



Evaluation of Lag Schedules and Rules on Persistent Response Variability With Preschoolers in a Group

Katie M. Wiskow¹ · Jasmine Torrecillas^{1,2} · Haide Rocha^{3,4} · Allison DaSilva^{3,4}

Published online: 6 October 2020

© Association for Behavior Analysis International 2020

Abstract

In the present study, experimenters evaluated the influence of lag schedules of reinforcement in combination with accurate and inaccurate (complete and incomplete) rules on the response variability of naming category items for typically developing preschoolers in a group format. Results showed that when lag schedules were introduced with 2 categories, response variability generalized to the third category. Furthermore, after participants experienced the lag schedule, variability persisted when the contingency no longer required variability. Participants continued to vary their responses unless the rule and contingency required them to repeat responses. We discuss potential clinical applications of using lag schedules in a group format and including rules during teaching, as well as directions for future research in this area.

Keywords Categories · Group instruction · Intraverbals · Rules · Variability

Learning to vary responses may provide individuals with multiple ways to access reinforcers in their environment and assist with creativity and problem-solving skills. One method to increase response variability is the use of lag schedules of reinforcement. In a lag schedule, a response is reinforced if it is different from x number of previous responses (Neuringer, 2002). For example, in a Lag 1 schedule, a response is reinforced only if it varies from the previous response. There are many situations in daily life that might require a person to vary his or her response to access reinforcement. If someone asks, “What do you want to drink?” and the person replies, “Lemonade,” the first person might respond, “I do not have lemonade; what else do you

✉ Katie M. Wiskow
kwiskow@csustan.edu

¹ Department of Psychology and Child Development, College of Science, One University Circle, California State University, Stanislaus, Turlock, CA 95382, USA

² Present address: Fresno Unified School District, Fresno, CA, USA

³ California State University, Stanislaus, CA, USA

⁴ Present address: Kadiant, Modesto, CA, USA

want?” The second person will then need to ask for a different drink. Similarly, someone may ask, “What do you like to do for fun?” the person responds, “I like to play soccer,” and the first person responds, “I do not play soccer; what else do you like to do?” The second person will then need to state another conversation topic. Experimenters have used lag schedules to increase the variability of socially meaningful behaviors, such as conversation topics (Lepper, Devine, & Petursdottir, 2017), answering social questions (e.g., Lee, McComas, & Jawor, 2002; Susa & Schlinger, 2012), mands (e.g., Silbaugh & Falcomata, 2019), and food consumption (Silbaugh & Falcomata, 2017). Experimenters have also used lag schedules to increase varied and novel responses in academic skills, such as naming category items (e.g., Wiskow, Matter, & Donaldson, 2018).

Naming category items requires divergent intraverbal responding. Divergent intraverbal responding consists of one stimulus that occasions multiple possible responses. For example, “fruit” may evoke the responses “strawberry,” “banana,” “kiwi,” and “watermelon.” There is a variety of games that require categorical response variability, such as Family Feud, Scattergories, and Hive Mind. An individual’s ability to participate in these games may provide access to social reinforcers. In addition, creativity assessments such as the Torrance Tests of Creative Thinking include measures of divergent intraverbal responding (Torrance, 1990). Creativity may be an important component of problem solving, well-being, and adult success (Plucker, Beghetto, & Dow, 2004). Without variability, creativity is low (Neuringer, 2002).

Most of the research on response variability has been implemented at an individual level, but it may be especially important to understand the effects of lag schedules in a group context, like what naturally occurs in classrooms. Children often learn academic skills in a group instructional format in school; however, many of our evaluations of teaching procedures have occurred in an individual instructional format. Wiskow and Donaldson (2016) extended research on lag schedules by evaluating the effects of Lag 0 (i.e., fixed ratio [FR] 1) and Lag 1 schedules of reinforcement applied to a group of young children on the levels of varied responses to naming category items. In contrast to individual-level lag schedules, group-based lag schedules require individuals to vary their response from x number of previous responses in the group. Depending on the size of the group and the lag criterion, this can include varying their response from their peers’ previous response(s) and their own. For example, in a group-based Lag 1 schedule, the individual must say a response that is different from a peer’s during the previous turn. In a group-based Lag 2 schedule, with a group of two individuals, the individual must say a response that is different from the previous two responses (his or her own and the peer’s).

Wiskow and Donaldson (2016) delivered contingency-specifying instructions (i.e., complete, accurate rules) before each session. Specifically, the experimenter stated that the participant could earn a token for a correct response (FR 1) or a response that was different from the previous response (Lag 1). Lag 1 also included specific feedback following a repeated response (i.e., “Nice try, but [participant] already said [response].”). Results showed that when the experimenters implemented the Lag 1 schedule, varied responses increased at the group level and influenced varied responding during individual testing sessions in which the schedule remained unchanged (i.e., FR 1). These results demonstrate the benefits of using a lag schedule during group intervention; however, a limitation of this study is that the Lag 1 criterion

was undemanding and permitted participants to repeat their own responses and still meet the criterion. Individual lag schedules show that higher lag criteria produce higher levels of variability (e.g., Susa & Schlinger, 2012; Wiskow et al., 2018), but it is unknown whether a higher lag criterion may produce higher levels of variability in a group. Furthermore, experimenters could not isolate the influence of the lag schedule from the accurate instruction or the specific feedback because they implemented all those components simultaneously. It is possible that the instruction exerted some control over responding.

Rule-governed behavior is behavior controlled by antecedent verbal stimuli that specify behavioral contingencies (Cooper, Heron, & Heward, 2020). Rules are verbal stimuli that exert control over responding following differential reinforcement of rule-following behavior. After the relationship between rule-following behavior and reinforcement is established, an individual may respond differentially to novel instructions without having directly contacted the contingencies specified in the rule (i.e., generalized rule-governed behavior). For example, Tarbox, Zuckerman, Bishop, and Olive (2011) taught three children with autism spectrum disorder to follow simple rules using multiple-exemplar training. The rules initially consisted of if/then statements in the presence of different picture cards (e.g., “If this is an apple, then stand up.”). Subsequently, the researchers altered the rule presentation (e.g., “Stand up if this is an apple.”). After learning several exemplars, all participants responded correctly to novel rules, demonstrating generalized rule-governed behavior. In applied research and clinical settings, rules may be implemented to promote rapid behavior change to establish discriminated responding across experimental conditions. For example, if a rule exerts stimulus control over responding, then we may avoid introducing more intrusive prompts to evoke behavior change; we can simply tell the individual how to respond to earn reinforcement (e.g., Joyce & Chase, 1990; Podlesnik & Chase, 2006). Therefore, it is possible that providing an accurate rule from the onset of teaching will promote response variability, at least initially, by specifying the behavioral contingency to earn reinforcement (Joyce & Chase, 1990). Additionally, tracking individual responses during group instruction to implement lag schedules may be tedious and impractical for teachers to do consistently; therefore, additional research on the influence of rules on levels of variability would be beneficial.

A few recent studies have further evaluated the effects of rules on response variability during individual instruction and found little to no effect of the rule (Radley, Dart, Helbig, & Schrieber, 2018; Wiskow et al., 2018). Radley et al. (2018) compared the effects of a complete, accurate rule (i.e., contingency-specifying rule) to those of an incomplete rule (i.e., “Say something different.”) in combination with a Lag 2 schedule on social skills variability. Results showed that neither rule produced significantly more variability compared to the Lag 2 schedule alone; however, the experimenters never evaluated the rules without the lag schedule. In another recent study, Wiskow et al. (2018) evaluated the effects of an inaccurate rule prior to the introduction of the lag schedule. During baseline, the experimenters instructed the participants to “tell me a [category]” and reinforced all correct responses (i.e., FR 1). After baseline, the experimenters implemented a phase in which they varied the instruction following repeat responses. For example, if the participant said a response that did not vary from the previous response, the experimenter asked a variation of “Can you think of a different [category]?” Regardless of whether the participant varied

his or her response after the instruction, the experimenter delivered reinforcement and initiated the next trial. Results showed that the varied instruction alone did not increase response variability for either participant. These results were likely due to participants' lack of contact with contingencies that required response variability, despite a rule to do so.

Unlike Wiskow et al. (2018), instructions are often excluded as a component of the initial intervention in applied research on lag schedules; however, one or more participants frequently require extra teaching or feedback components, in addition to the lag schedule, to evoke varied responding (e.g., Contreras & Betz, 2016; Susa & Schlinger, 2012). When additional teaching is needed, it often includes an accurate rule that is introduced simultaneously with lag schedules to produce discriminated responses. For example, one participant in Contreras and Betz (2016) required extra variability training that consisted of specific feedback to “try and tell me something different” (accurate rule), plus a picture prompt. In another study, Dracobly, Dozier, Briggs, and Juanico (2017) established stimulus control of response variability (repetitive and varied responding) of block placements by visual, schedule-correlated stimuli. In training, the experimenters provided contingency-specifying instructions (i.e., accurate rules) to either make the same response (repetitive, Lag 0) or make a different response (varied, Lag 1, and Lag 4). Results showed that all participants engaged in discriminated responding within seven training sessions; however, the experimenters did not test for the maintenance of variability in the absence of the lag schedule. A test for maintenance is important because it is unreasonable to expect a lag schedule to always be in effect in an individual's environment.

Studies that included tests for the maintenance of variability have produced mixed results. Lee et al. (2002) demonstrated increased response variability during the lag reinforcement phase, but response variability returned to baseline levels when the lag schedule was removed, whereas Heldt and Schlinger (2012) observed response maintenance of variability in tacting items in a picture following a Lag 3 schedule. At the group level, Wiskow and Donaldson (2016) observed generalization of varied responding to an individual testing environment, but only in the category that was exposed to the Lag 1 schedule with an accurate rule during group sessions; no generalization of variability to the other category in group or individual sessions occurred. It is possible that Wiskow and Donaldson (2016) observed generalization of variability to the individual testing sessions because the rule was the same as that of the Lag 1 group sessions when varied responses were required.

Thus, the ineffectiveness of a rule alone to produce initial response variability is likely due to a lack of differential reinforcement for varied responses. For example, Wiskow et al. (2018) varied the instruction while maintaining an FR 1 schedule of reinforcement for correct responding (i.e., inaccurate rule) and observed no effect, whereas previous research that implemented lag schedules of reinforcement simultaneously with the addition of accurate instructions or other prompting procedures observed increased response variability (e.g., Contreras & Betz, 2016; Dracobly et al., 2017; Wiskow & Donaldson, 2016). It would be beneficial to determine under what conditions an instruction alone may produce varied responding without requiring specific consequence-based strategies (e.g., lag schedule of reinforcement) in group teaching formats. If the rule alone produces response variability, it would negate the need to contact differential reinforcement for varied responses (i.e., lag schedule). This

might also be easier for teachers because they can simply read the rule versus track individual responses and determine when a particular response meets the lag criterion for reinforcement. In cases in which a lag schedule is necessary to increase varied responses, it is equally important to determine whether that contingency can be subsequently removed or implemented intermittently and high levels of variability maintain. Understanding the interaction between rules and lag schedules in a group format may provide guidelines for effective group-based instruction that produces persistent response variability, which has important implications for many types of learning (Neuringer, 2002). Furthermore, group-based lag schedules require participants to attend to the responses of their peers to discriminate what response(s) will be reinforced; this can promote observational learning and lead to efficient teaching arrangements.

Previous research has not specifically explored the influence of complete and accurate instructions (C) and incomplete and inaccurate instructions (I) on response variability in a group, or whether a rule may facilitate response generalization to other categories and maintenance. The literature on lag schedules to date primarily consists of children with autism because they often engage in repetitive behaviors. However, variability is a universally important dimension of behavior because it may increase divergent responding that is associated with creativity.

The purpose of this translational study was to extend the work of Wiskow and Donaldson (2016) and evaluate the influence of lag schedules of reinforcement on the response variability of intraverbal responses (i.e., naming category items) in a group format with typically developing children. Furthermore, the experimenters maintained the inaccurate rule in FR 1 Vary (I) during the introduction of lag schedules and the reversal to FR 1 Vary (I) to test for the maintenance of variability with the rule alone. Response variability maintained in the first reversal to the FR 1 Vary (I) phase after participants experienced lag schedules of reinforcement. Thus, the experimenters implemented a Lag 0 No Vary (C) phase to decrease variability and reintroduced FR 1 Vary (I) to determine if the rule would reestablish high levels of variability without lag differential reinforcement.

Method

Participants and Setting

Six typically developing children participated. The experimenters divided the children into groups based on their availability. Group 1 consisted of Ben and Aria. Ben was a 3-year-old boy, and Aria was a 5-year-old girl. Ben and Aria were siblings. Group 2 consisted of Megan and April. Megan was a 3-year-old girl, and April was a 4-year-old girl. Group 3 consisted of Miriam and Elis. Miriam was a 5-year-old girl, and Elis was a 4-year-old girl. Megan turned 4 years old and Elis turned 5 years old over the course of the study. Experimenters selected participants who acquired the category responses quickly, were able to maintain those responses, and were consistently present and willing to participate in sessions.

Experimenters conducted all sessions at an on-campus child development center in central California. Sessions took place in a small lab room at the center equipped with a

one-way mirror. During each session, participants sat next to each other at a rectangular or circular table. Each participant had an individual token board in front of him or her on the table, and the experimenter sat across from both participants.

Materials

Experimenters used token boards, tokens, picture cards, and rewards. Pictures corresponded with items in each of the categories as visual (tact) prompts during pretraining sessions. Each picture card showed one item. The rewards included tickles, high fives, dances, bubbles, smelly stickers, smelly stamps, kinetic sand, and soda- and candy-scented lip balm that the experimenter rubbed on the back of participants' hands. The experimenters created and color-coded token boards to serve as schedule-correlated stimuli during the initial FR 1 Vary (I) and lag schedule phases of the study to assist with discrimination between the contingencies without varying the rule. All pretraining and group session token boards had nine spaces for tokens.

Response Measurement

Experimenters measured participants' intraverbal responses. The experimenter instructed a participant to "tell me a [category]." The experimenter scored a response as *correct (independent)* if the participant emitted the response within 10 s of the experimenter's presentation of the discriminative stimulus (S^D) and the response matched the stated category (e.g., saying "pepper" when instructed to name a spice). The experimenter scored a response as *incorrect* if the participant emitted the response within 10 s of the S^D but the response did not match the category (e.g., saying "tomato" when instructed to name a spice), or in the absence of a response within 10 s of the S^D . The experimenters calculated the *average lag score* for each session (Wiskow et al., 2018) by first assigning a lag score to each correct response. Experimenters determined the lag score for each response by counting the number of different previous responses until a repeat response occurred. For example, if the responses were mozzarella, cheddar, swiss, cheddar, cheddar, and pepper jack, the lag score for each response would be mozzarella (0), cheddar (1), swiss (2), cheddar (1), cheddar (0), and pepper jack (3). The average was then calculated by obtaining the sum of the lag scores (7) and dividing by the number of correct responses with a lag score (5). In this example, the average lag score is 1.4. In extended sessions with Miriam and Elis, the experimenters converted the sum of the lag scores into a *percentage of maximum variability*. First, the experimenters determined the lag scores as described previously. Second, the experimenters determined the maximum variability score by giving each correct response an increasing number (i.e., 0, 1, 2, 3) and adding these values. Finally, they calculated the percentage of maximum variability by dividing the sum of the lag scores by the maximum variability score and multiplying by 100. For example, the maximum variability score for a session with six correct responses is 15 (0 + 1 + 2 + 3 + 4 + 5). Using the previous example, the sum of the lag scores (7) divided by the maximum variability score (15) equals 47% of maximum variability.

Interobserver Agreement

The experimenters calculated interobserver agreement (IOA) on a trial-by-trial basis during an average of 52% of sessions in each phase across groups. The experimenter scored an agreement if the primary and secondary observers recorded the same lag score within a trial and a disagreement if the lag scores were different. The experimenter calculated IOA by dividing the number of agreements by the total number of trials and multiplying by 100. Average IOA across phases for the average lag score was 99% (range 80%–100%) for Group 1, 99% (range 80%–100%) for Group 2, and 99% (range 40%–100%) for Group 3.

Procedure

The experimenters conducted six sessions per day, 2 to 4 days per week, for Group 1, except for 1 day when only three sessions were conducted. The experimenters conducted six or nine sessions per day, 1 to 2 days per week, for Group 2, except for 1 day when only three sessions were conducted. The experimenters conducted 1 to 18 sessions per day, 1 to 4 days per week, for Group 3. Participants in Group 3 attended the center full time; therefore, the experimenters were able to conduct sessions in the morning and afternoon. The experimenter randomized the order of the sessions so that each category was conducted no more than two times in a row. However, there were a few exceptions. Group 1 had one instance where the experimenter conducted a flowers category session three times in a row and another with the cheeses category three times in a row. Group 2 had one instance where the experimenter conducted sessions with the sports category three times in a row and another with the dogs category six times in a row. The experimenter conducted the dogs category six times in a row during the Lag 0 No Vary (C) phase. Group 3 had one instance where the experimenter conducted sessions with the birds category four times in a row.

The experimenters conducted the preference assessment, pretest, and pretraining sessions in an individual format. The experimenter conducted all subsequent sessions in groups. Following an initial FR 1 Vary (I) phase, the experimenters implemented lag schedules in the group. After the initial lag schedule phase, the experimenters conducted reversals between Lag 0 No Vary (C) and FR 1 with accurate or inaccurate (complete/incomplete) rules to compare the effects of schedules of reinforcement and accuracy of the instructions on response variability. Specifically, the experimenters introduced Lag 0 No Vary (C) to reduce variability and conducted reversals to FR 1 Vary (I) to determine whether the participants would again vary their responses after receiving praise and a token for repeat responses (i.e., history of differential reinforcement for repeat responses). For Groups 2 and 3, the experimenters also evaluated if varied responding would decrease when the experimenter provided an accurate rule (i.e., FR 1 No Vary [C]). Table 1 summarizes the conditions the experimenters evaluated across the groups. The experimenters made phase changes based on visual inspection.

Preference Assessment The experimenters implemented a multiple-stimulus without-replacement preference assessment with picture cards to identify a preference hierarchy of rewards. Experimenters conducted each preference assessment with one participant

Table 1 Conditions Evaluated With Each Group of Participants

Group 1 Ben and Aria	Group 2 Megan and April	Group 3 Miriam and Elis
FR 1 Vary (I) – red	FR 1 Vary (I) – red	FR 1 Vary (I) – red
Lag 0 No Vary (C) – red	Lag 0 No Vary (C) – red	Lag 0 No Vary (C) – red
Lag 1 Vary (I) – purple	Lag 2 Vary (I) – orange	Lag 1 Vary (I) – purple
Lag 2 Vary (I) – orange	Lag 3 Vary (I) – purple	Lag 3 Vary (I) – orange
	FR 1 No Vary (C) – red	FR 1 No Vary (C) – red
		FR 1 Vary (C) – red
		Lag 5 Vary (I) – orange

Note. FR = fixed ratio; I = incomplete, inaccurate instruction; C = complete, accurate instruction. Followed by color of token board

at a time. The experimenter arranged nine picture cards corresponding to tangible items and social interactions on the table in front of the participant, stated the name of the tangible item or social interaction while pointing to each picture, and then instructed the participant to “pick the one that is your favorite.” When the participant touched a picture card, the experimenter removed the remaining picture cards and provided the participant access to the reward. Then, the experimenter removed the previously chosen picture card, rotated the remaining picture cards, and conducted the next trial. The experimenter continued this process until the participant chose all items or the participant refused to select any remaining items.

The experimenters priced the tangible items and social interactions at one token, three tokens, six tokens, or nine tokens. Experimenters determined the prices by comparing the preference hierarchies of individual participants in each group and averaging the rank values. The price of the items remained constant across individual pretraining sessions and group sessions, except that the experimenters changed the price of stickers for Miriam and Elis at Session 13 because they consistently chose stickers over higher priced items when they earned all their tokens.

The experimenters used the tokens primarily to serve a discriminative function rather than a reinforcing function and maintained token delivery throughout all phases for consistency. Experimenters initially decided to use a token economy during sessions for four main reasons. First, experimenters paired the color of the token boards with different lag schedules during the initial lag schedule comparison phase to assist with the discriminability of the contingencies (e.g., Dracobly et al., 2017). Second, participants could compare their tokens to peer tokens to track their progress and enhance motivating operations (e.g., Wiskow & Donaldson, 2016). Third, the delivery of tokens may be more salient than praise alone because it provides visual feedback. Fourth, the use of tokens establishes initial motivation for responding during sessions (to access backup rewards).

Pretest The purpose of the pretest was to identify the categories for the category game. Experimenters conducted the pretest in an individualized format. During the pretest, the experimenter instructed the participant to “tell me a [category].” The experimenter provided praise following correct responses or said “That’s OK” following incorrect responses, and initiated the next trial with a new category. Experimenters selected

categories for which either (a) both participants in a group emitted zero correct responses or (b) a participant emitted up to three correct responses that the participant's partner did not emit. Experimenters then compared participant responses; if both participants in a group emitted the same response during the pretest, experimenters taught one or two different responses to each participant during pretraining sessions. Table 2 shows a summary of the correct responses each participant emitted during the pretest and the responses experimenters taught each participant during pretraining sessions.

Pretraining Sessions Experimenters conducted pretraining sessions separately with each participant. Pretraining sessions consisted of two trials, one trial for each response. Experimenters conducted two trials per session to avoid reinforcing repeated responses or within-session patterns. The experimenter terminated a session after the participant emitted two different responses independently or following prompts. During pretraining Sessions 1 and 2, the experimenter delivered the S^D (“Tell me a [category].”) and immediately provided a full vocal prompt while showing a picture of the target response. The experimenter delivered praise and a token for compliance with the prompt. During Session 3, the experimenter probed for independent correct responding. The experimenter delivered the S^D and waited 5 s for a response. The experimenter delivered tokens only for correct independent responses. Following a repeat response in Trial 2, the experimenter provided praise and asked if the participant could think of another item in the category. If the participant did not respond or repeated the response again after experimenter feedback, then the experimenter implemented the error-correction procedure and prompted the response the participant had not named. The error-correction procedure consisted of the experimenter delivering a partial vocal prompt, picture prompt, and full vocal plus picture prompt, with 5 s between prompt levels. If the participant emitted at least one correct answer independently in Session 3, then all subsequent pretraining sessions consisted of a 5-s prompt

Table 2 Responses Each Participant Emitted During Pretest or Learned During Acquisition

Group	Participant	Spices	Cheeses	Flowers	Dogs	Sports	Birds
1	Ben	Oregano Cinnamon	Mozzarella Blue cheese	Iris Sunflower	NT	NT	NT
	Aria	<i>Pepper</i> <i>Paprika</i>	<i>Cheddar</i> <i>String cheese</i>	<i>Daisy</i> <i>Rose</i>	NT	NT	NT
2	Megan	NT	Cheddar Mozzarella	NT	Dalmatian Pug	Soccer Baseball	NT
	April	NT	<i>String cheese</i> Pepper jack	NT	Poodle Boxer	<i>Gymnastics</i> Football	NT
3	Miriam	Nutmeg Oregano	<i>Cheddar</i> Mozzarella American	NT	NT	NT	<i>Blue jay</i> Parrot Dove
	Elis	<i>Salt</i> <i>Pepper</i>	<i>Cheddar</i> Swiss Pepper jack	NT	NT	NT	<i>Robin</i> <i>Blue jay</i> Pigeon

Note. NT = Not taught. Responses participants emitted during the pretest are in italics

delay. If the participant did not emit a correct response during Session 3, then the experimenter implemented a partial vocal prompt (Group 1) or partial vocal prompt plus picture (Groups 2 and 3) at a 0-s prompt delay. The experimenter provided praise and a token following correct prompted responses. When the participant emitted two correct responses with the partial vocal prompt, the experimenter conducted all subsequent sessions with a 5-s prompt delay. Experimenters continued pretraining sessions until both participants in a group reached mastery criteria with all categories. Mastery criteria were defined as a participant independently emitting two different responses in a session across two consecutive sessions of the same category.

Group Sessions Each session consisted of six trials of one category. Every session, the experimenter randomly selected who responded first and then continued to alternate turns within the session until all six trials were completed (i.e., three turns per participant). If a participant spoke out of turn, the experimenter reminded the participant to wait quietly; however, this rarely occurred. During group sessions, the experimenter presented a colored token board, delivered an initial instruction (see the specific condition that follows), and started the session. Across all conditions, the experimenter provided general praise (e.g., “Good!”) and a token for correct responses that met the schedule in effect. If the participant emitted an incorrect response, the experimenter said a variation of “Nice try.” If the participant did not respond after 10 s, the experimenter initiated the next trial.

The experimenter introduced a token exchange period every three sessions. Experimenters used contingency-correlated stimuli (colored token boards) for each condition (see Table 1). Each condition that follows is described first by the programmed schedule of reinforcement and then by the content of the rule (i.e., vary or no vary; complete [C] or incomplete [I]).

- **FR 1 Vary (I).** The experimenter stated an inaccurate and incomplete rule to vary responses and delivered praise and a token for all correct responses (FR 1). Prior to each session during FR 1 Vary (I), the experimenter said, “We are going to play the category game and see how many things you can think of. Wait quietly until it is your turn. Try to say something different than your friend. Let’s start with [category].” The experimenter provided praise (e.g., “That is a spice!” or “Great!”) and a token on an FR 1 schedule for correct responses. In this condition, the experimenter delivered praise and tokens for all correct responses (whether they were different or repeated).
- **Lag 1/2/3 Vary (I).** The experimenter delivered the same instruction as during FR 1 Vary (I); however, the experimenter differentially provided praise and tokens for varied responses according to the lag schedule. Experimenters arranged the lag schedules in a progressive manner within each session; not all six trials in a session could meet the lag criterion (Wiskow et al., 2018). In Lag 1 Vary (I), the experimenter provided a token only if the response was different from the previous correct response. For example, if the first participant said “robin” for birds, the second participant earned a token if he or she said any birds other than “robin.” In Lag 2 Vary (I), the experimenter provided a token only if the response was different from the previous two correct responses. For example, if the first participant said “robin” and the second participant said “blue jay,” the first participant had to say something

different from “robin” and “blue jay” to earn a token on his or her second turn. In Lag 3 Vary (I), the experimenter provided a token only if the response was different from the previous three correct responses. If a participant emitted a correct response that did not meet the lag criterion, the experimenter provided feedback: “Nice try, but [participant] just said that.”

- **Lag 0 No Vary (C).** The experimenter delivered a complete and accurate instruction: “We are going to play the category game. Wait quietly until it is your turn. You have to say the same thing to get a token.” The experimenter provided a token only if the response was identical to the previous correct response. For example, if Miriam said “robin” for birds, then Elis earned a token if she also said “robin” during her turn. If a participant emitted a correct response that was not identical to the previous correct response, the experimenter said, “Nice try, but that’s not what [participant] just said.” In this condition, the experimenter only delivered praise and a token for repeated correct responses.
- **FR 1 No Vary (C).** This condition was similar to FR 1 Vary (I) except the experimenter delivered a complete and accurate instruction. The experimenter said, “We are going to play the category game. Wait quietly until it is your turn. You do not have to say something different. You can say the same thing and still get a token.” The experimenter provided praise (e.g., “That is a spice!” or “Great!”) and a token on an FR 1 schedule for correct responses.
- **FR 1 Vary (C).** The experimenter stated an inaccurate and complete rule to vary responses and delivered praise and a token for all correct responses (FR 1). Prior to each session during FR 1 Vary (C), the experimenter said, “You have to say something different every time. You cannot say the same thing you said before, or what your friend said.” The experimenter provided praise (e.g., “That is a spice!” or “Great!”) and a token on an FR 1 schedule for correct responses.
- **Lag 5 Vary (C).** The experimenter said, “You have to say something different every time. You cannot say the same thing you said before, or what your friend said.” The experimenter provided a token only if the response was different from all other correct responses emitted during that session. If a participant repeated a correct response within a session, the experimenter said, “Nice try, but [participant] already said that.”

Results

Figures 1, 2, and 3 show the average lag score for Groups 1, 2, and 3, respectively. The gray shaded areas indicate when experimenters implemented a lag schedule in another category. For example, in Fig. 1, the gray shaded area in the spices category (during FR 1 Vary [I]) is aligned with the same point in time that experimenters implemented Lag 1 Vary (I) in flowers and Lag 2 Vary (I) in cheeses.

Figure 1 shows the average lag score for Ben and Aria across cheeses, flowers, and spices. In the initial FR 1 Vary (I) phase, the average lag score was low across the cheeses ($M = 0.67$), flowers ($M = 0.68$), and spices ($M = 0.79$) categories. In the next phase, experimenters implemented a Lag 2 Vary (I) schedule for cheeses and a Lag 1

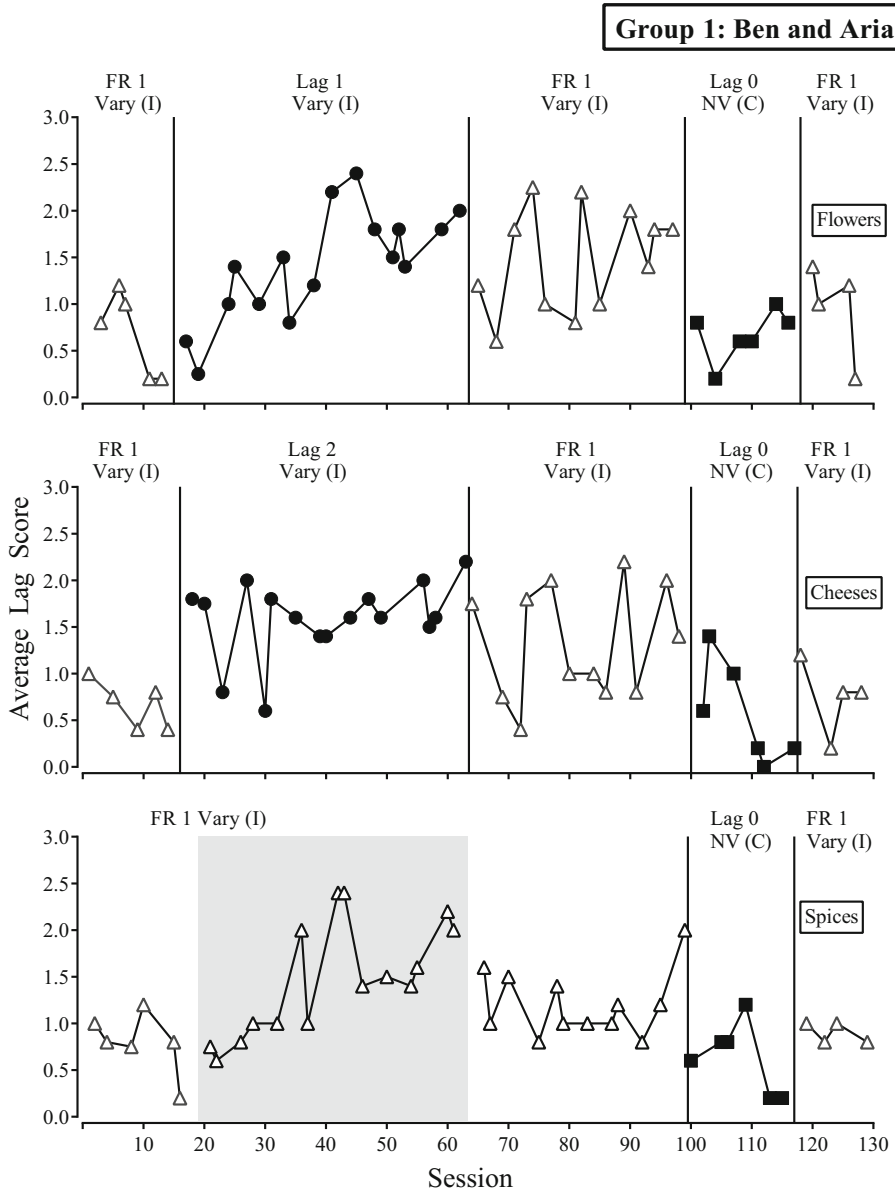


Fig. 1 Average lag score across categories for Group 1: Ben and Aria

Vary (I) schedule for flowers and observed a similar increase in both categories ($M = 1.59$ and $M = 1.42$, respectively). During this phase, experimenters maintained an FR 1 Vary (I) schedule for spices and observed a generalization of increased variability ($M = 1.47$). When experimenters returned to FR 1 Vary (I) across categories, there were more variable levels of responding with similar mean levels compared to the previous lag schedule phase, in the cheeses ($M = 1.33$) and flowers ($M = 1.49$) categories. The average lag score also decreased in the spices category ($M = 1.21$). When experimenters

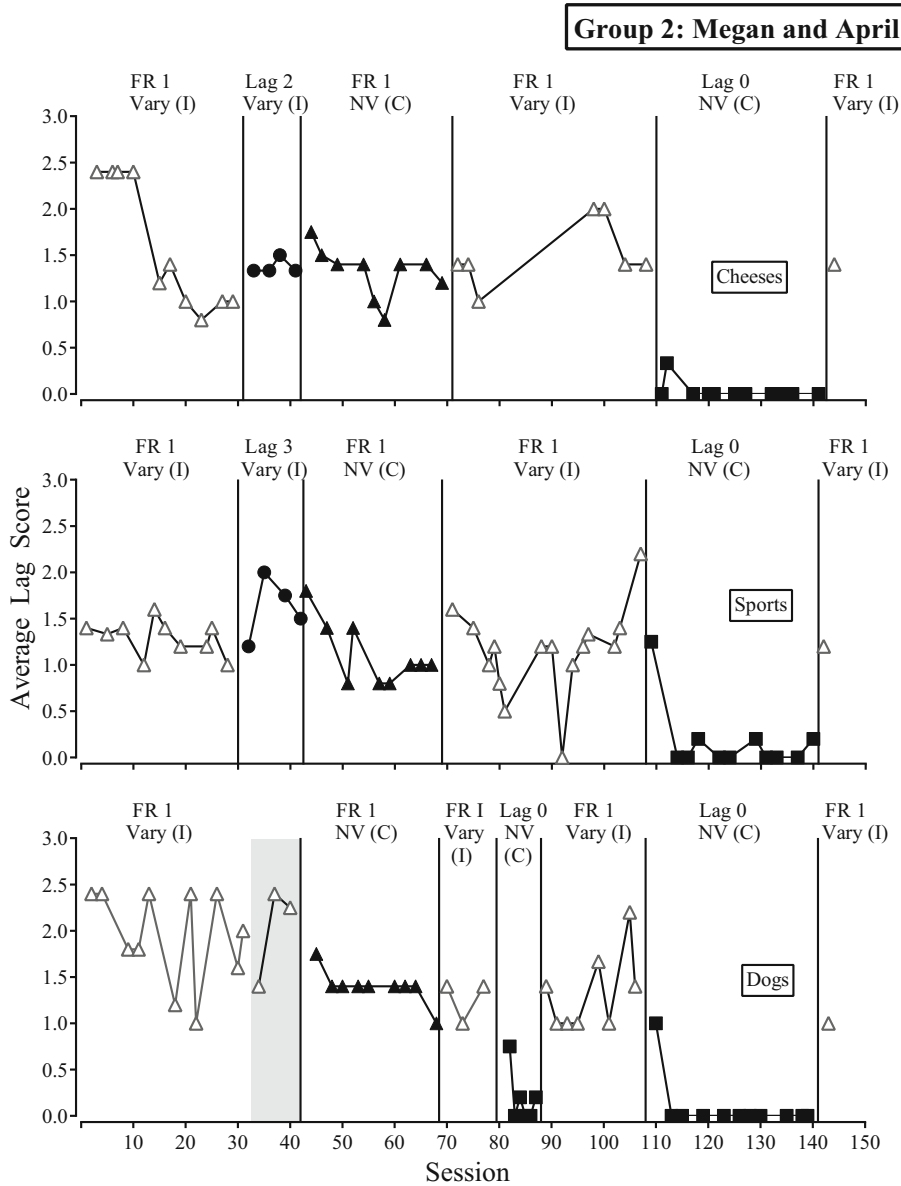


Fig. 2 Average lag score across categories for Group 2: Megan and April

implemented Lag 0 No Vary (C), variability decreased to low levels across the cheeses ($M = 0.57$), flowers ($M = 0.67$), and spices ($M = 0.63$) categories. In the final phase, experimenters reimplemented FR 1 Vary (I) and observed modest increases in variability across the cheeses ($M = 0.75$), flowers ($M = 0.95$), and spices ($M = 0.90$) categories.

Figure 2 shows the average lag score for Megan and April across the dogs, sports, and cheeses categories. Unlike Ben and Aria, Megan and April engaged in moderate to

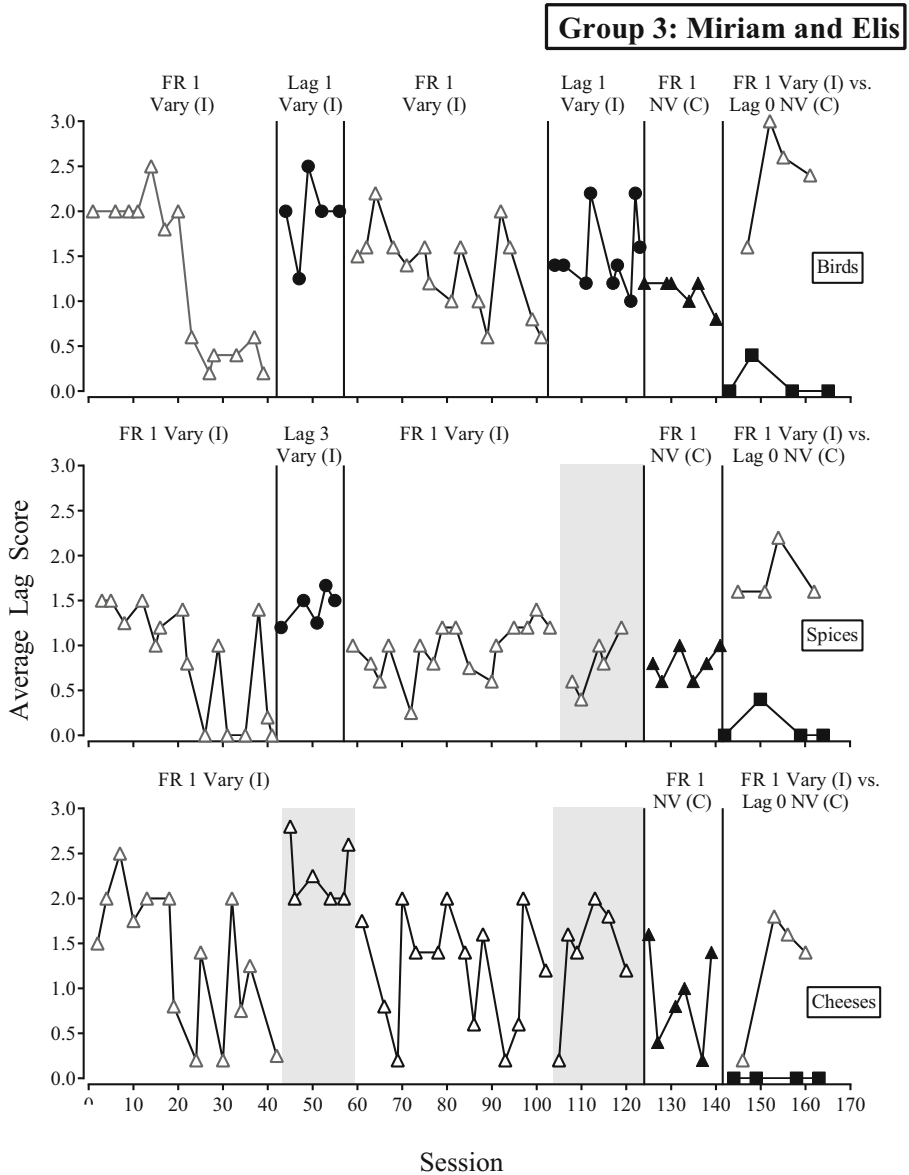


Fig. 3 Average lag score across categories for Group 3: Miriam and Elis

high variability during the initial FR 1 Vary (I) phase. Responding was variable in the dogs category ($M = 1.94$), stable in the sports category ($M = 1.29$), and started high and stable in the cheeses category, but significantly decreased to lower levels across sessions ($M = 1.6$). Because the average lag score was already moderately high across categories, experimenters implemented a Lag 3 Vary (I) schedule in the sports category ($M = 1.61$) and a Lag 2 Vary (I) schedule in the cheeses category ($M = 1.38$) and observed a modest increase, on average. Next, experimenters implemented FR 1 No Vary (C), meaning that the experimenter delivered praise and a token for any correct

response and told participants they did not have to say something different to earn a token; however, the average lag score remained unchanged across the dogs ($M = 1.39$), sports ($M = 1.11$), and cheeses ($M = 1.32$) categories. Experimenters reimplemented FR 1 Vary (I), and responding remained at similar levels. Because all participants' correct responses resulted in praise and a token to this point, experimenters changed the contingency to Lag 0 No Vary (C) in the dogs category and observed the average lag score immediately decrease, whereas responding persisted at similar (cheeses) and slightly lower (sports) levels. Experimenters reimplemented FR 1 Vary (I) with the dogs category and recaptured higher levels of variability. In the next phase, experimenters implemented Lag 0 No Vary (C) across all categories and observed responding decrease to zero or near-zero levels across categories. In the final session, experimenters implemented FR 1 Vary (I) and observed the average lag score immediately increase across categories.

Figure 3 shows the average lag score for Miriam and Elis across the birds, spices, and cheeses categories. During the initial FR 1 Vary (I) phase, there were moderate levels of variability that decreased (birds, $M = 1.28$) or became more variable (spices, $M = 0.85$; cheeses, $M = 1.33$) across sessions. When experimenters implemented a Lag 1 Vary (I) schedule for birds and a Lag 3 Vary (I) schedule for spices, variability increased ($M = 1.95$ and $M = 1.42$, respectively). Similar to Ben and Aria, this varied responding generalized to the cheeses category ($M = 2.28$). When experimenters reimplemented FR 1 Vary (I), there was a gradually decreasing trend in the birds category ($M = 1.35$), lower and variable levels in the spices category ($M = 0.95$), and extremely variable levels in the cheeses category ($M = 1.23$). Responding remained relatively unchanged when experimenters reimplemented Lag 1 Vary (I) in the birds category, and subsequently FR 1 No Vary (C) across all categories. In the final phase, there were high levels of variability in FR 1 Vary (I) sessions and low to zero variability in Lag 0 No Vary (C) sessions for birds ($M = 2.4$, $M = 0.1$), spices ($M = 1.75$, $M = 0.1$), and cheeses ($M = 1.25$, $M = 0$).

Figure 4 shows the percentage of maximum variability during subsequent analyses that experimenters conducted with Miriam and Elis across categories. The gray shaded areas in this graph indicate the same point in time when the experimenter implemented 18 trial sessions with another category. In the initial two phases, experimenters replicated previous conditions and observed similarly high levels of variability in the FR 1 Vary (I) phase ($M = 87\%$, 65% , 72%) and zero variability during the Lag 0 No Vary (C) phase across the birds, spices, and cheeses categories, respectively. When experimenters implemented a Lag 1 Vary (I) schedule, a similar level of variability as during the FR 1 Vary (I) phase occurred in the birds ($M = 84\%$) and cheeses ($M = 70\%$) categories, and slightly lower levels of variability in the spices ($M = 49\%$) category. Variability returned to zero when experimenters implemented Lag 0 No Vary (C). In the next phase, experimenters increased the variability requirement to Lag 5 Vary (C) and observed an increase in variability in the birds ($M = 100\%$) and spices ($M = 64\%$) categories, but similarly moderate levels of variability in the cheeses category ($M = 61\%$). Then, experimenters implemented FR 1 Vary (C) and observed maintenance in levels of variability despite the contingency not requiring variability. Because both participants learned additional responses in each category since the start of the study, experimenters implemented extended sessions consisting of 18 trials to increase the likelihood of participants contacting praise and a token for repeat responses across

Group 3: Miriam and Elis

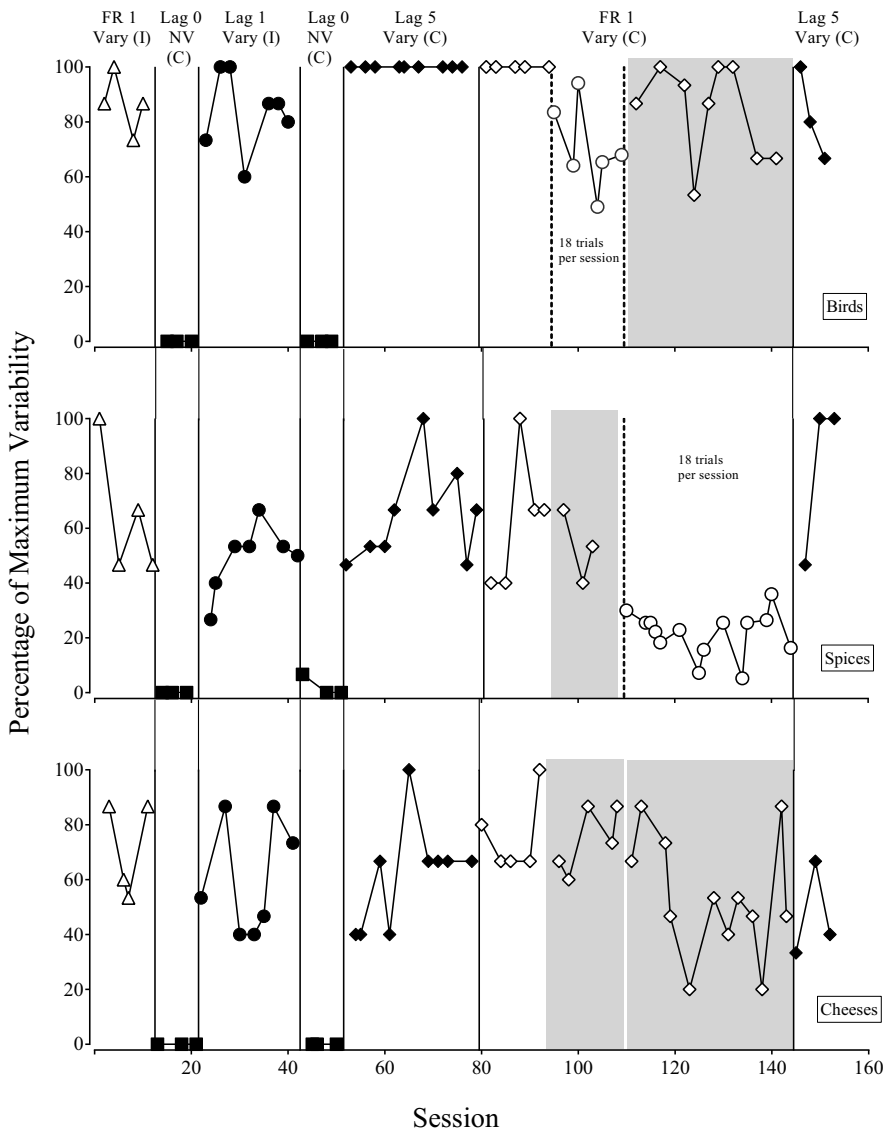


Fig. 4 Percentage of maximum variability across categories for Group 3: Miriam and Elis

successive trials during a single session. First, experimenters implemented the extended sessions with the birds category and observed a decrease in the percentage of maximum variability ($M = 71\%$). There was no change in the cheeses ($M = 75\%$) and spices ($M = 53\%$) categories. Then, experimenters implemented the extended sessions with the spices category and observed a decrease in the percentage of maximum variability ($M = 22\%$). Variability decreased in the cheeses category ($M = 53\%$) and more variability occurred in the birds category ($M = 84\%$). In the final phase, experimenters

reimplemented Lag 5 Vary (C) and observed high levels of variability in the spices category ($M = 82\%$) but relatively no change in variability in the birds ($M = 82\%$) and cheeses ($M = 46\%$) categories relative to the previous phase.

Discussion

The present study evaluated lag schedules in a group format and the influence of rules on persistent levels of variability. The results support previous research on the general utility of lag schedules to increase response variability (e.g., Contreras & Betz, 2016; Dracoby et al., 2017; Lee et al., 2002; Silbaugh & Falcomata, 2017, 2019; Susa & Schlinger, 2012; Wiskow et al., 2018). Furthermore, the results provide additional support for the effectiveness of group-based lag schedules to increase varied and novel responses (Wiskow & Donaldson, 2016). These results are promising, as group-based lag schedules may be an efficient way to teach a variety of category items. The present study provides three main findings that add to the applied literature on response variability in group formats: (a) maintenance of variability scores when experimenters removed the lag contingency and continued the same rule, (b) generalized variability to another category during the lag schedules, and (c) similar levels of variability when experimenters introduced different lag schedules. In subsequent manipulations, participants only repeated responses when the rule and contingency required it.

The lag schedules increased variability in two of three groups of preschoolers. Like Wiskow and Donaldson (2016), the application of group-based lag schedules increased the number of instances that participants varied their response from their peer's. The use of group-based lag schedules may lead to efficient teaching and requires individuals to attend to peers' responses, potentially enhancing the likelihood of observational learning. Experimenters extended Wiskow and Donaldson (2016) by comparing different lag schedules; however, there was no difference (data are available from the first author). This is inconsistent with research demonstrating that higher lag schedules may produce higher levels of variability during individual instruction (e.g., Susa & Schlinger, 2012; Wiskow et al., 2018). The similar levels of variability may be due to a lack of discriminability across sessions. Although experimenters used different colored token boards to signal changes in the contingency, the rule was the same across sessions and the schedules were similar (e.g., Lag 1 and Lag 2) and may have exerted more stimulus control over responding than the colored token boards.

In addition, due to the brevity of sessions, there was a limited number of opportunities for the participants to contact the contingency (Baron & Galizio, 1983), and participants earned tokens even if they exceeded the schedule in effect. For example, during Lag 1 Vary (I), participants could say all different responses to earn praise and a token, although they only had to vary from the previous one response. Experimenters may have punished all repeat responses within a session if the participant repeated the previous response and did not earn a token. Although individual participants sometimes earned all three tokens during a lag session, groups never earned all tokens during the lag schedules (Figs. 1, 2, and 3). Therefore, the items experimenters used as backup reinforcers may not have been reinforcing enough to consistently evoke all different responses within a session. Overall, these results suggest that clinicians may introduce a Lag 1 Vary (I) schedule first, and then increase the lag schedule progressively when

needed to evoke higher levels of variability. This is favorable because higher lag schedule criteria may be more difficult to track and it may be more difficult to deliver immediate feedback to participants.

Although participants may not have discriminated between the lag schedules, variability decreased when the rule and contingency required participants to repeat responses. Similar to previous research (e.g., Dracobly et al., 2017), participants learned to repeat their responses during Lag 0 No Vary (C); however, participants in the present study continued to vary their responses when experimenters reinstated the FR 1 Vary (I) condition that did not require participants to vary their responses. Recall that the rule was the same during the lag schedules phase (i.e., “Try to say something different.”); therefore, once the rule was paired with praise and a token contingent on varied responses, the rule continued to exert stimulus control over varied responding when experimenters removed the lag contingency. It is also interesting that this effect was more variable for Ben and Aria and decreased over time in the birds category for Miriam and Elis. Even in sessions where the participants contacted praise and a token for repeat responses (i.e., sessions with low lag scores), they did not continue to repeat responses in subsequent sessions. All participants repeated responses during Lag 0 No Vary (C) and varied their responses during FR 1 Vary (I). Participant statements suggested an understanding of the contingency and attentiveness to peer responses. For example, Elis stated, “But I did not hear her,” during one of her turns in Lag 1 Vary (I) and, “Wait. What did you say? I did not hear you,” during Lag 0 No Vary (C). These responses suggest that Miriam’s response was discriminative for Elis’s responses during her turn.

The lag criterion also may have influenced variability maintenance. For example, Lee et al. (2002) used a Lag 1 schedule to increase variability in answering social questions (e.g., “What do you like to do?”) and did not observe varied responding persist. In contrast, Heldt and Schlinger (2012) used a Lag 3 schedule to increase variability in tacting items in a picture (e.g., “What do you see?”) and observed that varied responding persisted 3 weeks after the final lag session. In the present study, all groups experienced a minimum of a Lag 2 schedule. However, in the present study, experimenters also continued to state the rule; therefore, the relative influence of the rule and the lag criterion on maintenance is unknown. Future experimenters may explore the influence of lag criteria on maintenance to determine if there is a point at which maintenance is likely, and whether rules enhance these effects.

Another interesting finding was that participants began to emit more varied responses in the spices category (Ben and Aria) and the cheeses category (Miriam and Elis) when experimenters implemented the lag schedules with the other two categories. Experimenters may have observed generalization because of carryover effects from the higher lag schedule to the lower lag schedule. To our knowledge, no applied study has evaluated different lag schedules within the same phase; experimenters typically introduce higher lag schedules sequentially (e.g., Contreras & Betz, 2016; Lepper et al., 2017; Susa & Schlinger, 2012). However, the consistency of the rule also may play a role. Wiskow and Donaldson (2016) used a multielement design in the group format; however, experimenters alternated FR 1 Vary (I) and Lag 1 Vary (I). Therefore, experimenters may have incidentally programmed common stimuli (the rule), utilized indiscriminable contingencies

(the lag schedules), or provided sufficient exemplars (two categories) that enhanced generalization to the third category (Stokes & Baer, 1977).

In the present study, we evaluated the relationship between rules and contingencies of reinforcement with typically developing children. An important consideration of these data is how they may translate to work with children with autism. For example, children with autism may require targeted training to establish rule-governed behavior (Tarbox et al., 2011) compared to their typically developing peers. A few recent studies demonstrated the development of stimulus control of response variability (repetitive and varied responding) by visual schedule-correlated stimuli (e.g., Brodhead, Higbee, Gerencser, & Akers, 2016) and visual and auditory schedule-correlated stimuli (Dracobly et al., 2017). Therefore, future research may further evaluate the relative effects of visual and auditory contingency-specifying instructions on response variability with a focus on assessing generalization and maintenance effects with children with autism.

There are limitations of the present study that warrant additional research. First, our group sessions consisted of six trials. Experimenters implemented brief sessions so that participants could contact the contingencies repeatedly across sessions; however, conducting a single, extended session may be more efficient such that the participants can contact the contingencies rapidly within a session. For example, experimenters saw the greatest change in responding when they extended the number of trials per session with Miriam and Elis. Miriam and Elis did not learn enough responses in the spices category to vary their responses each trial; therefore, when experimenters implemented the extended sessions with the spices category during FR 1 Vary (C), Miriam and Elis repeated more responses. As a result, when they contacted praise and a token for repeat responses, the percentage of maximum variability decreased in the spices category, and some of this effect generalized to the other categories. Similar to previous research, responding persisted regardless of the rule as long as participants continued to earn putative reinforcers (e.g., Baron & Galizio, 1983), and when rule-following behavior resulted in less reinforcement, the individuals stopped following the rule (Newman, Buffington, & Hemmes, 1995). Future experimenters could explore the influence of the number of opportunities a participant has to contact the contingency on levels of response variability. Experimenters could determine the number of opportunities by typical classroom response opportunities. Alternatively, experimenters could conduct a parametric analysis of the number of response opportunities and measure the rate of acquisition, acquisition via observational learning in the group, and generalization of responding to testing environments (e.g., Wiskow & Donaldson, 2016).

Second, the persistence of responding across phases weakens experimental control. Megan and April emitted similar levels of variability across all experimental conditions except when experimenters implemented Lag 0 No Vary (C). Therefore, high levels of variability occurred regardless of whether the contingency required variability or the rule changed. Variability only decreased when the rule and contingency required the participants to repeat responses. It is interesting to note that Megan and April typically stated the responses that the experimenters taught them individually but did not imitate the response(s) of the peer. The experimenters extended the Lag 0 No Vary (C) phase with this group because Megan only stated the responses the experimenters taught her. Therefore, in sessions that April started, Megan initially stated her response, did not earn a token, and then stopped responding. At Session 127, experimenters began

prompting Megan to imitate April (e.g., “Say [item].”). Experimenters observed severe rigidity in responding and unfortunately ran out of time before the end of the school year to conduct additional analyses of this phenomenon. It is possible that conducting a similar analysis with a response that can be physically prompted would be beneficial for individuals who are nonresponsive to less intrusive prompts. Alternatively, future research may conduct this type of analysis at an individual level and subsequently apply or test the contingencies in a group context. This process may more closely mirror targeted intervention with children in schools.

Third, the selection of categories may have influenced some responding across phases. To increase the likelihood the participants would learn from their peer, experimenters initially selected categories for which the participants emitted zero to two correct exemplars in the pretest; however, several participants learned responses outside of our sessions throughout the course of the study. For example, Elis stated that she “asked her mom” and played the category game at home. Participants learning responses outside of our sessions is a major limitation impacting the internal validity of the study, especially when analyzing the number of novel responses.

In addition, some categories may be easier (e.g., birds) to identify novel exemplars for than others (e.g., spices). Future experimenters may attempt to equate the categories by making a list of exemplars and including categories with a similar number of items. Alternatively, experimenters may select categories or target skills that are typically taught in classrooms and for which there is a definite set of correct responses (e.g., months, body parts, planets). This would permit the selection of a maximum lag value and the identification of a mastery criterion during group and individual sessions.

In sum, the results of the present study suggest that providing rules and using lag schedules in group instructional formats, at least initially, may lead to efficient acquisition rates. The results of the present study support previous research demonstrating that clinicians may use a rule to evoke immediate behavior change; however, the effects will diminish over time unless the contingency of reinforcement matches the rule (e.g., Group 3 in the present study). Our findings also provide preliminary support for using a rule to facilitate the maintenance of varied responding when clinicians remove the lag contingency or it is implemented inconsistently (i.e., integrity failures); however, more research is needed to determine these effects over time (e.g., development of rigid rule following).

The present study offers a novel evaluation of the effects of rule variations and lag schedules applied to groups of typically developing preschoolers on levels of response variability. Few studies have evaluated the use of lag schedules in a group format, despite demands on students to learn in large groups at school. Furthermore, increasing student response variability may be an important skill associated with learning, exploring, creating, and problem solving (Neuringer, 2002). This study provides a preliminary framework for additional applied research on observational learning and response variability with groups of students. Future research should continue to evaluate the effects of the interactions between rules and lag schedules on observational learning and on-task behavior (because of the peer observation requirement) and on the maintenance of responding in academic and social group learning contexts, and to extend group-based lag schedules to children with autism.

Acknowledgements We thank Maricela Rivera, Avery Urban-Wilson, Qingyuan Xie, and Lizbet Delgadillo for their assistance in data collection.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

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

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

References

- Baron, A., & Galizio, M. (1983). Instructional control of human operant behavior. *The Psychological Record*, 33(4), 495.
- Brodhead, M. T., Higbee, T. S., Gerencser, K. R., & Akers, J. S. (2016). The use of a discrimination training procedure to teach mand variability to children with autism. *Journal of Applied Behavior Analysis*, 49(1), 34–48. <https://doi.org/10.1002/jaba.280>
- Contreras, B. P., & Betz, A. M. (2016). Using lag schedules to strengthen the intraverbal repertoires of children with autism. *Journal of Applied Behavior Analysis*, 49(1), 3–16. <https://doi.org/10.1002/jaba.271>
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Hoboken, NJ: Pearson.
- Dracoby, J. D., Dozier, C. L., Briggs, A. M., & Juanico, J. F. (2017). An analysis of procedures that affect response variability. *Journal of Applied Behavior Analysis*, 50(3), 600–621. <https://doi.org/10.1002/jaba.392>
- Heldt, J., & Schlinger, H. D. (2012). Increased variability in tacting under a Lag 3 schedule of reinforcement. *The Analysis of Verbal Behavior*, 28(1), 131–136. <https://doi.org/10.1007/BF03393114>
- Joyce, J. H., & Chase, P. N. (1990). Effects of response variability on the sensitivity of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, 54(3), 251–262. <https://doi.org/10.1901/jeab.1990.54-251>
- Lee, R., McComas, J. J., & Jawor, J. (2002). The effects of differential and lag reinforcement schedules on varied verbal responding by individuals with autism. *Journal of Applied Behavior Analysis*, 35(4), 391–402. <https://doi.org/10.1901/jaba.2002.35-391>
- Lepper, T. L., Devine, B., & Petursdottir, A. I. (2017). Application of a lag contingency to reduce perseveration on circumscribed interests. *Developmental Neurorehabilitation*, 20(5), 313–316. <https://doi.org/10.3109/17518423.2016.1152612>
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, 9(4), 672–705. <https://doi.org/10.3758/BF03196324>
- Newman, B., Buffington, D. M., & Hemmes, N. S. (1995). The effects of schedules of reinforcement on instruction following. *The Psychological Record*, 45(3), 463–476. <https://doi.org/10.1007/BF03395155>
- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
- Podlesnik, C. A., & Chase, P. N. (2006). Sensitivity and strength: Effects of instructions on resistance to change. *The Psychological Record*, 56(2), 303–320. <https://doi.org/10.1007/BF03395552>
- Radley, K. C., Dart, E. H., Helbig, K. A., & Schrieber, S. R. (2018). An additive analysis of lag schedules of reinforcement and rules on novel responses of individuals with autism spectrum disorder. *Journal of Developmental and Physical Disabilities*. <https://doi.org/10.1007/s10882-018-9606-0>
- Silbaugh, B. C., & Falcomata, T. S. (2017). Translational evaluation of a lag schedule and variability in food consumed by a boy with autism and food selectivity. *Developmental Neurorehabilitation*, 20(5), 309–312. <https://doi.org/10.3109/17518423.2016.1146364>

- Silbaugh, B. C., & Falcomata, T. S. (2019). Effects of a lag schedule with progressive time delay on sign mand variability in a boy with autism. *Behavior Analysis in Practice*, 12(1), 124–132. <https://doi.org/10.1007/s40617-018-00273-x>
- 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>
- Susa, C., & Schlinger, H. D. (2012). Using a lag schedule to increase variability of verbal responding in an individual with autism. *The Analysis of Verbal Behavior*, 28(1), 125–130. <https://doi.org/10.1007/BF03393113>
- Tarbox, J., Zuckerman, C. K., Bishop, M. R., & Olive, M. L. (2011). Rule governed behavior: Teaching a preliminary repertoire of rule-following behavior to children with autism. *The Analysis of Verbal Behavior*, 27(1), 125–139. <https://doi.org/10.1007/BF03393096>
- Torrance, P. E. (1990). *Torrance tests of creative thinking*. Bensenville, IL: Scholastic Testing Services.
- Wiskow, K. M., & Donaldson, J. M. (2016). Evaluation of a lag schedule of reinforcement in a group contingency to promote varied naming of categories items with children. *Journal of Applied Behavior Analysis*, 49(3), 1–13. <https://doi.org/10.1002/jaba.307>
- Wiskow, K. M., Matter, A. L., & Donaldson, J. M. (2018). An evaluation of lag schedules and prompting methods to increase variability of naming category items in children with autism spectrum disorder. *The Analysis of Verbal Behavior*, 34(1–2), 100–123. <https://doi.org/10.1007/s40616-018-0102-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.