Despite its advantages, discrete-trial instruction (DTI) has been criticized for producing rote responding. Although there is little research supporting this claim, if true, this may be problematic given the propensity of children with autism to engage in restricted and repetitive behavior. One feature that is common in DTI that may contribute to rote responding is the prompting and reinforcement of one correct response per discriminative stimulus. To evaluate the potential negative effects of rote prompts on varied responding, we compared the effects of modeling rote versus varied target responses during the teaching of intraverbal categorization. We also evaluated the effects of these procedures on the efficiency of acquisition of any one correct response. For all four children, any increase in varied responding was fleeting, and for two participants, acquisition was slower in the variable-modeling condition.

Key words: autism, discrete-trial instruction, rote responding, variability

Despite the success of using discrete-trial instruction (DTI) to teach a variety of skills to children diagnosed with autism spectrum disorder (ASD), this teaching method has sometimes been criticized for producing rote responding (Cihon, 2007). Although there does not appear to be any research supporting this claim, if true, this may be problematic given the propensity of children with ASD to engage in restricted and repetitive behavior (American Psychiatric Association, 2013). One potential reason for this criticism may stem from the fact that DTI trials are often arranged such that one correct response is repeatedly prompted and reinforced for a given discriminative stimulus. For example, a therapist may prompt and reinforce “I’m fine” following the discriminative stimulus “how are you?” despite the fact that other responses such as “good, thanks” would also be socially acceptable. Thus, it is possible that the prompting and reinforcement of one correct response narrows the range of responses that (a) are emitted and (b) contact reinforcement, potentially contributing to the development of rote responding.

One approach to counter the potential negative effects of DTI procedures may be to
prompt and reinforce multiple correct responses during teaching, which is consistent with the response generalization strategies of teaching sufficient exemplars and training loosely described by Stokes and Baer (1977). However, simply broadening the contingency class by increasing the range of responses eligible for reinforcement may not be sufficient to promote varied responding. In fact, basic research has shown that contingencies that allow but do not require varied responding produce invariant (rote) responding over repeated exposure to the contingency (e.g., Antonitis, 1951; Neuringer, Kornell, & Olufs, 2001). For example, in Experiment 2, Neuringer et al. (2001) showed the importance of reinforcement contingencies on varied responding with rats that were required to complete a three-response sequence. For rats in the variability group (Var), low-probability response sequences were reinforced based on an equation that took into account the recency of the last instance of the response sequence. For rats in the yoked group (Yoke), the frequency and distribution of reinforcement was yoked to the Var group, thereby allowing but not requiring varied responding. Varied responding was significantly higher in the Var group relative to the Yoke group, suggesting that the reinforcement contingency and not extinction-induced variability was responsible for increases in varied responding. In Experiment 3 of the same study, there were two groups: Var (described above) or Repetition (Rep; in which reinforcement was contingent on a single response sequence). A comparison of varied responding across the Yoke group in Experiment 2 and the Rep group in Experiment 3 (Figures 4 and 7; Neuringer et al.) revealed comparable levels of varied responding. These results suggest that contingencies that simply allow but do not require varied responding are likely to result in invariant responding at comparable levels to when reinforcement is contingent on one correct response. Although some basic research suggests that varied responding will only increase when varied responding is differentially reinforced, other basic studies suggest that there are at least some conditions under which varied responding might be observed when contingencies allow but do not require varied responding, such as when such contingencies are preceded by a history of differential reinforcement of varied responding (e.g., Saldana & Neuringer, 1998). The ability to produce varied responding under contingencies that allow but do not require variability has practical implications if we presume that such contingencies are more commonly encountered in the natural environment than contingencies that require varied responding. For example, Rodriguez and Thompson (2015) suggested that natural consequences for variant or invariant responding (e.g., rolling one’s eyes after hearing a repeated joke) may be too subtle or intermittent to function as reinforcers or punishers for children diagnosed with ASD. Thus, research aimed at exploring the conditions under which varied responding will be observed or maintained under contingencies that allow but do not require variability is warranted. Like a history of reinforcement for varied responding (Saldana & Neuringer, 1998), it is possible that other supplemental strategies, such as modeling varied responding, may be sufficient to bring multiple responses in contact with reinforcement when reinforcement is provided for any correct response. To this point, several researchers have suggested the importance of including varied models to increase varied responding during social interactions (Weiss, LaRue, & Newcomer, 2009) and play (Jahr & Eldevik, 2002). However, to our knowledge only one study has evaluated the effects of modeling variability when prompting within DTI. Carroll and Kodak (2015) compared two procedures for modeling variability during the teaching of intraverbal categorization to two participants diagnosed with ASD. In the first condition, the experimenter prompted one of
20 different combinations of three of the four primary target exemplars (e.g., “train, bike, bus” in response to “tell me some vehicles”). The second condition was identical to the first condition except that the experimenter presented three of four secondary target exemplars following the delivery of reinforcement for a correct response (e.g., “That’s right! Car, boat, and airplane are vehicles too”). The delivery of secondary targets to which the participant is not required to respond during the consequent event of a trial is referred to as instructive feedback. Differential consequences were not provided for varied responding in either condition. The cumulative frequency of novel combinations of category exemplars (e.g., “bike, bus, train” or “bus, train, bike”) and novel exemplars (e.g., “camper, tank, bus” would be counted as two novel exemplars) across sessions increased across both variable-modeling and variable-modeling-plus-instructive-feedback conditions during an acquisition task for both participants. However, the variable-modeling-plus-instructive-feedback condition resulted in higher levels of varied responding across sessions, with a greater number of cumulative novel combinations for one participant and a greater number of cumulative novel exemplars for both participants. Although the main purpose of the study was to evaluate the effects of instructive feedback on varied and novel responses, the results also suggest that varied models alone can produce varied responding.

The current study extended Carroll and Kodak (2015) by comparing the effects of varied- and rote-model prompts on varied responding during the teaching of intraverbal categorization. As previously noted, DTI typically involves the repeated prompting and reinforcement of one response. A comparison of the effects of varied and rote modeling would allow for an evaluation of any potential undesirable effects of rote modeling on varied responding. Such a comparison could also reveal any desirable effects of variable modeling on varied responding. In other words, differential levels of varied responding with rote models versus variable models would suggest that the prompting of one response during DTI may contribute to rote responding, and providing variable models may be one way to address that limitation. However, because prompting multiple correct responses may interfere with acquisition, we also evaluated the efficiency of each modeling procedure.

To answer our experimental questions, we adopted a translational approach in which the skill we taught was relevant to our participants’ early intervention goals, but the response targeted for variability was also aligned with basic research. Specifically, we sought to teach our participants intraverbal categorization (e.g., responding “dog, cat, horse” when instructed, “tell me some animals”), while also targeting variability in the sequence in which the target exemplars were provided (e.g., “cat, horse, dog” vs. “horse, cat, dog”), with only one category targeted per condition. We targeted intraverbal categorization given the importance of varied vocal-verbal behavior for the development of strong vocal-verbal repertoires, and given the tendency for individuals with ASD to engage in rote responding (see Rodriguez & Thompson, 2015, for a review of research comparing levels of response variability across individuals with and without ASD). Targeting variability in the sequence of exemplars also aligned with the type of varied responding typically evaluated in basic research and allowed us to equate the number of exemplars (three) that participants were exposed to across trials and conditions, such that we could more easily isolate the effects of different types of prompts on varied responding and skill acquisition.

**METHOD**

**Participants, Setting, and Materials**

Participants included four children diagnosed with ASD between the ages of 3 and 8 years,
who were receiving early intervention services at a university-based center. Three participants received 2 to 3 years of early intervention services prior to the start of the study; Ronda received early intervention services for 4 months prior to her participation. Victor was a 5-year-old Caucasian boy, Ivan was an 8-year-old African American boy, Adam was a 5-year-old Arab American boy, and Ronda was a 3-year-old Caucasian girl. On the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP; Sundberg, 2008), Victor, Ivan, and Ronda were primarily Level 2 learners (i.e., learners with a developmental skill level comparable to a typically developing child aged 18 to 30 months), and Adam was a Level 3 learner (i.e., learner with a developmental skill level comparable to a typically developing child aged 30 to 48 months). All participants used short sentences (i.e., at least three to four words) to mand and tact, and had a previous history of intraverbal programming (e.g., feature, function, class) with multiple exemplars, although none had received DTI that specifically targeted varied responding.

Sessions were conducted in either a 2.5-m by 2.5-m private session room in which the participant received early intervention services (Victor, Adam, and Ronda) or a living area of the participant’s home (Ivan). The participant sat next to or across from the experimenter at a child-sized table. Each session consisted of 10 trials and was approximately 5 to 10 min. Two to four sessions were conducted per day, up to 5 days per week. Materials included data sheets and reinforcers.

Response Measurement and Interobserver Agreement

During each trial, paper-and-pen data were collected on the participant’s vocal response sequence verbatim (i.e., each word emitted by the participant was recorded in the order in which it was stated), and whether the vocal response occurred prior to or following the controlling prompt (i.e., independent or prompted responding, respectively). To evaluate differences in varied responding as well as the rate of acquisition, data were summarized to gather information on the following primary dependent variables for each session: (a) the percentage of trials with correct responding, and (b) the number of unique response sequences in a session. A correct response (independent or prompted) was defined as the participant providing three appropriate exemplars for the category, regardless of a particular response sequence (e.g., “dog, cat, horse” or “cat, dog, mouse” as exemplars of animals). The participant had 5 s to initiate responding, and a response was considered complete if the participant ceased responding for 5 s. For example, if the participant gave two exemplars, s/he had 5 s after providing the second exemplar to include a third and final exemplar, or the response would be counted as an error. A response was scored as independent or prompted depending on whether it occurred prior to or following the delivery of the prompt, respectively. For prompted responding, the sequence emitted by the participant did not need to match the sequence prompted by the experimenter to be scored as correct.

The number of unique response sequences was calculated by determining the number of trials in which there was a correct response and the order of exemplars differed from the order of exemplars in response sequences in all preceding trials in that session. Thus, if a participant alternated between two correct sequences (e.g., “dog, horse, cat”; “cat, horse, dog”; “dog, horse, cat”; “cat, horse, dog” and so on) in a session, the number of unique response sequences would be two. A response sequence containing a novel exemplar (i.e., an exemplar that was never prompted in the study, such as “mouse” in the previous example) was also scored as unique; this happened only once with Adam. We also calculated the number of
cumulative unique sequences across sessions per target category.

Finally, to further characterize the level of varied responding and assist with interpretation of the results, we conducted additional post-hoc data analyses of both independent and prompted responses. For independent responses, we determined the relative frequency of the three most common response sequences and the occurrence of generative responding for each category. The dominant response was defined as the response sequence that occurred most frequently across sessions. A generative response was defined as the occurrence of either of the two response sequences that were never modeled throughout the study (i.e., recombinative generalization; Axe & Sainato, 2010).

A second, independent observer collected data for 64% to 100% of sessions across participants. Interobserver agreement (IOA) was calculated for correct responding and response sequence on a trial-by-trial basis. An agreement for correct responding was defined as both observers scoring either correct or incorrect on a trial. An agreement for response sequence was defined as both observers scoring the same three exemplars in an identical order for both independent and prompted responses. Both data collectors had to score the same response(s) for correct responding, independent response sequence, and prompted response sequence for a trial to be scored as an agreement. We calculated IOA by dividing the number of trials with an agreement by the total number of trials (10), and then converting the quotient to a percentage. Mean IOA was 97% (range, 40% to 100%) for Victor, 99% (range, 90% to 100%) for Ivan, 99% (range, 70% to 100%) for Adam, and 99% (range, 90% to 100%) for Ronda.

Pre-assessment

We conducted a pre-assessment to identify unknown target responses to include in the comparison. Three categories (e.g., fruits, animals, and tools) were each presented four times in a session (resulting in 12 trials per session), during which the experimenter provided the discriminative stimulus, “tell me some [category].” If the participant responded incorrectly or did not respond within 5 s of the discriminative stimulus, we did not provide feedback and began the next trial. Any category for which a participant provided two or more exemplars (e.g., “peach” and “pear” for the category fruit) was excluded, and an alternate category was assessed. This process continued until we identified four categories for which no more than one exemplar was provided for two consecutive sessions. Three of the four participants (Victor, Adam, and Ronda) did not emit any correct exemplars for the categories selected for teaching. Ivan emitted one correct exemplar (i.e., football) for one category (i.e., outside sports). Exemplars that were emitted during the pre-assessment (or baseline) were not included in the three exemplars that were modeled during the experiment. All correct responses were reinforced with praise and a highly preferred edible or tangible (identified via a one-trial multiple-stimulus-without-replacement preference assessment similar to DeLeon et al., 2001).

Once we identified the target categories, we conducted a brief echoic pre-assessment to confirm that participants could echo each of the exemplars that would be modeled in the experiment. Twelve target exemplars (i.e., three exemplars across four categories) were interspersed with nontarget words. The order of presentation was randomized, and praise was delivered for correct responding. The echoic pre-assessment for Victor’s Set 3 was conducted after teaching Sets 1 and 2 and consisted of six target exemplars interspersed with nontarget words. Participants correctly echoed 100% of the target exemplars.

Finally, we assessed whether the exemplars, when presented by themselves, evoked speaker or listener responses (i.e., served as
discriminated operants) prior to the start of our study. This was done to ensure that responses prompted during the experiment proper were not being taught as a response chain of arbitrary stimuli, particularly in the rote-modeling condition. The majority of targets were evaluated as tacts. Specifically, a picture of each exemplar was presented at least two times within a six-trial session and was interspersed with at least two other nontarget stimuli or stimuli from different categories. Differential reinforcement was provided for correct responses, and no consequence was provided for incorrect responses. For exemplars for which we anticipated a picture might not exert precise control over the target (e.g., headache, suntan, meow), we evaluated listener repertoires (i.e., selecting the target from an array of three stimuli contingent on the spoken word) or intraverbals (e.g., “What does a cat say?”). All other procedures (e.g., number of trials per session, differential reinforcement) were the same as the procedures for testing exemplars as tacts. If participants did not respond correctly on the first two presentations, we introduced a 0-s prompt delay for two sessions. We included prompting to ensure that participants were exposed to each exemplar as an independent operant. Targets that required teaching were equally distributed across categories. Victor could tact all exemplars in Sets 1 and 2 but required exposure to prompting for all exemplars in Set 3; Ivan was exposed to prompting for 5 of the 12 exemplars (two exemplars for each category in Set 1 and one exemplar for one category in Set 2); Adam was exposed to prompting for 7 of the 12 exemplars (one to two exemplars per set); and Ronda was exposed to prompting for 4 of the 12 exemplars.

**Experimental Design**

We used an adapted alternating treatments design (Sindelar, Rosenberg, & Wilson, 1985) within a multiple probe design across sets of categories. There were two categories per set with three exemplars per category. Categories in each set were matched based on the number of syllables for each exemplar and then randomly assigned to either the rote- or variable-modeling condition. There were, however, two occasions in which the assignment of categories to conditions was not random but rather counterbalanced with the assignment of conditions for another participant. Table 1 lists the categories and exemplars for each condition and set per participant.

**Procedure**

Prior to each session, we conducted a one-trial multiple-stimulus-without-replacement preference assessment (based on procedures described by DeLeon et al., 2001). The item selected was used as a reinforcer throughout the session unless the participant requested an alternative item, in which case the requested item was delivered for the remainder of the session or until a different item was requested.

**Differential reinforcement baseline (DSR BL).** Each session consisted of 10 trials of one category. During each trial, the experimenter delivered the discriminative stimulus “tell me some [category]” and provided the participant the opportunity to respond. Contingent on a correct response (regardless of response sequence), praise and the reinforcer were delivered. Contingent on an incorrect response or no response within 5 s, the experimenter provided no feedback and proceeded to the next trial. The criterion for moving from baseline to the modeling phase was stable responding across both conditions using visual inspection.

**Progressive-prompt delay (PPD).** This phase was identical to baseline, except that (a) a PPD was used in both conditions and (b) each category was assigned to either the rote- or variable-modeling condition. The PPD phase started with at least two sessions at a 0-s
prompt delay, during which the experimenter immediately modeled a correct response following the discriminative stimulus during each trial. Following two consecutive sessions with correct prompted responding at or above 90%, the experimenter provided a 2-s delay following the discriminative stimulus before modeling the correct response. Thereafter, the criterion to increase the prompt delay was one session in which 50% or more of the errors consisted of a no response (omission errors; Kodak, Fuchtman, & Paden, 2012); the progression of the prompt delays was 2 s, 5 s, 10 s, and 20 s. The criteria for increasing the prompt delay were held constant across conditions such that each condition progressed through the changes in prompt delay independently. This was done to emulate how the PPD would be applied in practice when the conditions are not being compared in a formal evaluation.

Following a correct independent response, the experimenter provided praise and a reinforcer. If the participant provided an incorrect exemplar, did not respond, or provided an incomplete response, the experimenter provided a vocal model of a correct response. Following a prompt, the participant had 5 s to initiate responding. The trial ended when there was no response, a 5-s pause in responding, or three exemplars were provided and reinforced.

An additional manipulation was made for Victor. After meeting criteria to increase the prompt delay to 5 s, the majority of Victor’s errors in the variable-modeling condition shifted to errors of commission (e.g., exemplars from the rote-modeling condition). For this reason, we changed the contingencies for Set 1 of the variable-modeling condition starting with session 24 such that only praise was provided for correct prompted responding in this condition.

Table 1
List of Categories and Exemplars for each Condition and Set per Participant

<table>
<thead>
<tr>
<th>Participant</th>
<th>Set 1 Category</th>
<th>Exemplars</th>
<th>Set 2 Category</th>
<th>Exemplars</th>
<th>Set 3 Category</th>
<th>Exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victor</td>
<td>Little Animals</td>
<td>Frog</td>
<td>Things in the Fridge</td>
<td>Milk</td>
<td>Inside Sports</td>
<td>Bowling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dog</td>
<td></td>
<td>Apple</td>
<td></td>
<td>Wrestling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mouse</td>
<td></td>
<td>Cheese</td>
<td></td>
<td>Hockey</td>
</tr>
<tr>
<td></td>
<td>Big Animals</td>
<td>Bear</td>
<td>Things in the Bathroom</td>
<td>Sink</td>
<td>Outside Sports</td>
<td>Tennis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horse</td>
<td></td>
<td>Klenex</td>
<td></td>
<td>Fishing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whale</td>
<td></td>
<td>Soap</td>
<td></td>
<td>Sailing</td>
</tr>
<tr>
<td>Ivan</td>
<td>Outside Sports</td>
<td>Tennis</td>
<td>Tools in the Garden</td>
<td>Bake</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Baseball</td>
<td></td>
<td>Lawn-mower</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fishing</td>
<td></td>
<td>Shovel</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inside Sports</td>
<td>Bowling</td>
<td>Tools in the Kitchen</td>
<td>Micro-wave</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hockey</td>
<td></td>
<td>Blender</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wrestling</td>
<td></td>
<td>Spoon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adam</td>
<td>Rhymes with Snowflake</td>
<td>Earth-quake</td>
<td>Rhymes with Haystack</td>
<td>Backpack</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Headache</td>
<td></td>
<td>Kayak</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cupcake</td>
<td></td>
<td>Racetrack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhymes with Batman</td>
<td>Toucan</td>
<td>Rhymes with Allow</td>
<td>Snowplow</td>
<td>Eyebrow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ronda</td>
<td>Big Animals</td>
<td>Bear</td>
<td>Foods</td>
<td>Eggs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horse</td>
<td></td>
<td>Sandwich</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whale</td>
<td></td>
<td>Cheese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little Animals</td>
<td>Frog</td>
<td>Drinks</td>
<td></td>
<td>Milk</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dog</td>
<td></td>
<td>Juice</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mouse</td>
<td></td>
<td>Water</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The top category for each participant and set was assigned to the rote-modeling condition and the bottom category was assigned to the variable-modeling condition.
Mastery criterion was two consecutive sessions with correct responding at or above 90%. However, we continued to conduct sessions across both conditions until mastery criterion was met for both conditions for two reasons. First, we wanted to maintain the alternating fashion in which conditions were conducted to minimize the likelihood that switching to conducting only one condition would increase the speed of acquisition for that condition relative to when it is alternated with another condition. The only exception to this rule was at the end of Set 1 for Victor. Second, we began to notice a pattern of varied responding in which the number of unique sequences increased but only temporarily. For this reason, we sometimes conducted additional sessions after meeting the mastery criterion to allow for an evaluation of the persistence of varied responding across targets.

**Rote-modeling condition.** For the category in the rote-modeling condition, the experimenter modeled the same response during each trial (e.g., “apple, orange, banana” on every trial).

**Variable-modeling condition.** For the category in the variable-modeling condition, the experimenter modeled the same three correct exemplars but in a different sequence during each trial (e.g., “dog, cat, horse” for the first trial, “dog, horse, cat” for the second trial, and “cat, horse, dog” for the third trial). To ensure that participants were exposed to a range of possible sequences during the variable-modeling condition while also allowing for a test of recombinative generalization (i.e., the occurrence of a response sequence to which the participant had not been exposed; Axe & Sainato, 2010), four of the six possible sequences were modeled throughout this phase. The specific response sequence modeled was based on a predetermined list in which the four possible response sequences that were modeled during the experiment were pseudorandomized in sets of four to increase the probability of equal exposure to each of the four prompted sequences. Specifically, if a model was necessary, the experimenter modeled the next sequence on the list that differed from the participant’s response in the previous trial. For instance, if the child responded “dog, cat, horse” on the first trial and erred or did not respond on the second trial, the experimenter modeled the next sequence on the list that differed from the response on the preceding trial (e.g., “horse, dog, cat”; note that if “dog, cat, horse” was next on the list, it would have been skipped to ensure that the model was different from the previous response). Each programmed sequence was crossed off after it was modeled, and the experimenter continued with the remainder of the list.

To isolate the effects of modeling varied responding, differential consequences were not provided for varied responding across rote- and variable-modeling conditions (i.e., reinforcement was provided for any correct sequence independent of variability).

**RESULTS**

The first purpose of our study was to examine whether incorporating multiple models would lead to increased varied responding. The left panel of Figures 1–3, and 4 show the number of unique response sequences for Victor, Ivan, Adam, and Ronda, respectively. During the DSR BL, none of the participants provided a correct response to any of the intraverbal categories. Following the implementation of the PPD, an initial but temporary increase in the number of independent unique response sequences was observed for at least the variable-modeling condition of the first set for each participant. For Victor (Figure 1), the number of unique response sequences temporarily increased in the variable-modeling condition but not the rote-modeling condition for Set 1. Because varied responding was not replicated
Figure 1. Data for Victor. Left panels depict the number of independent unique response sequences (closed symbols) and unique response sequences emitted following prompts during prompted trials (open symbols). Right panels depict the percentage of correct independent responses (closed symbols) and correct prompted responses (open symbols; calculated out of the number of prompted responses). DSR BL = Differential reinforcement baseline. PD = Prompt delay. PPD = Progressive prompt delay. Asterisks denote when mastery criterion was met for each condition.
with Set 2 for Victor, we evaluated a third set; however, as in Set 2, Victor independently emitted only one response sequence across both conditions. Thus, the temporary increase in varied responding in the variable-modeling condition was not replicated across sets for Victor, potentially due to sequence effects in which the lack of differential contingencies for varied responding during Set 1 affected responding during subsequent sets. For Ivan (Figure 2), the number of independent unique response sequences temporarily increased, albeit only slightly, in only the variable-modeling condition for Set 1; however, the number of independent unique response sequences increased across both conditions in Set 2. For Adam (Figure 3), the number of independent unique response sequences increased across both conditions for both sets, with a larger and more persistent increase in varied responding in the variable-modeling condition for Set 2. For Ronda (Figure 4), an initial but temporary
increase in the number of independent unique response sequences differentially occurred in the variable-modeling condition, and only one independent unique response sequence was emitted across both sets in the rote-modeling condition. For all participants, any initial varied responding that occurred decreased to only one independent unique response sequence by the time the mastery criterion was met (asterisks in the figures denote when mastery criterion was met for each condition).

The second purpose of our study was to examine whether incorporating multiple models within prompts would compromise efficiency during acquisition. The right panels of Figures 1–3, and 4 show the percentage of correct responses for Victor, Ivan, Adam, and Ronda, respectively. Because we selected
intraverbal categories for which participants provided either no (Victor, Adam, Ronda) or one correct exemplar (Ivan for one category), there were no correct response sequences during DSR BL (also see Supporting Information 1). Following the introduction of the PPD, one of two general patterns of responding was observed across participants: (a) equal rates of acquisition across both phases, or (b) slower acquisition in the variable-modeling condition. For Victor (Figure 1), whereas the rate of acquisition was comparable across both conditions for Set 2, it took twice as many sessions to reach the mastery criterion in the variable-modeling condition in Set 1, and three additional sessions to reach mastery in the variable-modeling condition for Set 3. Following low levels of correct responding in the variable modeling condition for Set 1 after completing teaching for Set 2, we reintroduced
the 0-s prompt delay followed by the PPD to regain mastery level performance in the variable-modeling condition. For Ivan (Figure 2) and Adam (Figure 3, Set 2), acquisition rates were similar across conditions, suggesting that modeling varied responses in the variable-modeling condition did not interfere with acquisition. However, unlike Adam, Ivan’s continued correct responding was differentially lower following mastery in the rote-modeling condition for one set (Set 2, Figure 2). For Ronda (Figure 4), acquisition was notably slower in the variable-modeling condition, taking nearly twice (Set 1) or more than twice (Set 2) as many sessions to reach mastery criterion in the variable-modeling condition.

DISCUSSION

Our results suggest that although there may be some initial varied responding when providing variable modeling during teaching, such effects may be temporary. Thus, it seems unlikely that prompting and reinforcing one target response per discriminative stimulus, which is typical of DTI, uniquely contributes to rote responding in individuals with ASD. In other words, invariant responding may occur whether or not therapists provide variable models. Further, providing variable modeling led to slightly slower acquisition for some participants, suggesting that there can be a disadvantage to incorporating variable modeling compared to rote modeling, particularly when considered in light of the temporary effects of variable modeling on varied responding. Future research could explore the most efficient procedures for teaching multiple correct responses while also promoting the occurrence and maintenance of varied responding. For example, it is possible that additional intervention components (e.g., instructive feedback; Carroll & Kodak, 2015) could be included with variable modeling.

The temporary effects of the variable-modeling condition on varied responding may not be surprising in light of what we know about the effects of reinforcement. That is, reinforcement increases the future probability of the response that precedes it. A within-session analysis of the response sequences provided on each trial (see Supporting Information 2) revealed that one dominant response sequence emerged both within and across sessions for all participants, which supports the interpretation that the direct effects of reinforcement may have contributed to the development of invariant responding in the variable-modeling condition. Repeatedly engaging in one dominant response may also represent the least effortful pattern of responding. Thus, it seems unlikely that there would be a sustained increase in varied responding in the absence of reinforcement contingencies that require, rather than simply allow, varied responding, at least when prompting is no longer provided. However, additional research exploring the conditions under which varied responding will occur when reinforcement is provided for any correct response is needed, given the discrepancy between our results and those of Carroll and Kodak (2015) during the variable-modeling condition.

Whereas a temporary increase in the number of unique response sequences was observed only in the variable-modeling condition for at least one set for Victor, Ivan, and Ronda, there was an initial increase across both the rote- and variable-modeling conditions for Set 2 for Ivan and both Sets 1 and 2 for Adam. Despite efforts to reduce potential carryover effects by using an adapted alternating treatment design, the fact that (a) the conditions were similar in all regards except for the categories and exemplars targeted, and (b) the conditions were conducted in rapid alternation, may have led to indiscriminate responding. There may have also been carryover from the rote-modeling to the variable-modeling condition such that varied
responder responding may have persisted longer had variable-modeling and rote-modeling sessions not been intermixed.

Carryover effects may also explain the differences in the level of varied responding observed within the variable-modeling condition in Carroll and Kodak (2015) and in the current study (5 to 10 cumulative response sequences vs. 3 to 6 cumulative response sequences, respectively; see Supporting Information 1 for a summary of our participants’ cumulative unique sequences). Exposure to the variable-modeling plus instructive feedback condition may have increased the level of varied responding in the variable-modeling condition in Carroll and Kodak. Future research should assess the effects of variable modeling in isolation. However, it is worth noting that there were other differences between Carroll and Kodak and the current study that may have contributed to the differences in the level of varied responding during the variable-modeling condition across the studies. Examples include the size of the pool of exemplars, the number of varied sequences that were modeled, and the number of categories included in a session.

It is worth noting that, unlike other participants who met criteria for increasing the prompt delay from 0 s to 2 s within two to four sessions, Ronda required 12 sessions in the variable-modeling condition for Set 1. Initially, during the 0-s prompt delay of the variable-modeling condition, Ronda consistently echoed only a portion of the prompt correctly (e.g., “mouse, frog, mouse”). Because Ronda was able to echo three-part responses when the prompt remained constant in the rote-modeling condition, an inability to echo a three-part response cannot account for these errors. Instead, variations in the sequence of exemplars from trial to trial may have interfered with the stimulus control exerted by the prompt on a given trial (the loss of stimulus control due to competing responses to the same stimuli is consistent with a behavioral interpretation of forgetting; see Palmer, 1989). This alternative explanation is consistent with the types of errors she emitted. Specifically, repeating an exemplar (e.g., “dog, mouse, dog”) accounted for 51% of errors for prompted responses.

As previously noted, we adopted a translational approach to answer our experimental questions. Evaluating the level of varied responding in the sequence of exemplars allowed us to control the number of exemplars presented in each condition. Using response sequence as a measure of varied responding also allowed for measurement of recombinative generalization (Axe & Sainato, 2010). That is, after participants learned three exemplars for an intraverbal category, they could presumably emit the exemplars in any order without direct teaching. Toward this potential outcome, we modeled only four of the six possible response sequences. At least one generative response sequence occurred in at least one set for all participants.

We have depicted multiple measures of varied responding to highlight the different information provided by doing so. We used the number of unique response sequences as our primary dependent variable because it gave a summative depiction of varied responding that occurred initially while also showing that participants engaged in only a single response sequence over time. As a supplementary measure, we also depicted cumulative response sequences for a subset of sessions, starting with the introduction of the PPD until there were three consecutive sessions with only one dominant response sequence (Supporting Information 2). This was done to show the limited trial-by-trial varied responding and the emergence of one dominant response sequence. In the absence of data on socially acceptable levels of varied responding across types of responding (e.g., initiating a conversation), depicting varied responding in multiple ways may prove useful as researchers identify socially acceptable levels of varied responding, continue to refine
methods of teaching varied responding, and identify treatments that support the persistence of varied responding in the absence of direct reinforcement.

Several limitations of our study are worth noting. First, we did not start by teaching repertoires that have been shown to facilitate the emergence of intraverbal categorization, such as tacting exemplars, tacting categories, and listener-discrimination training by item and category (e.g., Grannan & Rehfeldt, 2012; Miguel, Petursdottir, & Carr, 2005). Several researchers have suggested that emergent responding is likely to be less rote. We chose not to start by teaching these related repertoires, as we were primarily interested in using the teaching of intraverbal categorization as a means of evaluating rote versus variable prompts. Second, each session consisted of 10 trials of the same category. We chose to include only one target question per condition to maintain consistency with previous studies that have evaluated the effects of differential reinforcement of variability on response variability (e.g., Contreras & Betz, 2016; Dracobly, Dozier, Briggs, & Juanico, 2017; Susa & Schlinger, 2012). In addition, we anticipated that supplementary interventions such as a lag schedule may be required to increase varied responding in subsequent evaluations if variable modeling was not sufficient. However, it is important to note that massed trials of one target do not allow one to ensure that responding is under appropriate stimulus control of the target question. This concern is somewhat mitigated in our study because we continued to conduct both conditions until mastery was achieved across both sets, and we observed correct first-trial performance across both sets.

Finally, we occasionally continued to conduct sessions beyond the mastery criterion, which may negatively impact the efficacy of contingencies such as lag schedules when subsequently applied. Given the importance of intraverbal categorization and variability in vocal verbal behavior, future researchers should evaluate whether varied responding could be achieved despite a relatively lengthy history of reinforcement of one rote response.

Given the propensity of individuals with ASD to engage in restricted and repetitive behavior and the results of the current study showing that invariant responding is likely to occur in the absence of reinforcement for varied responding, it would be useful to evaluate whether the temporary effects of variable modeling would also be observed in children of typical development. That is, it is unknown whether the fleeting effects of variable modeling during contingencies that allow but do not require variability are unique to individuals with ASD. A group comparison across individuals with ASD and peers matched for age and skill level would allow for an evaluation of the generality of our findings across populations. Future research should also be aimed at identifying the conditions under which varied responding will occur (and maintain) in the absence of differential reinforcement of varied responding.

REFERENCES
Contreras, B. P., & Betz, A. M. (2016). Using lag schedules to strengthen the intraverbal repertoires of


Received April 25, 2016
Final acceptance March 3, 2018
Action Editor, Tiffany Kodak

**SUPPORTING INFORMATION**

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