Journal of
Applied Behavior Analysis

Journal of Applied Behavior Analysis (Fall) 2020, **53,** 2024–2052 NUMBER 4

Assessment and treatment of response to name for children with autism spectrum disorder: Toward an efficient intervention model

Daniel E. Conine

Department of Learning Sciences, Georgia State University

Timothy R. Vollmer, Molly A. Barlow and Emma Grauerholz-Fisher

Department of Psychology, University of Florida

Cynthia M. Dela Rosa

Florida Autism Center

Abigail K. Petronelli

Department of Psychology, University of Florida and Florida Autism Center

Response to name (RTN) is an early developmental milestone, deficits in which are associated with autism spectrum disorder (ASD). This study extends previous research by evaluating an assessment and treatment model for RTN with 13 children with ASD. For all participants, phase 1 was a naturalistic social baseline. The 9 children who did not meet mastery criteria in phase 1 underwent a series of treatment conditions in phase 2. In phase 3, treatment components were removed, and generalization was assessed. Results indicated that tangible reinforcement procedures can produce rapid increases in discriminated RTN, sometimes without prompts. The total number of trials to mastery were reduced in the current study relative to previous research. Results also provide preliminary evidence to suggest that the phase 1 baseline condition may produce distinct patterns of RTN that could be used to predict treatment effects and further reduce trials to mastery in future work.

Key words: autism, discrimination training, early intervention, efficiency, response to name

Response to name (RTN) is a developmental milestone that is typically established during the first year of life (Gerber et al., 2010; Shelov & Altmann, 2009; Thomas et al., 2019). Deficits or delays in RTN are associated with a variety of developmental delays, especially autism spectrum disorder (ASD; Centers for Disease Control and Prevention, 2015; Miller et al., 2017; Nadig et al., 2007; Zhang et al., 2018). RTN is therefore included among many diagnostic criteria (Bryson et al., 2008; Lord et al., 2000; Lord et al., 1994; Robins et al., 2001) and early behavioral intervention curricula (e.g., Barbera, 2007; Maurice et al., 1996; Rogers & Dawson, 2010; Sundberg, 2008; Sundberg & Partington, 1998) for children with ASD. Despite the decades-long presence of RTN in intervention curricula, research on specific behavioral treatments for RTN in children with ASD has only recently emerged. Recent studies have demonstrated that it is possible to increase RTN for children with ASD using prompting and reinforcement strategies, with tangible reinforcement required in most cases (Conine et al., 2019; Cook et al., 2017; Rapp et al., 2019).

Abigail Petronelli is now at Florida Institute of Technology

We would like to thank the Florida Autism Center and Florida Autism Center staff for their collaboration and support of this project. We also thank Iser DeLeon, Brian Reichow, and Lisa Scott for their helpful comments on an earlier version of this manuscript.

Address correspondence to: Daniel E. Conine, College of Education & Human Development, Georgia State University, P.O. Box 3980, Atlanta, GA, 30302. E-mail: dconine@gsu.edu

doi: 10.1002/jaba.737

However, the exact intervention components required to increase RTN to mastery levels in this population have varied substantially across studies. A number of idiosyncratic intervention components have been necessary with individual participants, such as: prompts to remain seated (Rapp et al., 2019), modified prompts and prompt-fading strategies (Conine et al., 2019; Rapp et al., 2019), reinforcer variation (Conine et al., 2019; Cook et al., 2017; Rapp et al., 2019) and interruption of toy play or stereotypy (Conine et al., 2019). Perhaps as a result, the total number of treatment trials to mastery has varied widely across participants in previous studies, from as few as 130 trials to as many as 1720 trials (Cook et al., 2017). Given the importance often attributed to RTN as a critical target for early intervention (e.g., Miller et al., 2017), and given the costs associated with intensive behavioral interventions (Chasson et al., 2007; Jacobson et al., 1998), it remains necessary to develop RTN interventions that are more efficient for a greater number of learners.

One way to make treatment more efficient may be to identify the treatment components associated with the greatest likelihood of success in previous research. All three previous studies on increasing RTN for children with ASD have taken a progressive approach to introducing treatment components, starting with the most naturalistic or least intrusive treatment components and adding more intrusive or less naturalistic ones only when necessary (Conine et al., 2019; Cook et al., 2017; Rapp et al., 2019). In all three studies, the first tangible reinforcers evaluated were edible items. For many of the lengthier treatment cases, it was necessary to use other tangible reinforcers (e.g., toys, videos) instead of or in addition to the edibles. Thus, one strategy for reducing total trials-to-mastery may be to conduct thorough preference assessments at the start of intervention to identify all potential reinforcers, rather than arbitrarily using only edible items (Conine & Vollmer, 2019). A second lesson that may be learned from previous research is that if treatment is not effective after prompts and reinforcement for RTN are initiated, changes to reinforcement contingencies (e.g., reinforcer magnitude, interrupting access to competing reinforcement) may have a greater likelihood of success than changes to prompt-fading procedures (Conine et al., 2019). Thus, two changes seem likely to reduce overall trials-to-mastery: a) identifying potent reinforcers from the outset, and b) attempting reinforcement-based changes "earlier" than in previous treatment models.

In addition to the above changes, it may be possible to further reduce trials-to-mastery by arranging baseline as a screening assessment to predict the treatment condition that is most likely to be effective for each child on an individual basis. Such an approach is analogous to efforts made in the research literature on reducing problem behavior (Hagopian et al., 2015; Querim et al., 2013). To do this, an appropriate baseline condition must be selected, ideally one that captures RTN in a steady and natural state, before intervention. However, the no-consequence baseline used by Conine et al. (2019) produced some patterns consistent with extinction, and the contingent praise baseline used by Cook et al. (2017) and Rapp et al. (2019) produced some patterns of acquisition. Thus, the social condition used in Conine et al., in which naturalistic social comments were delivered contingent on RTN, may be the best suited for use as a preintervention screening to predict treatment effectiveness. However, that condition has never been evaluated as an initial baseline; it was implemented only after the no-consequence baseline in Conine et al. Thus, in the current study we piloted the use of this social condition as an initial baseline for all participants. To do so, we recruited children with ASD, with and without reported deficits in RTN. If this social condition is a valid baseline for RTN, it should capture steady-state, mastery-level RTN during the initial baseline with children who do not need intervention.

Furthermore, we evaluated whether other, nonmastery patterns of RTN observed in that condition would have predictive utility in determining treatment effects.

Finally, previous studies have extended the generality of RTN treatment by thinning reinforcement schedules after mastery and probing for generalization across people and settings (Conine et al., 2019; Cook et al., 2017; Rapp et al., 2019). It may also be possible to improve the efficiency of these parts of treatment. For example, all studies have conducted schedule thinning in an iterative manner (e.g., FR1, FR2, VR3, etc.) and Conine et al. (2019) conducted separate generalization probes across multiple contextual variables. A terminal probe design (e.g., LeBlanc et al., 2001; Slocum et al., 2018) may permit some schedule thinning steps to be skipped, and generalization probes may be further abbreviated by combining multiple contextual variables (e.g., people, location, distance) into a single probe condition.

Thus, the purpose of the current study was to improve the overall efficiency of treatments to increase RTN for children with ASD in three main ways. First, we modified the general treatment model used in previous studies to incorporate a thorough preintervention preference assessment and to focus on the treatment components that have been the most successful in previous studies. Second, we used the social condition from Conine et al. (2019) as an initial baseline and tested the utility of that condition as a screening to predict the treatment condition that is most likely to be effective on an individual basis. Third, we aimed to abbreviate treatment generalization by using a singular probe condition and a terminal probe design.

Method

Participants, Materials, and General Setting

Thirteen children with ASD participated in the current study: six females and seven males, between the ages of 3 and 11 years. Participant pseudonyms and demographic information are listed in the first three columns of Table 1. All participants were recruited from one of multiple clinics providing applied behavior analysis (ABA) services to children with ASD. To evaluate the social condition from Conine et al. (2019) as a naturalistic baseline, we aimed to assess a wide variety of children with ASD who did and did not require intervention. Thus, only eight of the 13 participants (Felicity, Jackson, Chloe, Kayla, Lucas, Lucian, Lydia, and Salvador) were referred to the study based on deficits in RTN reported by their Board Certified Behavior Analysts® (BCBAs). The remaining five participants were recruited for other reasons. Vera was recruited because her twin brother, Lucas, was recruited as described above. Bethany and Beckham were recruited as a means of assessing long-term RTN maintenance after their participation in Conine et al. (2019). Daniel and Ned were referred to the study because their BCBAs reported general deficits in attention and compliance and were interested in an assessment to determine if RTN merited intervention in pursuit of those broader behavioral goals.

Experimenters used a timer, pen, paper, and video camera to record data during sessions. All sessions were conducted at the clinic that each participant regularly attended, in one of two settings, termed the *generalization environment* and the *session room*. These settings are described in greater detail in the sections below.

Response Measurement

Data were collected on a trial-by-trial basis. The primary dependent variables were RTN and response to other names (RTO), as in Conine et al. (2019). RTN was defined as the participant looking at the experimenter's eye region for any duration within 5 s of the experimenter saying the participant's name. The

Table 1

Participant Demographic Characteristics, Experimental Phases Completed, Screening Type Assigned at the Conclusion of Phase 1, Prompt Used in Phase 2, and Effective Treatment in Phase 2

Participant	Age*	Sex	Phases Completed 1	2	3	Screening Type	Prompt Used	Effective Treatment
Bethany	11	F	X			Type 1	N/A	N/A
Daniel	9	M	X			Type 1	N/A	N/A
Ned	9	M	X			Type 1	N/A	N/A
Vera	3	F	X			Type 1	N/A	N/A
Beckham	7	M	X	X	X	Type 2	N/A	T1+
Felicity	6	F	X	X	X	Type 2	N/A	T1
Jackson	3	M	X	X		Type 3	N/A	N/A**
Chloe	4	F	X	X	X	Type 4	Hand to eyes	T2+
Kayla	4	F	X	X	X	Type 4	Hand to eyes + interruption	T4
Lucas	3	M	X	X		Type 4	Track fingers	T4
Lucian	5	M	X	X	X	Type 4	Hand to eyes	T4
Lydia	4	F	X	X	X	Type 4	Hand to eyes	T3
Salvador	4	M	X	X		Type 4	Hand to eyes + interruption	T2+

* Age at the start of study.
** Exited study prior to meeting mastery criteria.

experimenter's eye region included any part of the experimenter's face located vertically between the bottom of the nose and the top of the eyebrows, and horizontally between the outside corners of either eye. RTO was defined as the participant looking at the experimenter's eye region within 5 s of the experimenter saying any name other than the participant's name. As with RTN, any duration of RTO was scored. In some treatment conditions, data were also collected on *prompted* RTN. The definition for prompted RTN was the participant looking at the experimenter's eye region within 5 s of the experimenter saying the participant's name and delivering a prompt (described in Conditions below).

Experimental Design

The study progressed through three phases: screening (Phase 1), treatment (Phase 2), and generality (Phase 3). In all three phases, a multielement design was used in which differential contingencies were programmed for RTN and RTO (Conine et al., 2019). In Phase 2, treatment was staggered in keeping with a nonconcurrent multiple baseline across participants, although the multielement design remained the primary means of demonstrating experimental control.

Mastery required that three criteria were met across three consecutive sessions: a) 67% or greater RTN, b) a discrimination index (DI; Luczynski & Hanley, 2014) between RTN and RTO of 0.8 or greater, and c) no downward trend in RTN or upward trend in RTO. Discrimination indices were calculated by summing the percentage RTN across three sessions and dividing that sum by the sum of percentage RTN and percentage RTO for the same three sessions. During Phase 3, the mastery criterion for percentage RTN was modified to include either: a) the mastery criteria from the previous phases or b) two consecutive sessions with 100% RTN to accelerate schedule thinning in situations where RTN was maintaining at 100% (see Phase 3 below).

The 67% mastery criterion is lower than in previous studies and was selected for three reasons. First, in two studies (Conine et al., 2019;

Rapp et al., 2019), 70% RTN was used as a mastery criterion for a subset of participants. Given that our sessions included six RTN trials (see General Procedures below), 67% was the nearest percentage to this 70% precedent that was mathematically possible. Second, there is a relative paucity of normative data indicating how frequently people with or without ASD (e.g., typically developing children, adults) respond to their name under a variety of naturalistic conditions (see Beaulieu et al., 2012). Thus, we chose to use the lowest percentage RTN criterion used across previous studies. Finally, this mastery criterion for percentage RTN was supplemented by requirements that this response was highly discriminated (DI of 0.8 or greater).

Preassessment

For all participants, a preference assessment was conducted at the start of the study to identify the toys that would be used as distracter items during all sessions. Other preference assessments were later conducted to identify reinforcers for children who participated in Phase 2 (see Phase 2 below). For most participants, a multiple-stimulus-without-replacement (MSWO; DeLeon & Iwata, 1996) format was used. For Chloe and Lydia only, a response-restriction preference assessment (RRPA; Hanley et al., 2003) was conducted instead, due to inconsistent selection patterns and problem behaviors that were observed in previous MSWO assessments at their clinic. In either format, the initial preference assessment contained eight toys or leisure items, which were selected based on input from caregivers, clinical staff, or both, using an abbreviated RAISD form (Fisher et al., 1996) and brief supplemental interviews.

The four highest-preferred stimuli from this initial preference assessment were designated the *distracter items* for the course of the study (described below). If a screen-based media device (e.g., iPad) was among the four highest-preferred stimuli, it was not used as a distracter item, due to concerns that such devices might compete too strongly with reinforcement contingencies for RTN. The experimenter also verbally offered the participant the opportunity to choose one additional toy from the clinic before each session; a toy chosen in this manner was added to the distracter items for that session.

Settings

Sessions were conducted in two settings. One setting was a dedicated session room, which was only used for children who participated in Phase 2. The exact size of each session room varied according to the clinic in which sessions were conducted (approximately 3 m by 3.5 m). The session room included a table with two chairs placed side-by-side along one edge. The experimenter and the child sat in these chairs, approximately 2 ft apart, facing the same direction. The four or five distracter items described above were placed on the table. For Lucas only, a table and chairs were present, but the experimenter and child sat on the floor with the distractor items because Lucas did not regularly sit at a table during sessions at the clinic. All sessions in this setting were conducted by the same person, referred to as the *primary experimenter*. The primary experimenter was either a member of the research team who was novel to the participant (Chloe and Lydia), or a staff member from the clinic who worked regularly with the participant as a behavior technician (Beckham, Felicity, Jackson, Kayla, Lucas, Lucian, Salvador).

Sessions were also conducted in a *generalization* setting. The exact dimensions of the room used for this setting varied across clinic sites. For most participants (Bethany, Daniel, Ned, Beckham, Felicity, Chloe, Lydia), the generalization setting was a large playroom at the clinic (14.6 m by 7.0 m). For other participants (Vera, Jackson, Kayla, Lucas, Lucian, Salvador), the generalization setting was a shared classroom used by the participant and one to three additional children (approximately 6 m by 3 m). In either case, the room included at least one table and chairs, the distracter items, and a wide variety of other toys normally located in that room. At least one adult and one child in addition to the participant and experimenter were also present. Sessions in the generalization setting were conducted by three different experimenters, who alternated in a random order every three sessions. These three experimenters included: a) the primary experimenter, b) another therapist who frequently worked with the participant at the clinic and was likely to be present as a secondary observer during treatment in Phase 2, and c) another therapist from the clinic who did not work regularly with the participant and was thus unlikely to be present during sessions in Phase 2. Experimenters also stood in three different locations during sessions in this setting: 5 ft behind, 5 ft to the left of, and 5 ft to the right of the participant. These three locations were evenly distributed across all S^D and S^Δ trials (described below).

General Procedures

Across all phases of the experiment, each session included nine total trials: six S^D trials and three S^Δ trials. S^D trials began with the experimenter saying the participant's name (RTN); S^Δ trials began with the experimenter saying some other name (RTO). Other names were selected as in Conine et al. (2019): A list of other names was generated, which included the nine most popular, same-gendered baby names in the year of the participant's birth (Social Security Administration, 2018). Names that rhymed with the participant's name were excluded from this list and replaced with the next most-popular name. Three of these nine names were randomly selected for each session, such that all nine names were used every three sessions.

To begin each session, the experimenter positioned themself, the participant, and the distracter items as described in Settings (above), then presented the first trial. Trials were only presented if: a) at least 20 s had elapsed since the end of the last trial, and b) the participant had been engaged with one of the distracter items and not looking at the experimenter for the past 5 s. Engagement was defined as touching a toy, or looking at a toy with which the primary mode of engagement is looking (e.g., books). If the participant did not meet this engagement criterion, the experimenter waited until 5 s of consecutive engagement had occurred, then presented the trial. If the participant was not engaged for 20 s consecutively, the experimenter prompted engagement either verbally (e.g., "play with one of your toys") or by handing a toy to the participant. The experimenter did not look at or interact with the participant during any intertrial interval (ITI), except to prompt engagement if needed, according to the rules above.

When presenting S^D trials, the experimenter turned to face the participant as they said their name. When presenting S^Δ trials, the experimenter looked straight ahead (i.e., did not turn to face the participant). This difference was programmed to make the delivery of S^Δ trials more closely approximate the circumstances in which other people's names are likely spoken in the natural environment.

Phase 1: Screening

All Phase 1 sessions were conducted under the social condition (described below). A minimum of three sessions were first conducted in the generalization setting; one session with each of the three experimenters. Additional sessions were conducted if needed to establish stability for RTN or RTO. If a participant met mastery criteria during these sessions in the generalization setting, they did not participate in the remaining phases of the experiment (i.e., intervention was not necessary). If participants did not meet mastery criteria in the generalization setting, additional social sessions were conducted in the session room setting. These sessions were conducted until stability was established for both RTN and RTO. Trend was evaluated using data from both settings (i.e., including data from the generalization setting). Additional sessions were conducted in the session room with some participants after stability was achieved to accommodate the multiple baseline design planned for Phase 2. For Beckham and Felicity, a slightly different session pattern was used; Phase 1 sessions began in the session room, and only one session was conducted in the generalization setting (see Results). Any participants who did not meet mastery criteria by the end of Phase 1 moved to Phase 2. A visual depiction of these Phase 1 procedures is shown in the upper portion of Figure 1.

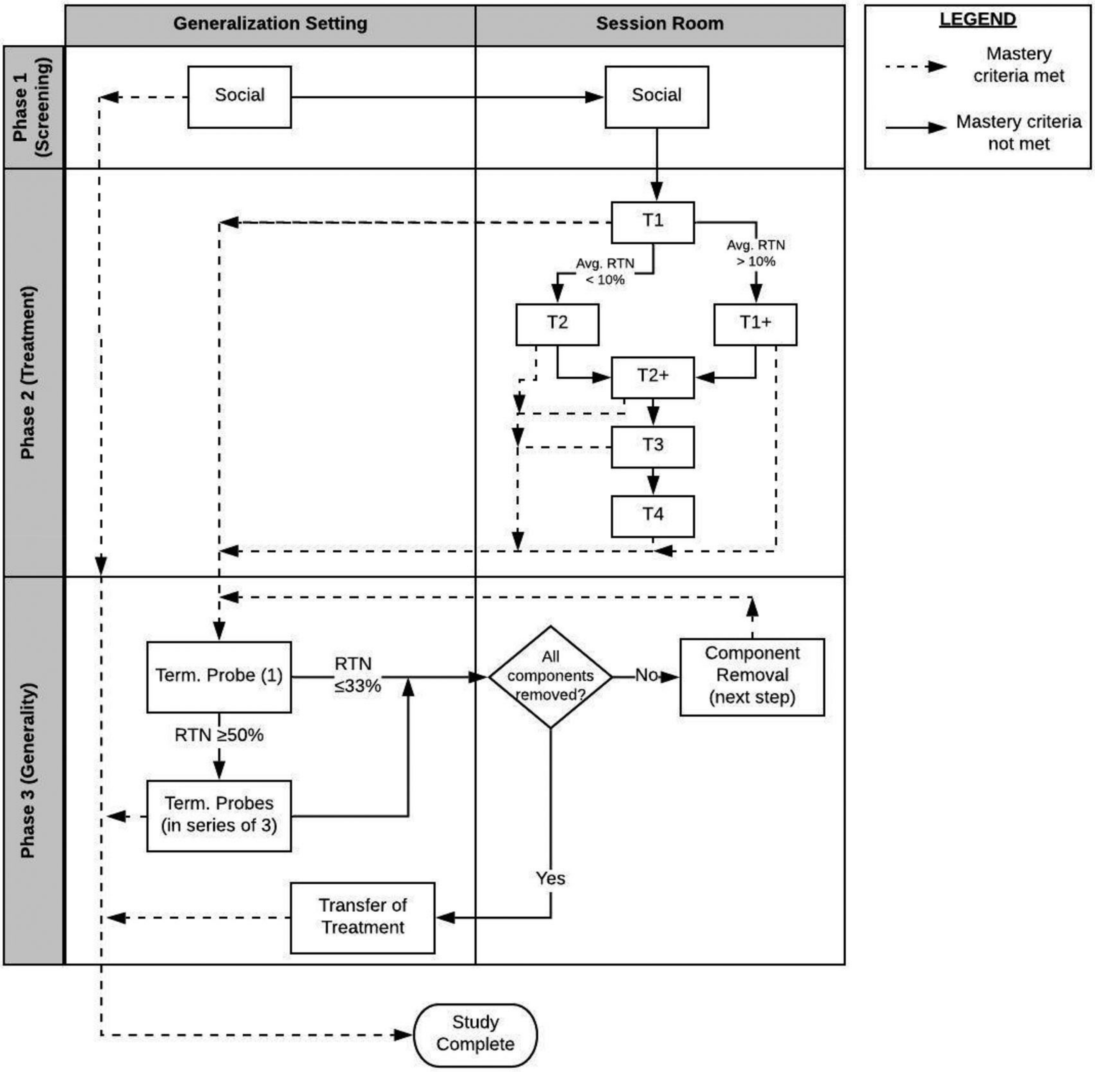
Social Condition. Procedures were identical to the social condition from Conine et al. (2019). Contingent on RTN the experimenter initiated a social interaction with the participant according to one of five categories: instruction, joint attention bid, "wh" question, yes/no question, and general comment (for definitions see Conine et al., Table 1). All social interactions started with a greeting (e.g., "hey," "hi") and were tailored to the current context (e.g., toy the participant was recently engaged with, interesting features of the room). Appropriate responses to social interactions were never prompted, but if they occurred were followed by praise related to that social interaction (e.g., "right, that is a ball!"). Incorrect or nonresponses to the social interaction were followed by a floorholding response (Hart & Risley, 1995; 1999). For example, if the experimenter asked, "What's this?" and the child did not respond, the experimenter might say "it's a puzzle" or "it's a red car" (see Conine et al., Table 1 for additional examples). If the participant engaged in RTO, the experimenter continued looking away from the participant for the remainder of the 5 s response interval, then began the next ITI.

Phase 2: Treatment

Only participants whose RTN did not meet mastery criteria during Phase 1 participated in Phase 2. Data from Phase 1 served as the baseline for Phase 2. All sessions in Phase 2 were conducted in the session room (with the primary experimenter) and progressed through the treatment conditions described below, in the order shown in the middle portion of Figure 1. Treatment conditions were arranged to progress from least intrusive and intensive to most intrusive and intensive. At least two sessions were conducted in each treatment condition, and sessions continued until both RTN and RTO were stable. If RTN and RTO were stable but did not meet mastery criteria, the next treatment condition was initiated. If participants met mastery criteria at any point during treatment, Phase 3 began (Fig. 1). Each treatment condition is described in greater detail below.

Two additional preference assessments were conducted at the start of Phase 2. First, an MSWO assessment containing eight edible items was conducted. Next, a combined edible-leisure MSWO (e.g., Conine & Vollmer, 2019) was conducted, using the top four items from the edible MSWO and the top four items from the leisure item assessment conducted at the start of the study. This combined MSWO was conducted in order to identify the highest preferred items overall (rather than within a single stimulus category) for use as reinforcers during treatment. However, the edible and combined preference assessments were not conducted for four participants (Chloe, Lydia, Kayla, and Lucian), because they had a history of food selectivity and consumed a very limited number of edible items, none of which were reported to function as reinforcers. Accordingly, results from either the leisure preference assessment at the start of the study (Chloe, Lydia, Kayla, and Lucian), or the combined preference

Figure 1
Order of Treatment Conditions across All Three Phases of the Study



assessment described above (all other participants) were used to select reinforcers as described in the treatment conditions below.

If an item that was previously used as a distracter item ranked among the three highest-preferred stimuli on the combined assessment (and would therefore be used as a reinforcer during treatment), that item was removed from the distracter items at the start of Phase 2 and replaced in the set of distracter items with the next highest-preferred leisure item that was not selected for use as a reinforcer.

Treatment 1: Differential Reinforcement (T1). Phase 2 began with T1 for all participants. Contingent on RTN, the experimenter gave a preferred tangible item to the participant and simultaneously initiated a social interaction (as in the social condition). If the reinforcer was a leisure item, 30 s access was provided. For edible items, one bite-sized portion was provided until it was consumed. If a participant requested a specific social interaction with the experimenter (e.g., tickles, picking up) during a reinforcement interval, that interaction was also delivered. If the request occurred during an ITI, that interaction was also delivered in addition to the tangible item contingent on RTN in the next trial. As in the social condition, the experimenter delivered no consequences (i.e., continued looking away from the participant) following RTO.

The tangible items used as reinforcers were chosen according to one of two rules. For most participants (Beckham, Felicity, Jackson, Lucas, and Salvador), the three highest-preferred stimuli from the combined preference assessment were designated as the reinforcers for T1. One of these three items was preprogrammed as the reinforcer for each S^D trial, randomly alternating in blocks of three. For example, for Felicity these three stimuli were juice, a ball, and a cookie. During T1, reinforcers may have been programmed as: juice for trial one, ball for trial two, cookie for trial three, ball for trial four, cookie for trial five, juice for trial six. However, if the participant requested any of the three items during an ITI, that requested item was used as the reinforcer on the next trial instead of the planned item. If the participant refused an item at any point when it was offered, that item was replaced on the list of reinforcers with one of the other two items for the rest of the session. For Lydia, Chloe, Kayla, and Lucian, the top-ranked item from the leisure item preference assessment at the start of the study was used as the reinforcer for all S^D trials in T1 (for these four participants, that highest-preferred stimulus was an iPad or computer that played videos). If any of the items used as a reinforcer was previously one of the distracter items in Phase 1, that item was not used as a distracter item in Phase 2.

If participants did not meet mastery criteria during T1, either T1+ or T2 was conducted next (Fig. 1). If the participant had engaged in RTN in greater than 10% of trials on average during T1, T1+ was initiated. If the participant emitted RTN in 10% of trials or fewer on average during T1, T2 was initiated.

Treatment 1 with Increased Reinforcer Magnitude (T1+). Procedures were identical to T1, except that the magnitude of tangible reinforcement contingent on RTN was increased in three ways. First, if three tangible items were used in T1, all three of those items were delivered simultaneously contingent on every instance of RTN in T1+. If only one tangible item was used during T1, the second- and third-preferred items from the combined MSWO (or the leisure MSWO, if a combined MSWO was not conducted) were delivered along with the highest preferred item after RTN. Second, the duration of access to all leisure items was increased from 30 s to 90 s (the quantity of each individual edible item was not changed). Finally, if any physical interactions (e.g., tickles, hugs) were reported by clinic staff to be highly preferred, access to these interactions was also offered to the participant for 90 s following RTN.

Prompt Assessment. If T1 or T1+ was unsuccessful, a prompt assessment was conducted, as all subsequent conditions involved prompts (Fig. 1). The prompts included in this assessment are listed and defined in Table 2, and are identical to those assessed in Conine et al. (2019), except that one prompt was added: tracking a highly preferred leisure item. One prompt assessment session was conducted for each of the prompts listed in Table 2.

All prompt assessment sessions took place in the session room and consisted of five trials. All distracter items that were included in T1 or T1+ were present. To start each trial, the experimenter turned toward the participant and delivered the prompt that was being assessed; the experimenter did not say any names. If the participant looked at the experimenter's eye region within 5 s of the prompt, the

Table 2

Prompts Used in Prompt Assessments

Prompt	Definition
Vocal	Experimenter says, "look at me."
Tap	Experimenter taps participant's shoulder with their index finger three times.
Track fingers	Experimenter holds outstretched index and middle finger in front of participant's eyes, and slowly moves them to point to the experimenter's eyes across a period of 5 s, or until target response occurs.
Track toy	Experimenter picks up the toy that the participant is currently engaged with, and slowly moves it to directly in front of and between the experimenter's eyes, across a period of 5 s, or until target response occurs.
Hand to eyes	Experimenter gently picks up the participant's hand and moves it to directly in front of and between the experimenter's eyes, and holds it there for 5 s, or until target response occurs.
Track HP edible*	Experimenter places a highly preferred edible item directly in front of and between the participant's eyes, and slowly moves it to directly in front of and between the experimenter's eyes, across a period of 5 s, or until target response occurs.
Track HP leisure item*	Experimenter places a highly preferred leisure item directly in front of and between the participant's eyes, and slowly moves it to directly in front of and between the experimenter's eyes, across a period of 5 s, or until target response occurs.
Interruption**	Experimenter removes any toy the participant is currently engaged with (e.g., by taking out of participant's hand), and blocks access to any other toys the participant attempts to engage with for 5 s.

* These prompts were only assessed if an edible item or leisure item, respectively, was among the highly preferred items that would be used as a reinforcer during Treatment 2.

** This prompt was never assessed in isolation, but rather was assessed in combination with each prompt listed above, and only if no prompts occasioned 100% response without interruption.

experimenter delivered social interaction and a tangible item as in T1. After one session was conducted with each prompt (Table 2), if any prompts had occasioned a correct response in 100% of trials, the prompt assessment ended. If not, the prompt assessment was repeated, but with an *interruption* procedure added to each prompt. When interruption was added, the experimenter began each trial by turning to face the participant, gently interrupting the participant's access to the distracter items and motor stereotypy, and simultaneously delivering the prompt.

Once the prompt assessment was complete, the least intrusive prompt which occasioned the highest percentage correct response was selected as the prompt to be used in all subsequent treatment sessions. The hypothesized order of intrusiveness for these prompts (from least to most) that was used to make this decision is listed in Table 2. This order is based on the order proposed in Conine et al. (2019), and is subjective (i.e., the intrusion level of these prompts was not empirically determined, and alternate orders of intrusiveness are possible). Any prompt combined with interruption was always considered more intrusive than any prompt without interruption.

Treatment 2: Reinforcement and Prompts (T2). Procedures were identical to T1 except that if the participant did not engage in RTN within 5 s of their name in an S^D trial, the experimenter repeated the participant's name and simultaneously delivered the prompt chosen from the prompt assessment (above). If the participant did not engage in RTN within 5 s of the prompt, the experimenter once again simultaneously said the participant's name and delivered the prompt, for up to a maximum of five total prompt presentations (though this rarely occurred). Contingent on RTN (prompted or unprompted), the experimenter delivered a preferred tangible item and a social interaction according to the same criteria outlined in T1.

Treatment 2 with Increased Reinforcer Magnitude (T2+). Procedures were identical to T2, except that the magnitude of tangible reinforcement described in T1+ was used.

Prompted and unprompted RTN were both followed by that same reinforcer magnitude.

For Lucian only, a brief presession preference assessment was also added to T2+ and all subsequent conditions, due to BCBA report that his preferences changed often. Lucian was given a choice between four or five highly preferred toys (not including the iPad) at the start of session, and the two that were selected first were used as reinforcers in addition to the iPad that was previously used in T1.

Treatment 3: Reinforcement, Prompts, and Differential Reinforcement of Unprompted RTN (T3). Procedures were identical to T2, except that different magnitudes of tangible reinforcement were provided for unprompted and prompted RTN. The large reinforcer magnitude used in T1+ and T2+ was delivered contingent on unprompted RTN. Contingent on prompted RTN, only the highest-preferred tangible item from the combined preference assessment was provided. If that item was a leisure item, it was provided for only 15 s; if it was an edible item, a portion was provided that was only one-half the size of the portion used in T1.

Treatment 4: Reinforcement, Prompts, Differential Reinforcement of Unprompted RTN, and Interruption (T4). Procedures were identical to T3 except that when initiating an S^D trial, the experimenter removed any toys that the participant was currently engaged with and placed them out of reach, then stated the child's name. The experimenter also blocked any attempt to reach for a distracter item or to engage in motor stereotypy for the remainder of the trial (excluding the reinforcement interval). During T4, the requirement that participants were engaged with a leisure item for 5 s prior to each trial was removed; however, trials were still only initiated if the participant had not been looking at the experimenter for 5 s.

Phase 3: Generality

Phase 3 was initiated after participants met mastery criteria in Phase 2. The purpose of Phase 3 was to assess whether participants would continue to engage in mastery-level RTN as treatment components were removed and as tangible reinforcement schedules were thinned. Phase 3 also assessed generalization across experimenters and settings.

The lower portion of Figure 1 outlines the order and decision-making process of Phase 3. This phase began with one terminal probe session (described below), conducted by the primary experimenter in the generalization setting. If the participant engaged in RTN in 33% of trials or fewer in this probe, the next step in treatment component removal (described below) was implemented in the session room. If the participant engaged in RTN in 50% of trials or greater on this first probe, terminal probe sessions continued in the generalization setting, rotating across the three generalization therapists assigned to that participant in a random order every three sessions, until either: a) mastery criteria were met, or b) RTN and RTO stabilized without meeting mastery criteria. If mastery criteria were met during terminal probes at any point, the study was concluded. If data stabilized in terminal probes without meeting mastery criteria, the next step in treatment component removal began in the session room. Terminal probes were periodically repeated in this same manner as a part of the treatment component removal process (described below).

Terminal Probes

All terminal probe sessions were conducted in the generalization setting. Contingencies for RTN were similar to T1, except that: a) only the single highest-preferred tangible item from T1 was used, and b) that item was delivered contingent on unprompted RTN according to a variable ratio 6 (VR6) reinforcement schedule (range, four to eight responses). Social interactions (as in the social condition) were delivered after unprompted RTN on all trials, including trials in which a tangible reinforcer was not delivered.

Component Removal

The purpose of the component removal subphase was two-fold: a) to evaluate whether RTN would maintain under less intensive, more naturalistic procedures if mastery-level RTN was not obtained in the first series of terminal probes and b) to evaluate whether generalized responding in the terminal probes was more likely to occur as components were gradually removed from treatment in the session room. All component removal sessions were conducted in the session room. We used a reverse-order treatment hierarchy from most- to least-intrusive as follows: T4, T3, T1+, T1, T1(VR2), T1(VR3), T1(VR6). As a note, T2 and T2+ were not a part of this hierarchy; this was done in order to evaluate maintenance of RTN without prompts as early as possible in Phase 3. For VR2 and VR3 sessions, contingencies were similar to the VR6 contingencies used in terminal probes except that the VR schedule requirement was set at two (range, 1 to 3) or three (range, 2 to 4) instances of unprompted RTN.

Component removal sessions began with the treatment condition that was one step down the above hierarchy from the condition in which RTN was mastered during Phase 2. For example, if RTN was mastered in T4, component removal began with T3; if RTN was mastered in T1+, component removal began with T1. Sessions were conducted in that condition until either: a) two consecutive sessions occurred with 100% RTN and a DI of .8 or greater, or b) three consecutive sessions occurred with 67% RTN or greater and a DI of .8 or greater. If these mastery criteria were met, terminal probes resumed as described above (Fig. 1). If component removal was repeated later in Phase 3 due to mastery criteria not being met in subsequent terminal probes (Fig. 1), sessions began with the next less-intrusive treatment than the component removal step that was conducted most recently. For example, if mastery criteria were most recently met in T3 during component removal, a subsequent return to component removal would begin with T1 + .

This alternation between component removal and terminal probes continued until one of four things occurred (Fig. 1). First, if mastery criteria were met during terminal probes at any point, the study ended. Second, if mastery criteria were met in the session room after all treatment components were removed, but the participant never met mastery criteria in the terminal probes, the transfer-of-treatment process began (described below). Third, if a participant did not meet mastery criteria in the session room at any step of treatment component removal, a reversal was conducted, to the component removal step that was most recently effective. After mastery-level RTN was established in this reversal, component removal resumed and continued until either: a) the same treatment condition had failed to support mastery-level RTN twice or b) one of the other above criteria for terminating Phase 3 were met (see Results for details on the application of these criteria across participants). Finally, for one participant (Felicity), when mastery criteria were met in terminal probes across one or two (but not all) of the generalization therapists, a generalization-across-therapists procedure was conducted (described below).

Transfer of Treatment

The purpose of this subphase was to conduct treatment in the generalization setting if the participant never met mastery criteria in terminal probes after completing the component removal process described above. Treatment began in the generalization setting with the treatment condition in which RTN was mastered in Phase 2. Treatment components were then removed in the generalization setting according to the hierarchy outlined for component removal until mastery criteria were met in the generalization setting under T1 (VR6) conditions, at which point the study ended. Alternatively, if the same condition failed to support mastery-level RTN twice, component removal ended and schedule thinning (to a maximum of VR6) was initiated under the treatment condition that was most recently successful in the generalization setting.

Generalization Across Therapists (Felicity only)

This subphase was conducted only if the participant met RTN mastery criteria during terminal probes for one or two, but not all, of the generalization setting therapists. Procedures were identical to terminal probes, except that each trial within a session was conducted by a different therapist who worked at the treatment center (for a total of nine therapists per session). These therapists ran trials as they entered and exited the generalization setting as a part of their regular daily clinic activities.

Interobserver Agreement and Treatment Integrity

Primary behavioral data in each session were collected by the experimenter conducting the session. A secondary observer collected interobserver agreement (IOA) data on a trial-by-trial basis for a subset of sessions across all phases and conditions of the study for all participants. The secondary observers recorded these data either in-vivo during session using pen and paper, or from session video recordings. In either case, the secondary observer or the person collecting video positioned themselves so that the experimenter's face and the participant's face were both visible throughout the session. Secondary observers were provided with a written response definition (attached to the data sheet) and the primary experimenter verbally reviewed this response definition with the secondary observers and answered any questions at the start of each session. For each session, the number of trials in which both observers agreed on the occurrence or nonoccurrence of RTN or RTO was divided by the total number of trials in the session and multiplied by 100 to yield a percentage. Means and ranges of IOA scores for all participants are shown in Table 3.

Treatment integrity was calculated from video records, when available. A sufficient number of video records was available at the conclusion of the study for four participants: Felicity, Beckham, Lydia, and Chloe. Treatment integrity was scored for all treatment conditions in Phase 2; Phases 1 and 3 were not scored. For each S^D trial (RTN), data collectors scored whether the therapist a) faced the child, b) spoke the name audibly, c) waited 5 s before turning away or delivering a prompt, d) delivered the correct prompt (if needed) or no prompt (if not needed), e) delivered the correct social interaction, and f) delivered the correct tangible item. If the participant went through T4 (described above) the data collector also scored whether the therapist interrupted stereotypy and toy engagement before delivering the S^D. For each S^Δ trial (RTO), the data collector scored whether the therapist a) did not face the child, b) spoke the name audibly, and c) delivered no consequences regardless of the child's response. For each trial (S^D or S^Δ), treatment integrity was calculated by dividing the number of correct components by the total number of components in the trial and multiplying by 100 to yield a percentage. For each session, an overall treatment integrity score was obtained by averaging the treatment integrity for all nine trials. Means and ranges of treatment integrity scores for these four participants are shown in Table 3.

Results

Phase 1 Results

We characterized the results of Phase 1 according to four patterns, or types, of RTN that were observed during screening. These

Table 3

Interobserver Agreement (IOA) and Treatment Integrity (TI) Data

Participant	Mean IOA (Range)	Percentage of Sessions with IOA	Mean TI (Range)	Percentage of Sessions with TI*
Bethany	100%	40%		
Daniel	99% (89-100%)	81%		
Ned	100%	33%		
Vera	96% (89-100%)	100%		
Beckham	98% (89-100%)	36%	100%	27%
Felicity	99% (89-100%)	31%	97% (96-98%)	27%
Jackson	98% (89-100%)	65%		
Chloe	100%	31%	90% (81-96%)	29%
Kayla	99% (78-100%)	52%		
Lucas	100%	44%		
Lucian	98% (89-100%)	48%		
Lydia	99% (89-100%)	35%	96% (89-100%)	26%
Salvador	98% (89-100%)	61%		

* For treatment integrity, percentage of sessions reflects Phase 2 only.

patterns were defined according to the level (percentage) of RTN and the discrimination (DI) between RTN and RTO. If data from the generalization setting and data from the session room were substantially different with respect to these features, we made this determination based on these features of the session room data (this determination was only necessary for Jackson and Salvador). The rationale for this determination was that if the screening was to predict treatment effectiveness, and if treatment was to be delivered in the session room in Phase 2, the session room screening may serve as a better predictor of treatment effects than the generalization setting screening.

Figure 2 shows representative graphs of the four screening types. Type 1 RTN (Figure 2, upper left panel) was RTN that met mastery criteria during screening for both percentage and DI. In other words, Type 1 RTN was both frequent and discriminated. Participants with Type 1 RTN screened out of participation in the treatment phases of the study. Type 2 RTN (Figure 2, upper right panel) met the DI mastery criterion, but not the percentage RTN criterion. In other words, Type 2 RTN was infrequent but highly discriminated. Type 3 RTN (Figure 2, bottom left panel) occurred to some extent during screening (greater than 10% on average) but did not meet the DI mastery criterion. In other words, Type 3 RTN can be characterized as undiscriminated. Finally, Type 4 RTN (Figure 2, bottom right panel) never or rarely occurred (less than 10% of trials on average) during screening. For these participants, RTN and RTO were both so infrequent that it was uninformative, even impossible, to calculate a DI. The second column from the right in Table 1 lists the types that were assigned to each participant's RTN at the end of Phase 1.

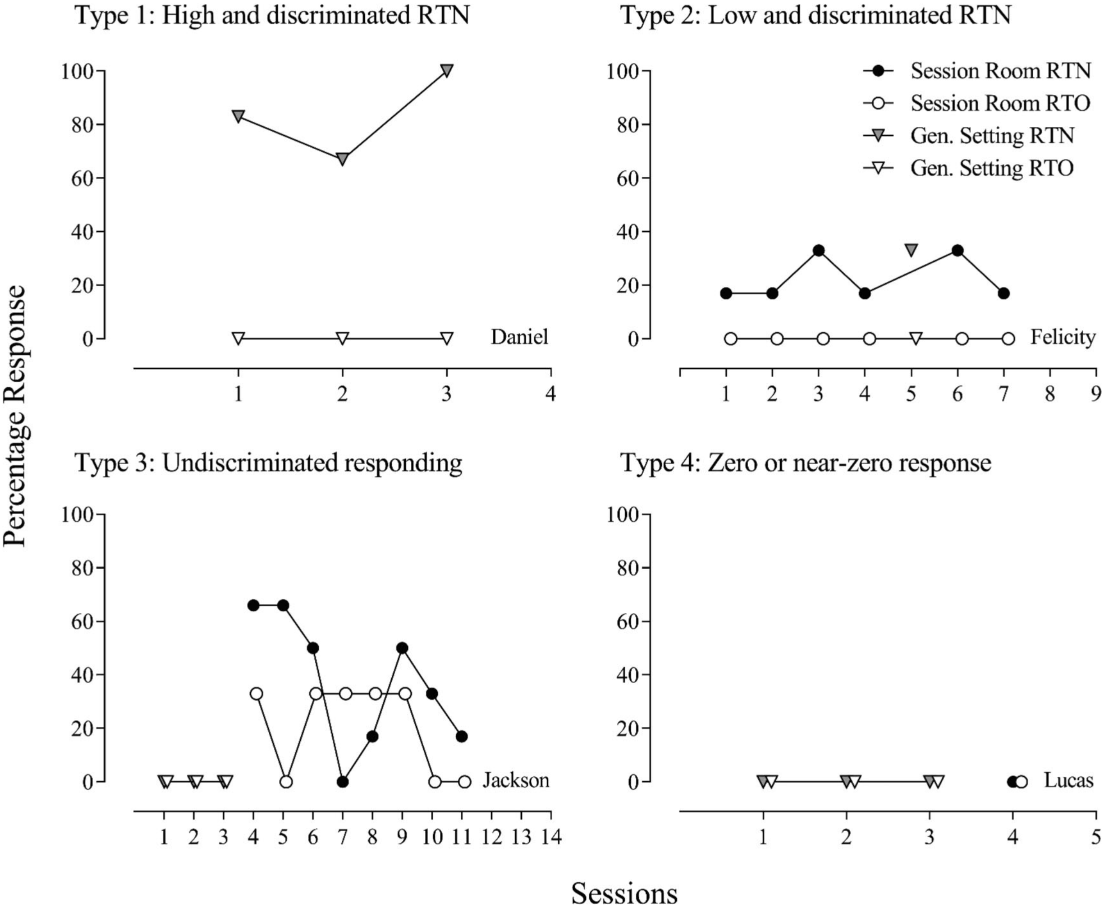
Figure 3 shows results from Phase 1 for the four participants with Type 1 RTN. Because these participants screened out of participation in subsequent phases, the data in this figure constitute all study data for these participants. All four of these participants met mastery criteria after the minimum three sessions of screening. Graphical depictions of Phase 1 results for all other participants are displayed as a part of Phase 2 figures, below.

Phase 2 Results

Figure 4 shows results for participants whose RTN during Phase 1 was characterized as Type 2 or Type 3. For Felicity (top panel), RTN was low (17-33%) but stable and highly discriminated

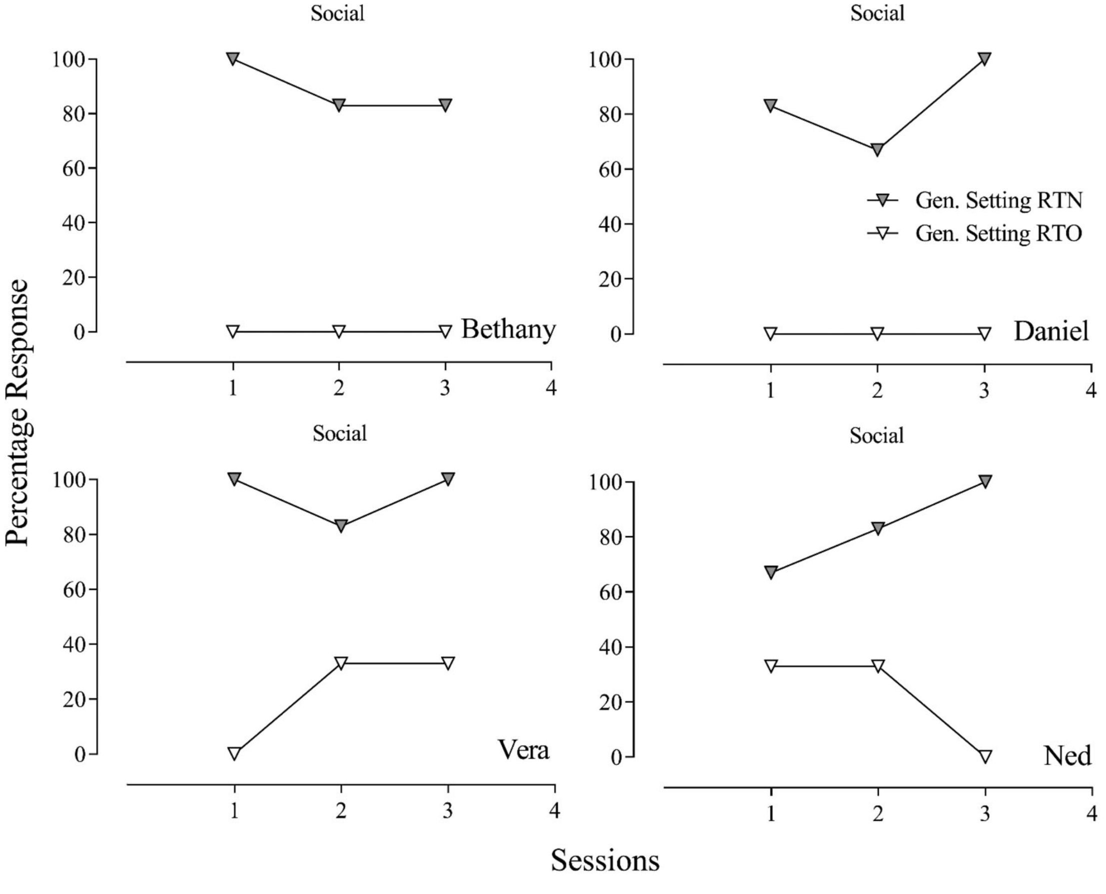
Figure 2

Representative Examples of the Four Screening Types Identified in Phase 1



from RTO during Phase 1. RTN immediately increased to mastery levels when T1 was introduced. However, RTO also began to increase by the third session, and T1 continued until RTO decreased. Felicity met mastery criteria in session 17. Jackson (middle panel), engaged in no RTN or RTO during Phase 1 in the generalization setting. During Phase 1 in the session room, RTN and RTO were both variable and low. Jackson's Phase 1 RTN was classified as Type 3 based on a DI below mastery criteria (0.64) in the session room. During T1, Jackson's RTN increased, but stabilized below mastery criteria (50% RTN, DI: 0.73). RTN increased substantially during T1+, with 83% RTN in the second and third sessions and a DI of 0.72 across those same two sessions. Jackson's participation in the study was suspended after session 25 due to a long period of absences from the clinic. Thus, Jackson did not meet mastery criteria before exiting the study, although T1+ appears to have been initially effective. For Beckham (bottom panel), RTN was discriminated (DI: 0.94) during Phase 1, but stabilized below 67%. When T1 was introduced, both percentage RTN and the DI remained largely unchanged. When T1+ was introduced, RTN immediately increased to 100%. Additional sessions were conducted until stability in RTN

Figure 3
Phase 1 Results for All Participants with Type 1 RTN



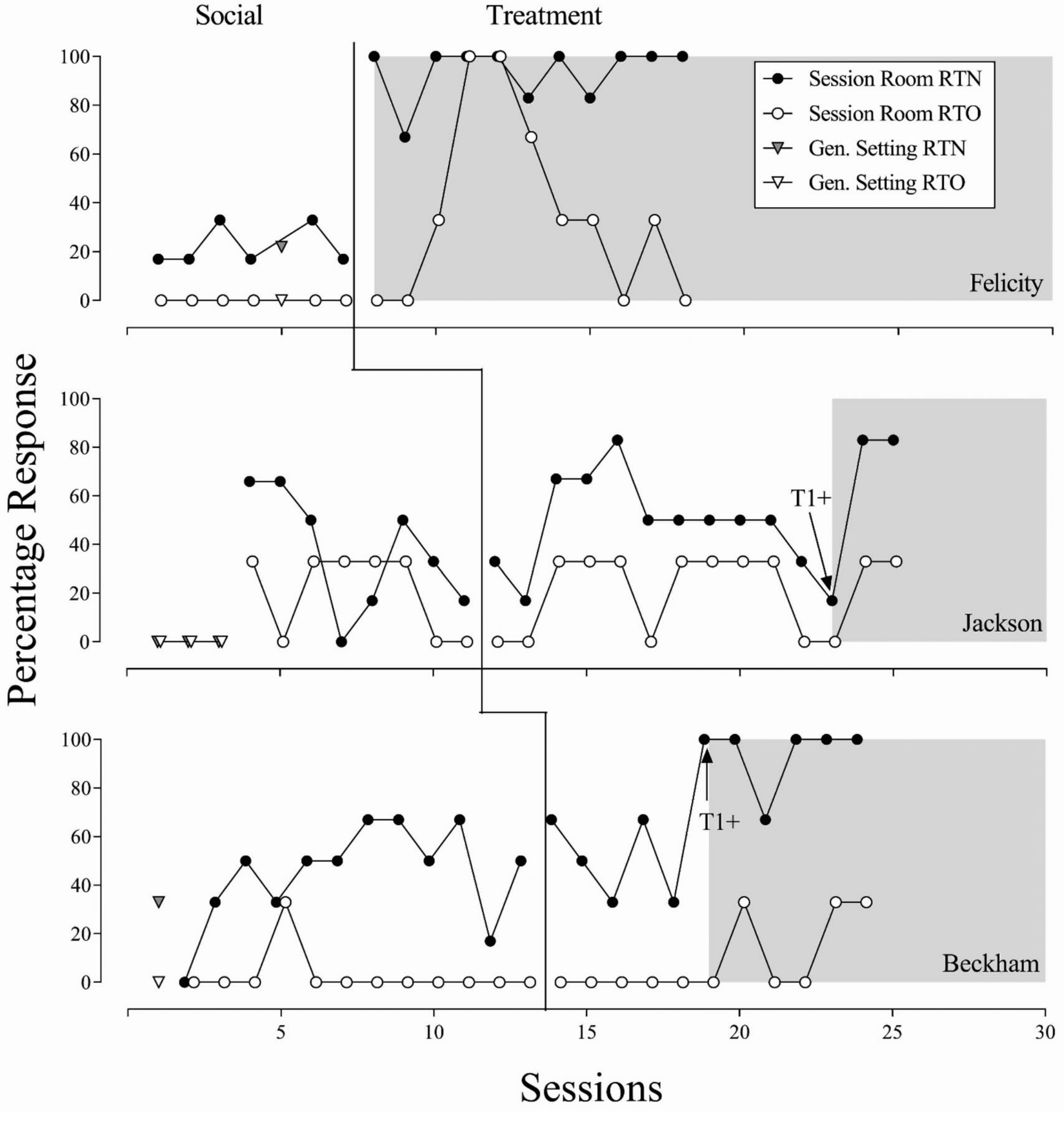
and RTO occurred, and Beckham met mastery criteria in session 24.

Figure 5 shows results for participants whose RTN was characterized as Type 4 during Phase 1. The second column from the right in Table 1 also indicates which prompts were used during treatment with these participants. Lucas (Panel 1), engaged in zero RTN and RTO during Phase 1 and during T1. Very minor increases in RTN and RTO were observed during T2, T2+, and T3. When T4 was introduced, RTN and RTO both increased in level and variability for the first four sessions (0-33%). RTN continued to increase while RTO decreased across the subsequent nine sessions, and Lucas met mastery criteria in session 36.

Lydia (Figure 5, Panel 2) engaged in RTN once during Phase 1. No RTN or RTO were observed during T1, and small but variable increases in both RTN and RTO were observed during T2. During T2+, discrimination increased (DI: 1.0), but percentage RTN remained below mastery. During T3, Lydia engaged in a pattern of responding that prompted a slight alteration in the definition of RTN. Lydia began "packing up" her distracter items (e.g., closing books, placing blocks back in a bucket), within 5 s of her name being presented, but did not engage in RTN within 5 s.

Figure 4

Phase 1 and Phase 2 Results for all Participants with Type 2 and Type 3 RTN

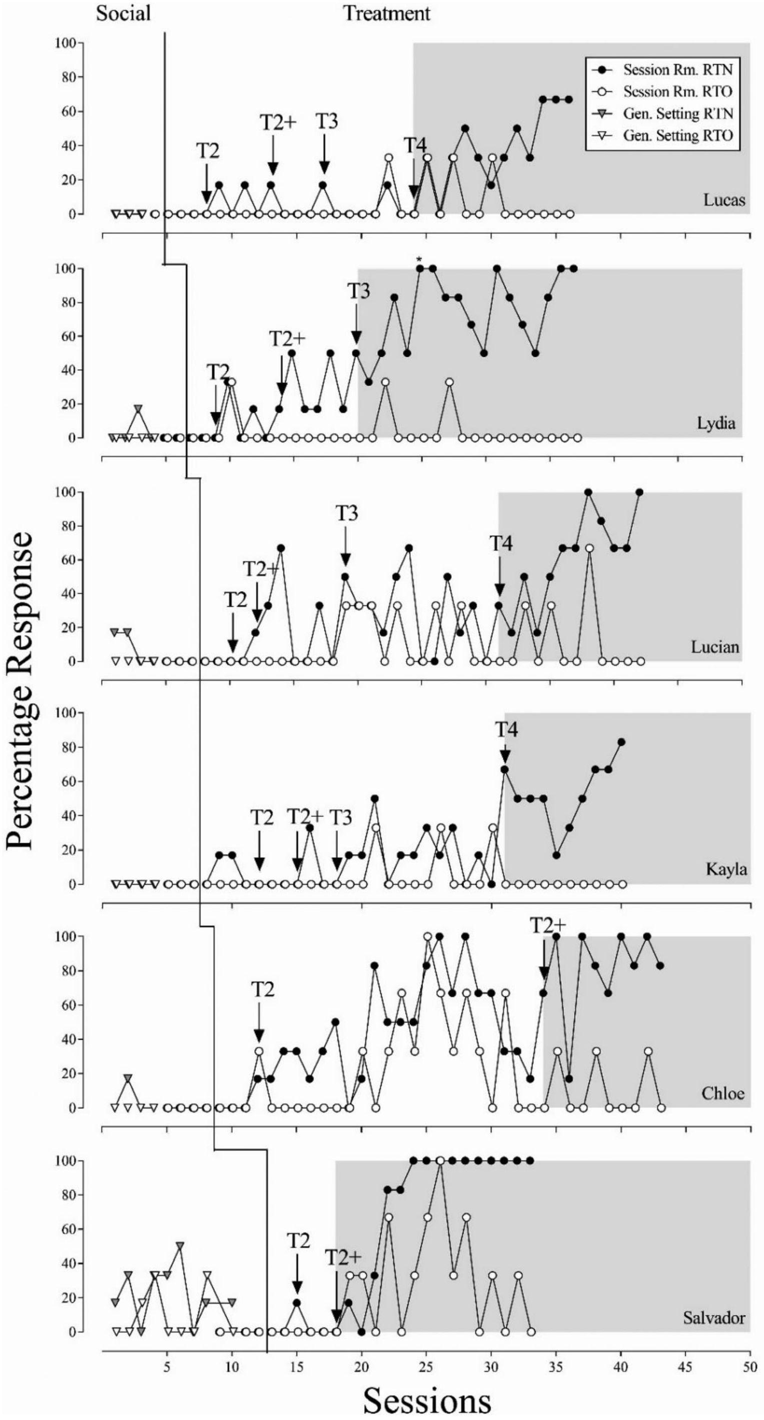


Note. Phase lines indicate the introduction of treatment (i.e., the start of Phase 2). Arrows indicate changes in treatment condition. Shaded backgrounds indicate sessions conducted in the treatment condition that was ultimately effective.

Thus, in session 25 (marked with an asterisk in Figure 5), RTN was scored as correct if it occurred within 5 s of this first "packing up" response, resulting in RTN being reinforced by the experimenter and scored as correct with latencies of up to 10 s, if a "packing up" response occurred first. Lydia's RTN gradually increased during T3, meeting mastery criteria in session 37 after a period of variability.

Lucian (Figure 5, Panel 3) engaged in RTN twice at the start of Phase 1, but zero RTN and RTO thereafter. RTN and RTO remained

Figure 5
Phase 1 and Phase 2 Results for Participants with Type 4 RTN



Note. Phase lines indicate the introduction of treatment (i.e., the start of Phase 2). Arrows indicate changes in treatment condition. Shaded backgrounds indicate sessions conducted in the treatment condition that was ultimately effective.

at zero during T1 and T2. Substantial but variable increases in RTN occurred in T2+. During T3, RTN and RTO both occurred at moderate levels, with high variability and low discrimination (DI: 0.64). RTN steadily increased during T4, and Lucian met mastery criteria in session 42. Kayla (Figure 5, Panel 4) did not engage in any RTN or RTO during Phase 1. During T1, Kayla only engaged in RTN on two occasions. RTN and RTO both returned to zero during T2 and remained near zero during T2+. RTN increased during T3 but remained below mastery criteria (DI: 0.67). When T4 was introduced, RTN immediately increased to 67%. After a period of variability, Kayla met mastery criteria in session 40. Chloe (Figure 5, Panel 5), only engaged in one instance of RTN during Phase 1, and RTN and RTO remained at zero during T1. During T2, RTN and RTO both increased and were highly variable, with moderate discrimination (DI: 0.64). During T2+, percentage RTN and DI both increased. Chloe met mastery criteria in session 43. Finally, for Salvador (Figure 5, Panel 6), RTN and RTO were moderate and undiscriminated during Phase 1 in the generalization setting. Salvador engaged in no RTN or RTO in the session room. Thus, Salvador's RTN was classified as Type 4 based on the Phase 1 session room data. RTN and RTO remained at zero during T1, and only one RTN occurred during T2. RTN and RTO both increased during T2+. RTO decreased after additional sessions, and RTN remained stable at 100%. Salvador met mastery criteria (DI: 0.92) in session 33.

The right-most column in Table 1 summarizes the treatment condition that was ultimately effective for increasing RTN to mastery criteria during Phase 2 for each participant (these data are not provided for Type 1 participants, for whom treatment was not conducted).

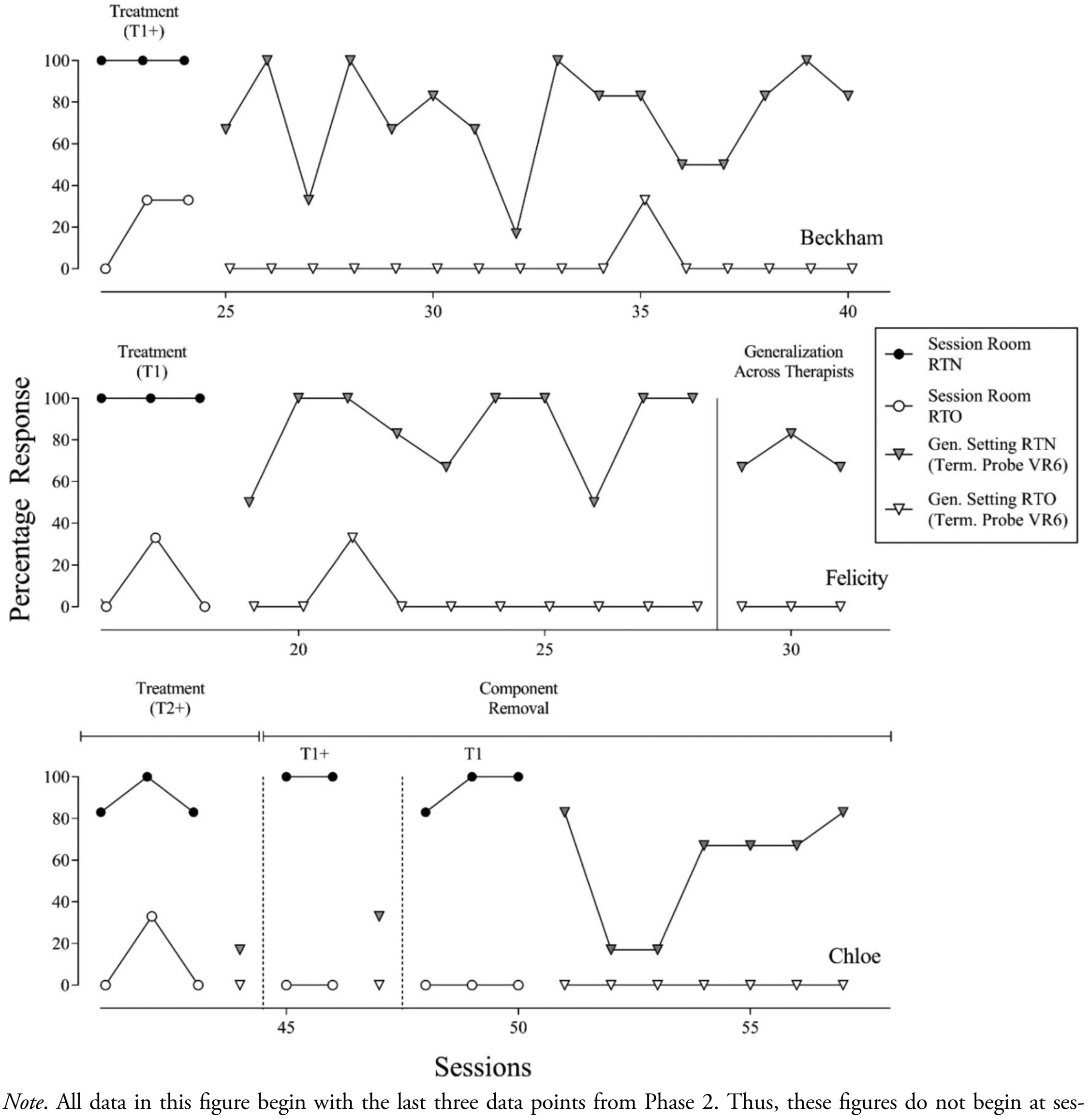
Phase 3 Results

Figures 6 and 7 show Phase 3 results for the six participants (Beckham, Felicity, Chloe, Lydia, Kayla, and Lucian) who completed Phase 3. Jackson did not participate in Phase 3 due to not completing Phase 2. Salvador and Lucas began Phase 3, but did not complete all planned steps of that phase before data collection for the study was ended (Table 1).

Figure 6 shows Phase 3 results for participants who mastered RTN during terminal probe sessions in Phase 3. For Beckham (top panel), RTN met mastery criteria during the first series of terminal probes, although RTN was variable across 16 total sessions before meeting mastery criteria. For Felicity (middle panel), RTN nearly met mastery criteria in terminal probes, but was below mastery criteria with only the novel experimenter assigned to the generalization setting (sessions 19, 23, and 26). Thus, we conducted the generalization-across-therapists subphase, and Felicity met mastery criteria after three of these sessions. For Chloe (bottom panel), RTN was mastered in terminal probes after some treatment components were removed in the session room. Chloe did not meet mastery criteria in the terminal probe conducted after Phase 2, nor in the probe conducted after the first step of component removal (T1+). However, Chloe met mastery criteria in the terminal probes that were conducted after the T1 step of component removal, after seven probe sessions.

Figure 7 shows Phase 3 results for Lucian, Kayla, and Lydia, for whom transfer of treatment to the generalization setting was necessary. For Lucian (top panel), the first terminal probe session after Phase 2 was unsuccessful, with 33% RTN and 100% RTO. Thus, component removal began with T3 in the session room. However, RTN dropped to zero in this condition, necessitating a return to T4. RTN and RTO were both variable during this return to T4, and mastery criteria were met again after 25 sessions. Anecdotally, these sessions were scheduled less frequently than during Phase 2, due to a change in staffing that resulted in Lucian being inconsistently scheduled for

Figure 6
Phase 3 Results in Which RTN Met Mastery Criteria During Terminal Probe Sessions

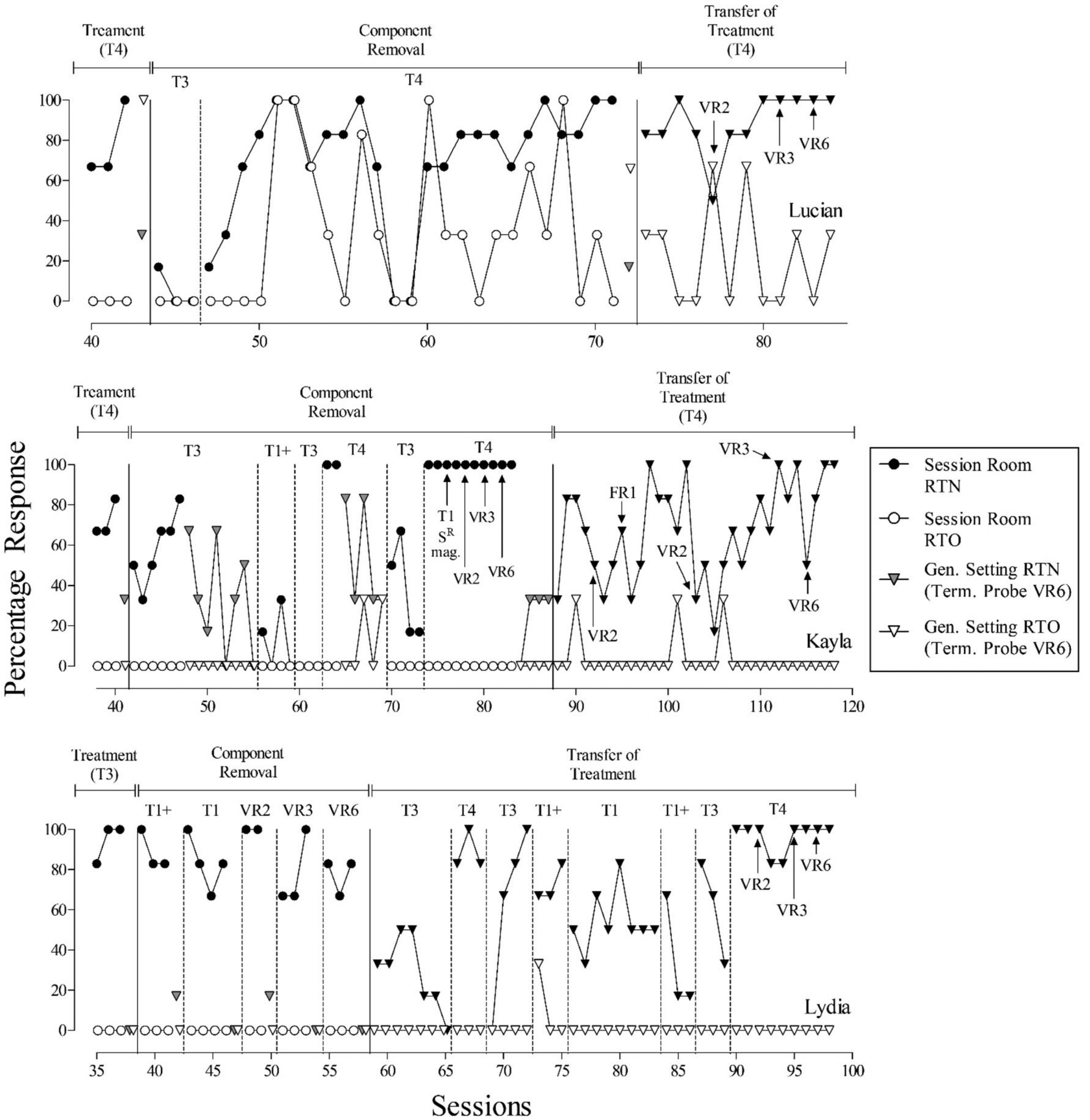


Note. All data in this figure begin with the last three data points from Phase 2. Thus, these figures do not begin at session number one, and some data overlap with Figures 4 and 5.

sessions with his primary experimenter. The terminal probe after T4 was unsuccessful, with RTO once again exceeding RTN. Thus, we determined that component removal (i.e., removing the interruption procedure of T4) was not likely to be successful for Lucian, and we began transfer of treatment to the generalization setting with T4. Lucian met mastery criteria after four T4 sessions in the generalization setting. We then thinned the schedule of tangible reinforcement to a VR6 schedule with T4 (i.e., interruption, prompts) in place, throughout which Lucian continued to meet mastery criteria.

Figure 7

Phase 3 Results in Which Transfer of Treatment Was Necessary



Note. All data in this figure begin with the last three data points from Phase 2. Thus, these figures do not begin at session number one, and some data overlap with Figure 5.

Kayla (Figure 7, middle panel) engaged in 33% RTN in the first terminal probe after Phase 2. Thus, component removal began with T3 in the session room. Kayla met mastery criteria after six T3 sessions. In the subsequent terminal probes, Kayla's RTN was highly variable and eventually dropped to zero. Component removal then progressed to T1+ in the session room, but RTN was very low (0-33%), necessitating a return to T3. However, RTN remained at zero upon this return to T3. Thus, we conducted another reversal to T4, during which Kayla immediately met mastery criteria. Kayla did not meet mastery criteria in the subsequent terminal probes, and therefore component removal progressed to T3 once again, but Kayla's RTN stabilized below mastery criteria. Thus, we conducted a final iteration of T4 and, after mastery criteria were met, thinned the schedule of tangible reinforcement to VR6 with T4 (i.e., interruption, prompts) in place, throughout which Kayla continued to meet mastery criteria. A final series of terminal probes was conducted, during which Kayla's RTN did not meet mastery criteria (0-33%). Thus, we began transfer of treatment in the generalization setting with T4. Kayla met mastery criteria after four T4 sessions in this setting, and reinforcement schedule thinning under T4 conditions was implemented. After some variability, including a reversal to an FR1 schedule, Kayla eventually met mastery criteria in the generalization setting under T4 conditions with a VR6 schedule of tangible reinforcement.

For Lydia (Figure 7, bottom panel), RTN continued to meet mastery criteria in the session room as all treatment components were removed, ending in T1 with a VR6 schedule of tangible reinforcement. However, Lydia engaged in little to no RTN in all terminal probes during component removal. Thus, transfer of treatment began with T3, the condition which was initially effective in Phase 2. RTN was highly discriminated during T3 in the generalization setting, but eventually dropped to zero. Thus, we implemented T4 in the generalization setting and RTN met mastery criteria in three sessions. We then began to remove treatment components in the generalization setting. RTN continued to meet mastery criteria during a second implementation of T3 and during T1 + in the generalization setting. However, RTN stabilized below mastery levels when T1 was conducted. Thus, a reversal to T1+ was conducted, but RTN stabilized below mastery levels during this reversal, and during a subsequent reversal to T3. Thus, T4 was implemented and RTN returned to mastery levels. Lydia's RTN continued to meet mastery criteria as the schedule of tangible reinforcement was thinned with T4 in place in the generalization setting. Lydia's RTN was ultimately maintained in the generalization setting under T4 conditions with a VR6 schedule of tangible reinforcement.

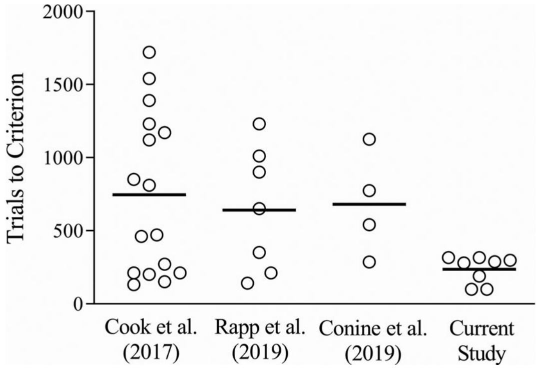
Discussion

Thirteen children with ASD were assessed for deficits in RTN. Four met mastery criteria during this assessment. For the remaining nine, RTN was increased to mastery criteria after participants underwent a series of treatment conditions that included the delivery of tangible reinforcement contingent on RTN. Schedules of tangible reinforcement were also thinned after treatment, and treatment effects were obtained across settings and across people.

Our primary goal was to develop a more efficient approach to the treatment of RTN than in previous research. Based on a comparison of trials-to-mastery, this goal generally appears to have been met. Figure 8 shows this comparison with respect to three previous studies (i.e., Conine et al., 2019; Cook et al., 2017; Rapp et al., 2019). Participants who did not meet mastery criteria (e.g., Jackson) in any study are not included in this figure. The mean number of trials to mastery was 746 (range, 130 to 1720) in Cook et al. (2017), 641 (range, 140 to 1230) in Rapp et al. (2019), 681 (range, 285 to 1125) in Conine et al. (2019), and 235 (range, 99 to 315) in the current study (note: comparing trials-to-criterion across these four studies required some computational decisions, which are described in the attached Supporting Information). It should be noted that S^{Δ} trials are included in the trials-to-mastery for the current study and for the Conine et al. study (Figure 8). Data from the

Figure 8

Trials to Criterion across Three Previous Studies and the Current Study



Note. Each data point depicts trials to criterion for an individual participant. Horizontal bars depict the mean trials to criterion in each study.

current study support the findings from Conine et al. that these trials are an important control procedure to include when teaching RTN. For example, Felicity (Figure 4) and Salvador (Figure 5) both engaged in 100% RTO when RTN initially increased to 100%. It was only with repeated exposure to reinforcement for RTN and extinction for RTO that these participants' responses became discriminated. Such patterns suggest that if discrimination is not explicitly programed by presenting stimuli other than the child's name, RTN may be acquired under faulty stimulus control.

The reduced trials-to-mastery in the current study may have resulted from the treatment conditions that we chose and the order in which we tested them. We based our six treatment conditions only on those components that were associated with some degree of success in Conine et al. (2019) (e.g., increasing reinforcer magnitude, interrupting access to competing reinforcers, differential reinforcement of unprompted RTN), and avoided treatment components that were not effective in that previous study (e.g., introducing new prompt-fading strategies). We also assessed preferences for both foods and toys before treatment, and made preferred social interactions available contingent on RTN. It was often necessary to change reinforcers before RTN was mastered in previous research (Conine et al., 2019; Cook et al., 2017; Rapp et al. 2019), and a thorough assessment of potential reinforcers at the start of the study may therefore have allowed us to avoid the need for such modifications with some participants in this study.

Despite this overall increase in treatment efficiency, the initial introduction of treatment (i.e., Phase 2) was not associated with an immediate increase in RTN for all participants. No singular treatment condition was effective with all participants; to the contrary, immediate treatment effects were only seen with one participant (Felicity, Figure 4). This is an important limitation of the current study and the treatment model used here. This limitation is reflected in the fact that the implicit multiple baseline design in Figures 4 and 5 does not show experimental control, and experimental control is only demonstrated by the multielement design arranged between RTN and RTO. To further improve the efficacy and efficiency of RTN intervention, it remains necessary to identify a means of introducing a treatment condition that produces immediate effects for most or all children with ASD.

To this end, we investigated whether the social condition could be used as a screening to predict the optimal treatment condition on an individual basis, and thereby further reduce total treatment trials in the future. The social condition was initially proposed in Conine et al. (2019) as a naturalistic condition that mirrors the real-world contingencies for RTN; however, it was not tested as such. If the social condition truly reflects natural contingencies for RTN, at least some children with ASD should engage in frequent and discriminated RTN during a baseline that includes those contingencies. This was the case for four children (Bethany, Daniel, Ned, and Vera) in the current study, whose RTN we characterized as Type 1. Data from these participants may lend some ecological validity to the use of the social condition as an initial baseline. This can be contrasted with no-consequence (Conine et al.) or contingent praise (Cook et al., 2017; Rapp et al., 2019) baselines from previous studies, which have shown downward or upward trends in baseline, respectively. Among the children who did not master RTN in baseline, we observed three other RTN patterns: low but discriminated (Type 2), undiscriminated (Type 3), and low to zero (Type 4). Some preliminary suggestions can be made regarding the treatments that were associated with mastery for each of these screening types.

For participants with Type 2 RTN (Beckham and Felicity), differential reinforcement without prompts was sufficient for RTN to reach mastery criteria. One way to interpret this finding may be that RTN was a *performance* deficit rather than a *skill* deficit for these children (e.g., Duhon et al., 2003; Noell et al., 2001). That is, discriminated RTN was in the behavioral repertoires of these children, but it did not occur frequently. Obtaining treatment effects without prompts represents a clinically important outcome. If prompts are never introduced, there is no risk of developing prompt dependency (e.g., Cividini-Motta & Ahearn, 2013; Karsten & Carr, 2009; Oppenheimer et al., 1993), nor do prompts need to be faded. Of note, increased reinforcer magnitude was needed for Beckham, but not for Felicity. Thus, it may be most efficient in future applications to use such large reinforcer magnitudes with children with Type 2 RTN from the outset of treatment if the overall goal is to produce treatment effects as quickly as possible. Of course, this larger magnitude constitutes another component to remove, and the relative merits of starting with either large or small reinforcer magnitudes may be worth investigation in future research and consideration by practicing behavior analysts.

In keeping with the conceptual framework mentioned above, Type 3 RTN may represent a skill deficit rather than a performance deficit; the target behavior (looking at the therapist) is in the repertoire but is not under appropriate stimulus control. However, we only identified one participant (Jackson) whose RTN was characterized as Type 3, and he left the study before finishing treatment. Thus, it is not possible to make suggestions regarding the most efficient treatment approach for children displaying this RTN pattern. Differential reinforcement at a high magnitude appeared to be initially effective for Jackson, but additional research is necessary.

For the six children (Chloe, Kayla, Lucas, Lucian, Lydia, and Salvador) whose RTN was characterized as Type 4, low-to-zero RTN and RTO made it difficult, even impossible, to calculate a DI in screening. Thus, it is also difficult to interpret whether the absence of RTN during screening may be thought of as a performance deficit or skill deficit for such children. Moreover, a variety of treatments were effective with these children: T2+ for two (Chloe and Salvador), T3 for one (Lydia), and T4 for three (Kayla, Lucas, and Lucian; Table 1).

All children with Type 4 RTN required, at a minimum, prompts and large reinforcer magnitudes before RTN was mastered. Thus, it may be advisable to start treatment with T2+ for children with Type 4 RTN in the future. Starting with T3 for Type 4 children may also be advisable, given that differential reinforcement of unprompted responses is often recommended for teaching new skills to children with ASD (Grow & LeBlanc, 2013; Vladescu & Kodak, 2010). However, starting with T4 may not align with ethical imperatives to implement the least-intrusive treatment necessary to obtain therapeutic outcomes (Behavior Analyst Certification Board, 2014), because T4 includes elements of blocking, which likely has aversive properties (e.g., Hagopian & Adelinis, 2001; Lerman & Iwata, 1996; Smith et al., 1999).

Given the variable treatment outcomes among this group of participants, future research is needed to further predict treatment effectiveness for Type 4 RTN.

Overall, the screening types identified in Phase 1 may be useful in future practice for bypassing treatment conditions that are likely to be ineffective for individual children, reducing the total trials to mastery. However, whether actual use of the screening in this predictive manner is effective requires future research. Sequence or practice effects could have played a role in the effectiveness of various treatment conditions in Phase 2. For example, it may be that T3 would not have been effective without prior exposure to T1 or T2. Moreover, using the screening to choose a treatment according to the above suggestions may entail implementing a treatment condition that was not evaluated with all participants of that screening type in the current study. For example, Felicity (Type 2) mastered RTN during T1; it is unclear whether starting with T1+ as recommended above would yield similar results. Future research is needed to evaluate practical use of the screening-to-treatment model proposed above.

The screening types observed in this study (Table 1) should not be taken as a prevalence estimate of these patterns of RTN among children with ASD. Most children (8 of 13) were recruited due to reported deficits in RTN. Thus, it should not be surprising that we obtained the greatest proportion of Type 4 screening outcomes, as these cases are likely to be the most salient to BCBAs and caregivers as needing intervention. Future studies could recruit all eligible children with ASD or include typically developing children as well, in order to characterize the prevalence of these four RTN patterns among children with and without ASD.

Six children completed Phase 3, which focused on extending the generality of treatment effects in a more efficient manner. We probed generalization across locations, people, and distance all at once, permitting fewer probe sessions to be conducted than in Conine et al. (2019). We also conducted terminal probes periodically throughout component removal and schedule thinning. This terminal probe strategy was highly efficient for some participants (Felicity, Beckham and Chloe), allowing us to bypass steps in component removal and in the transfer of treatment. However, this approach was less efficient with other participants (Kayla, Lucian, and Lydia), who initially required more intensive treatments in Phase 2. For these participants, we were unable to maintain mastery-level RTN when the interruption component of T4 was removed, although we were able to thin the schedule of tangible reinforcement under such conditions. This suggests that for these participants, the interruption procedure may have been a greater contributor to treatment effects than the tangible reinforcers. Future research should pursue additional strategies for removing this treatment component, or for achieving treatment effects without it.

We probed for generalization in Phase 3 using a VR6 schedule (i.e., one tangible reinforcer per session), given data from Conine et al. (2019) suggesting that maintenance in the absence of tangible reinforcers is unlikely to occur shortly after intervention (see Cook et al., 2017 and Rapp et al., 2019 for contrasting findings). However, Bethany and Beckham participated in Conine et al. and the current study, and both participants started the current study approximately 18 months after completing the previous one. RTN did not maintain at mastery levels in Conine et al. for either Bethany or Beckham when tangible reinforcers were completely withdrawn. For Beckham, RTN at the start of the current study occurred at similar percentages as during the social reversal in Conine et al. By contrast, Bethany mastered RTN in Phase 1 of this study, which marks an improvement from the social reversal conducted at the end of the prior study. Anecdotally, at the end of the prior study, Bethany and Beckham's treatment teams were asked to continue a maintenance treatment plan for RTN using a VR10 or VR6 schedule (for Bethany and Beckham, respectively), and to try thinning this schedule further over time. However, data were not collected on treatment fidelity after mastery or any further schedule thinning that was conducted with these participants by clinic staff in the gap between the two studies. Thus, future research is needed to evaluate variables and strategies that lead to long-term maintenance of RTN for some individuals (e.g., Bethany) and not for others (e.g., Beckham).

One limitation of the current study is that the terminal probes in Phase 3 tested for two changes at once: generalization that occurred as a result of treatment, and the introduction of a VR6 reinforcement schedule to the generalization setting. Treatment effects may have been possible with some participants if we had started treatment at this terminal goal: a VR6 reinforcement schedule in the generalization setting. Thus, one alternative approach for future research might be to start treatment in natural settings rather than a session room, at leaner schedules than FR1, or both. A related limitation is that Conine et al. (2019) suggested evaluating higher preferred social interactions (e.g., tickles, praise) as reinforcers for RTN before the introduction of tangible reinforcers. The current study took an alternative approach by starting treatment with a larger variety or magnitude of tangible reinforcers combined with the availability of social interactions, because the overall goal of this study was producing treatment effects as rapidly as possible (Figure 8). Tangible reinforcers were subsequently thinned during Phase 3 to produce a more socially valid treatment. An alternative approach that remains important to study is whether, for some children with ASD, treatment effects can be obtained by using highly preferred social interactions and not tangible items (Conine et al., 2019). To this end, a preference assessment that compares social, leisure, and edible reinforcers to one another at the outset of treatment may be useful (for discussion see: Conine & Vollmer, 2019, Morris & Vollmer, 2020). These divergent approaches (i.e., producing rapid treatment effects versus producing effects with the least intrusive intervention) are both important to take into consideration in future research on RTN.

Another major limitation of this study is that treatment effects were not extended to caregivers. Future studies should evaluate generalization to parents and other caregivers, conduct caregiver training, or consider training caregivers as the interventionists from the outset of treatment. Social validity measures were also not collected and remain important for future research. Various stakeholders (e.g., caregivers, teachers, peers) could be asked to evaluate the social importance and acceptability of the goals, procedures, and outcomes of RTN treatment (Wolf, 1978).

Finally, the broader social validity of interventions for RTN merits further research. For example, our mastery criteria were based on precedents established by previous studies, but the empirical data on which these criteria are based are currently limited (e.g., Beaulieu et al., 2012). This is true of both the percentage and the topography of RTN. As in previous studies, we targeted looking at the experimenter, but other topographies of RTN may be equally valid (e.g., saying "hi" but not looking at the speaker). Thus, future research could subject various mastery criteria for RTN to social validation by using formal surveys or by conducting descriptive analyses of RTN in children with and without ASD. Future studies could also collect data to test the notion that RTN is a developmental milestone and a behavioral cusp (e.g., Gerber et al., 2010; Miller et al., 2017). For example, future studies could evaluate whether increases in RTN are associated with increases in social skills, instruction-following, responses to safety instructions, or rates of learning in early intervention programs for children with ASD (e.g., Beaulieu et al., 2012; Kraus et al., 2012).

REFERENCES

Barbera, M. L. (2007). *The verbal behavior approach: How to teach children with autism and related disorders*. Jessica Kingsley.

Beaulieu, L., Hanley, G. P., & Roberson, A. A. (2012). Effects of responding to a name and group call on preschoolers' compliance. *Journal of Applied Behavior Analysis*, *45*(4), 687–707. https://doi.org/10.1901/jaba.2012.45-685

Behavior Analyst Certification Board. (2014). *Professional and ethical compliance code for behavior analysts*. Author.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The autism observation scale for infants: Scale development and reliability data. *Journal of Autism and Developmental Disorders*, *38*, 731–738. https://doi.org/10.1177/1362361308094500

Cividini-Motta, C., & Ahearn, W. H. (2013). Effects of two variations of differential reinforcement on prompt dependency. *Journal of Applied Behavior Analysis*, *46*(3), 640–650. https://doi.org/10.1002/jaba.67

Centers for Disease Control and Prevention (2015). Autism spectrum disorder (ASD): Signs and symptoms. https://www.cdc.gov/ncbddd/autism/signs.html

Chasson, G. S., Harris, G. E., & Neely, W. J. (2007). Cost comparison of early intensive behavioral intervention and special education for children with autism. *Journal of Child and Family Studies*, *16*, 401–413. https://doi.org/10.1007/s10826-006-9094-1

Conine, D. E., & Vollmer, T. R. (2019). Relative preferences for edible and leisure items in children with autism. *Journal of Applied Behavior Analysis*, *52*(2), 557–573. https://doi.org/10.1002/jaba.525

Conine, D. E., Vollmer, T. R., & Bolivar, H. A. (2019). Response to name in children with autism: Treatment, generalization, and maintenance. *Journal of Applied Behavior Analysis*, *53*(2), 744–766. https://doi.org/10.1002/jaba.635

Cook, J. L., Rapp, J. T., Mann, K. R., McHugh, C., Burji, C., & Nuta, R. (2017). A practitioner model for increasing eye contact in children with autism. *Behavior Modification*, *41*(3), 382–404. https://doi.org/10.1177/0145445516689323

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis*, *29*(4), 519–533. https://doi.org/10.1901/jaba.1996.29-519

Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2003). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review*, *33*(3), 429–443.

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment. *American Journal on Mental Retardation*, *101*(1), 15–25.

Gerber, R. J., Wilks, T., & Erdie-Lalena, C. (2010). Developmental milestones: Motor development. *Pediatrics in Review*, *31*(7), 267–277. https://doi.org/10.1542/pir.31-7-267

Grow, L., & LeBlanc, L. (2013). Teaching receptive language skills: Recommendations for instructors. *Behavior Analysis in Practice*, *6*(1), 56–75. https://doi.org/10.1007/BF03391791

Hagopian, L. P., & Adelinis, J. D. (2001). Response blocking with and without redirection for the treatment of pica. *Journal of Applied Behavior Analysis*, *34*(4), 527–530. https://doi.org/10.1901/jaba.2001.34-527

Hagopian, L. P., Rooker, G. W., & Zarcone, J. R. (2015). Delineating subtypes of self-injurious behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, *48*(3), 523–543. https://doi.org/10.1002/jaba.236

Hanley, G. P., Iwata, B. A., Lindberg, J. S., & Conners, J. (2003). Response-restriction analysis: I. Assessment of activity preferences. *Journal of Applied Behavior Analysis*, *36*(1), 47–58. https://doi.org/10.1901/jaba.2003.36-47

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.

Hart, B., & Risley, T. R. (1999). *The social world of children: Learning to talk*. Paul H. Brookes Publishing.

Jacobson, J. W., Mulick, J. A., & Green, G. (1998). Cost-benefit estimates for early intensive behavioral intervention for young children with autism – General model and single state case. *Behavioral Interventions*, *13*(4), 201–226. https://doi.org/10.1002/(SICI)1099-078X(199811)13:4<201::AID-BIN17>3.0.CO;2-R

Karsten, A. M., & Carr, J. E. (2009). The effects of differential reinforcement of unprompted responding on the skill acquisition of children with autism. *Journal of Applied Behavior Analysis*, *42*(2), 327–334. https://doi.org/10.1901/jaba.2009.42-327

Kraus, A., Hanley, G. P., Cesana, L., Eisenberg, D., & Jarvie, A. C. (2012). An evaluation of strengthening precursors to increase preschooler compliance. *Journal* *of Applied Behavior Analysis*, *45*(1), 131–136. https://doi.org/10.1901/jaba.2012.45-131

LeBlanc, L. A., Hagopian, L. P., Marhefka, J. M., & Wilke, A. E. (2001). Effects of therapist gender and type of attention on assessment and treatment of attention-maintained destructive behavior. *Behavioral Interventions*, *16*(1), 39–57. https://doi.org/10.1002/bin.73

Lerman, D. C., & Iwata, B. A. (1996). A methodology for distinguishing between extinction and punishment effects associated with response blocking. *Journal of Applied Behavior Analysis*, *29*(2), 231–233. https://doi.org/10.1901/jaba.1996.29-231

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule–generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Lord, C., Rutter, M., & Le Couteur, A. J. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Luczynski, K. C., & Hanley, G. P. (2014). How should periods without social interaction be scheduled? Children's preference for practical schedules of positive reinforcement. *Journal of Applied Behavior Analysis*, *47*(3), 500–522. https://doi.org/10.1002/jaba.140

Maurice, C., Green, G., & Luce, S. C. (Eds.). (1996). *Behavioral intervention for young children with autism: A manual for parents and professionals*. Pro-Ed.

Miller, M., Iosif, A., Hill, M., Young, G. S., Schwichtenberg, A. J., & Ozonoff, S. (2017). Response to name in infants developing autism spectrum disorder: A prospective study. *The Journal of Pediatrics*, *183*, 141–146. https://doi.org/10.1016/j.jpeds.2016.12.071

Morris, S. L., & Vollmer, T. R. (2020). A comparison of different methods for assessing preference for social interactions. *Journal of Applied Behavior Analysis*, *53*(2), 918–937. https://doi.org/10.1002/jaba.692

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatric and Adolescent Medicine*, *161*(4), 378–383. https://doi.org/10.1001/archpedi.161.4.378

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology*, *39*(4), 335–355. https://doi.org/10.1016/S0022-4405(01)00072-3

Oppenheimer, M., Saunders, R. R., & Spradlin, J. E. (1993). Investigating the generality of the delayed-prompt effect. *Research in Developmental Disabilities*, *14*(6), 425–444. https://doi.org/10.1016/0891-4222(93)90036-J

Querim, A. C., Iwata, B. A., Roscoe, E. M., Schlichenmeyer, K. J., Ortega, J. V., & Hurl, K. E. (2013). Functional analysis screening for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, *46*(1), 47–60. https://doi.org/10.1002/jaba.26

Rapp, J. T., Cook, J. L., Nuta, R., Balagot, C., Crouchman, K., Jenkins, C., … Watters-Wybrow, C. (2019). Further evaluation of a practitioner model for increasing eye contact in children with autism. *Behavior Modification*, *43*(3), 389–412. https://doi.org/10.1177/0145445518758595

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–144. https://doi.org/10.1023/A:1010738829569

Rogers, S. J., & Dawson, G. (2010). *The early start Denver model for young children with autism: Promoting language, learning, and engagement*. Guilford.

Shelov, S., & Altmann, T. R. (2009). *Caring for your baby and young child: Birth to age 5* (5th ed.). Bantam Books.

Slocum, S. K., Grauerholz-Fisher, E., Peters, K. P., & Vollmer, T. R. (2018). A multicomponent approach to thinning reinforcer delivery during noncontingent reinforcement schedules. *Journal of Applied Behavior Analysis*, *51*(1), 61–69. https://doi.org/10.1002/jaba.427

Smith, R. G., Russo, L., & Le, D. D. (1999). Distinguishing between extinction and punishment effects of response blocking: A replication. *Journal of Applied Behavior Analysis*, *32*(3), 367–370. https://doi.org/10.1901/jaba.1999.32-367

Social Security Administration (2018). *Popular baby names*. Retrieved from https://www.ssa.gov/oact/babynames/

Sundberg, M. L. (2008). *VB-MAPP: Verbal behavior milestones assessment and placement program*. AVB Press.

Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities*. Behavior Analysts.

Thomas, R. P., Wang, L. A. L., Guthrie, W., Cola, M., McCleery, J. P., Pandey, J., … Miller, J. S. (2019). What's in a name? A preliminary event-related potential study of response to name in preschool children with and without autism spectrum disorder. *PLoS One*, *14*(5), e0216051. https://doi.org/10.1371/journal.pone.0216051

Vladescu, J. C., & Kodak, T. (2010). A review of recent studies on differential reinforcement during skill acquisition in early intervention. *Journal of Applied*

Behavior Analysis, *43*(2), 351–355. https://doi.org/10.1901/jaba.2010.43-351

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*(2), 203–214. https://doi.org/10.1901/jaba.1978.11-203

Zhang, D., Roche, L., Bartl-Pokorny, K. D., Krieber, M., McLay, L., Bolte, S., … Marschik, P. B. (2018). Response to name and its value for the early detection of developmental disorders: Insights from autism spectrum disorder, Rett syndrome, and fragile X syndrome. A perspectives paper. *Research in Developmental Disabilities*, *82*, 95–108. https://doi.org/10.1016/j.ridd.2018.04.004

Received January 17, 2020
Final acceptance May 11, 2020
Action Editor, Bridget Taylor

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website.