; Boniecki & Moore, 2003; Drabman, Spitalnik, & Spitalnik, 1974; Nelson & Cone, 1979). In addition, token economies have shown to be effective across various settings (e.g., classrooms, psychiatric facilities, homes, work places, prisons, group homes, and communities; Ayllon & Azrin, 1968b; Holland & McLaughlin, 1982; Marholin & Gray, 1976). See Kazdin (1977) and Doll et al. (2013) for a detailed discussion of the different behaviors, populations, and settings for which token economies have found to be effective.

In a token economy, an individual (a) earns tokens (e.g., points, fake money, checkmarks, and marbles) for engaging in a target appropriate behavior or for the absence of inappropriate behavior, (b) loses tokens for an inappropriate behavior or the absence of appropriate behavior, or (c) both earns and loses tokens. Later, the individual is allowed to exchange these tokens for back-up reinforcers (e.g., preferred items, activities, privileges, or snacks). Tokens are initially neutral stimuli that become valuable due to their association (pairing) with back-up reinforcers during the token exchange (Carr, Frazier, & Roland, 2005; Hackenburg, 2009; Kelleher & Gollub, 1962; O'Leary & Drabman, 1971). Through this pairing process, the previously neutral tokens become conditioned reinforcers (see Kelleher & Gollub, 1962, for a detailed discussion on this process). Specifically, tokens are referred to as generalized conditioned reinforcers because they are often paired with many different types of back-up reinforcers, which is likely to result in tokens being effective under various establishing operations (Hackenburg, 2009; Kazdin, 1977; Miltenberger, 2008).

Researchers began studying token economies in the context of laboratory settings in the late 1930's (Hackenberg, 2009), and this body of literature has influenced current applied research on token economies. Applied researchers began to investigate token economies in the 1960's (Ayllon & Azrin, 1968b; Boerk & Reitman, 2011; O'Leary & Drabman, 1971). In fact, Ayllon and Azrin (1965) conducted one of the earliest published studies using a token economy in an applied setting. The authors showed in a series of experiments that the contingent delivery of tokens was effective for changing work behavior (i.e., increasing job performance and changing job choice) of psychiatric patients. Since the 1960's, hundreds of studies have been conducted showing the effectiveness of token economies (Doll et al., 2013; Kazdin, 1977), and the number of replications suggests a powerful technology that has external validity. Furthermore, token economies have been shown to have social validity in that teachers and caregivers rate token-economy procedures as favorable (e.g., Little & Kelley, 1989; McGoey & DuPaul, 2000; Theodore, Bray, Kehle, 2004). Finally, token economies have several advantages over other reinforcement procedure because (a) tokens are easily delivered and do not interfere with ongoing learning (Ayllon & Azrin, 1968b; Kazdin, 1977), (b) satiation is less likely to occur with tokens because they may be exchanged for a variety of reinforcers (Carr et al., 2005; Hackenberg, 2009), and (c) token economies can be individualized or implemented to change group behavior (Kazdin & Bootzin, 1972).

Although token economies have many advantages, they are complex systems that involve numerous components and considerations (see Boerk & Reitman, 2011 for a recent review). One important consideration in developing a token economy involves determining the contingencies that will be implemented to change behavior. The most common type of token economy involves earning tokens for engaging in target appropriate behavior (i.e., differential reinforcement of alternative behavior [DRA]; Ayllon & Azrin, 1968b; Betancourt & Zeiler (1971); Kazdin, 1977; Lahey & Drabman, 1974; O'Leary & Becker, 1967). However, other token economies have involved the delivery of tokens for the absence of problem behavior (differential reinforcement of other behavior [DRO]; Conyers, Miltenberger, Romaniuk, Kopp, & Himle, 2003; Cowdery, Iwata, & Pace, 1990; Walker & Buckley, 1968) or the removal of tokens for the occurrence of incorrect or inappropriate behavior (response cost [RC]; Burchard &

Barrera, 1972; Pace & Foreman, 1982; Sullivan & O'Leary, 1990). Furthermore, some token economies involve both the delivery of tokens for appropriate behavior and the removal of tokens for inappropriate behavior, simultaneously (Kazdin, 1977; Christophersen, Arnold, Hill, & Quilitch, 1972; Phillips et al., 1971).

Differential reinforcement of alternative behavior (DRA) in a token economy typically involves the delivery of tokens contingent on appropriate behavior. However, the contingencies by which the tokens are delivered vary. Previous research has shown that DRA token economies have been effective when tokens were delivered (a) following a behavior on a ratio schedule (e.g., one token is delivered after every X number of target responses; Boniecki & Moore, 2003; Brigham, Finfrock, Breunig, & Bushell, Jr., 1972) or interval schedule (e.g., one token is delivered for the first response that occurs after 1 min), (b) after some period of time for which the behavior has been occurring (e.g., token[s] delivered if target behavior occurs throughout a 10-min session; Mottram, Bray, Kehle, Broudy, & Jenson, 2002; O'Leary & Becker, 1967), or (c) at the moment an observation is conducted (e.g., token[s] delivered if target behavior is occurring at the exact moment of observation; Bushell, Wrobel, & Michaelis, 1968; Twardosz & Sajwaj, 1972).

Differential reinforcement of other behavior (DRO) in a token economy typically involves the delivery of tokens for the absence of target behavior(s) to decrease the occurrence of those behaviors. In order to implement DRO, an interval length is specified. The interval length is sometimes based on mean responding during baseline; however, interval lengths may be chosen for convenience (particularly if they are being implemented with a group). Once the interval is determined, the reinforcer is delivered at each interval, provided the target behavior has not occurred since the last reinforcer delivery. Previous research has suggested that DRO token economies that involve the delivery of a token for the absence of target behavior(s) during an entire interval (Cowdery et al., 1990; Didden, de Moor, & Bruyns, 1997; Walker & Buckly, 1968), as well as those procedures that involve the delivery of a token based on the absence of specific behavior(s) occurring at that moment (i.e., momentary DRO; Conyers et al., 2003; Drabman et al., 1974), have been effective for reducing various target behaviors.

In addition to the effectiveness of DRA and DRO in the context of token economies, reinforcement-based procedures such as DRA and DRO are often viewed as a positive approach to behavior change because they focus on increasing appropriate behavior and create a "positive environment" (e.g., Ayllon & Roberts, 1974; Cowdery et al., 1990; Tanol, McComas, & Cote, 2010; Twardosz & Sajwaj, 1972). Also, several authors have reported positive indirect effects (Boniecki & Moore, 2003; Kazdin & Bootzin, 1972; e.g., Ayllon & Roberts, 1974; Ayllon, Layman, & Kandel 1975; Boniecki & Moore, 2003) with both DR procedures. That is, behaviors—or dimensions of behavior—other than those targeted also improved with the implementation of these procedures.

Response cost (RC) procedures used within token economies involve the removal of conditioned reinforcers (e.g., points, tokens, or check marks) that an individual already possesses (Hackenburg, 2009; Kazdin, 1972) and function as a negative punishment procedure. Previous research has shown that RC token economies are effective for reducing various behaviors across different populations and settings (e.g., Gresham 1979; Holland & McLaughlin, 1982; Phillips et al., 1971; Staples, McGoey, Cowan, Crist, & Tankersley, 2006). However, with respect to the way in which tokens are removed, response cost procedures vary. That is, response cost may involve the removal of a token(s) (a) immediately following any instance of a target problem behavior (e.g., Pace & Foreman, 1982), (b) after some period of time (e.g., at the end of an interval, session, or day) if the target problem behavior occurred or an appropriate behavior did not occur during the observation period (e.g. Iwata & Bailey, 1974), or (c) at a specific moment of a scheduled observation if the target problem behavior was occurring or a target appropriate behavior was not occurring (e.g., Sullivan & O'Leary, 1990).

Although RC contingencies are considered punishment procedures, there are several possible advantages to the use of RC contingencies either alone (as compared to DR contingencies) or in combination with DR procedures within the context of a token economy. First, when token economies are highly effective for reducing problem behavior, few instances of problem behavior occur. Therefore, it may be easier to implement token economies with RC if token removals occur as compared to DRO contingencies that will involve many deliveries (Donaldson, DeLeon, Fisher, & Kahng, 2014; McGoey & DuPaul, 2000; O'Leary & Becker, 1967). Second, and likely related to the first advantage above, several researchers have shown that RC has been viewed positively by implementers (i.e., reported as acceptable or favorable by parents and teachers; Holland & McLaughlin, 1982; Little & Kelley, 1989; McGoey & DuPaul, 2000; Reynolds & Kelley, 1997).

Although there are several advantages of RC, there are also some possible disadvantages to consider when using RC within the context of token economies. One potential disadvantage of implementing RC alone within the context of a token economy is that if high levels of problem behavior occur, the individual may lose all tokens during a session or treatment period (Little & Kelley, 1989). If this occur, without any way for an individual to earn back any tokens, then the target problem behavior may emerge or the target appropriate behavior may cease to occur because there are no longer any consequences in place for behavior change. Another possible disadvantage of RC in the context of token economies is that it does not directly involve teaching or reinforcing appropriate behavior. Thus, RC procedures should be used if appropriate behavior is already occurring and does not need to be taught. However, if appropriate behavior does not increase to anticipated levels or if an appropriate behavior needs to be taught, RC should be combined with DR procedures for increasing appropriate behavior (e.g. Vollmer, 2002).

Although differential reinforcement and response cost within token economies have been shown to be effective independently, these procedures are sometimes combined, which provides the individual the opportunity to earn tokens for appropriate behavior (DRA) or the absence of inappropriate behavior (DRO) as well as lose tokens (RC) for problem behavior or the absence of appropriate behavior. Differential reinforcement-with-RC token economies are typically implemented in one of two ways. In some DR-with-RC token economies, tokens are delivered for one target behavior and removed for another target behavior that is incompatible (e.g., token delivery for correct responses and token removal for incorrect responses; LeBlanc, Hagopian, & Maglieri, 2000; Mclaughlin & Malaby, 1972; Mottram et al., 2002; Phillips, 1968; Rapport, Murphy, & Bailey, 1982; Reisinger, 1972). In other DR-with-RC token economies, tokens are delivered for one target behavior and removed for another unrelated target behavior (e.g., token delivery for correct responding and token removal for self-injurious behavior; Christophersen et al., 1972; De Martini-Scully, Bray, Kehle, 2000; Musser, Bray, Kehle, & Jenson, 2001; Phillips et al., 1971). Previous researchers have suggested that both procedures combined are effective for increasing appropriate behavior and decreasing inappropriate behavior.

Although the effectiveness of DR and RC within the context of token economies has been demonstrated in the literature, the conditions under which these procedures have been evaluated vary. Thus, it makes it difficult to compare the effects of these procedures within and across participants. Therefore, a few researchers have directly compared DR and RC procedures within and across participants under the same or similar conditions. Overall, results of these studies have been mixed. That is, out of 13 studies that compared the effects of these procedures on socially significant behaviors, eight have suggested that DR and RC token economies are similarly effective (Capriotti et al., 2012; Broughton & LaHey, 1978; Donaldsonet al., 2014; Humphrey & Karoly, 1978; Iwata & Bailey, 1974; McGoey & DuPaul, 2000; Panek, 1970; Sullivan & O'Leary, 1990), whereas the results of the five other studies have suggested differential effects. Out of these five studies, only one has shown DR to be more effective (Tanol et al., 2010), whereas four have shown RC to be more effective (Brent & Routh, 1978; Carlson et al., 2000; Conyers et al., 2004; Sindelar et al., 1982).

Among studies that have demonstrated similar changes in behavior across differential reinforcement and response cost, some have compared DRA to response cost (e.g., Broughton & LaHey, 1978; McGoey & DuPaul, 2000; Panek, 1970; Sullivan & O'Leary, 1990), whereas others have compared DRO to response cost (e.g., Capriotti et al., 2012; Donaldson et al., 2014; Iwata & Bailey, 1974). For example, Sullivan and O'Leary (1990) compared the effects of DRA and RC for increasing the on-task behavior of 10 children. During DRA, tokens (i.e., checkmarks on a chart) plus praise were delivered for on-task behavior (i.e., working quietly at desk). During RC, tokens were removed, and a reprimand was delivered for off-task behavior. The children could earn or lose up to four tokens throughout a session, and the tokens could be exchanged for extended recess play or stickers. Observations (opportunities for token delivery or removal) occurred every 5 min (on average) throughout a 20-min math or reading period. During the first evaluation, DRA and RC were assigned to one class; during the second evaluation, the procedures were switched between classes (e.g., DRA for reading and RC for math then DRA for math and RC for reading). Results showed that the procedures were similarly effective for increasing on-task behavior.

In another study, Iwata and Bailey (1974) compared DRO and RC for decreasing rule violations and increasing off-task behavior of 15 children in a classroom. During DRO, tokens were delivered at the end of a 3- to 5-min interval if no rule violations occurred during that interval. During RC, tokens were removed at the end of an interval if any rule violations occurred during that interval. The children could earn or lose up to 10 tokens throughout a 30-min math period, and the tokens could be exchanged for snacks and free time. Results showed that the procedures were similarly effective for reducing rule violations and off-task behavior.

As mentioned above, among the five studies that found differences in the effectiveness of DR as compared to RC, only Tanol et al. (2010) found that DR was more effective than RC. In this study, Tanol et al. evaluated the effects of DRA and RC for decreasing rule violations of six, 5- and 6-year-old children. Tanol et al. used an interdependent group-oriented contingency such that the behavior of each child affected the gain or loss of tokens (i.e., stars) for the group. During DRA, tokens were delivered (placed on a poster) when members of the team were following rules (but there was not an explicit schedule for token delivery). During RC, tokens were removed immediately when any child in the group broke a rule. The children could lose up to four tokens during RC. However, the authors did not specify the number of tokens that could be earned during DRA. The tokens remaining at the end of each 10-min session could be exchanged for edibles. Tanol et al. found that both procedures reduced rule violations; however, DRA was somewhat more effective than RC for decreasing rule violations for two participants. Because this study was conducted in a group setting using group contingencies, it might be that the behavior of peers differentially influenced responding. For example, if children were likely to respond positively (e.g., smile or make statements of exclamation, “We got a star!”), this consequence, combined with token delivery, might have enhance the efficacy of DRA.

Four out of 5 studies that showed a difference in effectiveness of the procedures showed that RC was more effective than DR (e.g., Brent & Routh, 1978; Carlson et al., 2000; Conyers et al., 2004; Sindelar et al., 1982). Of these four studies, one study found that RC was only slightly more effective (Sindelar et al., 1982), whereas three studies found a robust difference in the effectiveness of RC as compared to DR (e.g., Brent & Routh, 1978; Carlson et al., 2000; Conyers et al., 2004). For example, Brent and Routh (1978) compared DRA and RC for increasing correct sight-word reading with 30, 4th-grade students. One 45-min session was conducted for each group. During DRA, tokens (i.e., nickels) were delivered immediately for each sight word read correctly. During RC, tokens were removed immediately for each sight word read incorrectly. Each student could earn or lose up to 40 tokens for reading 40 sight words. Results showed that RC resulted in fewer errors as compared to DRA and control conditions. However, this study is limited because the participants only experienced the contingencies during one session.

In another study, Conyers et al. (2004) evaluated the effects of DRO and RC within an independent group-oriented contingency on disruptive behavior of 25 preschool children. During DRO, tokens were delivered (i.e., stars placed on a board) if problem behavior (e.g., screaming and throwing objects) had not occurred during a 1-min interval (on average). During RC, tokens were removed if problem behavior occurred during an interval. Children could earn or lose an equal number of tokens, and tokens could be traded for edibles at the end of a 15-min session. Conyers et al. initially observed lower levels of problem behavior during DRO; however, over time, RC resulted in less disruptive behavior than DRO. Because DRO was compared to RC, it could be that the immediate contingency involved in RC provided more salient consequences for problem behavior (e.g., immediate feedback), and was, therefore, more effective over time.

Given that different results have been shown in the comparison of DR and RC in token economies, continued study of the conditions under which the different procedures are more or less effective is warranted. There are several variables that may influence the efficacy of DR and RC in the context of a token economy, which may have resulted in mixed results in research that has compared the two procedures. First, different target behaviors (e.g., on task and sight-word reading) have been evaluated across studies, and the behavior for which the contingencies are placed may influence the effectiveness of a given procedure. For example, it is possible that acquisition tasks or tasks that are more difficult or effortful may influence the efficacy of the two procedures. Second, different types of back-up reinforcers have been used, and the type of reinforcer may influence motivation. Specifically, tokens exchanged for potent back-up reinforcers may be more valuable than tokens exchanged for low-preferred items or activities, resulting in drastic changes in behavior (i.e., increases in appropriate behavior or decreases in inappropriate behavior) regardless of the procedure(s) used within the token economy. Third, different experimental designs have been used to compare DR to RC across studies (e.g., multielement, reversal, and between subjects), and the type of design may influence results of the comparison. For example, group designs may mask individual differences, multielement designs may result in carryover effects due to rapid alternation of conditions, and reversal designs may result in history effects. Fourth, different contingencies have been used to compare DR to RC across studies (e.g., individual and group oriented), and the type of contingency may influence results of the comparison. Specifically, with respect to group-oriented contingencies, an individual may engage in an appropriate behavior because his peers are (a) engaging in an appropriate behavior, (b) prompting him to engage in an appropriate behavior, (c) providing reinforcers for him to engage in an appropriate behavior, or (d) implementing punishers for not engaging in an appropriate behavior (Salend & Kovalich, 1981). Fifth, different DR procedures have been used when comparing DR to RC across studies (e.g., DRA and DRO). For example, it is possible that the procedural differences across these DR procedures (i.e., delivery of a token for the occurrence of a behavior at a specific moment for DRA, and the delivery of a token for the occurrence of behavior during an entire interval) may result in different outcomes. Next, different types of tokens have been used when comparing DR to RC across studies (i.e., tangible and non-tangible), and the type of token may influence responding. For example, loss of a tangible token may be more effective during RC than loss of a non-tangible token (e.g., points) because tangible tokens are more salient stimuli. Finally, the literature has demonstrated that more dense schedules of reinforcement (or punishment) influence the efficacy of behavioral procedures (e.g., Herrnstein, 1961; Neef, Shade, & Miller, 1994); therefore, the opportunity for net tokens (i.e., number of tokens earned or kept) may influence the effectiveness of DR and RC.

To determine whether some of these variables may influence the outcomes of the comparison studies on DR and RC, I looked at each of these variables to determine if there seemed to be an influence. The results of this evaluation for all of the above variables except net tokens are listed in Table 1. Overall, this evaluation showed that across the 13 studies, the type of behavior (acquisition vs. non-acquisition), design, contingency, type of DR procedure, type of backup reinforcers, and population of participants did not seem to be associated with the outcome of the comparison studies. For example, studies that used between-subjects designs showed all results. That is, of the six studies that included a between-subjects design, four studies showed the two procedures to be equally effective and two studies showed RC to be more effective. In addition, in the five studies that used acquisition tasks as the target task, three showed that the procedures were equally effective and two showed that RC was more effective.

Although the above variables did not seem to influence the results of the studies comparing DR and RC, two other variables warrant mention. First, all of the studies that found differences in the effectiveness of the two procedures used tangible tokens, and it may be that tangible tokens (as compared to non-tangible tokens) influence the saliency of consequences. Specifically, 4 out of 5 studies found RC to be more effective than DR, which may suggest that the loss of tangible tokens is particularly aversive. Although, one might also argue that the saliency would affect both procedures equally. In addition, several studies that found the procedures to be equally effective also used tangible tokens; therefore, it seems unlikely that the type of token influences the effectiveness of a procedure. However, researchers have yet to evaluate this variable.

Second, an analysis of studies that reported the net tokens across conditions suggests that this variable may influence responding. Among the studies reviewed in the current paper, Iwata and Bailey (1974) and Donaldson et al. (2014) were the only authors to calculate net tokens across each condition and they found that the number of net tokens earned (or kept) by each child was approximately equal on average. The authors also found that the two procedures (e.g., DRO and RC) were equally effective. In addition, Capriotti et al. (2012) reported that the response cost procedure produced more net tokens compared to DRO and found that response cost was more effective. Thus, these studies suggest that the number of net tokens may influence the effectiveness of the procedures.

Although research in the basic literature has not made direct comparisons of reinforcement and response cost, some basic researchers have investigated whether decrements in responding during response cost are a function of token loss or net tokens (e.g., Pietras, Brandt, & Searcy, 2010; Pietras & Hackenberg, 2005; Raiff, Bullock, & Hackenberg, 2008). For example, Pietras et al. (2010) made two comparisons: the authors compared the effects of contingent token loss to (a) contingent delivery of tokens that were yoked to the number of net tokens in the token-loss condition and (b) noncontingent token loss that was yoked to the number of token losses in the token-loss condition. The authors found that responding decreased below baseline levels (contingent token delivery) when token removal was contingent and noncontingent on responding; however, responding did not decrease when the number of net tokens during token delivery was yoked to the number of net tokens during the token-loss conditions. Pietras and Hackenberg (2005) and Raiff et al. (2008) made similar comparisons and found the same results. These studies suggest that decreases in responding may be a function of token loss and not a result of fewer net tokens; therefore, based on these findings, net tokens may not influence efficacy. However, given the difference in the experimental arrangements between these and the applied studies, future research is warranted to evaluate the influence of net tokens on efficacy.

In addition to evaluating the effectiveness of token procedures, few researchers have empirically evaluated preference for DR and response cost within the context of token economies (e.g., Donaldson et al., 2014; Iwata & Bailey, 1974). Iwata and Bailey (1974) and Donaldson et al. (2014) both compared DRO to response cost with students in a classroom. Following a comparison of DRO and RC, the authors evaluated preference for these procedures by allowing students to select the procedure he or she would like to be implemented for that class period (or session). After each student made a selection, the selected procedure was then implemented for that student. The authors found that some students preferred the DRO procedure (selected DRO more than RC), some preferred RC (selected RC more than DRO), and others switched back and forth between the two procedures in their selections. However, the variables influencing preference are unknown.

Several variables may influence preference, such as characteristics of a procedure, peer influence, or net tokens. With respect to characteristics of a procedure that influence preference, a recent basic research study by Pietras et al. (2010) found that when net tokens were equated, participants selected away from the procedure that involved token loss, which may suggest that token loss as a procedural characteristic is aversive. In addition, Donaldson et al. (2014) provided anecdotal reports of preference. Specifically, participants who preferred DRA said they liked to earn, and participants who preferred RC made statements with respect to preference for losing few tokens. With respect to peer influence, Iwata and Bailey and Donaldson et al. conducted their preference evaluation in the classroom. That is, each student made a selection in the presence of other students. Thus, is it possible that student selections were partially influenced by the presence of others (e.g., one student picks RC because his best friend picked RC). Finally, with respect to the influence of net tokens on preference, Iwata & Bailey calculated the average number of net tokens for the class, and Donaldson et al. calculated individual net token averages, and both authors found that net tokens were similar across procedures. Although, the number of net tokens was similar, because some participants preferred one procedure over another, it could be that small differences may influence preference. However, the number of net tokens was not manipulated. Therefore, the influence of net tokens on preference is unknown, and research on this variable is warranted.

In summary, the results of studies comparing DR and RC in token economies are inconsistent; however, the majority of studies shows that these procedures are similarly effective. Several variables may influence the effectiveness of these procedures, but the variables that may contribute to the differences in these findings are unknown. Also, the majority of studies comparing DR to RC involved group data analyses; therefore, individual differences could not be evaluated. Next, many studies comparing DR to RC included other components such as feedback, which confounds the evaluation of token delivery and removal. In addition, few studies have empirically evaluated preference for these procedures, and several variables may influence preference. Therefore, the purposes of the current study were to (a) replicate research comparing differential reinforcement to response cost with individual children, (b) evaluate child preference for these procedures, (c) and evaluate whether varying the relative opportunity for net tokens influences the effectiveness of or preference for these procedures.

Study 1 Method: DRA versus RC

The purpose of Study 1 was to (a) directly compare the effects of DRA and RC within the context of a token economy for increasing on-task behavior during a table task (i.e., tracing shapes and letters on worksheets while seated at a table) for individual children and (b) determine which procedure was preferred by those children.

Participants and Setting

Thirteen typically developing preschool-aged children (3 years 7 months to 5 years 6 months) and one child with cerebral palsy (Brianna, 4 yrs 11 months) who were enrolled in the Edna A. Hill Child Development Center participated. Specific participant information including gender and age for all participants is displayed in Table 2. Although no systematic data were collected on participant skills that may be important for the procedures, anecdotal information from in-session and classroom observations suggested that all children could follow three-step instructions (e.g., “Go put your blanket in your cubby, get your shoes, and sit in the chair”), rote count up to 10 when instructed to count to 10 with no stimuli present, count up to 10 objects when asked to count how many objects were present, give up to 10 objects to another person when asked to give another person a specified number of objects, answer “What” questions (e.g., “What happens if you are tracing?”), describe what was happening or what they were doing when asked, “What are you doing?,” describe what happened when asked, “What happened when you were tracing?,” and name the colors red, green, and white. Sessions were conducted one to four times per day, 3-5 days per week in a session room that was adjacent to their classroom and contained tables, chairs, and relevant session materials.

Materials

During all sessions, worksheets with printed letters and shapes (see Appendix A) and markers or crayons were present on the table. A stack of a large number of worksheets was available such that all worksheets could not be completed within the duration of a given session. In addition, toys obtained from the preschool classroom (e.g., puzzles, dolls, toy cars, coloring book, and various colors of crayons) were present on the floor away from the work table. See Appendix B for the set-up of the room across sessions. During some sessions, tokens (i.e., pennies) were present that could be earned or lost. Tokens were attached to and removed from laminated strips of paper (approximately 4 in x 10 in) with 10 square pieces of Velcro® (see Appendix C). Access to a toy room with tangible items (e.g., stickers, plastic rings, spin tops, sticky hands, etc.), edibles (e.g., gummies, Smarties®, Skittles®, M&M’s®), and leisure activities (e.g., video games, DVDs) were earned via token exchange following some sessions (DRA and RC). Different colored materials (white, green, and red posters and token boards) were placed on the table during each of the different conditions to aid in discrimination between conditions.

Response Measurement and Interobserver Agreement

Trained graduate and undergraduate students collected data using handheld computers. The dependent variable during all sessions was percentage of on-task behavior. On-task behavior was defined as the first instance of walking to the work table, the first instance of removing the lid of the marker, tracing (moving the marker approximately within the boundaries of the printed letters or shapes on the worksheet), or grasping and turning over completed worksheets to access a new worksheet. All of the above behaviors were scored as on task because they were necessary for task engagement (i.e., tracing). On-task behavior was not scored if the child was scribbling (e.g., rapidly moving the marker back and forth across the worksheet), drawing pictures, making dots or other patterns inside or outside the boundaries of the letters or shapes, outlining a letter or shape, or turning over incomplete worksheets. In general, when children were not on task, they were either playing with toys or their behavior did not otherwise meet the definition of on task (i.e., non-examples listed above).

Sessions were partitioned into 5-s intervals, and data on on-task behavior were collected using partial-interval data collection. That is, on-task behavior was scored if it occurred during any portion of the 5-s interval. Data were converted to a percentage by dividing the number of intervals for which the child was on task by the total number of intervals in the session. Data were also collected on the frequency of token delivery (i.e., when the experimenter placed a token on the token board) and token removal (i.e., when the experimenter removed a token from the token board).

A second independent observer collected data during at least 30% of all sessions across participants. Interobserver agreement (IOA) was calculated for on-task behavior by dividing the number of 5-s intervals during which both observers agreed by the total number of intervals and multiplying by 100%. An agreement for on-task behavior was defined as both observers scoring (or not scoring) the occurrence of the behavior in a given interval. Interobserver agreement coefficients for token delivery or removal were calculated by dividing the session time into 5-s intervals and comparing observer data on an interval-by-interval basis. If exact agreement occurred (i.e., both observers scored a token delivery or removal within a 5-s interval), a score of 1 was given for that interval. For any disagreements, the smaller score in each interval was divided by the larger. The interval scores were summed, divided by the total number of observation intervals, and multiplied by 100%. Across participants, mean IOA was 93% (range: 70%-100%) for on-task behavior and 97% (range: 78%-100%) for token delivery or removal. IOA for Paul was 95% (range: 86%-100%) for on-task behavior and 98% (range: 86%-100%) for token delivery or removal. IOA for Frank was 92% (range: 82%-100%) for on-task behavior and 91% (range: 82%-100%) for token delivery or removal. IOA for Martin was 98% (range: 94%-100%) for on-task behavior and 96% (range: 94%-100%) for token delivery or removal. IOA for Emily was 86% (range: 75%-100%) for on-task behavior and 94% (range: 90%-100%) for token delivery or removal. IOA for Adrianna was 95% (range: 88%-100%) for on-task behavior and 98% (range: 92%-100%) for token delivery or removal. IOA for Elisa was 89% (range: 75%-100%) for on-task behavior and 96% (range: 78%-100%) for token delivery or removal. IOA for Adam was 92% (range: 73%-100%) for on-task behavior and 97% (range: 93%-100%) for token delivery or removal. IOA for Collin was 98% (range: 88%-100%) for on-task behavior and 98% (range: 95%-100%) for token delivery or removal. IOA for Anna was 95% (range: 70% -100%)

for on-task behavior and 98% (range: 95%-100%) for token delivery or removal. IOA for Caroline was 93% (range: 70%-100%) for on-task behavior and 97% (range: 83%-100%) for token delivery or removal. IOA for Zoey was 95% (range: 87%-100%) for on-task behavior and 99% for token delivery or removal. IOA for Brianna was 88% (range: 83%-91%) for on-task behavior and 99% (range: 93%-100%) for token delivery or removal. IOA for Mark was 92% (range: 87%-100%) for on-task behavior and 99% (range: 96%-100%) for token delivery or removal. IOA for Sam was 96% (range: 86%-100%) for on-task behavior and 93% (range: 80%-100%) for token delivery or removal. Although mean IOA was high for all participants, a few participants had lower IOA ranges (i.e., lower than 80%). These lower IOA values occurred during very few sessions, and immediately following these sessions, additional training was provided to all observers regarding operational definitions.

Procedures

All sessions were 5 min in length. As mentioned earlier, tracing worksheets were stacked on the table, and approximately eight activities typically found in a preschool classroom were available on the floor away from the work table (see Appendix B). Toys were available to provide an alternate activity in order to reduce the likelihood that participants would engage in the task because no other activity was available. However, preference of the toys available was not determined on an individual basis. Prior to the first session of each condition, the experimenter described the session contingencies and required the participant to practice engaging in related behaviors (i.e., tracing or playing with toys) to experience the consequences associated with each behavior. For example, the experimenter required the participant to practice tracing by providing a vocal and model prompt (i.e., the experimenter said, “Try tracing like this,” while demonstrating tracing) and used physical guidance as necessary. After the participant practiced tracing, the experimenter provided the relevant consequences and repeated the contingency for that particular phase (e.g., "Look, you got a token because you were tracing"). In addition, several researchers have suggested that it is important to train token economies (e.g. Kazdin, 1977; LeBlanc et al., 2000; O'Leary & Drabman, 1971; Pace & Forman, 1982), and given that the participants in the current study were young children, who could have had difficulty learning the rules and contingencies associated with token economies, the experimenter took the participant to the "token store" to practice exchanging the token(s) he or she earned or kept during pre-session exposure. The experimenter first showed the participant the items in the token store and pointed out the different values, then had the participant count the number of tokens he or she had (typically 1-3), and showed the participant from which bin he or she could choose an item. After the participant selected an item, the experimenter pointed to sign depicting the value of the item (e.g., 1 token) and asked, "How many tokens is this?" and prompted the child to state the value. Next, the experimenter asked the participant to give him or her said number of tokens and then gave the participant the item he or she had selected. Token training was only conducted prior to the first implementation of a token procedure.

Prior to the start of each subsequent session during a particular phase, the experimenter described the session contingencies and asked the participant to repeat or answer questions about the contingencies (see condition descriptions below). In addition, at the start of all sessions, the experimenter and participant stood away from the work table and play area. When the experimenter signaled the start of the session (e.g., "Ready, set, go!"), the participant could then move to the work table or play area. After the child moved to an area, the experimenter sat at the table, angled 90 degrees from the participant, such that the experimenter was not facing the work table or play area, and the experimenter remained seated for the remainder of the session.

During all sessions, the experimenter did not initiate any interactions with the participant. If the participant attempted to interact with or otherwise recruit attention (e.g., stating, "Look at what I made") from the experimenter, the experimenter either did not respond at all or responded briefly and neutrally (e.g., "Ok") with minimal eye contact. That is, the same type of responses were delivered by the experimenter regardless of what the participant said or whether the participant was playing with toys or engaging in the target task. However, anecdotal observations suggest that this rarely occurred during sessions, with the exception of one participant.

First, baseline sessions were conducted to determine the level of on-task behavior in the absence of programmed consequences. Next, the experimenter practiced token exchanging with the participant (described above). Then, DRA and RC were compared to determine their effects on on-task behavior. During DRA and RC sessions, the experimenter delivered or removed tokens according to a variable momentary schedule. For the current study, the experimenter observed the participant approximately every 30 s on average (ranging from 15 to 45 s) and either delivered a token for on-task behavior (DRA) or removed a token for the absence of on-task behavior (RC). Because the lines on which a participant could trace were not continuous, it was impossible for a participant to be continuously engaged in tracing at every moment because the participant had to transition from one tracing object (letter or shape) to another; and it was common for a child to be transitioning between tracing objects at the exact moment of an observation. In addition, other naturally occurring incompatible behaviors such as pushing hair away from face, scratching an itch, adjusting seating, and wiping nose also occasionally occurred during sessions. Therefore, when a brief pause (such as the examples above) occurred while a participant was oriented to the task, the experimenter waited up to 2 s following a scheduled observation for the participant to resume task engagement. However, if a child was engaged in an obvious off-task behavior, such as one of the non-examples mentioned above (e.g., scribbling, drawing, playing), the consequence occurred at that moment. The tokens were delivered or removed according to this schedule in order to equate the possible number of net tokens (i.e., 10 tokens) across conditions. Feedback was not provided during token delivery or removal so that the effect of token delivery or removal could be evaluated independently.

Following each DRA and RC session, the participant was taken to a room containing many different toys, leisure activities, edibles, and trinkets that were not found in the preschool classroom. The participant was given the opportunity to exchange tokens for edibles (i.e., small pieces of candy or other snacks), trinkets (e.g., small toys and stickers), or engagement with a toy or leisure activity (e.g., video game). Edibles and leisure items remained constant throughout the study; however, trinkets occasionally changed in order to replenish the token store. Edibles were one token each, trinkets were three tokens each, and engagement with a toy or leisure activity was one token per minute. The participant could spend the number of tokens he or she had for any combination of the above. For example, if a participant had seven tokens, he or she could exchange (a) seven tokens for seven edible items, (b) three tokens for one trinket and four tokens for playing a video game for 4 min, or (c) five tokens for playing a video game for 5 min and two tokens for two edibles. All participants were required to trade all tokens following each session; therefore, no participant could save tokens. Items for which tokens were exchanged varied across participants. That is, some participants mostly exchanged for edibles, some participants mostly exchanged for playing with a video game, and others exchanged for one or two trinkets and some edibles.

Baseline. Prior to the start of all baseline sessions, the participant was given rules describing the contingencies for the session, and a white board (with no tokens) and a white poster were present nearby (e.g., posted on the wall or on the side of the table). Rules were stated as follows: “Today you get the white board, and there are no tokens. When we start, you can either work on tracing or play with toys. If you are working, nothing will happen; if you are not working, nothing will happen.” In addition, the participant was asked to repeat the contingencies or was asked questions regarding the contingencies as previously mentioned (e.g., “What happens if you are working?” and “What happens if you are playing?”) and the participant was prompted to respond correctly if an incorrect or no response occurred. During the session, no programmed consequences were provided.

Differential reinforcement of alternative behavior (DRA). Prior to the start of all DRA sessions, the participant was given rules describing the contingencies for the session, and a green board (with no tokens) and a green poster were present. Rules were stated as follows: “Today you get the green board, and it doesn’t have any tokens on it. When we start, you can either work on tracing or play with toys. If you are working, you will get a token. If you are not working, you will not get a token. At the end, you can trade your tokens for prizes and snacks. If you don’t have any tokens, you don’t get anything.” In addition, the participant was asked to repeat the contingencies or was asked questions regarding the contingencies, as previously mentioned, and the participant was prompted to respond correctly if an incorrect or no response occurred. During the session, the experimenter followed the variable momentary observation schedule. At an observation time, if the participant was on task, then the experimenter placed a token on the token board. If the participant was not on task, no consequences were provided.

Response cost (RC). Prior to the start of all RC sessions, the participant was given rules describing the contingencies for the session, and a red board (with 10 tokens) and a red poster were present. Rules were stated as follows: “Today you get the red board and it has 10 tokens on it. When we start, you can either work on tracing or play with toys. If you are working, you will keep your tokens. If you are not working, you will lose a token. At the end, you can trade your tokens for prizes and snacks. If you don't have any tokens, you don't get anything." In addition, the participant was asked to repeat the contingencies or was asked questions regarding the contingencies, as previously mentioned, and the participant was prompted to respond correctly if an incorrect or no response occurred. During the session, the experimenter followed the variable momentary observation schedule. At an observation time, if the participant was on task, no consequences were provided. If the participant was not on task, the experimenter removed a token from the token board.

Experimental Design

A multielement design was used to compare the effects of the different procedures on on-task behavior for 10 participants. During the multielement design, session order was determined quasi randomly. A reversal design was also used for two of these 10 participants to rule out discrimination failure or carry over effects during the multielement comparison. However, because the reversal designs were conducted after the participants had a history of both procedures, a reversal design only was used with four participants, specifically to determine levels of responding during DRA prior to (and possibly after) a history of RC (discussed in detail below).

Preference Evaluation

Once stable levels of responding were observed in the DRA versus RC evaluation, a preference evaluation was conducted to determine the procedure that was most preferred by each participant (preference was evaluated for 10 participants). Prior to each session, the participant was presented with the token boards associated with each type of condition (i.e., baseline, RC, and DRA) and reminded of the contingencies associated with each set of materials. For example, the participant was reminded that the white board means that there are no tokens, the green board means that he or she can earn tokens if he or she is tracing, and the red board means that he or she could keep his or her tokens if he or she is tracing. In addition, the participant was asked questions regarding the contingencies to enhance the likelihood that the participants would respond to the programmed contingencies. The placement of the different sets of stimuli and materials were switched each session. For example, during one session, the stimuli associated with baseline would be in the middle, the stimuli associated with RC would be on the left, and the stimuli associated with DRA would be on the right. During the next session, these stimuli would be presented in different positions in front of the participant. After the experimenter reminded the participant of the contingencies associated with each set of materials, the experimenter asked the participant to pick (by verbal stating, pointing to, or touching a set of materials) which session he or she wanted to do. Once the participant made the selection, the experimenter or participant placed the materials of the selected procedure on the table, and then the experimenter explained the contingencies in place for the session (e.g., "You picked green; you will get a token when I see that you are working on tracing."). See Appendix D for the set-up of the preference evaluation. After the participant chose a procedure, the type of session that was chosen was implemented as described above. Sessions were conducted until a stable pattern of choice responding was observed. Preference was determined by counting the number of selections of each procedure, and the procedure that was selected most often was identified as the preferred procedure.

During the preference evaluation, IOA was calculated for selection of a procedure using a total agreement method. That is, an agreement was scored if both observers agreed which procedure was selected, and a disagreement was scored if the two observers disagreed. Thus, IOA for selection of a procedure for a particular session was either 100% (the two observers agreed) or 0% (the two observers disagreed). Interobserver agreement for selection was 100% for all participants.

Token Analysis

After all data were collected for the comparison phase (prior to the preference evaluation), a token analysis was conducted to determine the average number of net tokens for each procedure for each participant. The purpose of this analysis was to determine if the average number of net tokens in each procedure was different for a given participant. In order to calculate the average number of tokens per procedure, the number of tokens on the token board at the end of each session across all sessions of each procedure was summed and divided by the total number of sessions. In addition, differences in the average number and percent of net tokens were determined. In order to calculate differences in the average net tokens, the smaller number of average net tokens in one procedure was subtracted from the larger number of average net tokens in the other procedure. In order to calculate the percent difference in net tokens, the average net difference was divided by the smaller average net tokens and multiplied by 100%. For example, if a participant yielded an average of eight tokens in DRA and nine tokens in RC, the difference would be one token. Then, the difference (1) would be divided by the smaller average number of net tokens (8) and multiplied by 100%, which in this example is 12.5%. In this example, RC yielded 12.5% more net tokens than DRA. Percent difference in net tokens was calculated to weight token difference because the same number of an average token difference could be weighted more or less depending on a given participant's average tokens in each condition. For example, an average difference of one token for a participant who averaged five and six tokens would be weighted differently than an average difference of one token for a participant who averaged eight and nine tokens in each condition. With this information, I was able to evaluate whether the effectiveness of or preference for a procedure was correlated with the average difference in net tokens for each procedure. For example, if a participant averaged more net tokens during RC and the participant preferred RC, a correlation between average number of net tokens and preference would exist, which may suggest that net tokens influence preference.

In addition, the percentage of net tokens during each session was calculated in order to determine if the partial-interval method of data collection was representative of behavior occurring at other times during the session. Percent net tokens were calculated by dividing the number of net tokens (i.e., number of tokens earned or remaining at the end of a session) by the total number of possible tokens (i.e., 10) and multiplying by 100%.

Results and Discussion

Figures 1-8 display graphs for the 14 children who participated in Study 1. Results for the comparison between DRA and RC for Paul are in the top panel of Figure 1. During the initial baseline phase, Paul displayed 0% on-task behavior. During the DRA versus RC phase, Paul displayed high and stable levels of on-task behavior during both DRA (M = 97%; range: 95%-100%) and RC (M = 88%; range: 0%-98%) as compared to baseline (M = 0%). In addition, during the preference evaluation, Paul chose DRA (as denoted by the triangles) on every choice opportunity (10 out of 10 trials; 100% of trials) and on-task behavior remained high and stable (M = 94%; range: 88%-100%).

Results for the comparison between DRA and RC for Frank are in the bottom panel of Figure 1. During the initial baseline phase, Frank displayed moderate and variable levels of on-task behavior (M = 35%; range: 8%-83%). During the DRA versus RC phase, Frank displayed high levels of on-task behavior during both DRA (M = 95%; range: 85%-100%) and RC (M = 94%; range: 90%-100%) as compared to baseline (M = 19%; range: 0%-93%). In addition, during the preference evaluation, Frank chose DRA on every choice opportunity (6 out of 6 trials; 100% of trials) and on-task behavior remained high and stable (M = 95%; range: 88%-100%).

Results for the comparison between DRA and RC for Martin are in the top panel of Figure 2. During the initial baseline phase, Martin displayed variable levels of on-task behavior (M = 23%; range: 0%-62%). During the DRA versus RC phase, Martin displayed high and stable levels of on-task behavior during both DRA (M = 98%; range: 97%-98%) and RC (M = 95%; range: 93%-98%) as compared to baseline (M = 0%). In addition, during the preference evaluation, Martin chose DRA more (9 out of 11 trials; 82% of trials) than RC (2 out of 11 trials; 18% of trials) and on-task behavior remained high and stable across both DRA (M = 92%; range: 85%-98%) and RC (M = 93%; range: 87%-98%). Also, it is important to note that Martin chose DRA during the last nine trials.

Results for the comparison between DRA and RC for Emily are in the bottom panel of Figure 2. During the initial baseline phase, Emily displayed 0% on-task behavior. During the DRA versus RC phase, Emily displayed high levels of on-task behavior during both DRA (M = 94%; range: 80%-98%) and RC (M = 82%; range: 76%-100%) as compared to baseline (M = 0%); however, responding during DRA was consistently higher. In addition, during the preference evaluation, Emily chose DRA more (10 out of 17 trials; 59% of trials) than RC (7 out of 17 trials; 41% of trials) and on-task behavior remained high during both DRA (M = 73%;

range: 43%-100%) and RC (M = 73%; range: 65%-86%); however, responding during DRA was more variable. It should also be noted that Emily chose DRA during the last 10 trials.

Results for the comparison between DRA and RC for Adrianna are in the top panel of Figure 3. During the initial baseline phase, Adrianna displayed low levels of on-task behavior (M = 8%; range: 0%-23%). During the DRA versus RC phase, Adrianna displayed high and stable levels of on-task behavior during both DRA (M = 95%; range: 92%-97%) and RC (M = 94%; range: 90%-96%) as compared to baseline (M = 0%). In addition, during the preference evaluation, Adrianna chose DRA more (12 out of 17 trials; 71% of trials) than RC (5 out of 17 trials; 29% of trials) and on-task behavior remained high and stable during DRA (M = 94%; range: 82%-98%) and RC (M = 89%; range: 81%-98%).

Results for the comparison between DRA and RC for Elisa are in the bottom panel of Figure 3. During the initial baseline phase, Elisa displayed low levels of on-task behavior (M = 38%; range: 20%-86%). During the DRA versus RC phase, Elisa displayed variable, yet higher levels of on-task behavior during both DRA (M = 74%; range: 16%-90%) and RC (M = 74%; range: 43%-93%) as compared to the baseline condition (M = 2%; range: 0%-12%). In addition, during the preference evaluation, Elisa chose RC more (10 out of 13 trials; 77% of trials) than DRA (3 out of 13 trials; 23% of trials) and on-task behavior remained high and stable during DRA (M = 82%; range: 62%-93%) and RC (M = 89%; range: 78%-98%).

Results for the comparison between DRA and RC for Adam are in the top panel of Figure 4. During the initial baseline phase, Adam displayed variable levels of on-task behavior (M = 28%; range: 0%-60%). During the DRA versus RC phase, Adam displayed variable, yet somewhat higher levels of on-task behavior during both DRA (M = 47%; range: 0%-80%) and RC (M = 65%; range: 39%-85%) as compared to baseline (M = 27%; range: 0%-73%). In addition, during the preference evaluation, Adam chose RC more (13 out of 18 trials; 72% of trials) than DRA (5 out of 18 trials; 28% of trials) and on-task behavior remained at moderate levels during DRA (M = 43%; range: 32%-53%) and higher and more stable during RC (M = 51%; range: 28%-73%).

Results for the comparison between DRA and RC for Collin are in the bottom panel of Figure 4. During the initial baseline phase, Collin displayed low levels of on-task behavior (M = 5%; range: 0%-25%). During the DRA versus RC phase, Collin displayed high levels of on-task behavior during both DRA (M = 97%; range: 96%-98%) and RC (M = 97%; range: 90%-100%) as compared to baseline (M = 0%). In addition, during the preference evaluation, Collin selected RC on every choice opportunity (11 out of 11 trials; 100% of trials) and on-task behavior remained high and stable (M = 97%; range: 95%-100%).

Results for the comparison between DRA and RC (multielement and reversal) for Anna are in the top panel of Figure 5. During the initial baseline phase, Anna displayed low levels of on-task behavior (M = 16%; range: 8%-22%). During the DRA versus RC phase, Anna displayed increasing and high levels of on-task behavior during both DRA (M = 80%; range: 37%-100%) and RC (M = 76%; range: 25%-96%) as compared to baseline (M = 0%). During the preference evaluation, Anna chose RC more (14 out of 22 trials; 64% of trials) than DRA (3 out of 22 trials; 14% of trials) and on-task behavior remained high and stable during DRA (M = 85%; range: 68%-96%) and RC (M = 87%; range: 75%-95%). During the reversal, levels of on-task behavior during baseline were low (M = 18%; range 0%-90%), and levels of on-task behavior were high during DRA (M = 89%; range 75%-97%) and RC (M = 88%; range 67%-100%). Therefore, the reversal design allowed us to show that DRA and RC were equally effective procedures even when they were not implemented in a rapidly alternating fashion.

Results for the comparison between DRA and RC for Caroline are in the bottom panel of Figure 5. During the initial baseline phase, Caroline displayed low levels of on-task behavior (M = 5%; range: 0%-20%). During the DRA versus RC phase, Caroline displayed higher levels of on-task behavior during both DRA (M = 64%; range: 20%-83%) and RC (M = 65%; range: 0%-100%) as compared to baseline (M = 0%). In addition, during the preference evaluation, Caroline chose RC on every choice opportunity (9 out of 9 trials; 100% of trials) and on-task behavior remained at moderate levels. (M = 68%; range: 58%-83%). During the reversal, levels of on-task behavior during baseline were low (M = 38%; range 0%-97%), and levels of on-task behavior were high during DRA (M = 90%; range 78%-100%) and RC (M = 95%; range 88%-100%). As with Anna, the reversal design allowed us to show the effects of DRA and RC when they were implemented repeatedly within a phase rather than when those conditions were rapidly alternated. Therefore, I was able to reduce the possibility of carryover effects or lack of discrimination resulting in the similar effects across procedures, at least for Anna and Caroline.

Results for participants for whom DRA and RC were evaluated using a reversal design only are depicted Figures 6-7. I included this evaluation with several participants to attempt to determine whether the multielement design used with the other participants may have influenced the similar efficacy across the procedures due to treatment interference. Results for Brianna are in the top panel of Figure 6. Brianna engaged in high levels of on-task behavior during the first evaluation of DRA (M = 91%; range: 88%-95%) and RC (M = 88%; range: 75%-95%) and the second evaluation of DRA (M = 86%; range: 77%-95%) as compared to baseline (M = 20%). Results for Mark are in the bottom panel of Figure 6. Mark displayed high levels of on-task behavior during the first evaluation of DRA (M = 96%; range: 93%-98%) and RC (M = 93%; range: 85%-97%) and the second evaluation of DRA (M = 94%; range: 90%-97%) as compared to baseline (M = 0%). Results for Zoey are on the top panel of Figure 7. During DRA, Zoey engaged high levels of on-task behavior during the first evaluation of DRA (M = 95%; range: 88%-100%) and RC (M = 95%; range: 93%-98%) and the second evaluation of DRA (M = 95%; range: 92%-98%) as compared to baseline (M = 0%). Results for Sam are displayed in the bottom panel of Figure 7. During baseline sessions, on-task behavior occurred at low levels (M = 4%; range: 0%-31%). During the first DRA evaluation, on-task behavior occurred at variable, moderate levels (M = 63%; range: 35%-97%). During the first RC evaluation, on-task behavior occurred at high levels (M = 86%; range: 43%-98%). During the second evaluation of DRA, Sam displayed variable, moderate levels of on-task behavior (M = 60%; range: 77%-98%). During the second evaluation of RC, Sam displayed high levels of on-task behavior (M = 93%; range: 85%-100%).

Overall, the results from the comparison of DRA and RC showed that the procedures were equally effective for maintaining high levels of on-task behavior for 11 participants, DRA was more effective for one participant, and RC was more effective for two participants (see Table 3). The criterion by which a procedure was determined to be more effective was if the average level of on-task behavior was at least, or greater than, 10% more than the comparison procedure. With respect to effectiveness, the findings of the current study were similar to those reviewed in this paper. That is, the majority of studies comparing DR to RC found both procedures to be equally effective (8 out of 13; e.g., Iwata and Bailey, 1974; McGoey & DuPaul, 2000), and a few found one procedure to be more effective than the other (e.g., Conyers et al., 2004; Tanol et al., 2010). In addition, preference for DRA and RC varied among participants for whom it was evaluated, with five preferring DRA and five preferring RC (see Table 3), and these results were also similar to those reviewed in the current paper. That is, previous studies (e.g., Donaldson et al., 2014; Iwata & Bailey, 1974) also showed that preference varied among individuals.

Several reasons may explain similarities and differences in the effectiveness of these procedures. First, participants' level of on-task behavior in the current study may have been similar across procedures because the task was not effortful, which may not have been the case in other studies that found disparate results. That is, the tracing task used in the current study was a mastered task that was relatively simple for most participants to complete. Second, the participants in the current study may have engaged in similar levels of on-task behavior across both conditions because the value of the reinforcers available for token exchange was high and equal across conditions. That is, regardless of the condition, the motivation to earn or retain tokens was sufficient for maintaining on-task behavior.

Next, the similarity in effectiveness of the two procedures in the current study might be due to the experimental design used (i.e., multielement). Although there are many strengths to the use of the multielement design, such as efficiency and reducing the influence of history effects (Conners, Iwata, Kahng, Hanley, Worsdell, & Thompson, 2000; Hains & Baer, 1989), there are also some possible limitations. Two limitations of the design are the possibility of discriminative failure across conditions and carryover effects from one condition to the next due to the rapidity with which the conditions are alternated. Thus, it is possible that similar results were found across procedures due to the use of this design. Although it is possible that discrimination failure could occur with the rapid alternation of conditions given that the sessions across the two token conditions looked very similar (i.e., same task and therapist, same setup with presence of toys and task, and presence of token boards), I implemented various procedures to increase the likelihood of discrimination across conditions (e.g., different color token boards).

In addition to the discriminative stimuli used, in-session observations suggest that all participants could describe the contingencies associated with each condition. Finally, the data show that most of the participants' preferences were very strong for one condition over the other, which suggests that they were discriminating across the two conditions.

Because there was still the possibility of similar results due to carryover effects with the rapid alternation of the two token procedures, I attempted to rule this out by comparing the two procedures using a reversal design with two participants (Anna and Caroline). Results of this evaluation showed no difference in the effectiveness of the two procedures, which was similar to the results of the evaluation when using a multielement design. However, because the reversal design followed the multielement design, I could not rule out the influence of the participants' history with both procedures earlier in the study. Specifically, researchers have shown that the effects of punishment sometimes result in (temporary) behavioral suppression even after punishment is removed (e.g., Harris & Ersner-Hershfield, 1978; Mazur, 2006; Palen McGlynn & Locke, 1997). Therefore, it could be that that a history of response cost may influence the effectiveness of reinforcement. In order to evaluate DRA prior to exposure to RC, I conducted an evaluation using a reversal design only with five participants.

The results of these data showed that three participants had high levels of on task behavior during DRA (Zoey, Brianna, and Mark; see Figure 6 and 7), and one participant had variable levels (Sam; see Figure 8). Data for the four participants whose responding during DRA was moderate to high without exposure to RC suggest that DRA is an effective procedure for increasing on-task behavior. However, this conclusion is tentative given that the participant's history with such procedures prior to the current study is unknown. Alternatively, data for Sam suggest that RC may influence the effectiveness of DRA for some individuals because responding was highly variable during the initial evaluation of DRA, and somewhat more stable and high following RC. However, it should also be noted that among studies reviewed in the current paper that compared DR to RC using a reversal design, none showed differences in the effectiveness of the DR procedure following exposure to RC (e.g., McGoey & DuPaul, 2000; Sindelar et al., 1982; Tanol et al., 2010). Therefore, it seems unlikely that use of a multielement design is confounding; however, as previously mentioned, this is an area of research that warrants investigation.

In addition to the type of task, potency of the backup reinforcers, and the design used, there are other possible reasons why we observed similar effects of the token procedures. That is, participants' level of on-task behavior in the current study may have been similar across procedures because the procedures compared in the current study were similar. That is, procedurally, DRA and RC involve direct contingencies for behavior occurring at a specific moment, whereas, DRO involves a contingency based on the occurrence of behavior during an entire interval. Therefore, during DRO, a token may not be earned at a specific moment when on-task behavior is occurring because off-task behavior occurred some time during the DRO interval. In addition, it may be that the effectiveness of the two procedures is similar because the number of tokens that could be earned (in DRA) or kept (in RC) was the same across conditions. That is, I programed for equal opportunity for net tokens in order to enhance experimental control when evaluating the independent variable, and this method has been used in previous research.

With respect to preference, as previously mentioned, some children preferred DRA and others preferred RC. Several reasons may explain varied preference among individuals. Preference for DRA over RC may have occurred for some individuals because they like to earn

(McGoey & DuPaul, 2000) or because RC was aversive in that it may have produced “anxiety” or “fear of failure” (Capriotti et al. 2012; Hallahan, Tarver, Kauffman, & Graybeal, 1978; Humphrey & Karoly, 1978; McGoey & DuPaul, 2000). Although response cost procedures may involve aversive consequences, half of the children preferred RC. Preference for RC over DRA may be due to the presence of the tokens (Donaldson et al., 2014). That is, immediate selection of the RC procedure resulted in the “delivery” of all tokens. In addition, selection of RC over DRA may be because, from the child’s perspective, starting with tokens was viewed as not having to work for the tokens. That is, the procedure appears less effortful.

Finally, varied preference among children may be a result of the average number of net tokens across conditions. That is, if some children averaged more net tokens in one procedure over another, a child may choose the procedure that averaged more net tokens than the procedure that averaged less net tokens. For example, if a child yielded an average of seven tokens during DRA and nine tokens during RC, the child may prefer RC (i.e., select RC more often than DRA) due to a larger magnitude of back-up reinforcers. With respect to magnitude of net tokens, the results of the token analysis showed that 5 out of 10 participants preferred the procedures for which the net tokens were greater. However, it should be noted that the differences in net tokens were small for some of these participants. Because the current study and previous studies have not evaluated the influence of the opportunity for net tokens when comparing reinforcement to response cost, the influence of this variable on preference is unknown. In order to evaluate the relationship between average number of net tokens and effectiveness of and preference for the procedures, I calculated the average number of net tokens in each procedure for each participant.

The results of the token analysis are shown in the last two columns of Table 3. The token analysis included a determination of the average difference in the number and percent of net tokens yielded during DRA and RC phases (not including preference phases). For Frank, the procedures were equally effective and he preferred DRA. The token analysis suggested that the average net tokens for Frank was 9.6 (range: 9-10) during RC and 9 (range: 7-10) during DRA. The procedure that yielded more tokens was not the same as the procedure that was more preferred. For Emily, DRA was more effective and she preferred DRA. The token analysis suggested that the average net tokens for Emily was 8.6 (range: 8-10) during RC and 9.4 (range: 9-10) during DRA. The procedure that yielded more tokens was the same as the procedure that was more effective and more preferred. For Martin, the two procedures were equally effective and he preferred DRA. The token analysis suggested that the average net tokens for Martin was 9.7 (range: 9-10) during RC and 10 during DRA. The procedure that yielded more tokens was the same as the procedure that was more preferred. For Adrianna, the two procedures were equally effective and she preferred DRA. The token analysis suggested that the average net tokens for Adrianna was 9.8 (range: 9-10) during RC and 9.5 (range: 9-10) during DRA. The procedure that yielded more tokens was not the same as the procedure that was more preferred. For Paul, the two procedures were equally effective and he preferred DRA. The token analysis suggested that the average net tokens for Paul was 9.1 (range: 3-10) during RC and 9.5 (range: 7-10) during DRA. The procedure that yielded more tokens was the same as the procedure that was more preferred. For Anna, the two procedures were equally effective and she preferred RC. The token analysis suggested that the average net tokens for Anna was 8.2 (range: 4-10) during RC and 8.4 (range: 5-10) during DRA. The procedure that yielded more tokens was not the same as the procedure that was more preferred. For Elisa, the two procedures were equally effective and she preferred RC. The token analysis suggested that the average net tokens for Elisa was 7.1 (range: 3-10) during RC and 8.1 (range: 2-10) during DRA. The procedure that yielded more tokens was not the same as the procedure that was more preferred. For Collin, the two procedures were equally effective and he preferred RC. The token analysis suggested that the average net tokens for Collin was 9.8 (range: 9-10) during RC and 9.8 (range: 9-10) during DRA. The number of net tokens was equal across procedures. For Adam, RC was more effective than DRA and he preferred RC. The token analysis suggested that the average net tokens for Adam was 6.8 (range: 4-8) during RC and 5.8 (range: 2-9) during DRA. The procedure that yielded more tokens was the same as the procedure that was more effective and more preferred. For Caroline, the two procedures were equally effective and she preferred RC. The token analysis suggested that the average net tokens for Caroline was 7.9 (range: 0-10) during RC and 5.9 (range: 1-10) during DRA. The procedure that yielded more tokens was the same as the procedure that was more preferred. For Brianna, the two procedures were equally effective, and the average net tokens was 8.5 (range: 8-10) during DRA and 8.7 (range: 8-10) during RC. For Mark, DRA and RC were equally effective, and the average net tokens was 9.2 during DRA and 9.6 (range: 9-10) during RC. For Zoey, the two procedures were equally effective, and the average net tokens was 9.4 during DRA (range: 8-10) and 8.7 during RC (range: 7-10). For Sam, RC was more effective than DRA, and the average net tokens was 6.1 (range: 3-10) and 8.9 (range: 6-10). The procedure that yielded more tokens was the same as the procedure that was more effective.

Overall results of the token analysis showed that the average number of net tokens was 8.5 (range: 5.4-10) in DRA and 8.8 (range: 7.1-9.8) in RC. The difference between average tokens is similar to the findings of previous researchers who also found small differences in average net tokens across procedures (e.g., Donaldson et al., 2014; Iwata & Bailey, 1974). However, for three participants (Emily, Adam, and Sam), the procedure that yielded more net tokens was the same as the procedure that was more effective, and for five participants (Emily, Martin, Paul, Adam, and Caroline), the procedure that yielded more net tokens was the same as the procedure that was more preferred. It should also be noted that for one participant (Collin), the average net tokens across procedures was equal, and the procedures were equally effective. Although the token analysis provided some evidence that opportunity for net tokens may influence effectiveness of or preference for DRA and RC, the opportunity for net tokens was not directly manipulated—nor has this variable been manipulated in previous studies. Therefore, the influence of opportunity for net tokens is unknown.

Study 2 Method: Token Evaluation

Purpose

Previous research on the evaluation of the effectiveness of and preference for differential reinforcement and response cost used within token economies has been limited and inconsistent, and several variables have not been evaluated. One limitation of previous research is that the opportunity for net tokens across procedures has been equated (including the methods of Study 1 above). That is, the net tokens possible in each procedure was the same. Although equating the opportunity for net tokens allows for experimental control, it may be that equating the opportunity for net tokens resulted in similar effectiveness of procedures.

The purpose of Study 2 was to evaluate how the opportunity for net tokens influences the effectiveness of and preference for differential reinforcement and response cost within a token economy. Understanding how the opportunity for net tokens influences effectiveness of and preference for a procedure is important for several reasons. First, when opportunities for net tokens are many, the implementer must observe, and sometimes deliver a consequence, frequently. Therefore, if a procedure is just as effective when there are few opportunities for net tokens, the procedure becomes less effortful for the implementer. In addition, when the opportunities for net tokens are many, more back-up reinforcers are required, which increases monetary costs (if back-up reinforcers are tangible or edible) and time (time spent exchanging for and accessing back-up reinforcers such as extra time with an activity). Therefore, if a procedure is just as effective when there are few opportunities for net tokens, fewer resources are required.

In order to evaluate the influence of opportunity for net tokens, I first compared the effects of DRA and RC for increasing on-task behavior with individual children when the number of opportunities for net tokens was equal (as in Study 1). Next, I evaluated child preference for the different procedures when the number of opportunities for net tokens was equal (as in Study 1). Finally, I compared DRA and RC under conditions in which the opportunity for net tokens was manipulated during one procedure by changing the number of tokens that could be earned or kept while holding constant the number of tokens that could be earned or kept in the comparison procedure. Furthermore, I evaluated child preference for these different procedures when the number of opportunities for net tokens was different. This manipulation allowed for an evaluation of whether the relative opportunity for net tokens influenced the effectiveness of or preference for reinforcement and response cost.

Participants, Setting, and Materials

Six preschool-age (3 years 8 months to 4 years 11 months) children participated in Study 2. The gender and exact age of each participant is displayed in Table 4. Systematic data were collected on children's skills that were pertinent to the target task and token economy procedures. Results of these assessments showed that all children could follow three-step instructions (e.g., "Go put your blanket in your cubby, get your shoes, and sit in the chair"), rote count up to 10 when instructed to count to 10 with no stimuli present, count up to 10 objects when ask to count how many objects were present, give up to 10 objects to another person when asked to give another person a specified number of objects, answer “What” questions (e.g., “What happens if you are tracing?), describe what is happening or what they were doing when asked, “What are you doing?,” describe what happened when asked, “What happened when you were tracing?,” and name the colors red, green, and white.

The setting and materials were the same as in Study 1; however, the token store included a larger variety of items, and item values were one, three, and five tokens (See Appendix E). Leisure activities were one token for 1-min access (as in Study 1); however, trinkets and edibles could be one, three, or five tokens. At the start of the token store, items were initially assigned token values according to their monetary cost (i.e., most expensive items worth five tokens, and least expensive items worth one token). However, throughout the study, when novel items were added, these items often started at a price of five tokens because novel items tended to be highly valuable. However, over time, the token price of items that were not selected for several weeks by any participant at their current value was decreased. For example, if a spin-top toy was five tokens, and no participant exchanged for this item for 2-3 weeks, the item’s price was decreased to three tokens, and so on. Occasionally, the price of items increased if several participants exchanged for the same items within a few days. For example, several participants preferred pink- and purple-colored trinkets (e.g., glitter rings); therefore, the price of the pink and purple trinkets increased from three tokens to five. Participants rarely commented on an item’s value change.

Response Measurement and Interobserver Agreement

Trained graduate and undergraduate students collected data using handheld computers. The dependent variable during all sessions was percentage of on-task behavior. On-task behavior was defined and scored as in Study 1, and token delivery and removal was also scored as in Study 1. In addition, data were collected on participants' bids for attention (e.g., "Look at this") during a minimum of 15% of sessions; however, bids for attention rarely occurred for any participant (0 bids for Becky, Carly, and Nancy; 1 bid for Austin; 2 bids for Erinn and Imilia).

Two independent observers collected data during at least 30% of all sessions across participants. Interobserver agreement (IOA) was calculated for on-task behavior and token delivery and removal as in Study 1. Overall, interobserver agreement was 92% (range: 63%-100%) for on-task behavior and 96% (range: 74%-100%) for token delivery or removal. Individual IOA for Erinn was 92% (range: 77%-100%) for on-task behavior and 97% (range: 74%-100%) for token delivery or removal. IOA for Nancy was 91% (range: 75%-100%) for on-task behavior and 96% (range: 83%-100%) for token delivery or removal. IOA for Imilia was 90% (range: 85%-100%) for on-task behavior and 97% (range: 92%-100%) for token delivery or removal. IOA for Austin was 94% (range: 85%-100%) for on-task behavior and 90% (range: 88%-100%) for token delivery or removal. IOA for Becky was 90% (range: 63%-100%) for on-task behavior and 98% (range: 87%-100%) for token delivery or removal. IOA for Carly was 92% (range: 75%-100%) for on-task behavior and 95% (range: 85%-100%) for token delivery or removal. For some participants, one or two sessions had IOA lower than 80%. When this happened, I provided additional training on the operational definitions to all data collectors.

Treatment Integrity

During at least 25% of sessions, an observer used a paper and pencil to record child behavior (i.e., on or off task) and experimenter behavior (i.e., token delivery or removal) during a scheduled observation (every 30 s or 1 min, depending on the condition) by circling *on task* or *off task* for child behavior and *T*+ (token delivery), *T*- (token removal), or *n/a* for experimenter behavior (*n/a* was circled if there was no consequence) (see Appendix F). Comparing child and experimenter behavior during each observation determined treatment integrity.

DRA. During DRA, if an observer recorded on task and token delivery, a value of 1 was given for that trial. If an observer recorded on task and n/a, a value of 0 was given for that trial. If an observer recorded off task and token delivery, a value of 0 was given for that trial. If an observer recorded off task and n/a, a value of 1 was given for that trial. Token removal was never scored during these sessions. After all observations had been scored as correct (a score of 1) or incorrect (a score of 0), the scores were summed and divided by the total number of observations. For example, if a token was delivered (or not delivered) correctly during eight observations, but a token was delivered (or not delivered) incorrectly during two observation, treatment integrity would be calculated by dividing 8 by 10. This example would yield 80% accurate treatment integrity.

RC. During RC, if an observer recorded on task and n/a, a value of 1 was given for that trial. If an observer recorded on task and token removal, a value of 0 was given for that trial. If an observer recorded off task and n/a, a value of 0 was given for that trial. If an observer recorded off task and token removal, a value of 1 was given for that trial. Token delivery was never scored during these sessions. After all observations had been scored as correct (a score of 1) or incorrect (a score of 0), the scores were summed and divided by the total number of observations. For example, if a token was removed (or not removed) correctly during eight observations, but a token was removed (or not removed) incorrectly during two observation, treatment integrity would be calculated by dividing 8 by 10. This example would yield 80% accurate treatment integrity.

Overall treatment integrity was 98% (range: 60%-100%). Individual treatment integrity for Erinn was 98% (range: 90%-100%). Treatment integrity for Nancy was 99% (range: 90%-100%). Treatment integrity for Imilia was 95% (range: 60%-100%). It should be noted that the session during which treatment integrity for Imilia was 60% was during a five-token condition; therefore, two disagreements resulted in 60% accuracy. In addition, no other session was below 80%. Treatment integrity for Austin was 99% (range: 90%-100%). Treatment integrity for Becky was 98% (range: 90%-100%). Treatment integrity for Carly was 99% (range: 90%-100%).

Procedures

First, I replicated Study 1. That is, baseline, token-exchange training, DRA versus RC comparisons when 10 net tokens were available across both conditions, and preference for DRA versus RC when 10 net tokens were available. Next, I conducted the same evaluation (evaluating the effectiveness of and preference for DRA and RC); however, opportunity for net tokens was different between the two procedures. That is, in some phases, the opportunity for net tokens was 5 during DRA and 10 during RC. Finally, the opportunity for net tokens was switched between procedures (e.g., 5 during RC and 10 during DRA), and the effectiveness of and preference for the procedures was again evaluated. All other general procedures were the same as in Study 1.

Baseline. Prior to the start of all baseline sessions, the participant was given rules describing the contingencies for the session, and a white board with no tokens was present. Rules were stated as follows: “Today you get the white board, and there are no tokens. When we start, you can either work on tracing or play with toys. If you are working, nothing will happen, if you are not working, nothing will happen." During the session, no programmed consequences were provided.

DRA (10 tokens). This condition was the same as DRA in Study 1 except that observations of child behavior (to determine whether the child was on task) occurred using a fixed observation schedule (i.e., once every 30 s) rather than a variable observation schedule for the purpose of consequence delivery (token delivery or token removal).

DRA (5 tokens). Prior to the start of all sessions, the participant was given rules describing the contingencies for the session, and a green board with no tokens was present. The green token board was half the size of the token board used in DRA (10 tokens) and had five spaces on which tokens could be delivered. Rules were stated as follows: "Today you get the green board, and it doesn't have any tokens on it, but you can earn up to 5 tokens. When we start, you can either work on tracing or play with toys. If you are working, you will get a token. If you are not working, you will not get any tokens. At the end, you can trade your tokens for prizes and snacks. If you don't have any tokens, you don't get anything." During the session, the experimenter observed the participant once every minute. Therefore, the number of opportunities for token delivery was 5, as compared to 10 during DRA (10 tokens). At an observation time, if the participant was on task, then the experimenter placed a token on the token board. However, if the participant was not on task, no consequences were provided.

RC (10 tokens). This condition was the same as RC in Study 1 except that observations of child behavior occurred using a fixed observation schedule (i.e., once every 30 s).

RC (5 tokens). Prior to the start of all sessions, the participant was given rules describing the contingencies for the session, and a red board with five tokens was present. The red token board was half the size of the token board used during RC (10 tokens), and 5 tokens were attached to the board. Rules were stated as follows: "Today you get the red board, and it has five tokens on it. When we start, you can either work on tracing or play with toys. If you are working, you will keep your tokens. If you are not working, you will lose tokens. At the end, you can trade your tokens for prizes and snacks. If you don't have any tokens, you don't get anything." During the session, the experimenter observed the participant once every minute. Therefore, the number of opportunities for token removal was 5, as compared to 10 during RC (10 tokens). At an observation time, if the participant was on task, then no consequences were provided. However, if the participant was not on task, then the experimenter removed a token from the token board.

Preference Evaluation

After stable levels of responding were observed during each of the DRA versus RC comparisons, preference was conducted as in Study 1; however, the token boards that were present were those that were used during the previous comparison phase. For example, when DRA (10) was compared to RC (5), the token boards associated with these conditions were presented together (i.e., green token board with 10 spaces and red token board with five tokens), in addition to the token board used in baseline (i.e., blank white board).

Experimental Design

A multielement and reversal design was used to evaluate the effects of varying the opportunity for net tokens across conditions. That is, after comparing the effectiveness of and preference for DRA and RC under conditions in which the number of opportunities for net tokens was equal (10 in each condition), the number of opportunities for net tokens was changed to five during one procedure. Once stable responding was observed during this comparison, the number of opportunities for net tokens was changed to five in the other procedure, and the number of opportunities for net tokens during the first procedure was changed back to 10.

The order in which participants experienced the various conditions depended on patterns of responding during the comparison and preference phases. Participants who showed preference for DRA when the opportunity for net tokens was equal, first experienced DRA (5) versus RC (10) then DRA (10) versus RC (5). Participants who showed preference for RC when the opportunity for net tokens was equal, first experienced DRA (10) versus RC (5) then DRA (5) versus RC (10). All phases during which the effectiveness of or preference for a procedure changed due to net token manipulation were replicated using a reversal design (with the exception of one participant with whom I was unable to continue evaluation due to inconsistent attendance).

Token Analysis

A token analysis was conducted for all comparison phases as in Study 1.

Results and Discussion

Figures 9-11 display graphs for the six children who participated in Study 2. Results for the comparison between DRA and RC for Erinn are in the top panel of Figure 6. During the initial baseline phase, Erinn displayed low levels of on-task behavior (M = 9%; range: 0%-18%). During the comparison evaluation, when the opportunity for net tokens was the same (DRA [10] vs. RC [10]), Erinn engaged in high levels of on-task behavior during DRA (M = 82%; range: 88%-100%) and RC (M = 83%; range: 90%-100%), and responding during baseline probes was zero (baseline patterns of responding were the same during all subsequent evaluations). These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Erinn alternated her selection between DRA and baseline; therefore, I removed baseline as an option in order to evaluate preference for the two token procedures (see notation on graph for session 32). Before the baseline choice was removed, Erinn selected DRA on two choice opportunities (2 out of 7 opportunities; 29% of trials) and baseline on five choice opportunities (5 out of 7 opportunities; 71% of trials). After the baseline choice was removed, Erinn selected DRA on all choice opportunities (3 out of 3 opportunities; 100% of trials). In addition during the preference evaluation, on-task behavior remained high and stable (M = 92%; range: 88%-97%). These data suggest that when the opportunity for net tokens was the same, Erinn preferred DRA as compared to RC.

When the opportunity for token delivery for Erinn was changed to 5 for DRA, while the opportunity for token removal during RC remained at 10 (DRA [5] vs. RC [10]), Erinn engaged in high and stable levels of responding during DRA (M = 92%; range: 87%-100%) and RC (M = 97%; range: 95%-100%). These data suggest that the procedures remained equally effective despite fewer opportunities for net tokens during DRA. When evaluating preference under these same conditions (DRA [5] and RC [10]), Erinn alternated her selection of the two procedures; she selected DRA on 4 out of 9 opportunities (44% of trials) and RC on 5 out of 9 opportunities (56% of trials). In addition, levels of on-task behavior remained high during DRA (M = 95%; range: 95%-98%) and RC (M = 93%; range: 85%-100%). These data suggest that fewer opportunities for net tokens changed preference for DRA; that is, percent selection of DRA decreased when there were fewer opportunities for net tokens during DRA as compared to RC.

Next, when DRA (10) versus RC (10) was replicated, Erinn engaged in high and stable levels of responding during DRA (M = 97%; range: 95%-98%) and RC (M = 95%; range: 85%-100%). These data replicated the results of our initial evaluation demonstrating that both procedures are equally effective when the opportunity for net tokens is the same. When replicating preference under these same conditions, Erinn selected DRA on 7 out of 9 opportunities (77% of trials) and RC on 2 out of 9 opportunities (23% of trials), which replicated the results of our previous preference evaluation under these conditions. In addition, levels of on-task behavior remained high during DRA (M = 94%; range: 85%-100%) and RC (M = 94%; range: 90%-98%). These data provide evidence that the opportunity for net tokens influences preference for DRA. I was unable to conduct any further evaluations do to attrition (the participant was no longer enrolled in the Child Development Center).

Results for the comparison between DRA and RC for Nancy are in the bottom panel of Figure 9. During the initial baseline phase, Nancy displayed low levels of on-task behavior (M = 17%; range: 0%-58%). During DRA (10) versus RC (10), Nancy engaged in high levels of on-task behavior during DRA (M = 85%; range: 73%-92%) and RC (M = 86%; range: 67%-97%), and responding during baseline probes was zero (baseline patterns of responding were the same during all subsequent evaluations). These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Nancy alternated her selection between RC and baseline; however, she selected RC on 7 out of 11 opportunities (64% of trials) and baseline on 4 out of 11 opportunities (36% of trials). Although Nancy selected baseline on some trials, she chose RC more often and never selected DRA. In addition when she chose RC during the preference evaluation, on-task behavior remained high and stable (M = 80%; range: 66%-97%). These data suggest that when the opportunity for net tokens was the same, Nancy preferred RC as compared to DRA.

During DRA (10) vs. RC (5), Nancy engaged in high and stable levels of responding during DRA (M = 93%; range: 91%-98%) and RC (M = 89%; range: 85%-95%). These data suggest that the procedures remained equally effective despite fewer opportunities for net tokens during RC. When evaluating preference under these same conditions, Nancy selected DRA on 6 out of 6 opportunities (100% of trials). In addition, levels of on-task behavior remained high during DRA (M = 91%; range: 88%-95%). These data suggest that fewer opportunities for net tokens changed preference for RC; that is, percent selection of RC decreased to zero when there were fewer opportunities for net tokens as compared to DRA.

During DRA (5) vs. RC (10), Nancy engaged in moderate to high levels of responding during DRA (M = 77%; range: 0%-95%) and RC (M = 69%; range: 0%-93%). During this phase, responding decreased to zero for three consecutive sessions, and I hypothesized that the back-up reinforcers were no longer valuable and, therefore, added new items, after which I saw increased responding. These data suggest that the procedures remained equally effective despite fewer opportunities for net tokens during DRA. When evaluating preference under these same conditions, Nancy selected DRA on 5 out of 5 opportunities (100% of trials). In addition, levels of on-task behavior remained high during DRA (M = 88%; range: 78%-97%). Because preference for DRA continued despite fewer opportunities for net tokens during DRA, these data suggest a possible strong preference for DRA.

Next, I re-evaluated effectiveness and preference during DRA (10) vs. RC (10). Under these conditions I found that DRA (M = 85%; range: 82%-90%) and RC (M = 82%; range: 67%-93%) were equally effective, which replicated the results of the initial evaluation. In addition, Nancy selected DRA on all choice opportunities (5 out of 5 opportunities; 100% of trials), which suggests continued preference for DRA.

Because I did not observe a change in the effectiveness of DRA or RC under any condition, but did observe a change in preference, I next replicated previous preference evaluations. During DRA (5) versus RC (10), Nancy selected DRA on 1 out of 6 opportunities

(17% of trials) and RC on 5 out of 6 opportunities (83% of trials). Thus, when DRA was associated with less opportunity for net tokens, and RC was associated with more opportunity for net tokens, RC was more preferred. Next, I re-evaluated preference under DRA (10) versus RC (5), and Nancy selected DRA on all choice opportunities (5 out of 5 opportunities; 100% of trials). These data suggest that more opportunity for net tokens for DRA, as compared to RC, resulted in preference for DRA. Finally, I re-evaluated preference under DRA (5) versus RC (10), and Nancy selected RC on 5 out of 5 opportunities (100% of trials). These data suggest that more opportunity for net tokens during RC, as compared to DRA, resulted in preference for RC. In addition, during all preference evaluations, on-task behavior remained high and stable (DRA: M = 92%, RC: M = 87%; DRA: M = 93%; DRA: M = 90%, RC: M = 92%)

Overall, preference results for Nancy are difficult to interpret. That is, initially when the opportunity for net tokens across DRA and RC was equated, Nancy preferred RC; and when DRA was associated with more possible net tokens, she preferred DRA. However, in the next two evaluations in which RC was associated with more possible net tokens, and then net tokens were again equated, Nancy continued to prefer DRA suggesting that more exposure to the procedures changed her preference to DRA regardless of the net token opportunity. However, in the last three phases in which only preference was evaluated, Nancy's preference seemed to be associated with whichever procedure resulted in more net token opportunity. Thus, conclusions regarding the influence of net tokens on Nancy's preference are tentative.

Results for the comparison between DRA and RC for Imilia are in the top panel of Figure 10. During the initial baseline phase, although Imilia engaged in high levels of on-task behavior during two sessions; however, she engaged in low levels of responding during the other three sessions (M = 30%; range: 0%-93%). During DRA (10) vs. RC (10), Imilia engaged in high levels of on-task behavior during DRA (M = 89%; range: 59%-98%) and RC (M = 92%; range: 82%-98%), and responding during baseline probes in this and subsequent phases was low to zero (M = 25%; range: 0%-96%), with the exception of two sessions. These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Imilia chose DRA on all choice opportunities (100% of trials). In addition, during the preference evaluation, on-task behavior remained high and stable (M = 96%; range: 90%-100%). These data suggest that when the opportunity for net tokens was the same, Imilia preferred DRA as compared to RC.

During DRA (5) versus RC (10), Imilia engaged in high levels of responding during DRA (M = 92%; range: 85%-98%) and RC (M = 92%; range: 78%-100%). These data suggest that the procedures were equally effective despite fewer opportunities for net tokens during DRA as compared to RC. When evaluating preference under these same conditions, Imilia selected DRA on 6 out of 7 opportunities (86% of trials) and RC on 1 out of 7 opportunities (14% of trials). In addition, levels of on-task behavior remained high during DRA (M = 90%; range: 72%-100%) and RC (91%). Because preference for DRA continued despite fewer opportunities for net tokens in DRA, as compared to RC, these data suggest a strong preference for DRA.

During DRA (10) versus RC (5), Imilia engaged in high levels of responding during DRA (M = 95%; range: 90%-97%) and RC (M = 89%; range: 83%-92%). These data suggest that the two procedures remained equally effective despite fewer opportunities for net tokens in RC as compared to DRA. When evaluating preference under these same conditions, Imilia selected DRA on 5 out of 5 opportunities (100% of trials) and engaged in high levels of responding during DRA (M= 92%; range: 87%-95%). These data suggest that Imilia preferred the procedure with a greater opportunity for net tokens.

Results for the comparison between DRA and RC for Austin are on the bottom panel of Figure 7. During the initial baseline phase, Austin engaged in low levels of on-task behavior (M = 21%; range: 0%-73%). During DRA (10) versus RC (10), Austin engaged in high levels of on-task behavior during DRA (M = 91%; range: 87%-95%) and RC (M = 87%; range: 75%-93%), and responding during baseline probes in this and subsequent phases was low (M = 11%; range: 0%-75%), with the exception of one session. These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Austin chose DRA on all choice opportunities (100% of trials). In addition during the preference evaluation, on-task behavior remained at moderate levels (M = 59%; range: 2%-92%), with the exception of two sessions. These data suggest that when the opportunity for net tokens was the same, Austin preferred DRA as compared to RC.

During DRA (5) versus RC (10), Austin engaged in high levels of responding during DRA (M = 89%; range: 80%-92%) and RC (M = 91%; range: 80%-95%). These data suggest that the procedures were equally effective despite fewer opportunities for net tokens in DRA as compared to RC. When evaluating preference under these same conditions, Austin selected DRA on all choice opportunities (100% of trials). In addition, levels of on-task behavior remained at moderate levels during DRA (M = 71%; range: 37%-82%), with the exception of one session. Because preference for DRA continued despite fewer opportunities for net tokens during DRA as compared to RC, these data suggest a strong preference for DRA.

During DRA (10) versus RC (5), Austin engaged in moderate to high levels of on-task behavior during DRA (M = 84%; range: 68%-100%) and RC (M = 90%; range: 77%-98%). These data suggest that the two procedures remained equally effective despite fewer opportunities for net tokens during RC as compared to DRA. When evaluating preference, Austin selected DRA on all choice opportunities (5 out of 5; 100% of trials) and engaged in low levels of responding during DRA (M = 41%; range: 27%-78%). These data suggest that Austin preferred the procedure with a greater opportunity for net tokens.

Results for the comparison between DRA and RC for Becky are in the top panel of Figure 11. During the initial baseline phase, Becky engaged in zero levels of on-task behavior. During DRA (10) versus RC (10), Becky engaged in moderate to high levels of on-task behavior during DRA (M = 77%; range: 33%-93%) and RC (M = 70%; range: 32%-100%), and responding during baseline probes was zero (baseline patterns of responding were the same during all subsequent evaluations). These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Becky frequently selected baseline; therefore, I removed baseline as an option in order to evaluate preference for the two token procedures (see notation on graph for session 41). Before the baseline choice was removed, Becky selected DRA on two choice opportunities (2 out of 6 opportunities; 33% of trials) and baseline on four choice opportunities (4 out of 6 opportunities; 66% of trials). After the baseline choice was removed, Becky alternated her selection between DRA (5 out of 10 opportunities; 50% of trials) and RC (5 out of 10 opportunities; 50% of trials). In addition during the preference evaluation, on-task behavior remained high and stable for DRA (M = 92%; range: 78%-98%) and RC (M = 90%; range: 77%-98%). These data suggest that when the opportunity for net tokens was the same, Becky did not have a specific preference.

During DRA (10) versus RC (5), Becky engaged in high and stable levels of responding during DRA (M = 95%; range: 88%-98%) and RC (M = 96%; range: 92%-100%). These data suggest that the procedures remained equally effective despite fewer opportunities for tokens during RC as compared to DRA. When evaluating preference, Becky selected DRA on 6 out of

6 opportunities (100% of trials). In addition, levels of on-task behavior remained moderate to high during DRA (M = 76%; range: 58%-93%). These data suggest that fewer opportunities for net tokens during RC as compared to DRA changed preference for RC; that is, Becky preferred the procedure with greater opportunity for net tokens.

During DRA (5) versus RC (10), Becky engaged in increasingly high levels of responding during DRA (M = 84%; range: 62%-95%) and RC (M = 85%; range: 75%-95%). These data suggest that the procedures were equally effective despite fewer opportunities for net tokens during DRA as compared to RC. When evaluating preference, Becky selected RC on 5 out of 5 opportunities (100% of trials). In addition, levels of on-task behavior remained high (M = 96%; range: 88%-100%). These data suggest Becky preferred the procedure with greater opportunity for net tokens.

Because I did not observe a change in the effectiveness of DRA or RC under any condition, but I did observe a change in preference, I next replicated previous preference evaluations. First, I re-evaluated preference during DRA (10) versus RC (10), and Becky selected DRA on one of seven opportunities (14% of trials) and RC on 6 out of 7 opportunities (86% of trials). These data suggest preference for RC, which was different than preference during our initial evaluation. In addition, levels of on-task behavior remained high during DRA (95%) and RC (M = 92%; range: 82%-98%). Next, I re-evaluated preference during DRA (10) versus RC (5). Under these conditions Becky selected DRA on six out seven opportunities (86% of trials) and RC on 1 out of 7 opportunities (14% of trials), which suggests that Becky prefers the procedure with greater opportunity for net tokens. These data replicated those of our previous evaluation. In addition, levels of on-task behavior were variable during DRA (M = 75%; range: 25%-95%) and 80 % during RC.

Results for the comparison between DRA and RC for Carly are in the bottom panel of Figure 11. During the initial baseline phase, Carly engaged in low levels of on-task behavior (M = 6%; range: 0%-13%). During DRA (10) versus RC (10), Carly engaged in increasingly high levels of on-task behavior during DRA (M = 65%; range: 28%-98%) and RC (M = 69%; range: 48%-98%), and responding during baseline probes in this and subsequent phases was zero. These data suggest that when the opportunity for net tokens was the same, the procedures were equally effective. During the preference evaluation, Carly chose DRA on all choice opportunities (100% of trials). In addition during the preference evaluation, on-task behavior was variable (M = 66%; range: 28%-89%). These data suggest that when the opportunity for net tokens was the same, Carly preferred DRA as compared to RC.

During DRA (5) versus RC (10), Carly engaged in variable, moderate to high levels of responding during DRA (M = 75%; range: 47%-92%) and RC (M = 83%; range: 67%-97%). These data suggest that the two procedures were equally effective despite fewer opportunities for tokens during DRA as compared to RC. When evaluating preference under these same conditions, Carly selected RC on all choice opportunities (100% of trials). In addition, levels of on-task behavior were high (M = 94%; range: 90%-98%). These data suggest that Carly preferred the procedure with greater opportunity for net tokens.

During DRA (10) versus RC (5), Carly engaged in variable levels of responding during DRA (M = 75%; range: 23%-100%) and high and relatively stable levels during RC (M = 73%; range: 72%-91%). These data suggest that fewer opportunities for tokens during RC, as compared to DRA, result in more stable levels of responding. When evaluating preference under these same conditions, Carly selected DRA on all choice opportunities (100% of trials). In addition, levels of on-task behavior were high (M = 88%; range: 63%-100%). These data suggest that Carly preferred the procedure with greater opportunity for tokens.

Taken together, results from the comparison of DRA and RC when the opportunity for net tokens was the same across procedures replicated the findings of Study 1, as well as previous research. That is, the procedures were equally effective, and preference varied among individuals (see Table 5). However, more participants preferred DRA to RC (4 out of 6), and one participant had equal preference.

Results from the comparison of DRA (5) versus RC (10) showed that the procedures were equally effective when the opportunity for net tokens was less during DRA as compared to RC (see Table 6). These data suggest that despite fewer opportunities for net tokens during DRA, the procedure remained effective. These results have important implications with respect to implementation because these results suggest that only five opportunities for tokens during a 5-min work period is sufficient for maintaining on-task behavior. Therefore, implementers can provide fewer opportunities for tokens throughout a work period, which will result in less time and effort to implement. With respect to preference, three participants preferred DRA (Austin, Imilia, and Nancy), two preferred RC (Becky and Carly), and one participant had equal preference (Erinn). These data have several implications. First, because three participants preferred DRA (and one participant continued to select DRA occasionally) when there were fewer opportunities for net tokens during DRA as compared to RC, it could be that (a) some characteristic of DRA controls preference for the procedure (e.g., children like to earn; McGoey & DuPaul, 2000) or (b) some characteristic of RC is aversive; therefore, participants select DRA to avoid RC contingencies (e.g., Pietras et al., 2010). Alternatively, because two participants preferred RC when there was greater opportunity for net tokens during RC as compared to DRA, it could be that these participants preferred the potential for a greater magnitude of back-up reinforcers; however, it could also be that these participants selected RC due to the presence of the tokens (i.e., selection of RC results in immediate access to tokens; Donaldson et al., 2014).

Results from the comparison of DRA (10) versus RC (5) showed that the procedures were equally effective when the opportunity for net tokens was less during RC as compared to DRA (see Table 6). These data suggest that despite fewer opportunities for net tokens during RC, the procedure remained effective. These results have important implications with respect to implementation, as mentioned above, because these results suggest that only five opportunities for tokens during a 5-min work period is sufficient for maintaining on-task behavior. Therefore, implementers can provide fewer opportunities for tokens throughout a work period, which will result in less time and effort to implement. With respect to preference, all participants (4 out of 4) preferred DRA. These data suggest that (a) some characteristic of DRA is more preferred than RC, (b) some characteristic of RC is aversive, and therefore, participants selected DRA to avoid RC contingencies, or (c) participants preferred a potential for greater magnitude of back-up reinforcers.

The results of the token analysis for DRA (10) versus RC (10) are presented in Table 5. Token-analysis results for Austin show that the procedures were equally effective and he preferred DRA. The token analysis suggested that the average net tokens for Austin was 8.3 (range: 6-8) during DRA and 8.5 (range: 7-10) during RC. The condition that yielded more tokens was not the same as the procedure that was more preferred. For Becky, the procedures were equally effective and preference for the procedures was equal. The token analysis suggested that the average number of net tokens for Becky was 6.6 (range: 5-10) during DRA and 6.9 (range: 1-10) during RC. The condition that yielded more tokens was selected on some, but not all, opportunities during the preference evaluation. For Carly, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Carly was 5.9 (range: 1-10) during DRA and 6.5 (range: 4-9) during RC. The condition that yielded more net tokens was not the condition that was more preferred. For Erinn, during the first evaluation, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Erinn was 7.6 (range: 3-10) during DRA and 9.3 (range: 8-10) during RC. The procedure that yielded more net tokens was not the condition that was more preferred. For Imilia, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Imilia was 8.8 (range: 4-10) during DRA and 9.1 (range: 9-10) during RC. The condition that yielded more net tokens was not the condition that was more preferred. For Nancy, during the initial evaluation, the procedures were equally effective and she preferred RC. The token analysis suggested that the average number of net tokens for Nancy was 7.2 (range: 6-8) during DRA and 9.3 (range: 8-10) during RC. The condition that yielded more net tokens was the same condition that was more preferred.

During the second evaluation of DRA (10) versus RC (10) for Erinn, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens was 9.8 (range: 9-10) during DRA and 9.3 (range: 8-10) during RC. The procedure that yielded more net tokens was not the same as the procedure that was more preferred. For Nancy, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Nancy was 8.3 (range: 6-10) during DRA and 9 during RC (range: 8-10). The procedure that yielded more net tokens was not the procedure that was more preferred.

Token-analysis results from DRA (5) versus RC (10) are presented in Table 6. For Austin, the procedures were equally effective, and he preferred DRA. The token analysis suggested that the average net tokens for Austin was 4.5 (range: 4-5) during DRA and 9.6 (range: 9-10) during RC. The procedure that yielded more tokens was not the same as the procedure that was more preferred. For Becky, the procedures were equally effective and she preferred RC. The token analysis suggested that the average number of net tokens for Becky was 4 (range: 0-5) during DRA and 9 (range: 7-10) during RC. The procedure that yielded more tokens was the procedure that was more preferred. For Carly, the procedures were equally effective and she preferred RC. The token analysis suggested that the average number of net tokens for Carly was 3.5 (range: 2-5) during DRA and 8.5 (range: 6-10) during RC. The procedure that yielded more net tokens was same the procedure that was more preferred. For Erinn, the procedures were equally effective and preference was equal. The token analysis suggested that the average number of net tokens for Erinn was 4.7 (range: 4-5) during DRA and 9.6 (range: 9-10) during RC. The procedure that yielded more net tokens was selected on some, but not all, choice opportunities. For Imilia, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Imilia was 4.4 (range: 3-5) during DRA and 9.6 (range: 8-10) during RC. The procedure that yielded more net tokens was not the procedure that was more preferred. For Nancy, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Nancy was 4.8 (range: 0-5) during DRA and 9.2 (range: 8-10) during RC. The procedure that yielded more net tokens was the same procedure that was more preferred.

Token-analysis results from DRA (10) versus RC (5) are presented in Table 6. For Austin, the procedures were equally effective and he preferred DRA. The token analysis suggested that the average net tokens for Austin was 6.8 (range: 3-10) during DRA and 4.2 (range: 3-5) during RC. The procedure that yielded more tokens was the same as the procedure that was more preferred. For Becky, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Becky was 9.8 (range: 9-10) during DRA and 5 during RC. The procedure that yielded more tokens was the procedure that was more preferred. For Carly, the procedures were equally effective. The token analysis suggested that the average number of net tokens for Carly was 6.9 (range: 1-10) during DRA and 4 (range: 1-5) during RC. For Imilia, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Imilia was 9.3 (range: 9-10) during DRA and 4.8 (range: 4-5) during RC. The procedure that yielded more net tokens was the procedure that was more preferred. For Nancy, the procedures were equally effective and she preferred DRA. The token analysis suggested that the average number of net tokens for Nancy was 9.2 (range: 8-10) during DRA and 4.1 (range: 3-5) during RC. The procedure that yielded more net tokens was the same procedure that was more preferred.

Overall results of the token analyses (Tables 5 and 6) showed that when the opportunity for net tokens was 10 for DRA, the average number of net tokens across participants was 8.1 (range: 5.9-9.8), which was 81% of possible net tokens. When the opportunity for net tokens was 5 for DRA, the average number of net tokens across participants was 4.3 (range: 3.5-4.8), which was 86% of possible net tokens. When the opportunity for net tokens was 10 for RC, the average number of net tokens across participants was 8.8 (range: 6.5-9.6), which was 88% of possible net tokens. When the opportunity for net tokens was 5 for RC, the average number of net tokens across participants was 4.4 (range: 4.1-5), which was 88% of possible net tokens. These data show that across comparisons, participants were earning similar averages of possible net tokens. These data are important because I manipulated *opportunity* for net tokens, not *actual* net tokens (as in yoking). Therefore, it was possible for a participant to have similar numbers of net tokens in DRA and RC during DRA (10) versus RC (5) and DRA (5) versus RC (10). However, an analysis of the data (see Table 6) shows that all participants had more net tokens during procedures with the opportunity for 10 net tokens as compared to 5 (represented by a positive %, where percentages at 100 mean that twice as many tokens were earned in the 10-token procedure as compared to the 5). One interesting finding was that during DRA (5) versus RC (10), 5 out of 6 participants had an average of more than twice as many tokens in the 10-token condition (% difference of more than 100%) as compared to the 5, but during DRA (10) versus RC (5), only 1 out of 5 participants had an average of more than twice as many tokens in the 10-token condition as compared to the five.

General Discussion

The results of Study 1 and 2 showed that when the opportunity for net tokens was equal, the procedures were equally effective for 17 out of 20 participants, DRA was more effective for one participant, and RC was more effective for two participants. These findings replicated those of previous studies (e.g., Donaldson et al., 2014; Iwata & Bailey, 1974). In addition, during all token manipulations (DRA [5 or 10 token]) versus RC [5 or 10 tokens]), levels of responding were similar. These results suggest that the opportunity for net tokens did not influence efficacy. Therefore, it could be that other variables, such as potency of the back-up reinforcer and task difficulty, were responsible for the similarity in effectiveness of the procedures. That is, motivation for the back-up reinforcers alone may have been sufficient to evoke responding regardless of the effort involved in task engagement or the effort involved in the task was minimal; therefore, earning (or keeping) tokens was easy. In addition, the effectiveness of the procedures may have not been influenced by a change in the number of tokens due to a history of a greater number of tokens. That is, because all participants had a history of earning 10 tokens in each condition, it could be that I observed maintenance of behavior change when the number of tokens was changed from 10 to 5. This is interesting because researchers have recommended fading slowly; however, the current results suggest that the opportunity for net tokens can be decreased by half and continue to maintain efficacy. Although this may be a limitation to the study, it has clinical significance because providing fewer opportunities for net tokens means fewer consequences; thus, the procedures are less effortful for the implementer.

Although little to no difference was found with respect to the effectiveness of DRA and RC, an interesting finding was observed for some participants (Frank, Elisa, Martin, and to a lesser degree, Erinn, Nancy, Austin, and Carly). That is, the levels of on-task behavior during baseline sessions that were conducted during the DRA versus RC comparison decreased as compared to initial baseline levels. For example, during the initial baseline, Frank's responding was highly variable and his average level of on-task behavior was approximately 40%. However, after Frank experienced both token procedures, his level of responding decreased to zero. Although it is unknown why this pattern of responding occurred, several hypotheses warrant mention. First, responding may have decreased after exposure to the token conditions due to a negative contrast (or interaction) effect (i.e., decreases in responding below levels of the initial baseline following exposure to stimuli associated with different schedules of reinforcement). Behavioral contrast is said to occur when the rate of responding moves in one direction in the presence of one stimulus and in an opposing direction in the presence of a different stimulus (McSweeney & Norman, 1979; Reynolds, 1961b). Contrast effects are often observed during multiple schedules (e.g., Herrnstein & Brady, 1958; Reynolds, 1961a; Wilton &

Gay, 1969) due to the rapid alternation of schedules of reinforcement that are associated with different stimuli. The multielement design used in the current study is similar to a multiple schedule because the conditions were rapidly alternated and involved different stimuli (e.g., colored token boards) that were associated with the different procedures (e.g., DRA, RC, and EXT). Therefore, with respect to the current study, levels of behavior during baseline sessions conducted before the introduction of the token economy may have differed from levels of behavior during baseline sessions conducted after the introduction of the token economy because prior to the introduction of the token economy, there were no alternate conditions present (i.e., consecutive extinction sessions were conducted). Once the token economy was introduced, additional stimuli associated with different procedures were now alternated with the extinction condition (i.e., baseline). Therefore, levels of behavior moved in a direction away from the levels of responding during the token procedures during the initial baseline, specifically responding decreased; thus, the contrast was negative (McSweeney & Norman, 1979).

In addition to a possible contrast effect, responding may have decreased after exposure to the token conditions due to an overjustification effect (i.e., a decrease in responding below levels of the initial baseline following a history of rewards [for a review of the overjustification effect see Deci, 1971]). Overjustification suggests that the inherently reinforcing properties of engaging in a specific behavior decrease once the behavior contacts reward contingencies. That is, once a previously automatically reinforced behavior contacts social contingencies, the behavior may become maintained by social consequences. Therefore, when the social consequences are no longer available (i.e., baseline), decreases in behavior are sometimes observed, possibly due to extinction. Overjustification may suggest that the removal of behavior interventions should be faded slowly so that decrements in responding are less likely to be observed.

With respect to token value, several participants lost interest in the back-up reinforcers, as evidenced by temporary decreases in responding (see Austin and Nancy). In addition, other participants occasionally refused to come to session (Imilia, Carly, and Becky). Thus, throughout the study, items were added to the token store. Second, but related to the first point, the methods of the current study required that participants exchange all tokens at the end of each session so that they accessed back-up reinforcers immediately, which may be important for maintaining the effectiveness of token economies with young children. However, future researchers might also give children the opportunity to save tokens, which would allow for more valuable (e.g., more costly) back-up reinforcers to be earned. Third, baseline was a choice option during the preference evaluation, and although the majority of participants selected a token procedure when given the choice between baseline, DRA, and RC, a few participants selected baseline. In fact, two participants selected baseline frequently; therfore, I had to remove baseline as an option in order to evaluate preference for the token proceures. These data suggest, in part, that the value of the tokens did not outweight the effort of the task (or reinforcing value of playing).

With respect to preference, the results of Study 1 and 2 showed that preference varied among individuals when the opportunity for net tokens was equal. Nine out of 16 participants preferred DRA, six participants preferred RC, and one participant had equal preference (however, this participant preferred RC during the second evaluation). The results of Study 2 showed that during DRA (5) versus RC (10), three participants preferred DRA, two preferred RC, and one had equal preference (however; one participant who preferred DRA switched her preference during the second evaluation). During DRA (10) versus RC (10), results showed that 4 out of 4 participants preferred DRA. Evaluating preference is important because, as recent researchers (Dozier, Vollmer, Borrero, Borrero, Rapp, Bourret, & Gutierrez, 2007; Heal & Hanley, 2007) have suggested, using a procedure that is preferred by the participant may increase (a) the effectiveness of the procedure, (b) how often the participant seeks out learning opportunities, or (c) increase the likelihood that a caregiver (i.e., parent or teacher) will implement the intervention. In addition, the use of preferred procedures may decrease problem behavior. For example, researchers have demonstrated decreases in problem behavior when task, task material, or both task and task material choices are provided (e.g., Dyer, Dunlap, & Winterling, 1990; Romaniuk, Miltenberger, Conyers, Jenner, Jurgens, & Ringenberg, 2002; Vaughn & Horner, 1997). Because choices often reflect preference, the use of preferred procedures may be especially useful when working with children who engage in escape-maintained problem behavior. That is, children who engage in problem behavior to escape non-preferred tasks, for example, may be more likely to comply when a preferred procedure is used during a demand context.

One interesting finding during the preference evaluation warrants mention. Across both studies, 6 out of 16 participants (in addition to one participant whose preference switched during the second evaluation) preferred RC when the opportunity for tokens was the same during RC and DRA, and this finding was similar to that of Iwata and Bailey (1974) and Donaldson et al. (2014) who found that some participants preferred RC as compared to DRO. In fact, Donaldson et al. found that more than half of the participants preferred RC. In addition, Hanley, Piazza, Fisher, & Maglieri (2000) found that participants preferred punishment with FCT as compared to FCT alone. Several reasons may influence preference for punishment. With respect to RC, the presence of the tokens may suggest to the participant that he or she does not have to work for the tokens. In addition, selection of RC results in immediate delivery of tokens; therefore, the procedure may be selected to gain immediate access to tokens (Donaldson et al., 2014). However, regardless of the variables which influence preference for RC, because RC was a preferred procedure for several participants when the opportunity for net tokens was equal, the results of the current study, along with others, should lead one to consider what is "best practice," given that the current general recommendation is to use reinforcement-based procedures when possible (Bailey & Burch, 2005).

Two important aspects of evaluating preference also warrant discussion. First, choices of some children might be influenced by the choices of their peers (e.g., choosing the procedure that a friend chose; Donaldson et al., 2014; Salend & Kovalich, 1981). Therefore, the current study evaluated preference when individuals were not in the presence of others because evaluating preference in the context of a classroom, for example, may not be indicative of individual preference because other contingencies may be controlling choice behavior (Donaldson et al., 2014). However, results of the current study were similar to those of Iwata and Bailey (1974) and Donaldson et al. (2014); that is, preference varied among individuals. Therefore, the influence of peers on preference is unknown and warrants further investigation.

Next, it is important to analyze preference within individuals across different manipulations. With respect to the Study 2, two patterns of preference were observed. Three participants' preference remained constant, even when the opportunity for net tokens was fewer, two participants' preference changed to the procedure with greater opportunity for net tokens, and one participant's preference was inconsistent. Austin, Erinn, and Imilia continued to select DRA during DRA (10) versus RC (5). These results suggest that some variable inherent to the procedure influences preference (as previously mentioned). For example, participants may like to earn (McGoey & DuPaul, 2000), want to avoid losing (Pietras et al., 2010), or like feedback regarding session length. With respect to preference for earning, Imilia commented during one session, "I like this one (pointing to green) because I get to earn." With respect to feedback regarding session length, it could be that a 5-min work period is long for a child, and token delivery throughout a work period provides information on the length of time until session completion. Whereas during RC, if a child is on-task, there are no indicators of work-period time remaining. In fact, during one DRA session, once Imilia had earned four tokens, she commented, "I'm almost done." In addition, following baseline sessions, Imilia commented, "That was fast." This information provides subjective evidence that this inherent characteristic of DRA may influence preference. Therefore, researchers might evaluate preference for DRA under conditions during which tokens are delivered throughout a session to conditions under which total earned tokens are delivered at the end of a session.

In addition, previous researchers have suggested that when colors are associated with procedures, some children may select a specific procedure due to color bias (e.g., Heal, Hanley, & Layer, 2009); that is, choosing a procedure based on the associated color rather than the contingencies associated with the procedure. Therefore, it could be that because green was associated with DRA, participants who continued to select DRA despite fewer opportunities for tokens had a preference for the color green. However, we had no anecdotal evidence to suggest that this was the case. In addition, Austin chose baseline twice during the preference evaluation, and Imilia chose RC once, which provides some evidence that their selections were not based on color. However, given that consistent selection of DRA may have been a result of color preference, an additional preference evaluation was conducted with stimuli of all the same color.

During this evaluation, the blank, white board used in baseline, an empty white board with 10 Velcro® spaces, and a white board with 10 pennies were presented to the participant, and a preference evaluation was conducted as in Study 1 and 2. Results of this evaluation showed Austin selected DRA on the first choice opportunity, and although Imilia selected RC on the first choice opportunity, she selected DRA on three subsequent assessments. These data suggest that the participants' preference was for DRA and not the color green.

Alternatively, Becky and Carly were two participants who selected the procedure with greater opportunity for net tokens. That is, when the two procedures had equal opportunity for net tokens, Becky had equal preference; however, when the opportunity for net tokens during RC was changed to five, Becky selected DRA exclusively. In addition, when the opportunity for net tokens during DRA was changed to five, Becky selected RC exclusively. Although preference was not equal when I re-equated the opportunity for net tokens during DRA and RC (Becky preferred RC, which may have been a result of carryover from the previous phase), Becky selected DRA when the opportunity for net tokens was changed to five during RC. Similarly, when the two procedures had equal opportunity for net tokens, Carly preferred DRA, and when the opportunity for net tokens during DRA changed to five, Carly preferred RC. Taken together, these data suggest that opportunity for net tokens may influence preference (although results for Carly are tentative). If opportunity for net tokens is a variable that influences preference, this information is useful. Specifically, if preference for individuals who prefer DRA can be changed to RC, this is beneficial because the implementer can implement a procedure that is less effortful (i.e., RC; Donaldson et al., 2014) and preferred.

Although the results of the current study provided useful and interesting information with respect to how the opportunity for net tokens influences efficacy and preference, several limitations warrant discussion. Overall, participants' levels of responding were moderate to high across all conditions and were similar during DRA and RC, and increases in on-task behavior were immediate. With respect to the levels of responding across conditions, because on-task behavior was measured using partial-interval recording and was relatively high for most participants, it is possible that the method of data collection inflated levels of on-task behavior. For example, if a participant was on task for only 1 s during every 5-s interval, the percent intervals on task would be 100%, which would not be an accurate representation of behavior. However, this was unlikely given that the behavior of tracing a shape or letter took several seconds, and the intervals were small (5 s). As an additional measure by which to compare percent on task, percent net tokens were calculated (see gray bars on graphs). For all participants, percent net tokens were similar to levels of on-task behavior, which provides additional evidence that the unit of measurement was appropriate for the dependent variable. Another variable that may have influenced continued high levels of responding is that DRA and RC were first evaluated under conditions during which the opportunity for net tokens was equal for all participants; therefore, responding may have maintained under conditions during which the opportunity for net tokens was decreased due to a history of a greater opportunity for net tokens. Future researchers might determine the influence of this history by first evaluating responding under conditions during which the opportunity for net tokens is unequal (e.g., DRA [10] vs. RC [5]).

With respect to the similarity in responding across DRA and RC, regardless of changes in the opportunity for net tokens, it is possible that the schedules of reinforcement and punishment evaluated in the current study were not disparate enough. Future researchers might compare levels of behavior under conditions during which the difference in opportunity for net tokens is greater. With respect to participants' immediate increases in levels of on-task behavior during the token procedures, it may be that the participants had a history with tokens or money prior to the start of the study. Therefore, in order to determine if the tokens are in fact neutral stimuli, future researchers should evaluate the effects of contingent token delivery when tokens are not exchangeable for back-up reinforcers.

Also, it was interesting that although Austin's levels of on-task behavior during the comparison phases were similar, responding during the preference evaluations was lower than during the comparison phases. This pattern of responding may suggest that the rapid alternation of RC with DRA during the comparison phase was influencing responding during DRA; thus, I may have observed differences in the procedures had I first compared DRA and RC within a reversal design (as was observed with Sam in Study 1). Although the use of a multielement design may have affected responding of some participants, data from Study 1 showed high levels of responding during DRA for four participants prior to exposure to RC. In addition, previous research using reversal designs did not show differences in the procedures. However, because some of these studies showed group averages, data showing the difference between DRA prior to and after RC for individuals are not available. Therefore, only tentative conclusions may be made.

With respect to preference, results of the current study showed inconsistent preference across participants (i.e., participants each had different preference under different conditions). Three participants' preference changed when the opportunity for net tokens changed (Becky, Carly, Erinn, and Nancy), and several participants preferred RC when the opportunity for net tokens was equal. With respect to inconsistent preference across participants, it is possible that each participant's history with these procedures outside of this study may have influenced choice. Future researchers might conduct an evaluation similar to that of this study with younger children (e.g., toddlers) to decrease the likelihood of history influencing preference. However, the procedures used in the current study would require modification in order to be evaluated with a younger population. With respect to changes in preference when the opportunity for net tokens changed, two participants selected the procedure with greater opportunity for net tokens. Because changing the opportunity for net tokens also changed the potential for back-up reinforcers, the results are confounded. Therefore, future researchers could change the value of tokens in each condition, such that the total potential for back-up reinforcers is the same (i.e., the total value is the same). For example, one token in DRA (10) is worth 1, and one token in RC (5) is worth 2. With respect to preference for RC, one possible confound (as previously mentioned) is the potential influence of the presence of tokens. Future researchers might evaluate the influence of the presence of tokens by including the presence of the tokens for DRA and RC (i.e., a cup of tokens next to the DRA token board and tokens attached to the RC board) or removing the presence of the tokens during both conditions (i.e., placing colored strips of paper representing each procedure or asking the participant which procedure he or she would like to do).

Furthermore, although we equated the opportunity for net tokens in our initial comparison, we did not equate the actual number of net tokens across conditions, and we may have obtained different results had we controlled for this variable. Future researchers might investigate responding during reinforcement and response cost under conditions during which net tokens are yoked (see Pietras & Hackenberg, 2005; Pietras et al., 2010; Raiff et al., 2008).

One additional confound is that because the experimenter responded to participants' bids for attention, participants could drive the amount of attention received, which could potentially influence levels of on-task behavior. That is, if a participant frequently requested attention while working, the experimenter's responses could potentially reinforce on-task behavior. Although this is a possible limitation, it should be noted that the experimenter's responses were brief and neutral, and this type of attention was not likely to be appetitive. In addition, bids for attention occurred during on- and off-task behavior; therefore, if experimenter's responses were to influence behavior, responses would influence both on- and off-task behavior equally. Furthermore, in-session observations and data suggest that participants rarely requested attention; thus, it is unlikely that participant's responding in the current study was influenced by experimenter attention.

Finally, the current study only evaluated one potential variable that may influence the effectiveness of and preference for reinforcement and response cost within token economies, and several other variables warrant investigation. Therefore, future researchers might also evaluate some of the variables discussed in this paper, such as different types of reinforcers (e.g., high- vs. low-preferred items) or task difficulty (e.g., easy vs. difficult tasks).

Despite these limitation, there are several important implications of the current study. First, given that the two procedures were equally effective, regardless of the opportunity for net tokens, then during implementation of either procedure, fewer tokens could be delivered or removed, which would make either procedure less effortful. That is, a teacher (for example) could observe students less often, which would allow the teacher to complete other tasks (e.g., grading and creating lesson plans) because the time between consequence deliveries would be greater when using a leaner schedule of reinforcement or punishment. Future researchers might do a parametric analysis to compare effectiveness of and time required to implement token procedures with varying opportunities for tokens.

Second, preference changed for four participants when the opportunity for net tokens changed, and changing preference may be advantageous for several reasons. First, a child's preference could be changed to match that of a parent or teacher, such that everyone is more satisfied with the intervention. Second, preference of some children could be changed in order to match the procedure preferred by others such that the same procedure can be implemented for all individuals, which may be easier for the implementer. Third, preference could be changed to the procedure that is easiest to implement (e.g., RC; Donaldson et al., 2014).

In summary, evaluating whether the opportunity for net tokens influences the effectiveness of or preference for DRA and RC is an important area of research for several reasons. First, this variable has not yet been investigated, and evaluating variables that influence responding under various conditions furthers our understanding of human behavior. Specifically, researchers have recently suggested that more research is needed in the area of aversive control for both conceptual and practical reasons (Critchfield & Rasmussen, 2007; Lerman & Vorndran, 2002). Second, the information gained from the current evaluation provides further "best practice" for the use of these procedures. Because a difference in the opportunity for net tokens changes preference for a given procedure (in some cases), then clinicians may change what and how procedures are implemented. Finally, the data obtained from the current research contribute to a growing body of literature on token economies, which provides guidance for future evaluations of variables that influence the effectiveness of and preference for procedures used within token economies.

References

Ayllon, T., & Azrin, N. H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior, 8*, 357-383.

Ayllon, T., & Azrin, N. H. (1968a). Reinforcer sampling: A technique for increasing the behavior of mental patients. *Journal of Applied Behavior Analysis, 1*, 13-20.

Ayllon, T., & Azrin, N. H. (1968b). *The token economy: A motivation system for therapy and rehabilitation.* New York, NY: Appleton-Century-Crofts.

Ayllon, T., Layman, D., & Kandel, H. J. (1975). A behavioral-educational alternative to drug control of hyperactive children. *Journal of Applied Behavior Analysis, 8*, 137-146.

Ayllon, T., & Roberts, M. D. (1974). Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis, 7*, 71-76.

Bailey, J. S., & Burch, M. R. (2005). *Ethics for behavior analysts*. Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Balsam, P. D., & Bondy, A. S. (1983). The negative side effects of rewards. *Journal of Applied Behavior Analysis, 18*, 283-296.

Betancourt, F. W., & Zeiler, M. D. (1971). The choices and preference of nursery school children. *Journal of Applied Behavior Analysis, 4*, 299-304.

Boniecki, K. A.. & Moore, S. (2003). Breaking the silence: using a token economy to reinforce classroom participation. *Teaching of Psychology, 30*, 224-227.

Brent, D. E., & Routh, D. K. (1978). Response cost and impulsive word recognition errors in reading-disabled children. *Journal of Abnormal Child Psychology, 6*, 211-219.

Brigham, T. A., Finfrock, S. R., Breunig, M. K., & Bushell, Jr., D. (1972). The use of programmed materials in the analysis of academic contingencies. *Journal of Applied Behavior Analysis,5*, 177-182.

Broughton, S. F., & Lahey, B. B. (1978). Direct and collateral effects of positive reinforcement , response cost, and mixed contingencies for academic performance. *Journal of School Psychology, 16*, 126-136.

Burchard, J. D., & Barrera, F. (1972). An analysis of timeout and response cost in a programmed environment. *Journal of Applied Behavior Analysis, 5*, 271-282.

Bushell Jr., D., Wrobel, P. A., & Michaelis, M. L. (1968). Applying “group” contingencies to the classroom study behavior of preschool children. *Journal of Applied Behavior Analysis, 1*, 55-61.

Capriotti, M. R., Brandt, B. C., Ricketts, E. J., Espil, F. M., & Woods, D. W. (2012). Comparing the effects of differential reinforcement of other behaior and resonse-cost contingencies on tics in youth with tourette syndrome. *Journal of Applied Behavior Analysis, 45*, 251-263.

Carlson, C. L., Mann, M., & Alexander, D. K. (2000). Effects of reward and response cost on the perfrmance and motivation of children with ADHD. *Cognitive Therapy and Research, 24*, 87-98.

Carr, J. E., Fraizer, T. J., & Roland, J. P. (2005). Token economy. In A. M. Gross & R. S. Drabman (Eds.), Encyclopedia of behavior modification and cognitive behavior therapy - Volume 2: Child clinical applications (pp. 1075-1079). Thousand Oaks, CA: Sage.

Christophersen, E. R., Arnold, C. M., Hill, D. W., & Quilitch, R. (1972). The home point system: Token reinforcement procedures for applications by parents of chilren with behavior problems. *Journal of Applied Behavior Analysis, 5*, 485-497.

Critchfield, T. S., & Rasmussen, E. R. (2007). It's aversive to have an incomplete science of behavior. *Mexican Journal of Behavior Analysis, 33*, 1-5.

Conners, J., Iwata, B. A., Kahng, S., Hanley, G. P., Worsdell, A. S., & Thompson, R. H. (2000). Differential responding in the presence and absence of discriminative stimuli during multielement functional analyses. *Journal of Applied Behavior Analysis, 33*, 299-308.

Conyers, C., Miltenberger, R., Maki, A., Barenz, R., Jurgens, M., Sailer, A., …Kopp, B. (2004). A comparison of response cost and differential reinforcement of other behavior to reduce disruptive behavior in a preschool classroom. *Journal of Applied Behavior Analysis, 37*, 411-415.

Conyers, C., Miltenberger, R., Romaniuk, C., Kopp, B., & Himle, M. (2003). Evaluation of DRO schedules to reduce disruptive behavior in a preschool classroom. *Child and Family Behavior Therapy, 25*, 1-6.

Cowdery, G. E., Iwata, B. A., & Pace, G. M. (1990). Effects and side effects of DRO as treatment for self-injurious behavior. *Journal of Applied Behavior Analysis, 23*, 497-506.

De Martini-Scully, Bray, M. A., & Kehle, T. J. (2000). A packaged intervention to reduce disruptive behaviors in general education students. *Psychology in the Schools*, 37*,* 149-156.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105-115.

Dickerson, F. B., Tenhula, W. N., & Green-Paden, L. D. (2005). The token economy for schizophrenia: Review of the literature and recommendations for future research. *Schizophrenia Research, 75*, 405-416.

Didden, R., de Moor, J., & Bruyns, W. (1997). Effectiveness of DRO tokens in decreasing disruptive behavior in the classroom with five multiply handicapped children. *Behavioral interventions, 12*, 65-75.

Donaldson, J. M., DeLeon, I. G., Fisher, A. B., & Kahng, S. (2014). Effects of and preference for conditions of token earn versus token loss. *Journal of Applied Behavior Analysis, 47*, 1-12.

Dozier, C. L., Vollmer, T. R., Borrero, J. C., Borrero, C. S., Rapp, J. T, Bourret, J., & Gutierrez, A. (2007). Assessment of preference for behavioral treatment versus baseline conditions. *Behavioral Interventions, 22*, 245-261.

Drabman, R. S., & Lahey, B. B. (1974). Feedback in classroomo behavior modifications: Effects on th etarget and her classmates. *Journal of Applied Behavior Analysis, 7*, 591-598.

Drabman, R., Spitalnik, R., & Sptalnik, K. (1974). Sociometric and disruptive behavior as a function of four types of token reinforcement programs. *Journal of Applied Behavior Analysis, 7*, 93-101.

Dyer, K., Dunlap, G., & Winterling, V. (1990). Effects of choice making on the serious problem behaviors of students with severe handicaps. *Journal of Applied Behavior Analysis, 23*, 515-524.

Gresham, F. M. (1979). Comparison of response cost and timeout in a special education setting. *The Journal of Special Education, 13*, 199-208.

Hackenburg, T. D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior, 91*, 257-286.

Hains, A. H., & Baer, D. A. (1989). Interaction effects in multielement designs: Inevitable, desirable, and ignorable. *Journal of Applied Behavior Analysis, 22*, 57-69.

Hallahan, D. P., Tarver, S. G., Kauffman, J. M., & Graybeal, N. L. (1978). A comparison of the effects of reinforcement and response cost on the selective attention of learning disabled children. *Journal of Learning Disabilities, 11*, 39-47.

Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis, 38*, 51-65.

Haring, T. G., & Kennedy, C. H. (1990). Contextual control of problem behavior in students with severe disabilities. *Journal of Applied Behavior Analysis*, *23,* 235-243.

Harris, S. L., & Ersner-Hershfield, R. (1978). Behavioral suppression of seriously disruptive behavior in psychotic and retarded patients: A review of punishment and its alternatives. *Psychological Bulletin, 85*, 1352-1375.

Heal, N. A., & Hanley, G. P. (2007). Evaluating preschool children's preference for motivational systems during instruction. *Journal of Applied Behavior Analysis*, *40,* 249-261.

Heal, N. A., Hanley, G. P., & Layer, S. A. (2009). An evaluation of the related efficacy of and children's preferences for teaching strategies that differ in amount of teacher directedness. *Journal of Applied Behavior Analysis*, *42,* 123-143.

Herrnstein, R. J., (1961). Relative and absolute strength of a response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*, 267-272.

Herrnstein, R. J., & Brady, J. V. (1958). Interaction among components of a multiple schedule. *Journal of the Experimental Analysis of Behavior, 1*, 293-300.

Holland, E. L., & McLaughlin, T. F. (1982). Using public posting and group consequences to manage student behavior during supervision. *The Journal of Educational Research, 76*, 29-34.

Humphrey, L. L., & Karoly, P. (1978). Self-management in the classroom: self-imposed response cost vs. self-reward. *Behavior Therapy, 9*, 592-601.

Iwata, B. A., & Bailey, J. S. (1974). Reward versus cost token systems: An analysis of the effects on students and teacher. *Journal of Applied Behavior Analysis, 7*, 567-576.

Kazdin, A. E. (1977). The token economy: A review and evaluation. New York, NY: Plenum Press.

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis, 5*, 343-372.

Kelleher, R. T., & Gollub, L. R. (1962). A review of positive conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 5*, 543-597.

Lahey, B. B., & Drabman, R. S. (1974). Facilitation of the acquisition and retention of sight-word vocabulary through token reinforcement. *Journal of Applied Behavior Analysis, 7,* 307-312.

LeBlanc, L. A., Hagopian, L. P. & Maglieri, K. A. (2000). Use of a token economy to eliminate excessive inappropriate social behavior in an adult with developmental disabilities. *Behavioral Interventions, 15*, 135-143.

Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment: Implications for treating behavior disorders. *Journal of Applied Behavior Analysis, 35*, 431-464.

Little, L. M., & Kelley, M. L. (1989). The efficacy of response cost procedures for reducing children's noncompliance to parental instructions. *Behavior Therapy, 20*, 525-534.

Marholin II, D., & Gray, D. (1976). Effects of group response-cost procedures on cash shortages in a small business. *Journal of Applied Behavior Analysis, 9*, 25-30.

Matson, J. L., & Boisjoli, J. A. (2009). The token economy for children with intellectual disability and/or autsim: A review. *Research in Developmental Disabilities, 30*, 240-248.

Mazur, J. E. (2006). Learning and Behavior. (6th ed.; p. 185). Upper Saddle River, NJ: Pearson Education, Inc.

McGoey, K. E., & DuPaul, G. J. (2000). Token reinforcement and response cost procedures: Reducing the disruptive behavior of preschool children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly, 15*, 330-343.

McLaughlin, T., & Malaby, J. (1972). Reducing and measuring inappropriate verbalizations in a token classroom. *Journal of Applied Behavior Analysis, 5,* 329-333.

McSweeney, F. K., & Norman, W. D. (1979). Defining behavioral contrast for multiple schedules. *Journal of the Experimental Analysis of Behavior, 32*, 457-461).

Miltenberger, R. G. (2008). *Behavior modification*. (4th ed.; pp. 507-513). Belmont, CA: Thomson Wadsworth.

Mottram, L. M., Bray, M. A., Kehle, T. J., Broudy, M., & Jenson, W. R. (2002). A classroom-based intervention to reduce disruptive behavior. *Journal of Applied Psychology, 19*, 65-74.

Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review, 30*, 294-304.

Neef, N. A., Shade, D., & Miller, M. S. (1994). Assessing influential dimensions of reinforcers on choice in students with serious emotional disturbance. *Journal of Applied Behavior Analysis, 27,* 575-583.

Nelson, G. L., & Cone, J. D. (1979). Multiple-baseline analysis of a token economy for psychiatric inpatients. *Journal of Applied Behavior Analysis, 12*, 255-271.

O'Leary, K. D., & Becker, W. C. (1967). Behavior modification of an adjustment class: A token reinforcement program. *Exceptional Children, 33*, 637-642.

O'Leary, K. D., Becker, W. C., Evans, M. B., & Saudargas, R. A. (1969). A token reinforcement program in a public school: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 2*, 3-13.

O'Leary, K. D., & Drabman, R. (1971). Token reinforcement programs in the classroom: A review. *Psychological Bulletin, 75*, 379-398.

Pace, D. M., & Forman, S. G. (1982). Variables related to the effectiveness of response cost. *Psychology in the Schools, 19*, 365-370.

Palen McGlynn, A., & Locke, B. J. (1997). A 25-year follow-up of a punishment program for severe self-injury. *Behavioral Interventions, 12*, 203-207.

Panek, D. M. (1970). Word association learning by chronic schizophrenics on a token economy ward under conditions of reward and punishment. *Journal of Clinical Psychology, 26*, 163-167.

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement place: Modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis, 4*, 45-59.

Pietras, C. J., Brandt, A. E., & Searcy, G. D. (2010). Human responding on random-interval schedules of response-cost punishment: The role of reduced reinforcement density. *Journal of the Experimental Analysis of Behavior, 93*, 5-26.

Pietras, C. J., & Hackenberg, T. D. (2005). Response-cost punishment via token loss with pigeons. *Behavioural Processes, 69*, 343-356.

Raiff, B. R., Bullock, C. E., & Hackenberg, T. D. (2008). Response-cost punishment with pigeons: Further evidence of response suppression via token loss. *Learning & Behavior, 36,* 29-41.

Rapport, M. D., Murphy, H. A., & Bailey, J. S. (1982). Ritalin vs. response cost in the control of hyperactive children: A within-subject comparison. *Journal of Applied Behavior Analysis, 15*, 205-216.

Reisinger, J. J. (1972). The treatment of "anxiety-depression" via positive reinforcement and response cost. *Journal of Applied Behavior Analysis, 5*, 125-130.

Reynolds, G. S. (1961a). An analysis of interactions in a multiple schedule. *Journal of the Experimental Analysis of Behavior, 4*, 107-117.

Reynolds, G. S. (1961b). Behavioral contrast. *Journal of the Experimental Analysis of Behavior, 4*, 57-71.

Reynolds, L. K., & Kelley, M. L. (1997). The efficacy of a response cost-based treatment package for managing aggressive behavior in preschoolers. *Behavior Modification, 21*, 216-230.

Romaniuk, C., Miltenberger, R., Conyers, C., Jenner, N., Jurgens, M., & Ringenberg, C. (2002). The influence of activity choice on problem behaviors maintained by escape versus attention. *Journal of Applied Behavior Analysis, 35*, 349-362.

Rowbury, T. G., Baer, A. M., & Baer, D. M. (1976). Interactions between teacher guidance and contingent access to play in developing preacademic skills of deviant preschool children. *Journal of Applied Behavior Analysis, 9,* 85-104.

Salend, S. J., & Kovalich, B. (1981). A group response cost system mediated by free tokens: An alternative to token reinforcement. *American Journal of Mental Deficiency, 86*, 184-187.

Sindelar, P. T., Honsaker, M. S., & Jenkins, J. R. (1982). Response cost and reinforcement contingencies of managing the behavior of distractible children in tutorial settings. *Learning Disability Quarterly, 5*, 3-13.

Staples, L., McGoey, K. E., Cowan, R. J., Crist, N., & Tankersley, M. (2006). *Journal of Evidence-based practices for Schools, 7,* 142-164.

Sullivan, M. A., & O'Leary, S. G. (1990). Maintenance following reward and cost token programs. *Behavior Therapy, 21*, 139-149.

Tanol, G., Johnson, L., McComas, J., & Cote, E. (2010). Responding to rule violations or rule following: A comparison of two versions of the Good Behavior Game with Kindergarten students. *Journal of School Psychology, 48*, 337-355.

Theodore, L. A., Bray, M. A., & Kehle, T. J. (2004). A comparitive study of group-oriented contingencies and randomized reinforcers to reduce disruptive classroom behavior. *School Psychology Quarterly, 19*, 253-271.

Twardosz, S., & Sajwaj, T. (1972). Multiple effects of a procedure to increase sitting in a hyperactive, retarded boy. *Journal of Applied Behavior Analysis, 5,* 73-78.

Vaughn, B. J., & Horner, R. H. (1997). Identifying instructional tasks that occasion problem behaviors and assessing the effects of student versus teacher choice among these tasks. *Journal of Applied Behavior Analysis, 30*, 299-312.

Vollmer, T. (2002). Punishment happens: Some comments on Lerman and Vorndran's Review. *Journal of Applied Behavior Analysis, 35*, 469-473.

Walker, H. M., & Buckley, N. K. (1968). The use of positive reinforcement in conditioning attending behavior. *Journal of Applied Behavior Analysis, 1*, 245-250.

Wilton, R. N., & Gay, R. A. (1969). Behavioral contrast in one component of a multiple schedule as a function of the reinforcement conditions operating in the following component. *Journal of the Experimental Analysis of Behavior, 12*, 239-246.

Tables

Author	Effect	Behavior	Design	Contingency	Procedure	Back-up Sr	Token	Population
Tanol et al. (2010)	DR > RC	NON-ACQ (follow rules)	Reversal	Interdep group	DRA	Edible	TANG (stars)	Kindergarten
Capriotti et al. (2012)	DR = RC	NON-ACQ (tics)	Multi	Individual	DRO	Money	NON-TANG (points)	Tourettes
Broughton & LaHey (1978)	DR = RC	ACQ (correct answers)	BS w/ Rev	Individual	DRA	Activities	NON-TANG (points)	Low performance
Donaldson et al. (2014)	DR = RC	NON-ACQ (disruptive behavior)	Rev & Multi	Individual	DRO	Edible	NON-TANG (checkmarks)	Bx problems
Humphrey & Karoly (1978)	DR = RC	ACQ (correct answers)	BS w/ Rev	Indep group	DRA	Various	TANG (plastic chips)	Elementary
Iwata & Bailey (1974)	DR = RC	NON-ACQ (on task & rule violations)	BS w/ Rev	Indep group	DRO	Edible & free time	TANG (tokens)	Elementary
McGoey & DuPaul (2000)	DR = RC	NON-ACQ (follow rules)	Reversal	Indep group	DRA	Stickers & hand stamps	TANG (buttons)	ADHD
Panek (1970)	DR = RC	ACQ (common word association)	BS	Individual	DRA	Various	TANG (tokens)	Schizophrenia
Sullivan & O'Leary (1990)	DR = RC	NON-ACQ (on task)	Rev & Multi	Indep group	DRA	Extra recess & stickers	NON-TANG (checkmarks)	Bx problems
Brent & Routh (1978)	RC > DR	ACQ (sight-word reading)	BS	Individual	DRA	Money	TANG (tokens)	Elementary
Carlson et al. (2000)	RC > DR	ACQ (correct & complete math problems)	BS	Indep group	DRA	Money	TANG (tokens)	ADHD & Non-ADHD
Conyers et al. (2004)	RC > DR	NON-ACQ (disruptive behavior)	Rev & Multi	Indep group	DRO	Edibles	TANG (stars)	Preschool
Sindelar et al. (1982)	RC > DR	NON-ACQ (look at task)	Reversal	Individual	DRO	Extra recess & stickers	TANG (abacus)	Learning disability

Table 1. This table depicts a summary of information for studies comparing differential reinforcement to response cost. The first column lists the author(s) and year. The second column lists which procedure was more effective. The third column lists the target behavior and whether it was an acquisition task (ACQ) or non-acquisition (NON-ACQ) behavior. The fourth column lists the design used including reversal (rev), multielement (multi), or between subjects (BS). The fifth column lists the type of contingency including interdependent group (interdep), individual, or independent group (indep). The sixth column lists the type of differential reinforcement procedure used including differential reinforcement of alternative behavior (DRA) and differential reinforcement of other behavior (DRO). The seventh column lists the type of back-up reinforcer(s) used. The eighth column lists the type of token used and whether it was a tangible (TANG) or non-tangible (NON-TANG) token. The final column lists participant population. If the participants did not have a diagnosis, the grade level is listed; if the participants did have a diagnosis, the diagnosis is listed.

Participant	Gender	Age
Adam	M	4 yrs 4 mo
Adrianna	F	4 yrs 8 mo
Anna	F	4 yrs 5 mo
Brianna	F	4 yrs 11 mo
Collin	M	4 yrs 5 mo
Caroline	F	4 yrs 11 mo
Elisa	F	3 yrs 7 mo
Emily	F	4 yrs 6 mo
Frank	M	5 yrs 4 mo
Mark	M	4 yrs 10 mo
Martin	M	5 yrs 1 mo
Paul	M	4 yrs 8 mo
Sam	M	5 yrs 6 mo
Zoey	F	4 yrs 8 mo

Table 2. This table depicts each participant's name (first column), gender (second column), and age (third column).

Participant	Effectiveness	Preference	Avg. # Different	% Difference
Frank	=	DRA	0.6 (RC)	6.7%
Emily	DRA	DRA	0.6 (DRA)	6.8%
Martin	=	DRA	0.3 (DRA)	3.1%
Adrianna	=	DRA	0.3 (RC)	3.2%
Paul	=	DRA	0.4 (DRA)	4.3%
Anna	=	RC	0.2 (DRA)	2.4%
Elisa	=	RC	1.0 (DRA)	14.0%
Collin	=	RC	0.0	0%
Adam	RC	RC	1.1 (RC)	20.4%
Caroline	=	RC	2.0 (RC)	33.8%
Brianna	=	N/A	0.2 (RC)	2.4%
Mark	=	N/A	0.4 (RC)	4.3%
Zoey	=	N/A	0.7 (DRA)	8.0%
Sam	RC	N/A	2.8 (RC)	45.1%

Table 3. This table depicts which procedure was more effective (an equal sign denotes that the difference between the procedures was less than 10%) or preferred and the average difference in the number of net tokens between differential reinforcement of alternative behavior (DRA) and response cost (RC) procedures for each child as well as the average percent difference. The procedure in the parentheses is the procedure for which more tokens were earned.

Participant	Gender	Age
Austin	M	3 yrs 8 mo
Becky	F	3 yrs 9 mo
Carly	F	4 yrs 3 mo
Erinn	F	4 yrs 11 mo
Imilia	F	4 yrs 9 mo
Nancy	F	4 yrs 5 mo

Table 4. This table depicts each participant's name (first column), gender (second column), and age (third column).

Participant	Effectiveness	Preference	Avg. # Different	% Difference
Austin	=	DRA	0.0 (DRA)	0.0%
Becky	=	=	0.6 (RC)	9.0%
Carly	=	DRA	0.6 (RC)	10.2%
Erinn	=	DRA	1.7 (RC)	22.4%
Imilia	=	DRA	0.3 (RC)	3.4%
Nancy	=	RC	2.1 (RC)	26.4%

Table 5. This table depicts which procedure was more effective (an equal sign denotes that the difference between the procedures was less than 10%) or preferred (an equal sign denotes that the participant alternated selections of the procedures) and the average difference in the number of and percent net tokens between differential reinforcement of alternative behavior (DRA) and response cost (RC) procedures during each evaluation when the opportunity for net tokens was equal (DRA [10] vs. RC [10]).

	2nd DRA (10) vs. RC (10)				DRA (5) vs. RC (10)				DRA (10) vs. RC (5)			
Participant	Effect	Pref	Avg. # Diff	% Diff	Effect	Pref	Avg. # Diff	% Diff	Effect	Pref	Avg. # Diff	% Diff
Austin	N/A	N/A	N/A	N/A	=	DRA	5.1 (RC)	113%	=	DRA	2.6 (DRA)	62%
Becky	N/A	N/A	N/A	N/A	=	RC	5.0 (RC)	125%	=	DRA	4.8 (DRA)	96%
Carly	N/A	N/A	N/A	N/A	=	RC	5.0 (RC)	143%	=	N/A	2.9 (DRA)	73%
Erinn	=	DRA	0.5 (DRA)	5.4%	=	=	4.9 (RC)	104%	N/A	N/A	N/A	N/A
Imilia	N/A	N/A	N/A	N/A	=	DRA	5.2 (RC)	118%	=	DRA	4.4 (DRA)	92%
Nancy	=	DRA	0.7 (RC)	8.4%	=	DRA	4.2 (RC)	92%	=	DRA	5.1 (DRA)	124%

Table 6. This table depicts which procedure was more effective (Effect) or preferred (Pref) and the average difference in the number (Avg. # Diff) and percent (% Diff) of net tokens between differential reinforcement of alternative behavior (DRA) and response cost (RC) procedures during the second evaluation of DRA (10) vs. RC (10), and the evaluations of DRA (5) vs. RC (10) and DRA (10) vs. RC (5).

Figures

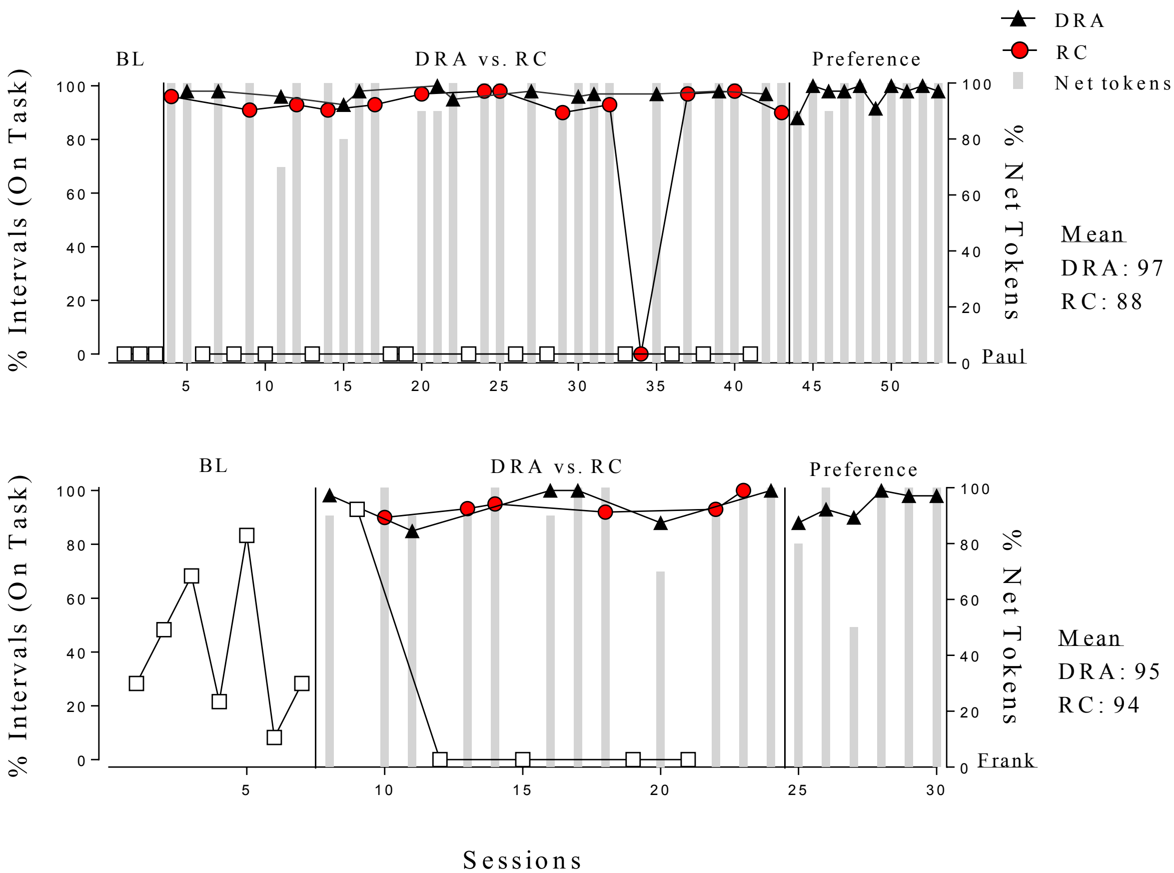


Figure 1. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Paul (top panel) and Frank (bottom panel) during baseline, differential reinforcement, and response cost sessions in the comparative analysis (intervention) and preference phase (scaled to the left y-axis). The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed to the right of the right y-axis are the means for each procedure during the comparison phase.

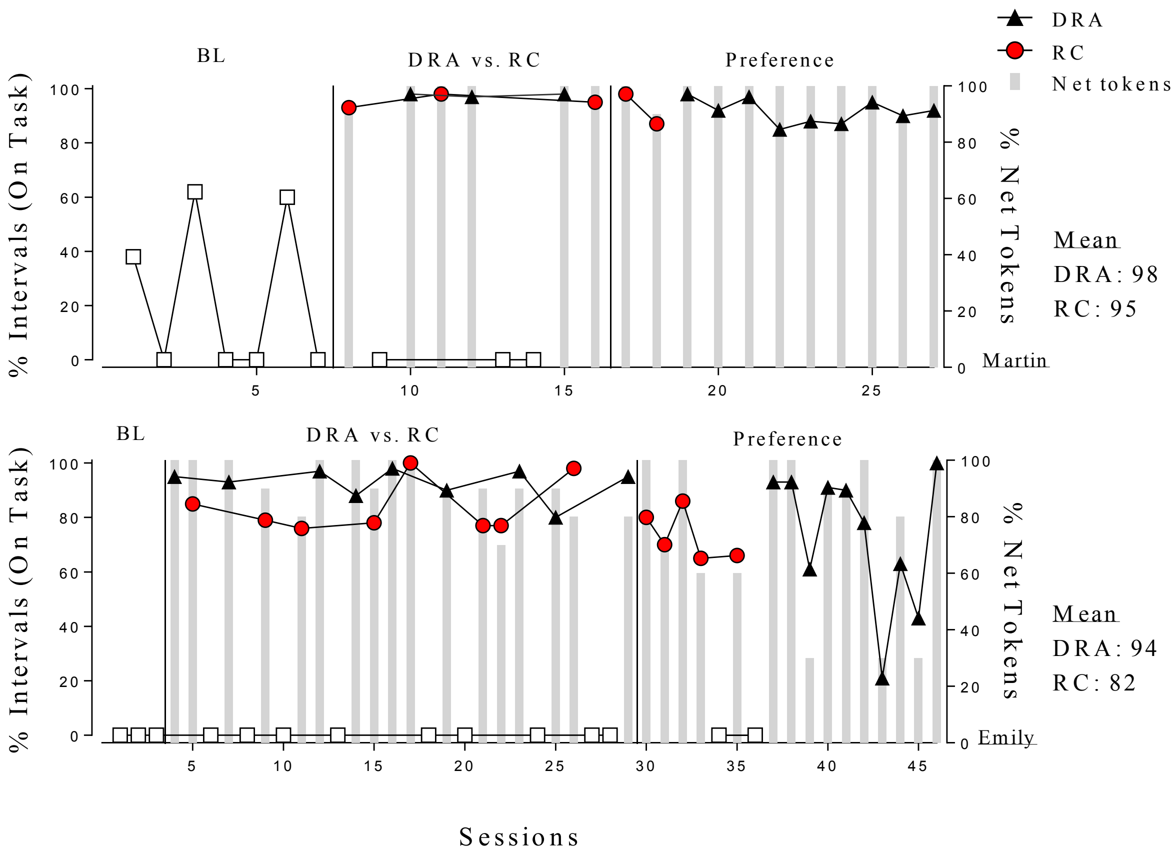


Figure 2. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Martin (top panel) and Emily (bottom panel) during baseline, differential reinforcement, and response cost sessions in the comparative analysis (intervention) and preference phase. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed to the right of the right y-axis are the means for each procedure during the comparison phase.

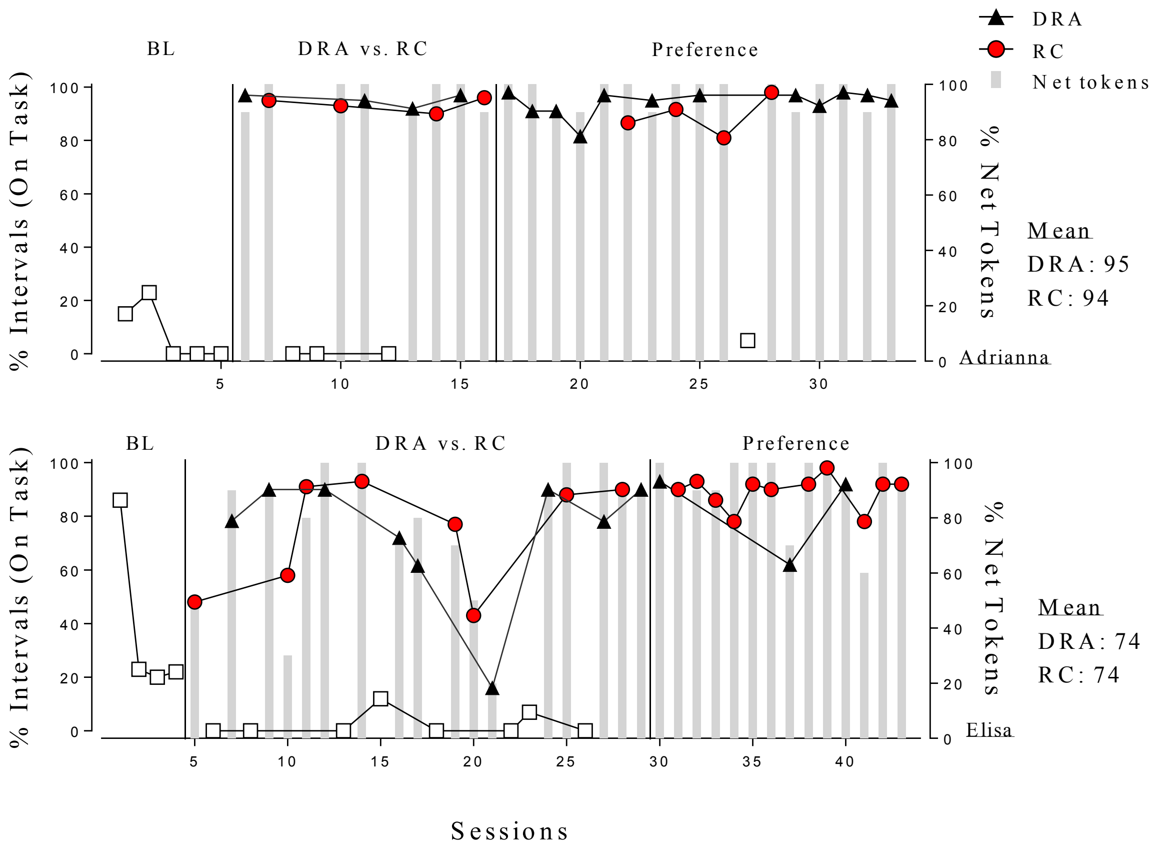


Figure 3. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Adrianna (top panel) and Elisa (bottom panel) during baseline, differential reinforcement, and response cost sessions in the comparative analysis (intervention) and preference phase. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed to the right of the right y-axis are the means for each procedure during the comparison phase.

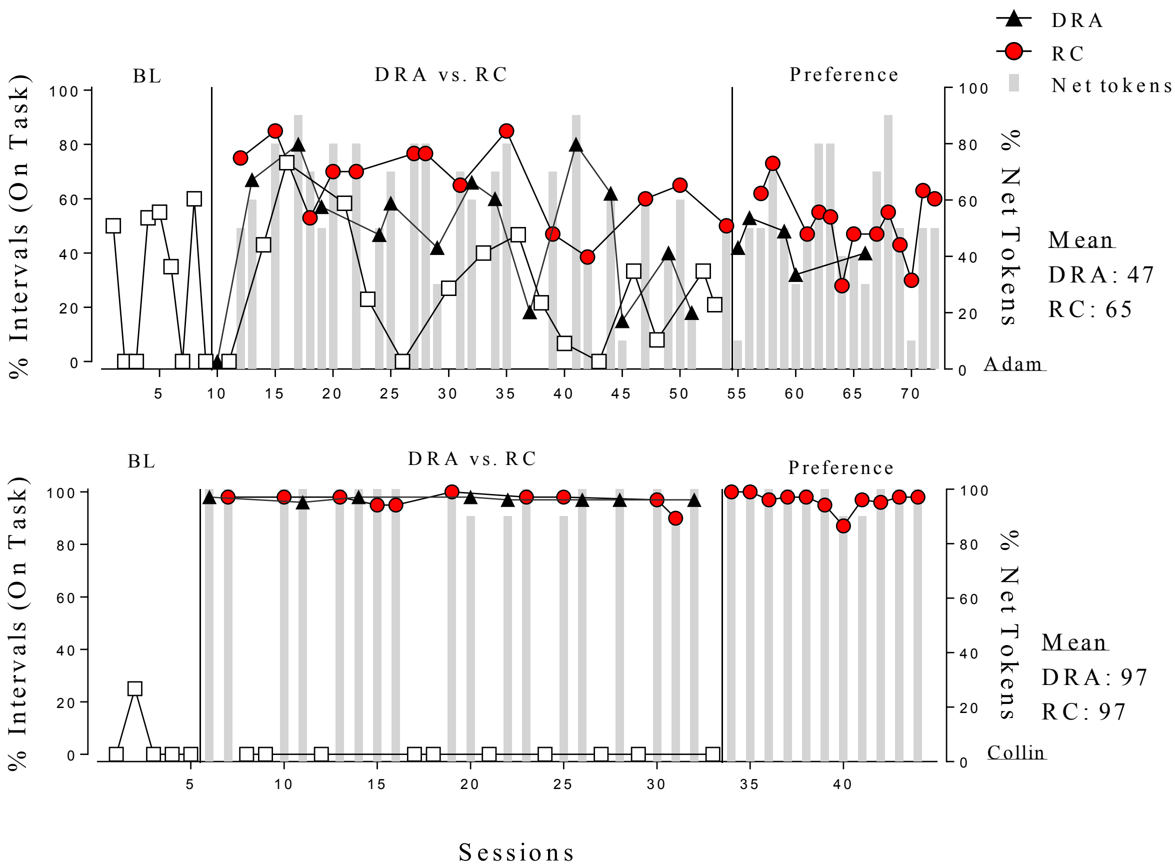


Figure 4. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Adam (top panel) and Collin (bottom panel) during baseline, differential reinforcement, and response cost sessions in the comparative analysis (intervention) and preference phase. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed to the right of the right y-axis are the means for each procedure during the comparison phase.

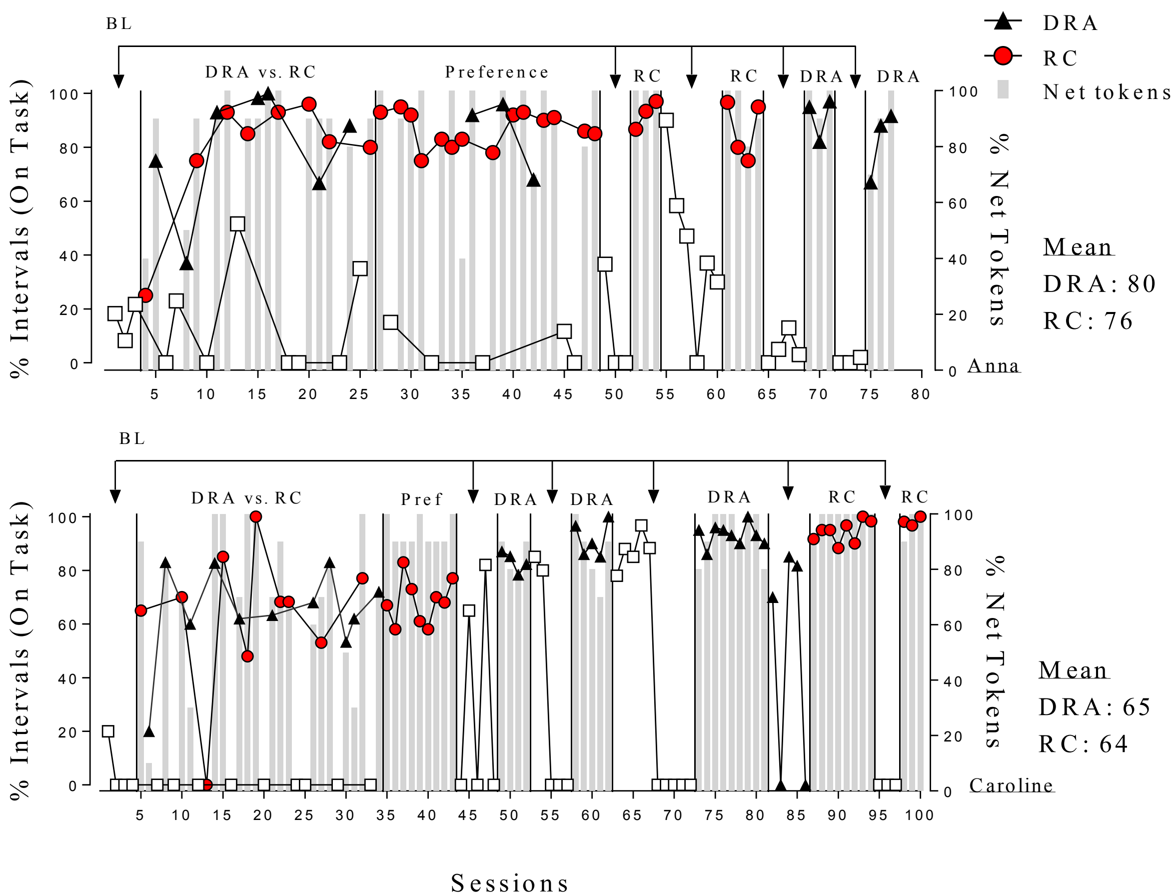


Figure 5. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Anna (top panel) and Caroline (bottom panel) during baseline, differential reinforcement, and response cost sessions in the comparative analysis (intervention) and preference phase. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed to the right of the right y-axis are the means for each procedure during the comparison phase.

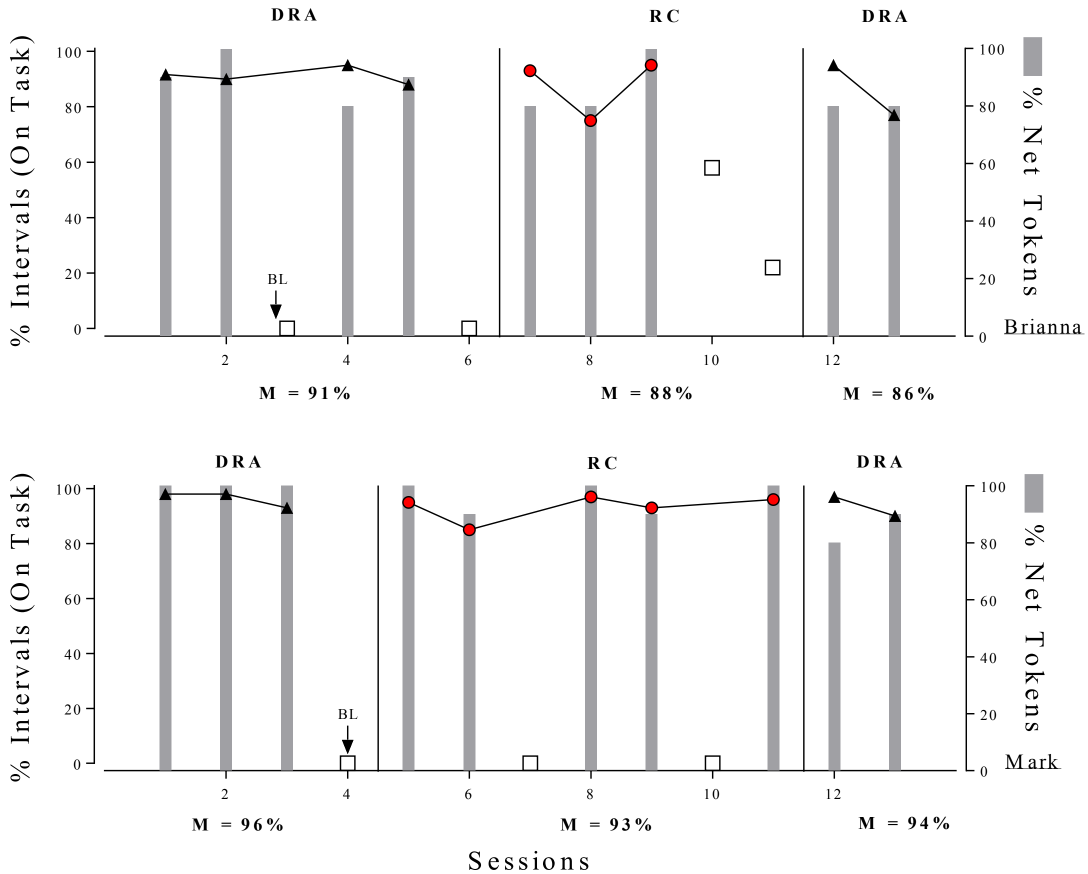
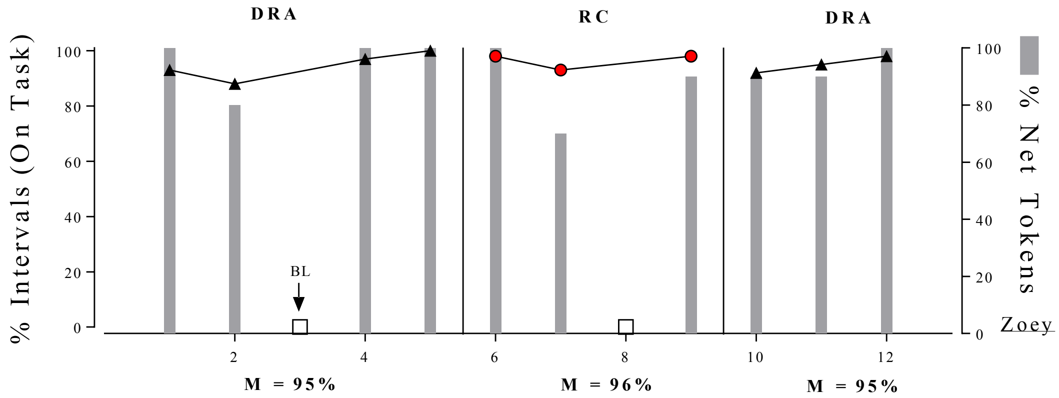


Figure 6. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Brianna (top panel) and Mark (bottom panel) during baseline (open squares) and differential reinforcement (closed triangles). The means listed below the x-axis are the means for each procedure.



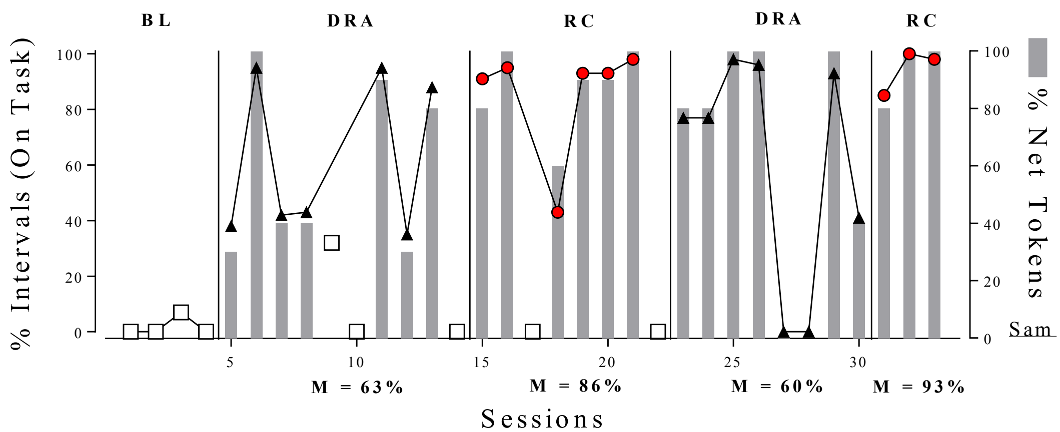


Figure 7. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Brianna (top panel) and Mark (bottom panel) during baseline (open squares), differential reinforcement (closed triangles), and response cost (red circles). The means listed below the x-axis are the means for each procedure.

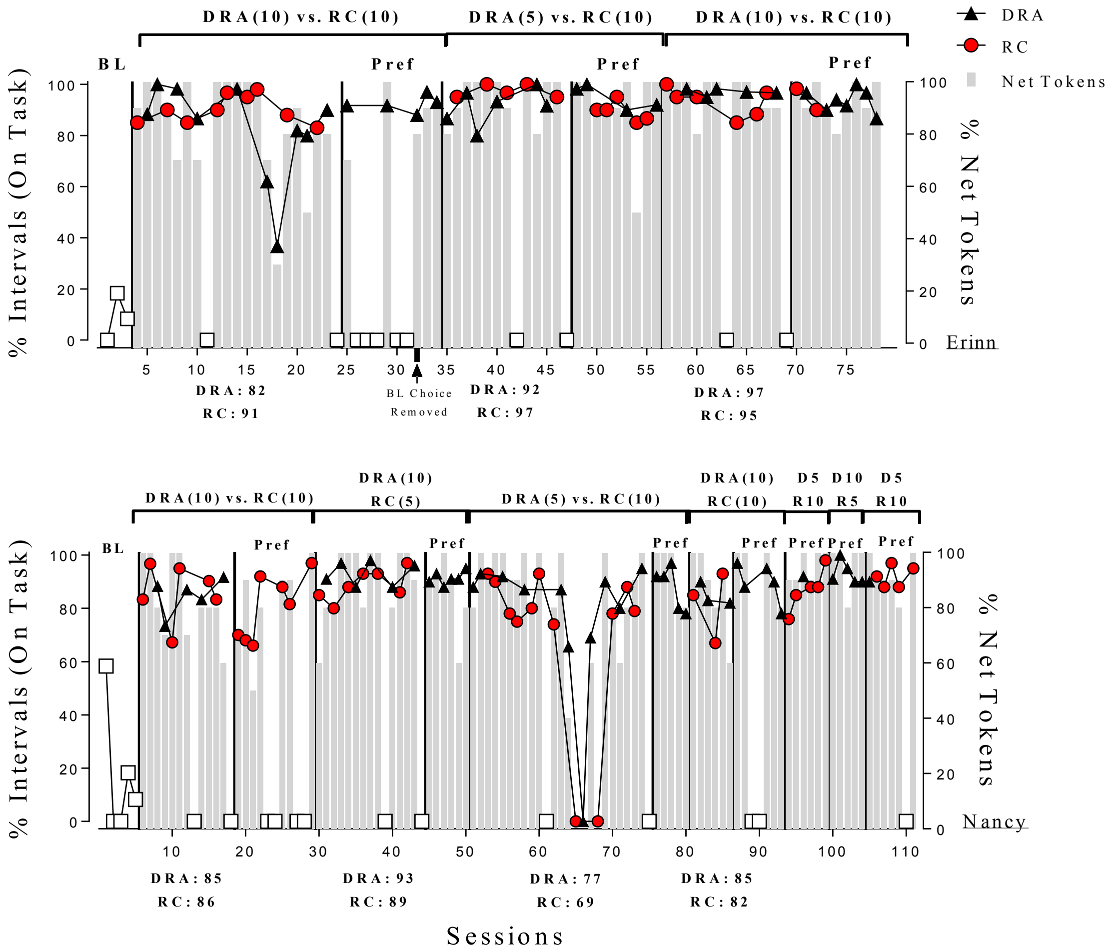


Figure 9. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Erinn (top panel) and Nancy (bottom panel) during baseline, differential reinforcement, and response cost sessions in all comparative analyses (intervention) and preference phases. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed under the x-axis are the means for each procedure during each comparison phase.

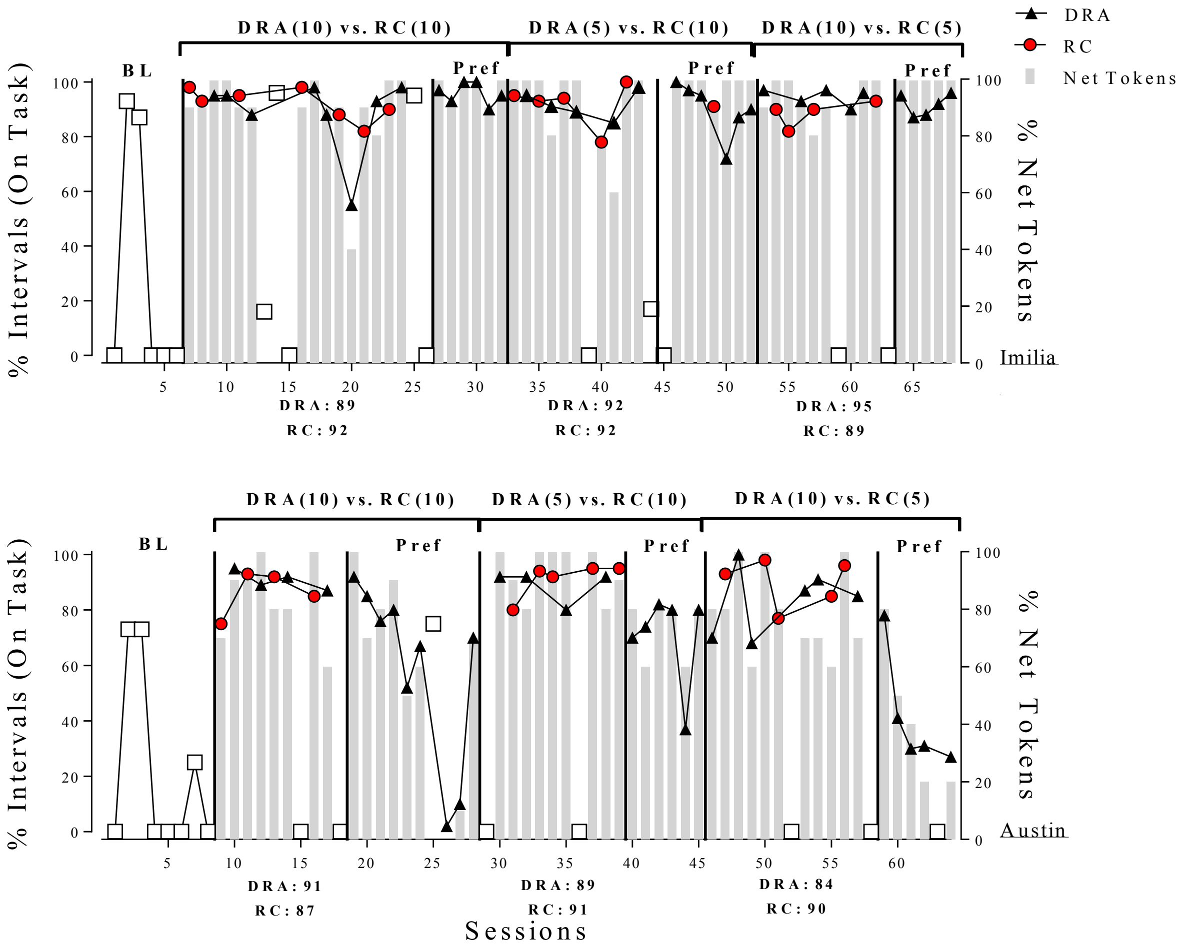


Figure 10. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Imilia (top panel) and Austin (bottom panel) during baseline, differential reinforcement, and response cost sessions in all comparative analyses (intervention) and preference phases. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed under the x-axis are the means for each procedure during each comparison phase.

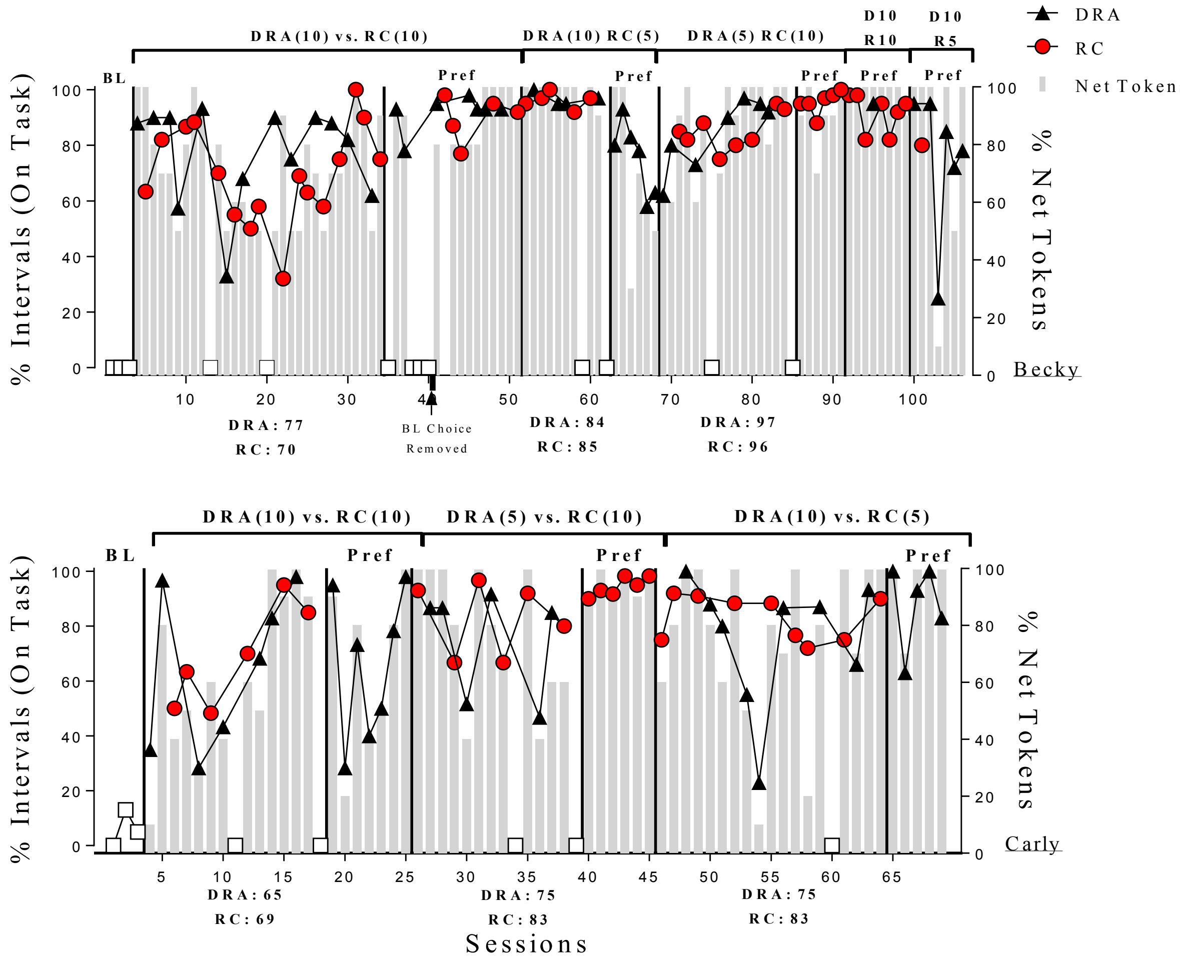


Figure 11. These graphs show the percentage of on-task behavior (left y-axis) and net tokens (right y-axis) for Becky (top panel) and Carly (bottom panel) during baseline, differential reinforcement, and response cost sessions in all comparative analyses (intervention) and preference phases. The type of data point graphed for each session in the preference phase represents the procedure selected by the participant. The means listed under the x-axis are the means for each procedure during each comparison phase.

Appendix A. Tracing Worksheet (letters)

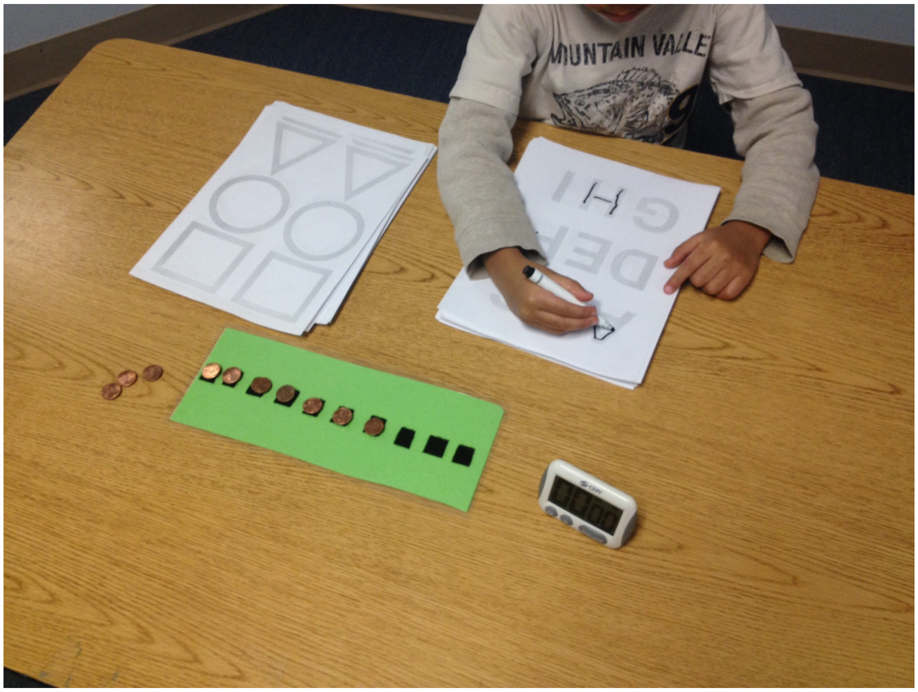
A B C

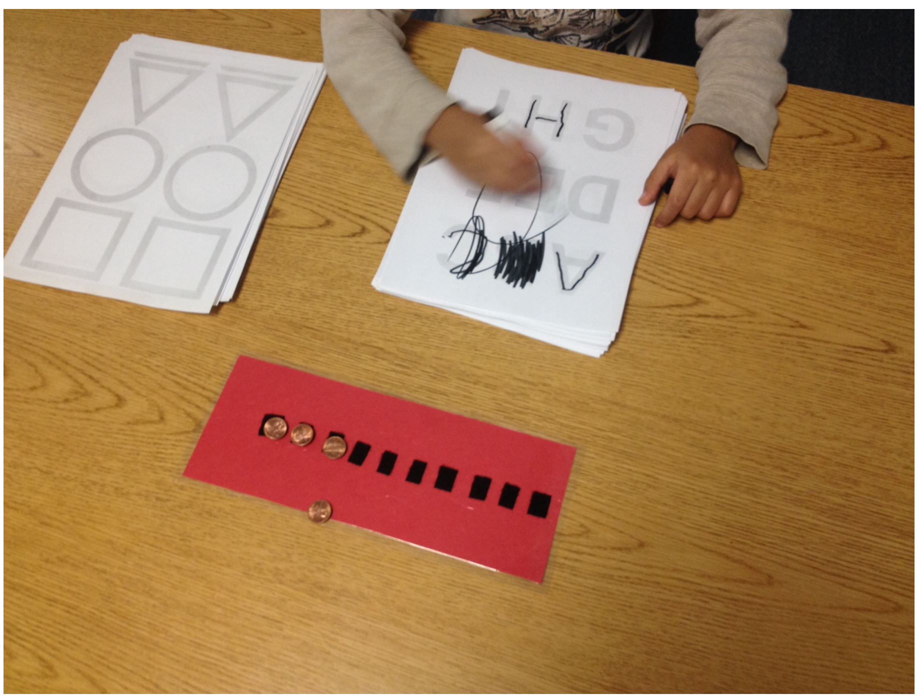
D E F

G H I

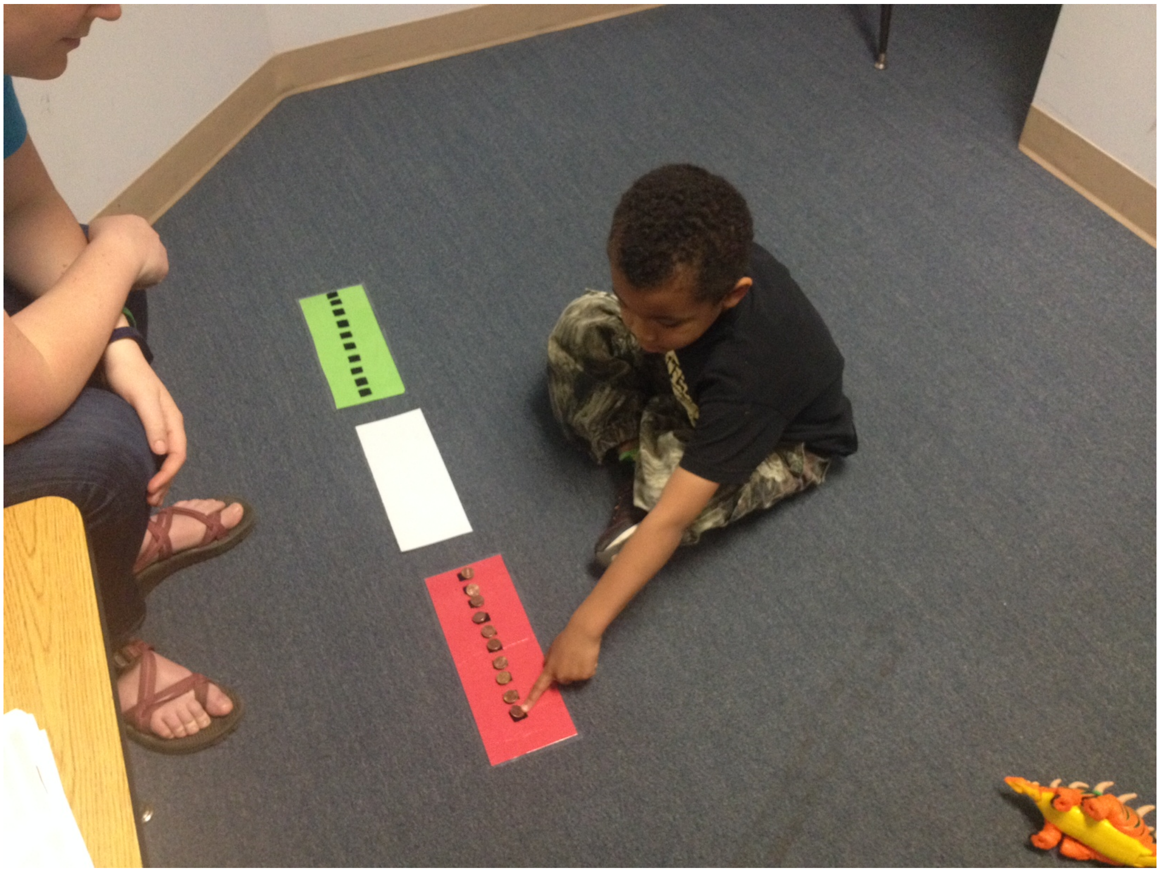
Appendix B. Research Room

Appendix C. Token Boards





Appendix D. Preference Arrangement



Appendix E. Token Store

Appendix F. Treatment Integrity Data Sheet

Date: 6/28
Session: 101

Observer: JC
Condition: DRA 10

Participant: Johnny
Therapist: EJ

Time	Child Behavior		Therapist Behavior			Score
0:30	OnTask	OffTask	T+	T-	n/a	1
1:00	OnTask	OffTask	T+	T-	n/a	1
1:30	OnTask	OffTask	T+	T-	n/a	0
2:00	OnTask	OffTask	T+	T-	n/a	1
2:30	OnTask	OffTask	T+	T-	n/a	1
3:00	OnTask	OffTask	T+	T-	n/a	1
3:30	OnTask	OffTask	T+	T-	n/a	1
4:00	OnTask	OffTask	T+	T-	n/a	1
4:30	OnTask	OffTask	T+	T-	n/a	1
5:00	OnTask	OffTask	T+	T-	n/a	0

Treatment
Integrity = 80%