Educ. Treat. Child.
https://doi.org/10.1007/s43494-022-00084-4

ORIGINAL ARTICLE

A Preliminary Evaluation of Conventional and Progressive Approaches of Discrete Trial Teaching for Teaching Tact Relations with Children Diagnosed with Autism

Christine M. Milne · Justin B. Leaf · Mary Jane Weiss · Julia L. Ferguson · Joseph H. Cihon · Matthew S. Lee · Ronald Leaf · John McEachin

Accepted: 1 September 2022

Abstract This study compared the effects of a conventional approach and a progressive approach to discrete trial teaching when teaching tact relations to 12 children diagnosed with autism spectrum disorder (ASD). The conventional approach was informed by best practice guidelines for teaching receptive and expressive language (Green *Focus on Autism and Other Developmental Disabilities, 16*(2), 72–85, 2001; Grow & Leblanc *Behavior Analysis in Practice, 6*(1), 56–75, 2013) and other relevant research (e.g., Graff & Karsten *Behavior Analysis in Practice, 5*(2), 37–48, 2012; León & Rosales *Journal of Behavioral Education, 27(*1), 81-100, 2018; Majdalany et al. *Journal of Applied Behavior Analysis, 47*(3), 657–662, 2014). The progressive approach was informed by guidelines outlined by Leaf, Cihon, Leaf, et al. *International Electronic Journal of Elementary Education, 9*(2), 361–372, (2016a) and Leaf, Leaf, McEachin, et al. *Journal of Autism & Developmental Disorders, 46*(2), 720–731, (2016c). In this study, highly vocal-verbal participants diagnosed with ASD were randomly assigned to the conventional or progressive condition and received 20 sessions of intervention. Following intervention, statistical analysis was used to evaluate and compare the effects of each condition. Although both were effective in teaching the participants tact relations, there was a statistically significant difference in the number of tact relations learned in the progressive condition when compared to the conventional condition.

Keywords tacts · expressive labels · discrete trial teaching · progressive · conventional

Discrete trial teaching (DTT) is a teaching methodology that has been used to teach a variety of skills (e.g., language, social, academic) for individuals diagnosed with autism spectrum disorder (ASD; Smith, 2001). DTT provides a structured approach to teaching that typically involves concentrated practice through an increased number of trials to maximize the number of learning opportunities for a targeted skill (Eikeseth et al., 2014; Smith, 2001). DTT involves four main components: (1) an interventionist-delivered

The first and second authors conceptualized the original paper and wrote the original draft. All other authors made major contributions and changes throughout the process. The first, second, fourth, fifth, sixth, and seventh author have commercial products available about behavioral intervention, progressive ABA, and autism.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43494-022-00084-4.

C. M. Milne · J. B. Leaf · M. J. Weiss · J. H. Cihon
Endicott College, Beverly, MA, USA

C. M. Milne · J. B. Leaf (✉) · J. L. Ferguson · J. H. Cihon · M. S. Lee · R. Leaf · J. McEachin
Autism Partnership Foundation, Justin B. Leaf., 200 Marina Drive, Seal Beach, CA 90740, USA
e-mail: Jblautpar@aol.com

Published online: 07 October 2022

instruction; (2) a response from the learner; (3) a consequence provided by the interventionist; and (4) an intertrial interval. An optional fifth component, a prompt, may occur simultaneously with the interventionist-delivered instruction, or between the delivered instruction and the response from the learner, to increase the likelihood of an accurate response. Given the opportunity for variation across these many components, it is not surprising that varying approaches to the implementation DTT exist (e.g., Grow & LeBlanc, 2013; Leaf, Cihon, Leaf, et al., 2016a).

The differing approaches to the implementation of DTT have been conceptualized across a continuum (Leaf, Cihon, Leaf, et al., 2016a). In particular, Leaf, Cihon, Leaf, et al. (2016a) termed the two ends of the spectrum as *conventional* and *progressive*. Although these terms may be disputed among practicing behavior analysts, we use these two terms to facilitate a discussion of the differing approaches to the implementation of DTT. As we begin to recognize and better understand the differences between these two approaches, we may better recognize the advantages and disadvantages to each approach, as well as potential differing outcomes.

A Conventional Approach to DTT

A primary difference between conventional and progressive approaches to DTT is the type of structure and extent of variations used within the protocols of each procedure. More specific differences can be seen across broad categories such as reinforcer identification, reinforcers used, rate of reinforcement, instructions used, trial order, prompting methods, feedback used, use of previously acquired skills, and data collection (see Table 1). Within a conventional approach, the main source of control for the interventionist's

Table 1 Components of a Conventional and Progressive Approach to DTT

Component	Conventional (C)		Progressive (P)		Main difference between approaches
	General Recommendations	Selected for this study	General Recommendations	Selected for this study	
Reinforcer Identification	Formal preference assessment	Paired Stimulus Preference Assessment	No formal preference assessment	In-the-moment Reinforcement Analysis (IMRA)	C–Formal P–Not Formal
Reinforcer Modality	Based on formal preference assessment	Praise, tokens, tangibles, edibles	Varied	Praise, tokens, tangibles/edibles, socials	C–Limited P–Not limited
Rate of Reinforcement	FR1	FR1	Varied	Interventionist Assessment	C–Fixed rate P–Variable rate
Instruction Type	Same	Same (e.g., "Who is it?")	Varied	Varied	C–No Variation P–Variation
Instruction Complexity	Simple	Simple (e.g., "Who")	Simple to Complex	Simple to Complex (e.g., "Who does this one look like?")	C–Simple P–Varied
Trial Order	Counterbalanced	Counterbalanced	Varied	Interventionist Assessment	C–Predetermined P-Not predetermined
Prompting Method	Errorless learning	Progressive Prompt Delay with Model	Flexible prompting	Flexible Prompt Fading (FPF)	C–Fixed type P–Varied types
Feedback	Error Correction	Error Correction plus retrial with echoic prompt (0s delay)	Varied	Varied (e.g., instructive, error correction, no feedback)	C–Fixed type P–Varied types
Maintenance	Interspersed or 2 trials per target prior to teaching	2 trials per target prior to teaching	Varied	Interventionist Determined	C- Predetermined P–Not determined
Data Collection	Trial by trial	Trial by Trial	On a spectrum	Estimation	C–More P–Less

behavior, or what informs how the interventionist may respond, is commonly a protocol. The interventionist adheres to what is written in the protocol (e.g., instruction to use, when to prompt, what to target) and there is little variation from the protocol regardless of other variables (e.g., learner affect, environmental context; Leaf, Cihon, Leaf, et al., 2016a). The underlying rationale for this approach is to help ensure the development of responses under the desired stimulus conditions, rather than learning to respond to erroneous stimuli (Green, 2001; Grow & Leblanc, 2013). A conventional approach to DTT may be beneficial, especially with newer interventionists. That is, the use of a protocol outlining how the interventionist should behave could help ensure the interventionist refrains of unwanted instructional behavior leading to undesired stimulus control (Grow & LeBlanc, 2013). However, evaluations of the components of a conventional approach to DTT have found that this approach may be less efficient than a progressive approach to DTT (Cihon et al., 2020; Ferguson et al., 2020; Leaf et al., 2015; Leaf et al., 2019; Leaf, Cihon, Ferguson, et al., 2018a; Leaf, Dale, et al., 2014a; Leaf, Leaf, et al., 2014b; Leaf, Leaf, Alcalay, et al., 2016b; Leaf, Leaf, Leaf, et al., 2018b; Soluaga et al., 2008; Taubman et al., 2013; Wong et al., 2020).

A conventional approach to DTT typically involves the interventionist conducting a formal preference assessment to identify reinforcer(s) prior to teaching. Formal preferences assessments may occur monthly or up to several times a day to determine items to be used as reinforcers throughout intervention (Graff & Karsten, 2012; Love et al., 2009). A commonly used formal preference assessment is the paired stimulus (PS) preference assessment (Graff & Karsten, 2012). The stimuli commonly selected for use as potential reinforcers include praise, edibles, tokens, and other tangible items (Graff & Karsten, 2012). Related to this, within a conventional approach, the interventionist commonly adheres to a fixed ratio-1 schedule of reinforcement (each correct response results in the delivery of a reinforcer), especially when in the early stages of skill acquisition (Green, 2001).

A conventional approach to DTT also involves instructing interventionists on instruction form and trial design. The interventionist typically provides the same, simplest instruction across all trials (e.g., "What is it?" not "What is the name of this?"; Green, 2001; Grow & LeBlanc, 2013). When teaching receptive or expressive language, the interventionist counterbalances the location of stimuli by manipulating the placement of stimuli in the stimulus array and the order of trials (Green, 2001; Grow & LeBlanc, 2013). Each stimulus, target and nontarget, occurs in each position in the stimulus array an equal number of times across trials. In addition, each target stimulus is presented on an equal number of trials across a trial block, with no stimulus targeted on more than two consecutive trials. Finally, recommendations regarding maintenance of mastered targets include interspersing mastered targets with those in acquisition (Koegel & Koegel, 1986) and conducting maintenance trials prior to intervention for skills in acquisition (Sam & AFIRM team, 2016).

Attempts to prevent (i.e., errorless prompting strategies) or respond to errors also differ between a conventional and progressive approach to DTT. The interventionist typically adheres to a prompting system (e.g., most-to-least prompting) that prescribes when to prompt, when to fade prompts, and what prompt type to provide. Echoic model prompts within a progressive prompt delay system are the most common within a conventional approach when implementing tact training (e.g., Kodak et al., 2012). This is likely due to suggestions for the use of errorless teaching within this approach (Green, 2001; Grow & LeBlanc, 2013).

A final variable that distinguishes a conventional approach to DTT from a progressive approach is the way in which the interventionist collects data. The use of continuous measurement (e.g., trial-by-trial) is often recommended and implemented within a conventional approach (Cummings & Carr, 2009; Lerman et al., 2011; Love et al., 2009) in which learner responding and prompt levels are recorded on each trial. This "approach allows a comprehensive, ongoing account of the learner's performance across all programmed learning opportunities" (Cummings & Carr, 2009, p. 57).

A Progressive Approach to DTT

On the other end of the continuum, a progressive approach to DTT involves interventionists using several strategies and variations of those strategies within the same teaching session to accommodate ongoing contextual changes (e.g., environmental

changes, learner motivation, learner responding). Within a progressive approach to DTT, the main sources of control for the interventionist's behavior are the learner's behavior and other contextual variables (Leaf, Cihon, Leaf, et al., 2016a; Leaf, Leaf, & McEachin, 2018c; Leaf, Leaf, McEachin, et al., 2016c). Interventionists have the flexibility to make changes based on in-the-moment assessment, or clinical judgement. Commonly assessed variables include, but are not limited to, learner motivation, learner attentiveness, learner health, nonverbal behavior, changing function(s) of behavior, and recent and past learner performance (Leaf, Cihon, Leaf, et al., 2016a; Leaf, Leaf, & McEachin, 2018c; Leaf, Leaf, McEachin, et al., 2016c). A progressive approach to DTT may allow for more individualized and immediate changes to intervention resulting in more effective and efficient teaching. This approach likely requires a highly skilled interventionist, as it requires constant analysis (Leaf, Leaf, McEachin, et al., 2016c). In addition, the flexibility and responsiveness of a progressive approach to DTT creates difficultly for replication efforts within research.

With respect to the provision of reinforcement, a progressive approach to DTT differs from a conventional approach in several ways. First, to identify potential reinforcers, a suggested guideline is the use of in-the-moment reinforcement analysis (IMRA; Leaf, Cihon, Leaf, et al., 2016a). When using IMRA, the interventionist does not conduct a formal preference assessment prior to sessions, but rather identifies potential reinforcers through in-the-moment assessment of several different variables related to reinforcement (e.g., learner age, verbal and nonvocal behavior, change in behavior). Second, although a progressive approach to DTT uses similar stimuli as potential reinforcers as a conventional approach, there is a strong emphasis on the use and development of social interactions as reinforcers, reducing the use of edible reinforcers, and expanding stimuli that function as reinforcers through conditioning (Leaf et al., 2012; Leaf, Cihon, Leaf, et al., 2016a). Finally, the schedule of reinforcement is determined based on in-the-moment assessment of a variety of relevant contextual variables (e.g., current and past responding, rate of responding; Leaf, Cihon, Leaf, et al., 2016a). This typically results in the use of a flexible earning requirement (Cihon, Ferguson, Leaf, et al., 2019a) in which the interventionist can more closely match consequences to learner performance. That is, the interventionist has more flexibility to provide reinforcement differentially based on the learner's current and past responding.

There are also differences with respect to the interventionist's instructions and trial design within a progressive approach to DTT. For instance, the interventionist is instructed to assess learner responding and use the outcomes to adjust their instructions to what may be most effective for the learner during each individual session. The overall aim is to use natural language and vary instructions when possible (e.g., "Who is this?," "What do we call him?," "His name would be?"). The interventionist is also instructed to select the target as well as the placement of the target based on learner responding, not a priori based on a protocol (Leaf, Cihon, Leaf, et al., 2016a). Variables assessed to determine stimulus location include location biases, accuracy of responding, and attending. As the learner responds to a specific target correctly over consecutive trials, the interventionist may intersperse new targets immediately to maximize rate of acquisition and shift targets into maintenance. Finally, although the progressive approach does not provide a direct recommendation for maintenance targets, the recommendation of analyzing relevant factors in-the-moment could be extended to the maintenance of mastered skills (Leaf, Cihon, Leaf, et al., 2016a). Some of these relevant factors might include attending, accuracy of responding, time between maintenance trials, and number of other targets simultaneously in acquisition.

The methods designed to attempt to prevent or respond to errors during progressive DTT also differ from a conventional approach to DTT. Within a progressive approach to DTT, the interventionist typically implements flexible prompt fading (FPF; Soluaga et al., 2008). FPF involves the interventionist determining when to prompt, when to fade prompts, and what type of prompts to provide based upon various environmental variables (e.g., responses on previous trials, attending, aberrant behavior) on a trial-by-trial basis rather than adhering to a formal prompting protocol. Furthermore, the interventionist has the discretion to allow errors to happen and uses a variety of error correction procedures (Leaf, Cihon, Leaf, et al., 2016a).

Within a progressive approach, it is recommended that interventionists collect data in a way that

corresponds to the level of detail needed to make programmatic decisions (Leaf, Cihon, Leaf, et al., 2016a; Leaf, Leaf, McEachin, et al., 2016c). There are times when a high level of precision is needed (e.g., trial-by-trial data collection) and times in which discontinuous or estimation data collection yields adequate precision and predictiveness of mastery (Ferguson et al., 2020). As a result, it is common for estimation data to occur within a progressive approach to DTT.

Component Comparisons

The differences across the components of a progressive and conventional approach have been highlighted throughout the literature (Green, 2001; Grow & Leblanc, 2013; Leaf, Cihon, Leaf, et al., 2016a; Leaf, Leaf, McEachin, et al., 2016c). There have also been several studies evaluating and comparing these components. For example, Leaf et al. (2015) compared PS to IMRA for reinforcer identification to assess the effects on rate during a sorting task and Leaf, Leaf, Leaf, et al. (2018b) compared PS to IMRA to assess learner acquisition during tact relation task. In both studies, one interventionist was assigned to the PS condition and a second was assigned to the IMRA condition, both of whom were kept blind with respect to the other condition. The interventionist assigned to the PS condition ran a PS preference assessment to identify the top three stimuli to be used as reinforcers. The interventionist assigned to the IMRA condition did not conduct a formal preference assessment but had access to all 10 stimuli evaluated in the PS preference assessment. The results across both studies showed that the PS preference assessment and IMRA were effective, but IMRA was more efficient with respect to sessions to mastery.

To date, there has been one study comparing conventional and progressive approaches when it comes to counterbalancing of stimulus placement (i.e., Leaf, Cihon, Ferguson, et al., 2018a) and counterbalancing trial order (i.e., Wong et al., 2020). Leaf, Cihon, Ferguson, et al. (2018a) evaluated three methods of stimulus rotation to teach listener behavior (i.e., receptive labels) for five children diagnosed with ASD. The counterbalanced condition consisted of the interventionist rotating stimuli based on Grow and LeBlanc's (2013) recommendations. The fixed condition involved the interventionist never rotating the stimuli in the array, and the in-the-moment assessment of placement condition involved the interventionist rotating the stimuli based on in-the-moment assessment of participant responding. The results were idiosyncratic across participants and provided preliminary evidence that counterbalancing may not be the best practice for every learner or context. Wong et al. (2020) compared three approaches to target trial order when teaching listener behavior for three children diagnosed with ASD. The stimulus array was counterbalanced based on Grow and Leblanc's recommendation across all three conditions. In the predetermined condition, the target and nontarget stimuli were equally distributed across a trial block and each stimulus was targeted three times each in a counterbalanced order. The constrained condition consisted of targeting each stimulus a total of three times without a predetermined or counterbalanced order. In the unconstrained condition, the interventionist used in-the-moment assessment to determine the order and number of times a stimulus was targeted. The results indicated that all three approaches were effective, but the unconstrained condition was typically more efficient with respect to the number of teaching sessions to reach mastery criteria.

Several studies have compared FPF to conventional prompting strategies (e.g., Cihon et al., 2020; Leaf et al., 2019; Leaf, Leaf, et al., 2014b; Leaf, Leaf, Alcalay, et al., 2016b; Soluaga et al., 2008). In general, the results of these studies have found FPF to be as effective, or more effective than other prompting strategies when used within a progressive approach to DTT. In a more recent study, Cihon et al. (2020) compared constant time delay, most-to-least, and FPF using a randomized clinical trial when teaching tact relations to individuals diagnosed with ASD. The results indicated that all three prompting strategies were effective; however, participants in the FPF condition generally reached the mastery criterion in fewer sessions compared to the other two prompting strategies.

Suggestions for data collection within a progressive approach to DTT involve collecting data on a sliding scale (Leaf, Cihon, Leaf, et al., 2016a; Leaf, Leaf, McEachin, et al., 2016c). That is, select a data collection procedure that provides data that will inform changes, if any, to the teaching procedures and that does not interfere with teaching. This often involves the option to use estimation data involving the use of

a Likert scale that correlates with learner performance. There have been mixed results in terms of the accuracy of estimation data (e.g., Taubman et al., 2013), but recent research has found estimation data to be just as accurate with respect to predicting mastery as trial-by-trial data. In particular, Ferguson et al. (2020) compared trial-by-trial data collection to estimation data collection when using DTT to teach three children diagnosed with ASD tact relations. Both data collection procedures were examined for accuracy, efficiency (i.e., number of trials delivered per session), and rate of acquisition. The results of an adapted alternating treatment design replicated across participants and targets indicated that estimation data collection was as accurate as trial-by-trial data collection in determining mastery. Furthermore, estimation data was also accurate when compared to post-hoc trial-by-trial data.

When examining the literature, it seems clear there have been several studies comparing components of conventional and progressive approaches to DTT. Although many of these studies found differences across a wide variety of variables, it is possible that the overall effects of these approaches may differ when the components are combined. To date, the authors are unaware of any studies that have done so. Therefore, the purpose of this study was to compare a conventional to a progressive approach to DTT when teaching tact relations for 12 children diagnosed with ASD.

Method

Participants

Twelve children diagnosed with ASD participated in the study. All participants resided in the western United States and were recruited from the same agency in which the study was conducted. Participant ages ranged from 3 to 9 years. The following assessments were conducted with each participant within 14 months prior to the study: IQ assessment (i.e., Wechsler Intelligence Scale for Children-5th edition, Wechsler Preschool and Primary Scale of Intelligence-4th edition), Vineland Adaptive Behavior Scales-3rd edition (Vineland-3; Sparrow et al., 2016), Peabody Picture Vocabulary Test-4th edition (PPVT; Dunn & Dunn, 2007), and the Expressive One-Word Picture Vocabulary Test-4th edition (EOWPVT-4; Martin & Brownell, 2011). The grouped and averaged results of these assessments (see Table 2) revealed that all participants had moderate to high language skills (e.g., could independently communicate their wants and needs, conversational skills) and had average to elevated levels of maladaptive behavior based on the results of the Vineland-3 assessment. Although some participants had elevated levels of maladaptive behavior, the rates of challenging behavior during research sessions were low and did not interfere with learning.

The participants were randomly divided into the conventional or progressive group using Random.org. A *t*-test conducted prior to intervention across all assessments resulted in no statistically significant differences across the two treatment groups. No a priori power analysis was conducted to determine group size.

The diagnosis and intervention history are displayed in Table 3. The participants in this study were primarily white, Asian American, or a combination of the two. The age in which participants received their diagnosis of ASD ranged from 1 year, 7 months to 4 years, 4 months. The age in which ABA-based services began for participants ranged from 1 year, 3

Table 2 Participant Ages and Assessment Scores across Groups

Demographic Category	Conventional Group Average	Progressive Group Average	*p* Value
Age (Months)	76.7 (Range: 42–111)	77 (Range: 49–100)	0.978
IQ	96 (Range: 72–113)	105 (Range: 89–122)	0.344
Vineland-3	75 (Range: 61–84)	75.5 (Range: 60–85)	0.924
PPVT-4	97.2 (Range: 75–114)	103.5 (Range: 77–119)	0.483
EOWPVT-4	103.8 (Range: 84–119)	109.2 (Range: 83–127)	0.528

Note. *P*-values are provided to indicate any statistically significant differences between the two groups on each demographic category prior to intervention. *P*-values greater than .05 indicate there was no statistically significant difference. Scores on IQ, Vineland-3, PPVT-4, and EOWPVT-4 are reported as standard scores. Standard scores have an average (mean) of 100. Typically scores falling between 90 to 110 are considered to be average for a child's age.

months to 3 years, 6 months. For participants C1, C3, and P1, ABA-based services started prior to a formal diagnosis because they were considered at risk prior to the age of 3. Some participants (C1, C2, C6, P1, P2, P4, P5) started ABA-based services prior receiving services at their current clinic, whereas some (C3, C4, C5, P3, P6) had received ABA-intervention solely from their current clinic. The study occurred during the school year transition into summer, so the hours of behaviorally based intervention the participants received throughout the study varied. As a result, the clinic and school-based hours provided for each participant is specified using a range. Some participants received additional services (e.g., occupational

Table 3 Participant Demographic Information

Participant	Age at Start of Study	Race/ Ethnicity	Age of Diagnosis	Started Services	ABA Clinic Based Hours Per Week	ABA School Based Hours Per Week	Other Services	Previous History with Interventionist	Research Settings
C1	7 years	Asian American and White	3 years, 2 months	2 years, 9 months	0	0-32.5	OT, SP	No history	School, Clinic
C2	5 years	White	2 years, 7 months	2 years, 8 months	0-4	0-32.5	OT	No history	School, Clinic, Home
C3	6 years	Asian American	2 years, 3 months	2 years, 3 months	4-21.5	3.5-32.5	OT, SP	Research studies	School, Clinic
C4	3 years	Asian American	1 year, 7 months	1 year, 3 months	28	4.5	None	No history	Clinic
C5	9 years	White	4 years, 4 months	3 years	19-35.5	0-20.25	OT, SP	No history	Clinic
C6	5 years	White	1 years, 6 months	2 years, 4 months	0-13.5	0-5	None	No history	School, Clinic
P1	7 years	Asian American and White	3 years, 2 months	2 years, 9 months	0	0-32.5	OT, SP	No history	School, Clinic
P2	6 years	White	1 years, 8 months	1 years, 8 months	0-4	30-32.5	OT	<1 month of supervision Research studies	School, Clinic
P3	6 years	White	3 years, 1 months	3 years, 5 months	4-16	19.-32.5	NS	Research studies	School, Clinic
P4	5 years	White	3 years, 4 months	3 years, 6 months	4-22.75	4.5-30	OT, SP	Research studies	School, Clinic
P5	4 years	Asian American and White	1 year, 2 months	1 year, 10 months	14-26	6.75-8	None	No history	Clinic
P6	8 years	Asian American and White	2 years, 6 months	2 years, 6 months	8	0	OT, SP	No history	Clinic, Home

Note. The study took place when participants were transitioning from the end of the school year into the summer break, therefore a range of ABA service (clinic and school based) hours are provided. OT = Occupational Therapy; SP = Speech Therapy; NS = Not Specified

therapy, speech therapy) at some point, but not necessarily during the time of the study. Those services are also listed in Table 3.

The participants in each group were paired with a participant assigned to the other group based upon their availability for sessions each week. For example, if a participant from the conventional group was available for three sessions a week, they were paired with a participant from the progressive group who also was available for three sessions per week. If a participant was absent for a session, their paired participant also did not have a session. This was done to control for long absences and fluctuating durations between sessions so a fair comparison of the two teaching conditions could occur. For the ease of discussion, participants assigned to the progressive condition are identified as P1, P2, P3, P4, P5, and P6, whereas the paired participant from the conventional condition are identified as C1, C2, C3, C4, C5 and C6. Participants with matching numbers were the ones paired together (e.g., P1 was paired with C1).

Interventionist

The first author served as the interventionist for all sessions across conditions. She had 9.5 years of experience providing behavior-analytic intervention for individuals diagnosed with ASD. In particular, she had 9.5 years of experience implementing a progressive approach to DTT and 4 years of experience implementing a conventional approach to DTT. In addition, the interventionist had a master's degree in applied behavior analysis and was a board certified behavior analyst. Although she worked in the same clinic as the participants, she did not provide any intervention with most of the participants (i.e., C1, C3, C4, C5, C6, P1, P5, P6). She had a previous history with 17% of participants in the conventional group and 50% of participants in the progressive group. For one of the participants (i.e., P4), she provided case supervision for less than 1 month, ending the supervisory relationship 15 months prior to the start of the study. For some participants (i.e., C2, P2, P3, & P4), she had a history of running research sessions unrelated to this study.

Setting

Sessions occurred in one of three settings based on the participants' therapeutic schedules and the interventionist's schedule (see Table 3). First, all participants had some sessions occur in a research room located in a private clinic that provides behavior-analytic intervention for individuals diagnosed with ASD. The research room contained child-sized tables, child-sized chairs, adult desks and chairs, and couches. Other researchers may have been present in the research room, though they were preoccupied with other tasks. Second, sessions took place in a school classroom for seven of the participants. The classroom contained multiple tables for group work, individual tables for students, materials for learning, and a play area with shelves filled with a variety of toys and activities. Teaching sessions took place at a table in the back of the classroom, with bookshelves used to divide the area where research sessions took place from the rest of the classroom. Other students were sometimes present in the classroom during intervention, though, if present, they remained away from the intervention area. However, no other children were present for pre- and post-assessment probes. Finally, some sessions were conducted in the homes of two participants. Sessions in the home typically occurred in the participant's room, a playroom, or the kitchen. In each room, various play materials and intervention materials were present and family members remained in different parts of the home. Teaching occurred on the floor (one session) or at a table (all other sessions) in the home setting. Across all settings, teaching occurred in a one-to-one instructional format.

Target Stimuli

The interventionist selected 100 pictures of cartoon characters (contact first author for a complete list) to potentially teach during the intervention sessions. One hundred pictures were selected because it was perceived as a number that could not be taught during the limited number of sessions, therefore decreasing the likelihood of a ceiling effect. By having more than enough targets to teach, the researchers aimed to obtain better outcome measures of the differences between the two approaches, if any were present. The target stimuli were characters from various television shows, movies, and video games. The researchers selected less popular cartoon characters (e.g., Felix the Cat, Woody Woodpecker) because it was deemed (1) less likely that participants would come in contact with these characters outside of the study; (2) age appropriate; and (3)

would not be taught within the participants' current clinical ABA intervention. The pictures were colored laminated photos (7.5 cm x 10 cm).

Dependent Measures

The primary dependent measure was participant independent correct responses during pre- and post-assessment probes (described below). An *independent correct* response was defined as the participant engaging in a vocal response that corresponded with the presented picture within 5 s of its presentation (e.g., saying "Felix" within 5 s of presentation of the Felix picture). An *independent incorrect* response was defined as the participant engaging in a vocal response that did not correspond with the presented picture (e.g., saying "Garfield" within 5 s of presentation of the Felix picture), or not engaging in a vocal response within 5 s following the presentation of a picture. If the participant started to respond incorrectly but changed to the correct response before completing the incorrect response and within 5 s of the start of the trial (e.g., saying "Bugs . . . I mean Porky Pig"), it was considered correct. This allowed for the participant to self-correct. If the participant completed an incorrect response, then changed to the correct response (e.g., saying "Bugs Bunny, I mean Porky Pig"), it was considered incorrect. This was to reduce continuous guessing and inadvertently reinforcing a chain of incorrect responses. If this occurred, feedback typically occurred just after the incorrect response. If an incorrect response was initiated but not completed, and the correct response was attempted after the 5 s, the trial was scored as incorrect. This was to reduce the participant from guessing and reduce the latency between the instruction and response. Responses for pre- and post-assessment probes were measured using trial-by-trial data collection.

There were also seven secondary dependent measures: total number of trials, responses during teaching, number of targeted stimuli, stimuli used as reinforcers, instructions used, trial type, and reinforcement rate. Researchers measured the total number of trials per session in both teaching conditions. The researchers calculated the total trials per session across all participants within a teaching condition to determine the total number of trials per condition. Because the purpose of DTT is to provide intensive practice through an increased number of trials to maximize the number of learning opportunities (Eikeseth et al., 2014; Smith, 2001), it would be beneficial to determine the difference, if any, across approaches in terms of the number of trials. This could be especially important for those with limited hours of intervention.

Researchers evaluated participant responding using trial-by-trial data during teaching across both conditions. Participant responding was categorized into four different responses. Independent correct and independent incorrect responses were defined identical as in pre- and post-assessment probes. *Prompted correct* responses were defined as the participant engaging in the vocal response that corresponded with the presented picture following a prompt. *Prompted incorrect* responses were defined as the participant engaging in the vocal response that did not correspond with the presented picture or no vocal response following a prompt.

To identify any differences between the number of targeted stimuli each participant was exposed to, the researchers measured the number of targeted stimuli the participant was exposed to during each teaching session across the two conditions. The researchers evaluated the type of stimuli provided as presumed reinforcers by tallying the number of times a stimulus was used or selected as a reinforcer, and the duration of access to reinforcement across the two conditions. Finally, the researchers counted the type of instructions provided, the type of trial (described later), and rate of reinforcement provided.

Design

The two approaches to DTT were assessed using a randomized clinical trial (RCT). Participants were randomly assigned to one of two conditions (i.e., conventional or progressive). Each group received treatment based on the assigned condition and comparisons were made on the relative effectiveness of each treatment (Abbott, 2014).

Pre- and Post-Assessment Probes

Two pre-assessment and two post-assessment probes were conducted to determine participant

performance prior to and following intervention. Each target was presented twice per probe, resulting in 200 trials during each probe. Each trial started with the interventionist holding up the target picture while providing an instruction (e.g., "Who's this?"). The participant was given 5 s to respond. Following a response or 5 s, the interventionist provided neutral feedback (e.g., "Thanks") and moved to the next item. Participants were given 30-s to 2-min breaks throughout pre- and post-assessment probes as needed or requested by the participant. These breaks occurred in an effort to prevent extinction effects because there was no programmed reinforcement for responding during pre- and post-assessment probes. During these breaks, toys (e.g., cars, books) were available and selected by the participant prior to pre- and post-assessments probes. No other persons were present, except for an occasional additional researcher to collect interobserver agreement (IOA) data.

General Teaching Overview

For both teaching conditions, a total of 20 sessions occurred, with each session lasting 15 min. Only one teaching session occurred per day. Both conditions used DTT in a one-to-one instructional format. The general structure of the teaching trial was the interventionist presenting an instruction while holding up a target stimulus with or without the use of a prompt, providing the participant 5 s to respond, and providing feedback (i.e., praise, corrective feedback, or instructive feedback) based on the participant's response. Nine tokens were used in both conditions and were presented by drawing shapes on a whiteboard. Once all nine tokens were obtained, the participant was given access to a presumed reinforcer.

Conventional Approach

Table 1 provides an overview of the components of each approach within this study. Efforts were made to ensure that both approaches reflected commonly used practices that appear in the published literature. Every participant in the conventional approach condition started their first session with a 10-item PS preference assessment. The interventionist selected the items based on each participant's supervisor's responses on the supervisor interview survey (Willis & LaVigna, 1991). The top three items from the PS preference assessment were used as presumed reinforcers throughout teaching. After the PS preference assessment, teaching using the conventional approach to DTT began. During reinforcement breaks, the interventionist would attempt to interact with the participant. If the participant showed positive affect (e.g., smiling, continuing the conversation), then the interventionist continued to provide social interaction throughout the reinforcement break. If the participant did not show positive affect (e.g., ignored the interventionist, asked interventionist to stop talking), then the interventionist stopped engaging until the reinforcement break was over.

A modified data sheet recommended by Grow and LeBlanc (2013) was used to determine the trial order (see Supplemental Materials). The data sheet helped ensure that the trial order was counterbalanced across a trial block, as well as inform the interventionist of which prompt to provide. The data sheet was continuously filled out throughout the teaching session to track the appropriate prompt, remove stimuli as they were mastered, and add new stimuli to teaching. A trial block consisted of nine trials, three trials for each target stimulus. Each trial in this condition consisted of teaching only tact relations. Each trial started with the interventionist presenting the targeted stimulus while simultaneously providing a simple instruction (i.e., "Who is it?") on every trial. The interventionist took trial-by-trial data, based on the responses for each trial.

All trials within the conventional approach were tact trials. Therefore, all trials required the same type of response (i.e., naming the stimulus) with the intention of teaching the names of the stimuli. If a participant incorrectly articulated the stimulus name but remained within a close approximation of the target stimulus, it would still be considered correct. However, no trials were used to improve articulation. In addition, if interfering behaviors occurred during teaching (e.g., attending to events other than teaching, making perseverative comments about specific stimuli), trials were not manipulated to specifically address these behaviors. Rather, corrective feedback was provided specific to the behavior, as well as praise for engaging in appropriate learning behaviors (e.g., sitting, looking at relevant materials, comments on topic).

The interventionist used a progressive time delay prompting system with a model prompt (e.g., "Who

is it? Bugs Bunny;" Kodak et al., 2012; León & Rosales, 2018; Majdalany et al., 2014; Taylor & Harris, 1995). The time delay increased across 0 s, 2 s, 4 s, and 10 s before the interventionist provided the model prompt. Upon first presentation of a target stimulus, the interventionist provided a model prompt with a 0-s delay following the instruction (e.g., "Who is it? Bugs Bunny."). Following three consecutive correct responses, the prompt delay then progressed to a 2-s delay between the instruction and the prompt (e.g., "Who is it?" Followed by 2 s with no response from participant and the interventionist then saying "Bugs Bunny"). This continued until the maximum delay was reached. If the participant responded incorrectly to the same target stimulus on two consecutive trials, the prompt level reduced to the lowest prompting delay (i.e., 0 s).

Following every correct or prompted correct response, the interventionist provided a token and praise (e.g., "Good job!"). Following any incorrect response, the interventionist used an error correction procedure (i.e., saying, "No, it's [correct response]"), followed by a retrial with an echoic prompt (León & Rosales, 2018) with a 0-s delay until the participant responded correctly. If the participant responded correctly following the error correction procedure, the next trial was started without providing a token or praise.

If the participant engaged in an independent correct response for one target across six consecutive trials, that target stimulus was considered mastered and was moved to maintenance for the next session. A new target stimulus then replaced the position of the mastered target on the data sheet. Once a target stimulus reached the mastery criterion, the interventionist implemented maintenance trials for those targets. Maintenance trials occurred during the first 5 min of each teaching session or until all maintenance trials occurred. Maintenance trials consisted of two trials for each mastered target. Each trial began with an instruction (i.e., "Who is it?") while the interventionist presented the target stimulus. If the participant engaged in a correct response, the interventionist provided praise and moved to the next trial. If the participant engaged in an incorrect response, the interventionist provided error correction (i.e., saying, "No, it's [correct response]"). If the participant engaged in an incorrect response for the same target stimulus on two consecutive trials, the target moved from maintenance into teaching. The order of trials was randomly determined using random.org. No tokens or prompts were provided during maintenance trials (De Boer, 2018; Sam & AFIRM Team, 2016).

Progressive Approach

Although there may be debate about the components of a progressive approach to DTT, efforts were made to incorporate elements that have been consistently associated with the implementation of a progressive DTT based on the published in the literature. No formal preference assessment was conducted during any session. Rather, the interventionist used IMRA (Leaf et al., 2015; Leaf, Leaf, Leaf, et al., 2018b) to select potential reinforcers.

To select the item(s) used as reinforcers, the interventionist was instructed to consider the participant's age, verbal and nonvocal behavior, level of engagement, behavior observed during work or previous reinforcement periods, and any change in behavior following the use of a presumed reinforcer. The tangible stimuli available were any used during the PS preference assessments for participants in the conventional condition; however, additional activities were introduced as sessions progressed based on comments made by the participant during sessions or based on observations of how the participant interacted with stimuli, how the participant engaged with the interventionist, and how the participant responded to the environment. For example, one participant frequently commented on different things occurring in the classroom. The next time the participant earned nine tokens, the interventionist introduced a game of "iSpy," and the participant and interventionist took turns commenting on things occurring in the classroom. In addition, tangible stimuli identified as reinforcers were often paired with social interaction, with the intention of making the interaction a more potent reinforcer than the item itself, with the goal of eventually fading the use of the tangible stimuli and using social interaction (e.g., conversation, games) alone as a reinforcer.

The order of trials introduced throughout each session was determined via in-the-moment assessment, similar to Wong et al. (2020). That is, the interventionist had the flexibility to present a target stimulus on one or several consecutive trials within a session. Some variables that influenced the trial order included, but were not limited to, participant

attentiveness, response on the previous trial, and affect. In addition, the interventionist had the flexibility to determine when to introduce a new target stimulus. Some variables that influenced whether a new stimulus would be introduced included, but were not limited to, responses to stimuli from previous sessions, attentiveness, rate of acquisition within a session/across sessions, complexity of the targets, number of stimuli in acquisition, number of stimuli mastered, and how the participant responds to new stimuli. The data sheet the interventionist used was an estimation data sheet (see Supplemental Materials). The interventionist filled out the data sheet at the end of the session with the names of the stimuli targeted during the session. The interventionist then used a 0–4 Likert scale to score the accuracy of participant responding for each target with the numbers 0, 1, 2, 3, and 4 representing 0%–20%, 21%–40%, 41%–60%, 61%–80%, and 81%–100% independent correct responses throughout the session, respectively.

Each teaching trial began with an instruction. The interventionist assessed participant behavior in-the-moment to determine the type and complexity of instructions used on each trial. Some variables that influenced the type and complexity of instruction used by the interventionist included, but were not limited to, participant comprehension, attending, and affect. For instance, if the participant was consistently responding correctly in the presence of the target stimulus and looking at the interventionist and relevant materials, the interventionist was likely to use a novel or complex instruction for upcoming trials to generalize a response to other vocal instructions. If introducing novel and complex instructions resulted in continued accurate responding and positive affect, the interventionist continued to use novel and complex instructions. The type of instructions varied within and across sessions, and included "Who is it?" and "What's his name?" In addition, the complexity of the instruction varied from presenting the target stimulus with no vocal instruction, to presenting the target stimulus and saying, "What's the name of this guy?," to more complex instructions that included a reference to general knowledge (e.g., "I'm thinking of a guy with a baby named Sweet Pea").

The interventionist also had the flexibility to introduce different trial types within a session based upon their analysis of participant behavior. These included tacting, listener behavior (i.e., receptive labels), articulation, general knowledge questions, yes/no, I don't know, or a trial that focused primarily on aberrant behavior. For example, if the participant was engaging in incorrect responses following the fading of a verbal prompt and the participant's affect became flat, the interventionist may have chosen to shift to trial types that involve listener behavior to gain behavioral momentum. Then, based on the participant's affect, accuracy, and attentiveness, the interventionist would slowly integrate tacting trials for a specific target. Responses for varying trial types only affected estimation scores if the participant was required to provide a vocal response that directly corresponded the target stimulus. For example, for an articulation trial, the participant may incorrectly articulate the stimulus name, but would be considered correct if it was a close approximation of the target stimulus.

Listener behavior trials occurred when the interventionist placed two or more target stimuli on the table, then provided an instruction for the participant to find a particular target stimulus. The instructions provided for these trials included the interventionist stating the target stimulus (e.g., "Casper"), to more complex instructions (e.g., "Which one is Casper?"), to more complex instructions that may have referenced general knowledge (e.g., "Find the picture of Casper, it's the white ghost.").

Although all participants engaged in low rates of aberrant behavior, some aberrant behavior occasionally occurred. For example, some participants engaged in repetitive behaviors following a trial with a specific target (e.g., commenting and asking questions about the photo, arm flapping). Therefore, the primary objective of some trials was to target aberrant behavior rather than correct responses to the target stimuli. If aberrant behavior occurred, the interventionist provided corrective feedback (e.g., "You didn't keep a calm body," "You got stuck"), then presented the previously completed trial again. During the representation of the trial, or the aberrant behavior trial, the feedback following the response was directed toward the behavior (e.g., "You said it with a cool body!") rather than the correctness of the response.

Articulation trials focused primarily on accurate articulation. For example, if the target was "Cindy," and a participant pronounced it "Cid-ney," the interventionist said the word more slowly, while adding an emphasis on the part of the word said incorrectly (e.g., "CiN" pause "dey").

General knowledge trials were used to assess the acquisition of general knowledge regarding the target stimulus provided within instructive feedback. Though this was not a dependent measure, the interventionist hypothesized it would be beneficial to include these types of trials to determine if the participant was learning from instructive feedback.

Yes/no trials were used for two purposes. First, as another way to assess attending and general knowledge. For example, while holding a picture of Woodstock, the interventionist would ask "Is this Tweety?," with the expectation that the participant would respond with "no" and provide the correct response (e.g., "No, it's Woodstock!"). If the participant responded with the correct yes/no response, but did not provide the character name, the interventionist would prompt the additional response (e.g. "it's___"). A second use of yes/no trials was to improve articulation. If the interventionist found that the participant was struggling with articulating a name, the interventionist used a yes/no trial to highlight the part of the name being said incorrectly. For example, while holding up Mr. Crocker, the interventionist might ask "Is this Mr. Crocket?," with the expectation that the participant would respond with "no" and provide the correct response (e.g., "No, it's Mr. Crocker").

When presented with a stimulus not yet targeted, some participants did not respond following the instruction, sat silently, and stared at the target stimulus. *I don't know* trials were used to teach the participant a vocal response to unknown stimuli as a replacement behavior. The interventionist used instructive feedback to inform participants they could say "I don't know" for unknown stimuli. When presented with an unknown target stimulus in succeeding trials and the participant stated, "I don't know," the interventionist provided praise. The interventionist hypothesized these trials would be beneficial for participants outside of the experimental conditions. Once the participant would successfully say "I don't know" when presented with new stimuli, the interventionist would then either provide instructive feedback or prompt the next trial.

The interventionist used FPF (Soluaga et al., 2008) throughout all sessions. FPF permitted the interventionist to assess learner behavior to determine on which trials to provide a prompt, the type of prompt to provide, and how to fade prompts. When the interventionist provided a prompt, it was based on the assessment of a variety of variables including, but not limited to, how the participant responded in a previous trial, the delay between trials for a specific target stimulus, and participant effort (e.g., sitting appropriately, looking at the stimulus, using filler words "um"). For instance, if the participant had previously responded accurately and independently for a target stimulus but was responding inaccurately in the moment while also slow to respond and not looking at the relevant teaching materials, the interventionist was not likely to prompt the next trial, but rather provide feedback (e.g. "you're not paying attention"). If the participant had previously responded accurately and independently to about 60% of trials for a target stimulus but was currently responding incorrectly while looking at the relevant materials and responding to other mastered stimuli correctly, the interventionist was more likely to provide either a partial vocal or multiple alternative prompts. Prompts included full vocal prompts or model prompts, partial vocal prompts, and multiple alternative prompts.

The interventionist delivered a variety of consequences following correct and incorrect responses. Correct responses (prompted or independent) were followed by provided praise (e.g., "Nice job."), a token, no feedback, and/or a social cue (e.g., head nod). Independent correct responses for target stimuli were more likely to result in differential consequences (e.g., token and praise) than those stimuli in maintenance (which were more likely to result in vocal statements like "good"). Prompted correct responses were also candidates for differential consequences if the prompt was faded to a less assistive prompt (e.g., full vocal prompt to partial vocal prompt). Incorrect responses (prompted or independent) were followed by a simple "no," error correction (e.g., "No, it's [correct response]"), and/or a social cue (e.g., a scrunched face with a raised eyebrow). Incorrect responses for stimuli in maintenance were more likely to occasion differential consequences (e.g., blunt "no") than incorrect responses for target stimuli (e.g., "no" with upspeak or a social cue). The interventionist determined a target reached mastery if the target had a score of 4 for at least two consecutive sessions, as well as responding with fluency (e.g., participant responding correctly within 2 s of the instruction). The interventionist also determined when to introduce new target stimuli within a session. Therefore, the interventionist could introduce no new

target stimuli within a session or all 100 target stimuli within a given session. The decision of which maintenance targets to review as well as the number of maintenance trials for each target in maintenance was determined through in-the-moment assessment by the interventionist; though, maintenance was commonly interspersed between trials in acquisition throughout the session.

Interobserver Agreement and Treatment Fidelity

A researcher recorded responding during every pre- and post-assessment probe and an independent observer recorded responding during 61.13% of sets for pre-assessment probes and 33.33% of sets for post-assessment probes. Interobserver agreement was calculated by totaling the number of agreements (i.e., trials in which both observers scored the same participant behavior) on each trial and dividing the number of agreements plus disagreements (i.e., trials in which the two observers scored a different participant behavior) and converting this ratio to a percentage. IOA was 100% and 99.62% (range: 90%–100%), for pre- and post-assessment, respectively.

Treatment fidelity was taken for 25% of pre- and post-assessment probes. The following components were scored for assessment probes: (1) provides an instruction (e.g., "Who is it?," "Who's this?"); (2) does not provide a prompt; (3) allows student 5 s to respond; and (4) provides neutral feedback (e.g. "thanks"). Treatment fidelity was scored by marking the occurrence or nonoccurrence of each component throughout the session. Therefore, if a prompt was provided for any of the trials within the session, component b was marked as "nonoccurrence." The number of occurrences was divided by the total number of components, then multiplied by 100 to get the percentage. Treatment fidelity was 100% for pre- and post-assessment probes.

Treatment fidelity was taken for 35% of sessions for both teaching conditions (see Supplemental Materials). The following components were scored for the conventional condition: (1) instruction: "Who is it?"; (2) use of progressive prompt delay; (3) one token and praise provided for prompted and independent correct responses; (4) error correction plus re-trial with echoic prompt (0-s delay) provided for incorrect responses (prompted and independent); (5) returning to 0-s prompt delay for incorrect responses to the same stimulus twice in a row; (6) stimulus was removed from teaching following independent correct responses for three consecutive responses for the same stimulus across two blocks of trials; (7) trial order was counterbalanced; and (8) data was taken on each trial. The components for scoring treatment fidelity during maintenance included: (1) instruction: "Who is it?"; (2) no prompt provided; (3) allow participant 5 s to respond; (4) providing praise (no token) following correct responses; (5) providing error correction following incorrect responses; and (6) data was taken on each trial. The treatment fidelity components scored during the preference assessment included: (1) presented two items at a time; (2) each item presented with every other item; (3) location of item rotated (left and right); and (4) participant had 10–30 s to engage with item after choosing between the two items presented. Similar to treatment fidelity for pre- and post-assessment probes, treatment fidelity was scored by marking the occurrence or nonoccurrence of each component throughout the session. The component had to be observed implemented correctly throughout the entire session in order to be marked as an occurrence. The number of occurrences was divided by the total number of components, then multiplied by 100 to get the percentage. Treatment fidelity for the conventional approach was 99.64% (range; 98.92%–100%).

Treatment fidelity for the progressive condition was scored for the following components observed within each session: (1) instruction provided (verbal or presented picture); (2) more than one type of instruction used; (3) when a prompt was provided, the interventionist varied the type of prompt; (4) used a whiteboard to draw nine shapes as tokens; (5) provided terminal reinforcer after nine tokens were earned; (6) provided varied feedback following participant response (e.g., praise, corrective feedback, neutral feedback); and (7) estimation data was taken. Treatment fidelity percentage was calculated similarly to treatment fidelity scores for pre- and post-assessment probes and the conventional condition. Treatment fidelity for the progressive approach was 99.66% (range: 98%–100%).

Statistical Analyses

The Wilcoxin signed-ranked test with continuity correction was used to compare the performance from

pre- to post-assessment probes for participants in each condition. Wilcoxon signed-rank test is a nonparametric statistical test that can be used to compare two related samples, matched samples, or repeated measurements on a single sample when the distribution of the differences between the two samples cannot be assumed to be normally distributed. A continuity correction is an adjustment used to correct for small sample sizes. To compare performance on post-assessment probes across the two conditions, a Wilcoxin rank-sum test with continuity correction was used. A Wilcoxin rank-sum test is a nonparametric statistical test used to compare two independent samples. All statistical tests used a significant level of $p < 0.05$.

Results

Pre- and Post-Assessment Probes

Figure 1 displays the average responding per each participant across the two conditions during post-assessment probes. No pre-assessment data is depicted as all participants scored 0%, which was planned to ensure no prior learning history. On the x-axis are the participants in the two teaching conditions and on the y-axis is the average percentage of correct responses across the two post-assessment probes. Following conventional DTT, the mean percentage of correct responding across all participants and targets was 35.7%. The results of a Wilcoxin signed-ranked test indicated the change from pre-assessment to post-assessment was statistically significant; $V = 21$, $p = 0.03552$. Following progressive DTT, the mean percentage of correct responding across all participants and targets was 90.4%. The results of a Wilcoxon signed-rank test indicated the change was statistically significant; $V = 21$, $p = 0.03401$. Finally, the Wilcoxin rank-sum test indicated there was a significant difference when comparing post-test probes between the two conditions; $W = 2$, $p = 0.01228$.

Total Trials

Figure 2 displays the number of total trials per session in each condition. The conventional condition had a total of 7,037 trials presented across sessions and participants, with an average of 58.6 trials per session (range; 0–78 trials across sessions). The progressive condition had a total of 15,445 trials presented across sessions and participants, with an average of 136.7 trials per session (range: 77–208 trials).

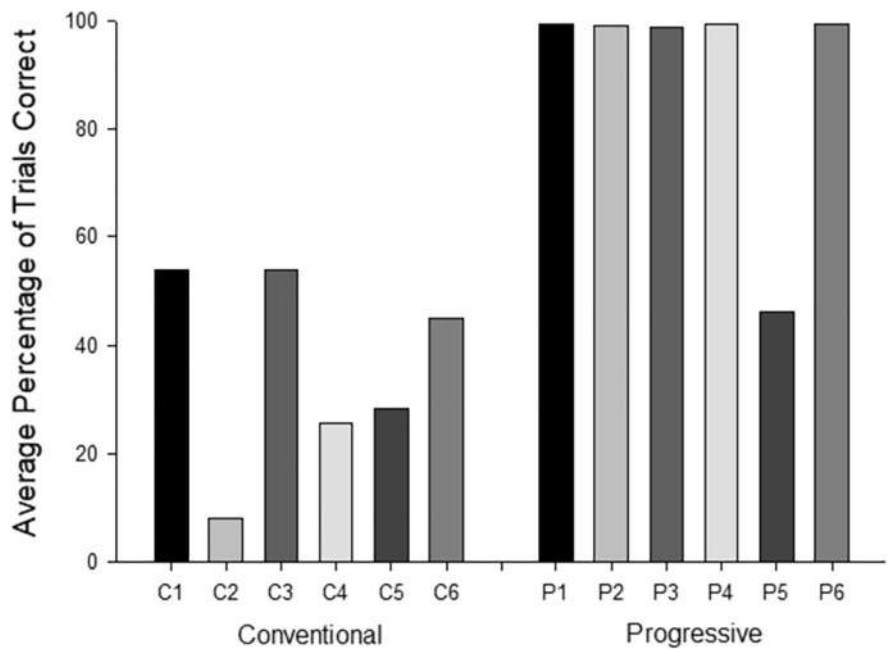

Fig. 1 Post-Assessment Score Average across Participants. Note. Each bar represents a participant within the condition

Participant Responding During Teaching

Table 4 displays participant responding during teaching. The mean percentage of independent correct responses was 76.47% and 84.10% for the conventional and progressive conditions, respectively. The results of a *t*-test indicated that there was no statistically significant difference ($t = 1.728$, $p = 0.115$). The mean percentage of independent incorrect responses was 3.89% and 8.26% for the conventional and progressive conditions, respectively. The results of a *t*-test indicated that there were significantly more errors in the progressive condition ($t = 2.229$, $p = 0.0499$). The mean percentage of prompted correct responses was 19.45% for the conventional condition and 6.69% for the progressive condition. The *t*-test revealed there were significantly more prompted responses in the conventional condition ($t = -5.991$, $p = < 0.001$). The mean percentage of prompted incorrect responses was 0.18% for the conventional condition and 0.95% for the progressive condition. The Mann-Whitney U statistic indicated that there were significantly more prompted incorrect responses in the progressive condition ($U = 3$, $p = 0.015$).

Stimuli Exposed Per Session

Figure 3 displays the number of target stimuli the participants were exposed to during each session. For the conventional condition, an average of 21.59 (range: 0–36) target stimuli were presented per session. For the progressive approach, an average of 54.63 (range: 4–100) target stimuli were presented per session.

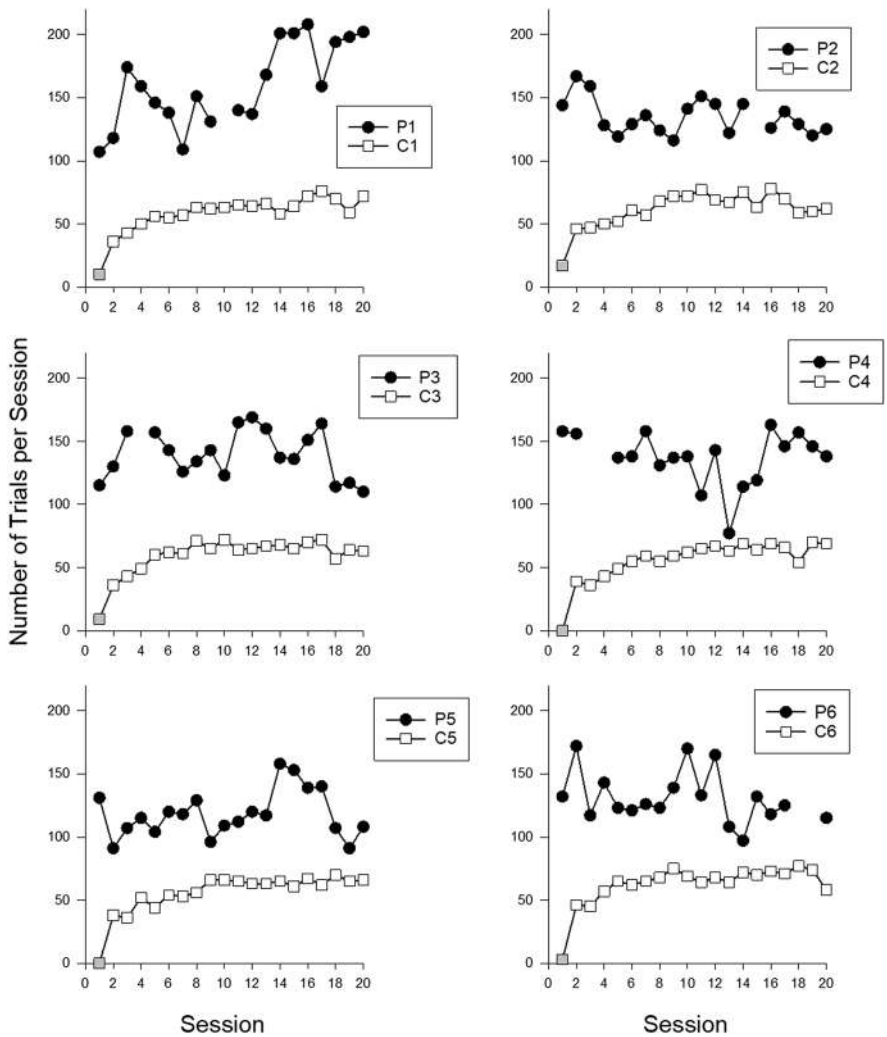

Fig. 2 Total Trials across Sessions. Note. Missing data points are due to video recordings lost. Grey data points indicate the first session when a formal preference assessment occurred and used a majority if not the entire session

Table 4 Percentage of Responses Correct and Incorrect across Participants and Conditions

Participant	Independent Correct	Independent Incorrect	Prompted Correct	Prompted Incorrect
C1	80.96%	1.29%	17.66%	0.09%
C2	74.06%	4.42%	21.11%	0.41%
C3	81.83%	1.44%	16.65%	0.08%
C4	78.80%	4.22%	16.71%	0.27%
C5	68.44%	7.82%	23.56%	0.18%
C6	74.84%	4.33%	20.95%	0.08%
Mean	76.47%	3.89%	19.45%	0.18%
P1	93.62%	3.45%	2.66%	0.26%
P2	83.82%	9.24%	6.43%	0.51%
P3	87.22%	5.66%	6.79%	0.34%
P4	82.38%	7.35%	9.26%	1.02%
P5	67.61%	16.49%	13.74%	2.16%
P6	86.99%	8.99%	2.29%	1.74%
Mean	84.10%	8.26%	6.69%	0.95%

Stimuli Selected for Reinforcement

Results of the stimuli selected as presumed reinforcers were analyzed across the groups (see Supplemental Materials). Participants in the conventional condition had the option of picking one of three stimuli that were selected from the initial PS preference assessment. The top three items selected for terminal reinforcers in the conventional condition were gummy bears, Oreos and the iPad.

For participants in the progressive condition, the interventionist could select a variety of reinforcers and was not bound to reinforcers included in or identified by a formal preference assessment. The top three selected items used as terminal reinforcers in the progressive condition were conversation, gummy bears, and chips. In addition, reinforcing stimuli may have been paired with a social interaction (social or conversation). During the progressive condition, an average of 7.33 (range: 5–10) different stimuli were provided as presumed reinforcers per participant across sessions. Edibles were used the most across both conditions. Tangible items and social stimuli were the second most frequently used stimuli for the conventional approach and progressive approach, respectively.

Duration of Reinforcement

The duration of access to reinforcement for participants in the conventional and progressive condition is displayed in Figure 4. Each participant from the conventional approach is graphed with the participant they were paired with from the progressive condition. Of the 113 videotaped sessions of the conventional approach, the average duration spent with access to presumed reinforcement was 5.36 min (range: 0–9.08 min) per session. For the 113 videotaped sessions of the progressive approach, the average duration spent with access to presumed reinforcement was 3.82 min (range: 1.3–7.28 min) per session.

Instruction Analysis

The number of novel instructions used per session were evaluated across groups (see Supplemental Materials). For the conventional condition, there was an average of 1.14 (range: 1–4) different types of instructions used per session. These instructions exclusively included "Who is it?," "Who's that?," "Who is that?," "Who's this?," and "Who is this?" Of the 113 videotaped sessions, there was an average of 27.99 (range: 6–51) different types of instructions used per session within the progressive condition. Examples of instructions included holding up the stimulus, asking "Who's this?" and "Who's the squirrel with the blue goggles?"

Types of Trials

The variations of trials used across sessions were tracked for each participant within each condition (see Supplemental Materials). For the conventional condition, most trials were tacting, though some aberrant behavior trials (three trials across all 20 sessions) were used for one participant. For the progressive condition, the most frequent trial types included tacting and listener behavior, which were used for all six participants, whereas the less frequent type was "I don't know," which was used for only one participant.

Rate of Reinforcement

Figure 5 displays the average number of correct responses before a token was earned for each participant across sessions. In the conventional approach, a token was given for every correct response (prompted or independent) across participants and

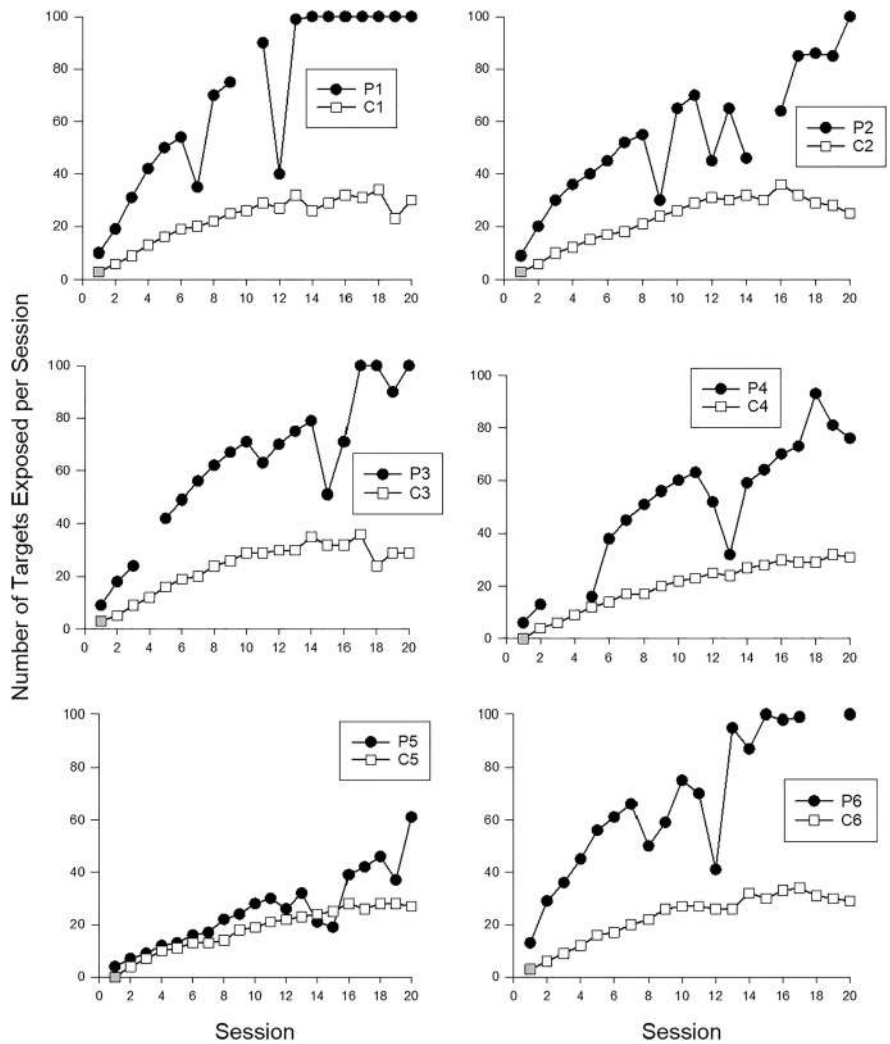

Fig. 3 Target Stimuli Exposed per Session. *Note.* Missing data points are due to video recordings lost. Grey data points indicate the first session when a formal preference assessment occurred and used a majority if not the entire session

sessions. For the progressive approach, a token was given for an average of 5.12 (range: 1.79–9.85) correct responses, prompted or unprompted, across participants and sessions.

Discussion

Over the past decade there have been numerous studies that have compared components of a conventional approach to DTT to components of a progressive approach to DTT as they relate to individuals diagnosed with ASD. Although the names of the approaches might be disputed, the terms were chosen for clarity; conventional refers to commonly used procedures, whereas progressive refers to deviations from these procedures. The purpose of this study was to expand upon previous research by comparing the two approaches more wholly, as opposed to their component parts, to teach tact relations to 12 children diagnosed with ASD and with moderate to high language skills. The results of this study indicated that both approaches were effective as evident by statistically significant differences when comparing pre- and post-assessment probes. This is an important, central finding that supports the efficacy of both approaches. However, when comparing the post-assessment probes across conditions, participants in the progressive condition scored significantly higher than participants in the conventional condition, and

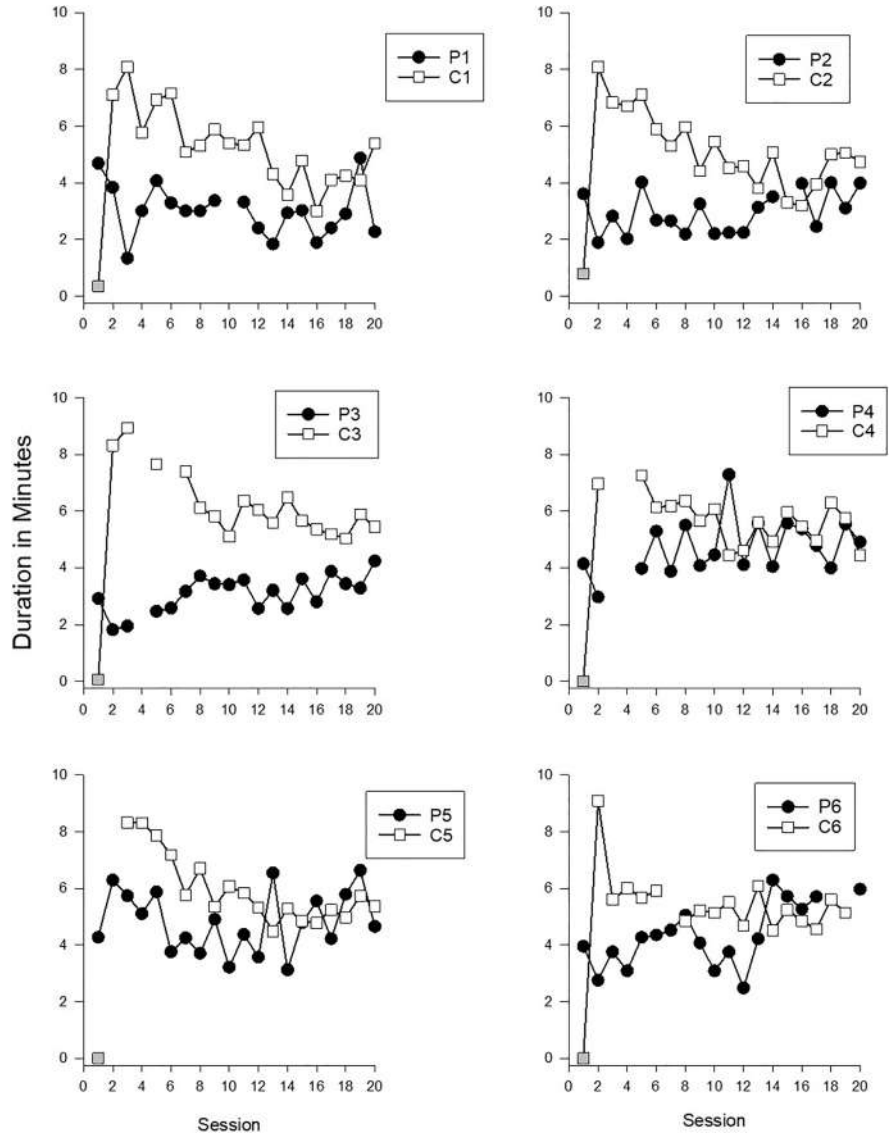

Fig. 4 Duration of Reinforcement. Note. Missing data points are due to video recordings lost. Grey data points indicate the first session when a formal preference assessment occurred and used a majority if not the entire session

the effect size was large ($d = 2.99$). Analysis of participant responding during teaching yielded additional differences across the two conditions. The conventional approach to DTT, on average, resulted in more prompted correct responses across all participants and sessions, which is not surprising given the prompting methods employed within this condition. The progressive approach to DTT, on average, resulted in more independent correct responses, independent incorrect responses, and prompted incorrect responses across all participants and sessions. Each of these differences was statistically significant except for independent correct responses. That is, although there was a difference (i.e., 7.63%) in the average percentage of independent correct responses between the two conditions, this difference was statistically significant. Taken collectively, the results of this study demonstrated a progressive approach to DTT was more effective than a conventional approach when teaching tact relations to children diagnosed with ASD between ages 3 and 9 years with moderate to high language skills. These results are consistent with previous research comparing components of a conventional and progressive approach to DTT (e.g., Cihon et al., 2020; Ferguson et al., 2020; Leaf et al., 2015; Leaf et al., 2019; Leaf, Cihon, Ferguson, et al.,

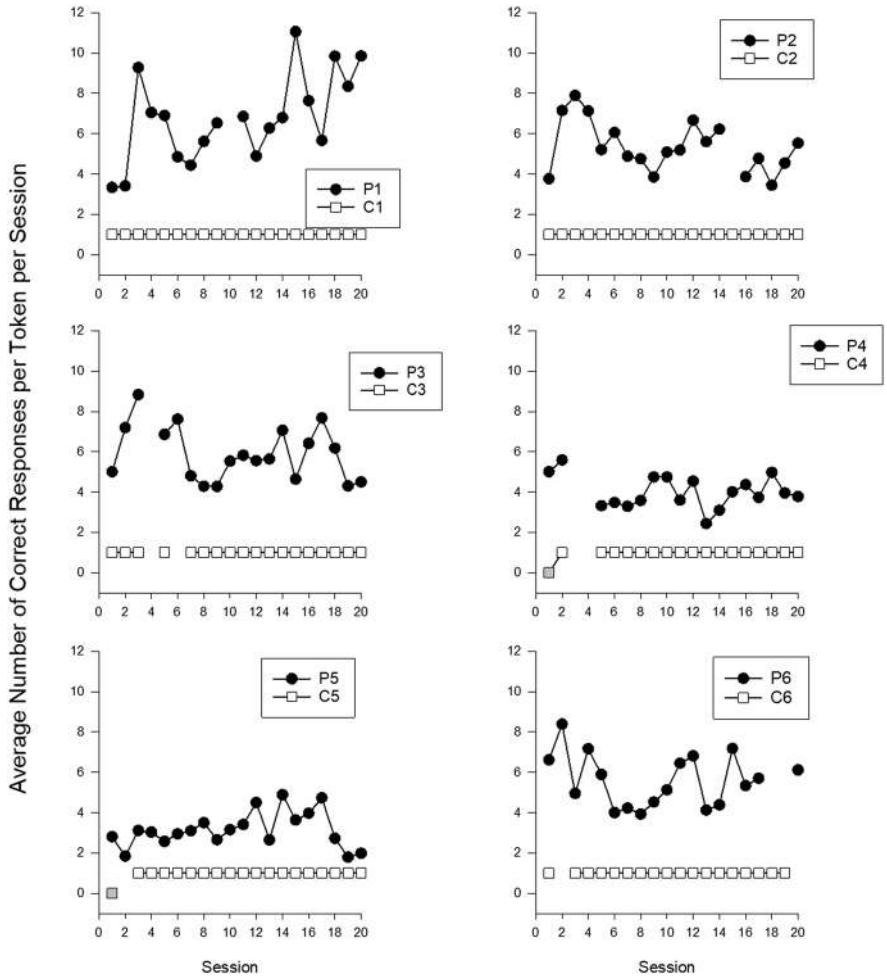

Fig. 5 Average Number of Correct Responses per Token. *Note.* Missing data points are due to video recordings lost. Grey data points indicate a session occurred, but no trials were conducted

2018a; Leaf, Dale, et al., 2014a; Leaf, Leaf, et al., 2014b; Leaf, Leaf, Leaf, et al., 2018b; Leaf, Leaf, McEachin, et al., 2016c; Soluaga et al., 2008; Taubman et al., 2013; Wong et al., 2020).

There are several variables that may have contributed to the differences in the results of the two conditions. One potential reason the progressive approach to DTT was more effective could have been the number of trials that occurred in that condition. Even though the session duration in both conditions was limited to 15 min, overall there were twice the number of trials provided in the progressive condition. It is likely that more trials in the same amount of instructional time resulted in the increased acquisition in the progressive condition (Greer & McDonough, 1999). Furthermore, the increased number of trials more closely aligns with the underlying philosophy of DTT focused on the rapid presentation of trials to increase teaching opportunities (Eikeseth et al., 2014; Smith, 2001).

There are several possible explanations as to why the progressive condition resulted in more trials, even though both conditions were limited to the same session duration. First, there was no time spent on formal preference assessments: rather, potential reinforcers were identified in-the-moment by the interventionist. Second, multiple instructions and varied consequences occurred in the progressive condition. Some trials consisted of holding up the photo with no vocal instruction, which could have reduced the time required to complete a trial. This time saved could have created the opportunity to present more trials. It should be noted, however, there are many other potential benefits of multiple instructions including the development of novel behavior (LaFrance & Tarbox, 2020), reducing

learner boredom (Leaf & McEachin, 1999), and promoting generalization (Stokes & Baer, 1977). The participants in this study were able to respond to multiple forms of instruction without additional teaching, however the use of varied instructions should be evaluated with learners of varying language skills in order to gain a better understanding of the effects on acquisition using highly varied instructions. Similar to the use of multiple instructions, the variation in consequences in the progressive condition may have also allotted time for more trials (e.g., the time it takes to say "no" vs. "no" with remedial trial). Third, the interventionist faded prompts flexibly (using FPF) opposed to a solely time-based prompting system, which could have allowed for more trials. That is, the time required by the progressive prompt delay in the conventional condition may have reduced the number of trials. Finally, the interventionist was not required to collect data on every trial. The use of estimation data most likely permitted the interventionist to include more instructional trials (Ferguson et al., 2020). All in all, these differences may point to an efficiency advantage to the progressive approach.

A second variable that could have contributed to the large differences in the results of the two conditions was the use of FPF. There are several reasons FPF may have contributed to better acquisition in the progressive condition. First, FPF affords the interventionist the discretion to individualize prompting. This individualization likely led to more effective prompts for each participant in varied contexts. Second, within FPF the interventionist could adjust the level of assistance included in the prompt in more subtle and learner-driven ways (e.g., partial vocal prompts). Progressive prompt delay, on the other hand, only allows for complete assistance (i.e., a full vocal prompt, the time delay just varies). Third, some prompts within FPF help to promote attending and engagement, as evident in yes/no trials (e.g., asking "Is this Tweety?," in the presence of Woodstock). This potential increase in attending and engagement could have resulted in better acquisition. Finally, and relatedly, the use of multiple alternative prompts within FPF resulted in multiple learning opportunities on the same trial. For example, asking the learner, "Is this Tweety or Woodstock?" could result in a correct response in the presence of Woodstock and increase the likelihood of a correct response in the presence of Tweety.

A third variable that may have affected the differences between the two conditions relates to the trial order and different types of trials used within the progressive condition. The participants in the progressive condition were exposed to multiple types of trials. By varying the type of trials, the interventionist could individualize teaching to adapt to learner errors as well as provide multiple exemplars, both of which have been documented to improve learning (LaFrance & Tarbox, 2020; Leaf & McEachin, 1999; Stokes & Baer, 1977). Also, in the progressive condition, the interventionist could manipulate the trial order as opposed to following a prespecified trial order outlined on a datasheet. This allowed the interventionist to intersperse maintenance with acquisition trials potentially creating more behavioral momentum than in the conventional condition (see Lipschultz & Wilder, 2017 for a review). Manipulation of trial order also allowed the interventionist to repeat trials with one or more target stimuli and select target stimuli to teach together or separately. In addition, within each session of the conventional condition, time needed to be allocated toward preparing the data sheet for the next round of teaching trials, which potentially took away time from instructional time.

A fourth potential reason the progressive approach to DTT was more effective relates to the flexibility in consequences provided in the progressive condition. The interventionist had the freedom to select the schedule of reinforcement appropriate for the learner, as well as introduce a variety of reinforcers throughout teaching. This opened the possibility of a flexible earning requirement (Cihon, Ferguson, Milne, et al., 2019b) and a more intermittent schedule of reinforcement for participants in the progressive condition. Although the conventional and progressive approach had similar durations of reinforcement, the number of responses needed to access reinforcement differed. The increased number of responses required to access reinforcement, and the flexibility to use more neutral feedback during teaching may have promoted better performance in post-assessment probes. That is, post-assessment probes occurred under extinction-like conditions, and thinning the schedule of reinforcement prior to these probes may have reduced the contrast effect when no reinforcement was provided. By continuously embedding a variety of reinforcers and conditioning new social-based reinforcers, this also could influence rapport and overall responding

under extinction-like conditions. Further analysis on the effects of utilizing a variety of reinforcers may be compelling.

Across these findings, there are emergent themes. It appears that the progressive approach may have been more efficient, as noted especially in the difference in learning opportunities. It also may have been more individualized, particularly in the use of prompt strategies and selected reinforcers. These differences may have allowed for more precise targeting of skills and more nuanced instruction.

This study expands upon previous research in at least two ways. First, this study represents the first to compare the two approaches to DTT in a comprehensive way. Although the previous research has compared components of a progressive approach to DTT to components of a conventional approach to DTT, these comparisons do not represent a comprehensive approach to DTT. It may be the case that small, if any, differences occur when evaluating only one component (e.g., counterbalancing of comparison array), and overall outcomes remain relatively the same. This study, on the other hand, identified significant differences when all components were evaluated in concert. Second, previous research comparing the two approaches included a limited number of targeted stimuli for acquisition (Leaf et al., 2019; Wong et al., 2020), which is likely not representative of what occurs during the course of intervention for individuals diagnosed with ASD. This study, on the other hand, involved 100 target stimuli, which may align more closely to typical intervention for individuals diagnosed with ASD.

The results of this study have several implications for clinicians. The components of and approaches to DTT could be conceptualized as occurring across a continuum, and this study evaluated what could be considered the two ends of that continuum. Given this continuum, it will be important for clinicians to consider the conditions under which each approach, or certain elements of each approach, would be feasible and most advantageous for their clients. One important consideration in adopting a progressive approach to DTT should be the potential training required. The interventionist in this study had years of training and experience that exceeds that of many direct line therapists. However, it is currently unknown how much training is required in order to competently apply a progressive approach to DTT in order to obtain similar results. Given the training of the interventionist in this study, if clinicians adopt a progressive approach to DTT, the level of training that is necessary may require going beyond existing minimum standards (e.g., Behavior Analyst Certification Board, 2019).

Another consideration relates to providing the most effective and efficient treatment possible. If effectiveness and efficiency are the goal of intervention, based on the results of the present study, as well as previous component comparisons, it is likely that the methods employed in the progressive condition should be selected. Although many variables should be considered prior to selecting one approach over the other, clinicians should actively work toward arranging contingencies to permit the selection of a progressive approach to DTT. Doing so may accelerate the learning for our clients and in turn improve their overall quality of life. Replications of this work need to occur in order to make broader conclusions.

The results of this study have implications for researchers, especially in replication contexts. First, all sessions across conditions were conducted by an interventionist with many years of experience and training in implementing both approaches. As such, it is not known what the results would be for interventionist with less training (e.g., a registered behavioral technician). As such, future researchers should attempt to replicate these results with interventionists with differing experience and training. Second, all the participants in this study had average to high language and IQ scores as well as minimal aberrant behavior. Thus, it is not known if similar results would be obtained with participants with lower IQ scores or higher rates of aberrant behavior. Future researchers should attempt to replicate the results of the present study with a population of individuals with lower language scores and/or higher rates of aberrant behavior. Third, this study compared the two approaches with respect to teaching tact relations. However, in clinical intervention, individuals diagnosed with ASD are taught a variety of other behaviors (e.g., receptive labeling, social behavior). As such, future researchers should continue evaluating the two approaches across variety of other behaviors, in isolation and combination.

This study did not go without its limitations that warrant discussion. First, although all pre- and post-assessment probes and teaching sessions for every participant in both conditions were scored (288 recordings total), 13 recordings were lost due to technical issues. Six videos in the conventional condition

(C3: 2, C4: 2, C5: 1, C6: 1) and seven in the progressive condition (P1: 1, P2: 1; P3: 1; P4: 2; P6: 2) were lost. Thus, a complete analysis of every moment of teaching could not be analyzed. However, enough of the videos remained (i.e., 275 of 288; 95%) to provide a representative picture of what occurred. Second, the conventional DTT condition involved the use of a PS preference assessment. Although this is the most cited preference assessment within the literature, it can take longer than other formal preference assessments. Future comparisons could address this limitation by selecting a more streamlined formal preference assessment. Third, the conventional method used in this study tested for maintenance at the beginning of sessions rather than interspersing maintenance trials in teaching. Future studies should continue to compare the effects of a conventional to progressive approach if both methods intersperse maintenance. Fourth, research in evaluating a progressive approach to DTT has primarily come from one research lab. It will be important for other research labs and clinical settings to attempt to replicate the results of studies evaluating a progressive approach to ensure its replicability and scalability. Fifth, only one interventionist implemented the procedures used across the participants and conditions. Future researchers should consider evaluating the effects of each procedure implemented across multiple interventionists to determine if similar outcomes occur. Sixth, as noted in previous studies (e.g., Cihon, Ferguson, Leaf et al., 2019aa; Cihon, Ferguson, Milne, et al., 2019b), the use of procedures that require the interventionist to analyze contextual variables and change their behavior based on the outcomes of that analysis creates challenges with respect to replication for future researchers. Although replicability challenges may occur for researchers, it may be the case that these conditions more closely align to those in which practitioners find themselves working. Future research is necessary to identify the variables that occasioned the interventionist's behavior as well as the necessary interventionist behavioral repertoires to ensure replication. Next, it should be noted that there is limited diversity in the sample of participants (i.e., Asian American and white). There is a discrepancy in access to diagnostic services and treatment for Black and Latinx children diagnosed with autism (Liptak et al., 2008), and limited research evaluating interventions and outcomes for these populations. The findings of this study are meaningful for the children included; however, additional research is needed to evaluate the impact of these interventions for other children with diagnosed with ASD.

Finally, no a priori power analysis was used to determine if the sample size would be large enough to yield a valid effect. This limitation must be considered when interpreting the results. In the setting in which this study occurred, the participants included represented a convenience sample (i.e., the individuals most accessible to the researchers) and, given previous research (e.g., Cihon et al., 2020; Soluaga et al., 2008; Taubman et al., 2013; Wong et al., 2020), there was a high probability of differences with 12 participants (i.e., 12 participants is a large enough sample size to detect intervention effects). Furthermore, visual inspection of the results indicates large differences between the two groups in pre- and post-probe assessments. Nonetheless, these data should be viewed as preliminary given the lack of an a priori power analysis. Future research evaluating and comparing conventional and progressive approaches to DTT using an RCT may benefit from incorporating criteria from organizations such as Cochrane or What Works Clearinghouse.

Given the importance of early intervention, and the limited hours of intervention that some individuals with ASD receive, it is important to maximize the learning opportunities during these limited times. If there are procedures that allow sessions to result in better treatment effects, it is pertinent to identify these advantages. Some may argue that the training required prohibits the feasibility of a progressive approach; however, if the results of research continue to be this significant, perhaps we should reevaluate our standards of intervention, as well as the training required to meet those standards, to work with such a sensitive population. It may also be that more efficient instructional methods could be developed to teach clinical decision making in time-efficient and effective ways.

Funding No funding was received for this study.

Declarations

Ethical Approval All procedures performed in studies involving human participants were in accordance with ethical standards of the institutional research committee and with 1964 Helsinki decoration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from the parents of all individual participants included in the study.

References

Abbott, J. H. (2014). The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: What they are and are not. *Journal of Orthopaedic & Sports Physical Therapy, 44*(8), 555–558. https://doi.org/10.2519/jospt.2014.0110PMID-25082389

Behavior Analyst Certification Board (2019). *Board certified behavior analyst.* https://www.bacb.com/bcba/

Cihon, J. H., Ferguson, J. L., Leaf, J. B., Leaf, R., McEachin, J., & Taubman, M. (2019a). Use of a level system with flexible shaping to improve synchronous engagement. *Behavior Analysis in Practice, 12*(1), 44–51. https://doi.org/10.1007/s40617-018-0254-8

Cihon, J. H., Ferguson, J. L., Milne, C. M., Leaf, J. B., McEachin, J., & Leaf, R. (2019b). A preliminary evaluation of a token system with a flexible earning requirement. *Behavior Analysis in Practice, 12*(3), 548–556. https://doi.org/10.1007/s40617-018-00316-3

Cihon, J. H., Ferguson, J. L., Leaf, J. B., Milne, C. M., Leaf, R., & McEachin, J. (2020). A randomized clinical trial of three prompting systems to teach tact relations. *Journal of Applied Behavior Analysis, 53*(2), 727–743. https://doi.org/10.1002/jaba.617

Cummings, A. R., & Carr, J. E. (2009). Evaluating progress in behavioral programs for children with autism spectrum disorders via continuous and discontinuous measurement. *Journal of Applied Behavior Analysis, 42*(1), 57–71. https://doi.org/10.1901/jaba.2009.42-57

De Boer, S.R. (2018). How to do discrete trial training: Overview of the new 2nd Edition [Conference Session]. First Annual Richard L. Simpson Conference on Autism, Overland Park, KS, USA. https://mslbd.org/file_download/6acd2b45-af54-4833-89e8-c9dbf4618ca1

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Eikeseth, S., Smith, D. P., & Klintwall, L. (2014). Discrete trial teaching and discrimination training. In J. Tarbox, D. R. Dixon, P. Strumey, & J. Matson (Eds.), *Handbook of early intervention of autism spectrum disorders (pp. 229–253)*. Springer. https://doi.org/10.1007/978-1-4939-0401-3_12

Ferguson, J. L., Milne, C. M., Cihon, J. H., Dotson, A., Leaf, J. B., McEachin, J., & Leaf, R. (2020). An evaluation of estimation data collection to trial-by trial data collection during discrete trial teaching. *Behavioral Interventions, 35*(1), 178–191. https://doi.org/10.1002/bin.1705

Graff, R. B., & Karsten, A. M. (2012). Assessing preferences of individuals with developmental disabilities: A survey of current practices. *Behavior Analysis in Practice, 5*(2), 37–48. https://doi.org/10.1007/BF0339182

Green, G. (2001). Behavior analytic instruction for learners with autism advances in stimulus control technology. *Focus on Autism & Other Developmental Disabilities, 16*(2), 72–85. https://doi.org/10.1177/108835760101600203

Greer, R. D., & McDonough, S. H. (1999). Is the learn unit a fundamental measure of pedagogy? *The Behavior Analyst Today, 22*(1), 5–16. https://doi.org/10.1007/BF03391973

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

Kodak, T., Fuchtman, R., & Paden, A. (2012). A comparison of intraverbal training procedures for children with autism. *Journal of Applied Behavior Analysis, 45*(1), 155–160. https://doi.org/10.1901/jaba.2012.45-155

Koegel, L. K., & Koegel, R. L. (1986). The effects of interspersed maintenance tasks on academic performance in a severe childhood stroke victim. *Journal of Applied Behavior Analysis, 19*(4), 425–430. https://doi.org/10.1901/jaba.1986.19-425

LaFrance, D. L., & Tarbox, J. (2020). The importance of multiple exemplar instruction in the establishment of novel verbal behavior. *Journal of Applied Behavior Analysis, 53*(1), 10–24. https://doi.org/10.1002/jaba.611

Leaf, R., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. DRL Books.

Leaf, J. B., Oppenheim-Leaf, M. L., Leaf, R., Courtemanche, A. B., Taubman, M., McEachin, J., Sheldon, J. B., & Sherman, J. A. (2012). Observational effects on the preferences of children with autism. *Journal of Applied Behavior Analysis, 45*(3), 473–483. https://doi.org/10.1901/jaba.2012.45-473

Leaf, J. B., Dale, S., Kassardjian, A., Tsuji, K. H., Taubman, M., McEachin, J. J., & Leaf, R. B. (2014a). Comparing different classes of reinforcement to increase expressive language for individuals with autism. *Education & Training in Autism Developmental Disabilities*. https://doi.org/10.2307/24582349

Leaf, J. B., Leaf, R., Taubman, M., McEachin, J., & Delmolino, L. (2014b). Comparison of flexible prompt fading to error correction for children with autism spectrum disorder. *Journal of Developmental & Physical Disabilities, 26*(2), 203–224. https://doi.org/10.1007/s10882-013-9354-0

Leaf, J. B., Leaf, R., Alcalay, A., Leaf, J. A., Ravid, D., Dale, S., Kassardjian, A., Tsuji, K., Taubman, M., McEachin, J., & Oppenheim-Leaf, M. (2015). Utility of formal preference assessments for individuals diagnosed with autism spectrum disorder. *Education & Training in Autism & Developmental Disabilities, 50*(2), 199–212.

Leaf, J. B., Cihon, J. H., Leaf, R., McEachin, J., & Taubman, M. (2016a). A progressive approach to discrete trial teaching: Some current guidelines. *International Electronic Journal of Elementary Education, 9*(2), 361–372.

Leaf, J. B., Leaf, J. A., Alcalay, A., Kassardjian, A., Tsuji, K., Dale, S., Ravid, D., Taubman, M., McEachin, J., & Leaf, R. (2016b). Comparison of most-to-least prompting to flexible prompt fading for children with autism spectrum disorder. *Exceptionality, 24*(2), 109–122. https://doi.org/10.1080/09362835.2015.1064419

Leaf, J. B., Leaf, R., McEachin, J., Taubman, M., Ala'i-Rosales, S., Ross, R. K., Smith, T., & Weiss, M. J. (2016c). Applied behavior analysis is a science and, therefore, progressive. *Journal of Autism & Developmental Disorders*, *46*(2), 720–731. https://doi.org/10.1007/s10803-015-2591-6

Leaf, J. B., Cihon, J. H., Ferguson, J. L., McEachin, J., Leaf, R., & Taubman, M. (2018a). Evaluating three methods of

stimulus rotation when teaching receptive labels. *Behavior Analysis in Practice, 11*(4), 334–349. https://doi.org/10.1007/s40617-018-0249-5

Leaf, J. B., Leaf, R., Leaf, J. A., Alcalay, A., Ravid, D., Dale, S., Kassardjian, A., Tsuji, K., Taubman, M., McEachin, J., & Oppenheim-Leaf, M. L. (2018b). Comparing paired-stimulus preference assessments with in-the-moment reinforcer analysis on skill acquisition: A preliminary investigation. *Focus on Autism & Other Developmental Disabilities, 33*(1), 14–24. https://doi.org/10.1177/1088357616645329

Leaf, R., Leaf, J. B., & McEachin, J. (2018c). *Clinical judgment.* Different Roads to. *Learning.*

Leaf, J. B., Cihon, J. H., Ferguson, J. L., Leaf, R., & McEachin, J. (2019). Comparing no-no prompt to flexible prompt fading to teach expressive labels to individuals diagnosed with autism spectrum disorder. *Education and Training in Autism & Developmental Disabilities, 54*(3), 274–287.

León, A. L., & Rosales, R. (2018). Effects of bilingual tact instruction for a child with communication disorder. *Journal of Behavioral Education, 27*(1), 81–100. https://doi.org/10.1007/s10864-017-9272-9

Lerman, D. C., Dittlinger, L. H., Fentress, G., & Lanagan, T. (2011). A comparison of methods for collecting data on performance during discrete trial teaching. *Behavior Analysis in Practice, 4*(1), 53–62. https://doi.org/10.1007/BF03391775

Lipschultz, J., & Wilder, D. A. (2017). Recent research on the high-probability instructional sequence: A brief review. *Journal of Applied Behavior Analysis, 50*(2), 424–428. https://doi.org/10.1002/jaba.378

Liptak, G. S., Benzoni, L. B., Mruzek, D. W., Nolan, K. W., Thingvoll, M. A., Wade, C. M., & Fryer, G. E. (2008). Disparities in diagnosis and access to health services for children with autism: Data from the National Survey of Children's Health. *Journal of Developmental & Behavioral Pediatrics, 29*(3), 152–160.

Love, J. R., Carr, J. E., Almason, S. M., & Petursdottir, A. I. (2009). Early and intensive behavioral intervention for autism: A survey of clinical practices. *Research in Autism Spectrum Disorders, 3*(2), 421–428. https://doi.org/10.1016/j.rasd.2008.08.008

Majdalany, L. M., Wilder, D. A., Greif, A., Mathisen, D., & Saini, V. (2014). Comparing massed-trial instruction, distributed-trial instruction, and task interspersal to teach tacts to children with autism spectrum disorders. *Journal of Applied Behavior Analysis, 47*(3), 657–662. https://doi.org/10.1002/jaba.149

Martin, N. A., & Brownell, R. (2011). *Expressive one-word picture vocabulary test-4 (EOWPVT-4)*. Academic Therapy Publications.

Sam, A., & AFIRM Team. (2016). *Discrete trial training*. National Professional Development Center on Autism Spectrum Disorder, FPG Child Development Center, University of North Carolina. http://afirm.fpg.unc.edu/discrete-trial-training

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism & Other Developmental Disabilities, 16*(2), 86–92. https://doi.org/10.1177/108835760101600204

Soluaga, D., Leaf, J. B., Taubman, M., McEachin, J., & Leaf, R. (2008). A comparison of flexible prompt fading and constant time delay for five children with autism. *Research in Autism Spectrum Disorders, 2*(4), 753–765. https://doi.org/10.1016/j.rasd.2008.03.005

Sparrow, S. S., Cicchetti, D. C., & Saulnier, C. A. (2016). *Vineland adaptive behavior scales, (Vineland-3)*. Pearson.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Taubman, M. T., Leaf, R. B., McEachin, J. J., Papovich, S., & Leaf, J. B. (2013). A comparison of data collection techniques used with discrete trial teaching. *Research in Autism Spectrum Disorders, 7*(9), 1026–1034. https://doi.org/10.1016/j.rasd.2013.05.002

Taylor, B. A., & Harris, S. L. (1995). Teaching children with autism to seek information: Acquisition of novel information and generalization of responding. *Journal of Applied Behavior Analysis, 28*(1), 3–14. https://doi.org/10.1901/jaba.1995.28-3

Willis, T. J., & LaVigna, G. W. (1991). Reinforcement inventory. In T. J. Willis, G. W. LaVigna, & A. M. Donnellan (Eds.), *Behavior assessment guide* (pp. 84–91). Institute for Applied Behavior Analysis.

Wong, E., Ferguson, J. L., Milne, C. M., Cihon, J. H., Leaf, J. B., McEachin, J., Leaf, R., Schulze, K., & Rudrud, E. (2020). Evaluating three methods of the presentation of target stimuli when teaching receptive labels. *Behavioral Interventions, 35*(4), 542–559. https://doi.org/10.1002/bin.1744