



A Systematic Synthesis of Lag Schedule Research in Individuals with Autism and Other Populations

Bryant C. Silbaugh¹ · Clodagh Murray² · Michelle P. Kelly³ · Olive Healy⁴

Received: 29 July 2019 / Accepted: 13 March 2020 / Published online: 15 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Lag schedules increase operant variability. Several researchers have explored their clinical and educational applications, especially to address repetitive behavior or limited repertoires in individuals with autism spectrum disorder. In the current study, we provide the first comprehensive synthesis and appraisal of lag schedule research in humans. A multistep search strategy was employed to identify all experimental studies of lag schedules in humans published in peer-reviewed journals. We identified 38 studies that met inclusion criteria, summarized the study and participant characteristics, and evaluated evidential certainty. The results suggest that lag schedules are emerging as a promising applied behavioral technology for increasing operant variability, especially in individuals with autism spectrum disorder. We conclude with preliminary practice guidelines based on evidential certainty provided by the studies and identify future avenues of research.

Keywords Lag schedules · Operant variability · Evidential certainty · Autism

Complex behaviors such as communication, learning through observation, reading, writing, and play develop over time as “progressive changes in interactions between the behavior of an individual and the people, objects, and events in the environment” (Bijou 1993, p. 12). Such changes rely in part on natural and contrived differential reinforcement of successive approximations to a target response, called shaping, but an organism must first exhibit variations in behavior the environment can reinforce. Operant variability, which can be conceptualized as “a continuum ranging from repetitive at one end to stochastic at the other” (Neuringer 2002, p. 672), may consist of variation in topography (e.g., Pryor et al. 1969), dimensional quantities (e.g., Blough 1966), stimuli selected (e.g., Goetz and Baer 1973), and other characteristics of responding over

time as a function of the environment (Skinner 1938). When operant variability enables an organism to solve problems, maximize reinforcement, or minimize contact with aversive stimuli, especially when contingencies change, it may be considered adaptive (e.g., Sidman 1960). Researchers have started to explore the possibility that reinforcing operant variability itself, as opposed to any particular response variation, may promote a more typical developmental trajectory in individuals with autism spectrum disorder (ASD) by using differential reinforcement (e.g., Miller and Neuringer 2000) to replace characteristic repetitive and stereotyped behavior (i.e., an invariance problem) with adaptive variability (e.g., Rodriguez and Thompson 2015) susceptible to shaping.

While most behavior analysts have focused almost exclusively on the repetition-selective effects of reinforcement since the 1930s, studies examining the potential variability-selective effects of reinforcement have gained momentum since the 1960s. For example, Blough (1966) demonstrated that pigeon pecking approximated stochastic-like responding as a result of differential “reinforcement of least frequent interresponse times” (p. 582). In 1969, Pryor et al. reported delivering reinforcement to porpoises contingent on novel performances (i.e., not previously demonstrated during training). Coinciding with training, the porpoises began to demonstrate a broad range of varied and complex response topographies, including responses not previously observed in the

✉ Bryant C. Silbaugh
bryant.silbaugh@utsa.edu

¹ Department of Interdisciplinary Learning and Teaching in the College of Education and Human Development, The University of Texas, San Antonio, TX, USA

² National University of Ireland, Galway, Ireland

³ Emirates College for Advanced Education, Abu Dhabi, United Arab Emirates

⁴ Trinity College, The University of Dublin, Dublin, Ireland

species (although see Holth 2012a, b for an alternative interpretation). Similarly, Goetz and Baer (1973) increased the variety of block forms produced by typically developing 4-year-old girls by delivering praise contingent on making block forms that differed from all previous block forms within a session.

Page and Neuringer (1985) reported more direct and conclusive evidence for the reinforcement of variability using lag schedules to reinforce sequence variability in pigeons in six rigorous experiments. Under a lag schedule, a response or response sequence is reinforced if it differs from N responses or response sequences, with N equal to the value of the lag. For example, under a Lag 1 schedule, every response is reinforced when it differs from the preceding response within a session (e.g., Lee et al. 2002). Under a Lag 5 schedule, a response is reinforced if it differs from the preceding five responses within a session (e.g., Falcomata et al. 2018). The major findings from Page and Neuringer's experiments were the following: (a) response sequence variability (i.e., variability in fixed-length sequences of left and right key presses) was shown to be reinforced in pigeons; (b) increasing the lag value increased sequence variability; (c) pigeons' response variability increased, as the required response sequence length increased; (d) high levels of variability required a contingency between response variability and reinforcement; and (e) sequence variability was brought under discriminative control.

Reinforcement of operant variability has now been replicated many times over, across multiple species (e.g., Neuringer 2002), leading to the conclusion of many behavior analysts (although not all; Palmer 2012) that variability is a dimension of operant behavior that can be reinforced. Recent studies in the experimental analysis of behavior have led some behavior analysts to suggest that differential reinforcement of "various combinations of switches and repeats" may ultimately produce a generalized, higher-order response class of "reinforced variation" in a manner similar to multiple exemplar trainings (Doughty and Galizio 2015).

Stokes (1999) conducted the first basic experiment on lag schedules in humans in which participants earned points on a computer game by pressing up or down directional keys to move a white square to the lower right corner of a grid. Twenty female undergraduate students were randomly assigned to two groups. In one group, key pressing was initially reinforced on a Lag 0 schedule, followed by a gradual increase to Lag 2, Lag 10, and Lag 25. In the other group, the order of the conditions was reversed. In both groups, key press variability systematically increased across lag schedule values, and the group exposed to the higher lag schedule first exhibited overall higher levels of variability. Stokes (1999) interpreted the results to suggest that variability is a learned behavior, lag schedules teach individuals how much to vary their behavior, and higher variability levels result from reinforcing variability early on in an organism's training.

In the first applied lag schedule study, Lee et al. (2002) evaluated the effects of lag schedules on responses to social questions (e.g., "How are you?"; i.e., intraverbal behavior) in male participants with ASD. A Lag 1 schedule of positive reinforcement increased intraverbal variability and the cumulative number of novel and appropriate intraverbals in two out of three participants. Multiple researchers have since extended this work by evaluating the effects of lag schedules in isolation and combined with other treatment components on verbal and nonverbal behavior in multiple domains such as intraverbal (i.e., conversational) responding to interview questions in adults (e.g., O'Neill and Rehfeldt 2014), tacts (i.e., labeling the environment; e.g., Heldt and Schlinger 2012), mands (e.g., Brodhead et al. 2016), manding (i.e., requesting reinforcement) during functional communication training (e.g., Adami et al. 2017), activity selection (e.g., Ivy et al. 2019), play skills (e.g., Baruni et al. 2014), feeding (e.g., Silbaugh and Falcomata 2017), and multiple skills targeted within a manualized social skills curriculum (e.g., Radley et al. 2017a). Findings from these studies converge to suggest that lag schedules can increase variability in verbal and nonverbal behavior in the otherwise repetitive behavior of individuals with ASD. Given their potential for clinical and educational benefit to individuals with ASD, it is imperative that researchers continue to seek a better understanding of how lag schedules work in humans generally (e.g., effects on developmental trajectories and adaptive behavior) and generate practice guidelines for the appropriate use of current lag schedule-based technology.

One study to date reviewed the published literature on lag schedule research but also included studies of other interventions for increasing variability too. Specifically, Wolf et al. (2014) reviewed the studies of lag schedules and other interventions used to increase operant variability in individuals with ASD and provided practice guidelines based on their findings. Their review suggested that lag schedules were effective alone or in combination with antecedent manipulations such as script training, prompting, and prompt fading, for nearly 90% of participants across all 14 included studies. Wolf et al. (2014) provided a practical and valuable initial framework for increasing variability in individuals with ASD using current available technology. However, the guidelines were not specific to lag schedules and may be outdated due to recent growth in the lag schedule literature, they omitted findings from basic lag schedule research in humans, and there was no direct link to the evidential certainty provided by studies critical for implementing lag schedules in the evidence-based practice of applied behavior analysis (ABA; Slocum et al. 2014).

Therefore, the aims of the current synthesis were to (a) synthesize participant and study characteristics; (b) assess what we know and do not know about the effects of lag schedules in individuals with ASD and other populations from the

basic, translational, and applied literature; (c) determine for whom and when lag schedule-based interventions are appropriate; (d) update existing preliminary practice guidelines based on the evidential certainty provided by the studies; and (e) identify future avenues of lag schedule research.

Method

Search Procedures

A multistep search and screening process was conducted to identify studies for inclusion. In step 1, the first author conducted an electronic database search using the EBSCO host through the University of Texas at San Antonio library by searching the Academic Search Complete database for relevant peer-reviewed journal articles written in English and published no earlier than 1985 (i.e., publication of the seminal lag schedule study; Page and Neuringer 1985). The following search terms were utilized: “lag schedule*” or “operant variability” or “lag reinforcement” or “lag reinforcement schedule*” or “lag schedule of reinforcement,” which yielded 40 studies. Seventeen of these studies met the inclusion criteria. In step 2, the first author conducted an ancestral search for additional studies by screening the references of studies included from step 1. After removing duplicates, of those 250 references, the first author identified an additional 12 studies that met the inclusion criteria. In step 3, the first author identified seven additional studies that met inclusion criteria by entering the titles of each included study singly into Google Scholar in a forward search. In total, the search process identified 36 studies for inclusion. After this search, two additional studies were published which met the inclusion criteria and brought the total to 38.

At each step of the search process, the first author provided the search yield to the second author, who repeated the screening. The first author then assessed interrater reliability by calculating the percentage of the total search yield at that step of the search process that the raters agreed. Agreement at steps 1, 2, and 3 were 100%, 100%, and 86% (6/7 studies), respectively.

Inclusion and Exclusion Criteria

Empirical (i.e., experimental and nonexperimental) evaluations of lag schedules (i.e., “lag schedule” as explicitly stated by the authors of the study, even if we later concluded that the study did not use a lag schedule) employing any methodology such as group design or single-subject design (including experimental or AB designs) in human participants, in any environment such as homes, schools, community, or laboratory settings, were included. All other studies were excluded.

Data Extraction and Intercoder Agreement

Data extraction procedures were similar to other recent syntheses of behavior analytic research (e.g., Verschuur et al. 2014; Silbaugh et al. 2016). We developed a coding guide which we used to extract data on participant and study characteristics from included studies. Any features of studies considered potentially pertinent to the goals of the synthesis were incorporated into the coding guide to minimize post hoc selection biases (Cooper 2010). Participant and study coding guide sections included operational definitions of dependent variables and recording guidelines. The first and second authors tested a draft of the coding guide by independently coding three randomly selected applied lag schedule studies to clarify any ambiguous operational definitions or recording guidelines and revised the coding guide as needed. We updated the data in the coding guides for these lag schedule studies according to any revisions to the coding guide. Then, the authors discussed and agreed on a final coding guide and initiated data extraction. To facilitate high intercoder reliability and replication of the data extraction process (Cooper 2010), the first author coded all studies, and each coauthor independently coded a different subset of the studies. Then, the first author assessed intercoder agreement for all dependent variables, including the questions related to evidential certainty, coded for each study, and the mean agreement was 88%. Then, the authors discussed any disagreements until consensus was reached on the entire data set.

Dependent Variables

The coding guide enabled data collection on 34 dependent variables. Operational definitions and a coding guide template are available upon request. For participant characteristics, data were collected on age, sex, ethnicity, education or treatment history (excluding basic experiments), and diagnoses or conditions. For study characteristics, data were collected on the journal, manualized interventions (excluding basic experiments), setting (excluding basic experiments), methodology, dependent variables in single-subject design studies, dependent variables in group design studies, type of single-subject design (including further classification as a demonstration, parametric analysis, component analysis, or sequential analysis; Kennedy 2005), data collection methods, invariance assessments (excluding basic experiments), social validity assessment (excluding basic experiments), procedural fidelity (excluding basic experiments), interobserver agreement (excluding basic experiments), prompts (excluding basic experiments), self-monitoring procedures (excluding basic experiments), reinforcer identification (excluding basic experiments), stability criteria for single-subject design studies only, condition signaling

for single-subject design studies only, generalization assessment, maintenance assessment, treatment components (excluding basic experiments), unwanted effects, rules and instructions, lag schedule value, verbal operants (excluding basic experiments), reinforcement type, type of contingency, lag schedule contingency, academic and functional curricular areas (excluding basic experiments), and evidential certainty (excluding basic experiments).

Evidential Certainty Classification

Certainty of evidence provided by each experiment (i.e., some studies contained multiple experiments) was classified as suggestive, preponderant, or conclusive (Verschuur et al. 2014). The evidential certainty provided by an experiment was classified as *suggestive* if the experiment did not use a true experimental design or did not meet the criteria for a preponderant classification. Evidential certainty was classified as *preponderant* if the experiment used a true experimental design, assessed at least 30% of sessions with at least 80% agreement and fidelity, operationally defined dependent variables, and provided sufficient procedural details to enable replication. The evidential certainty provided by an experiment was classified as *conclusive* if the study included all the attributes of a preponderant classification and the design provided some control for alternative explanations for treatment outcomes (i.e., threats to internal validity could be ruled out by the design).

Results

General

We use the term “study” to refer to individual articles as some studies consisted of multiple “experiments.” The

cumulative publications per domain are displayed in Fig. 1. The publication of applied studies ($n = 30$) greatly outpaced other studies (i.e., basic and translational; $n = 9$) as indicated by a steep increased rate of applied study publication beginning in 2013.

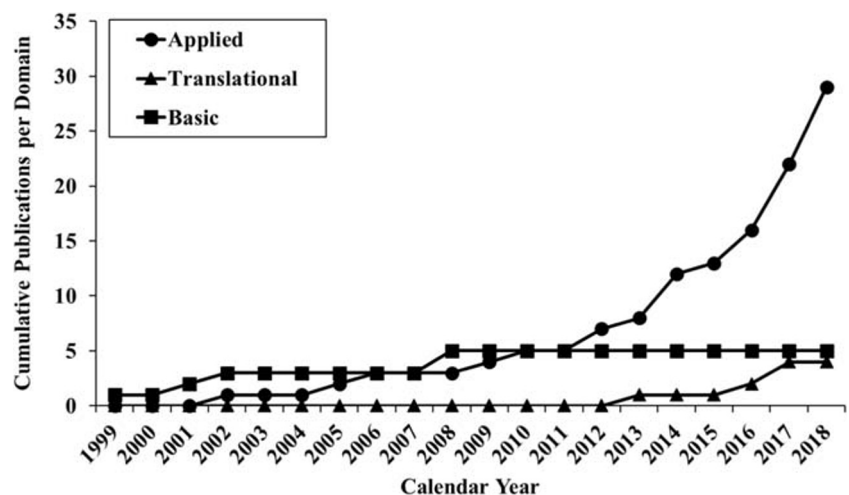
Applied Studies were published in ten journals, listed here from most to least: *The Analysis of Verbal Behavior*, the *Journal of Applied Behavior Analysis*, *Developmental Neurorehabilitation*, the *Journal of Developmental and Physical Disabilities*, *Behavior Analysis: Research and Practice*, *Research in Autism Spectrum Disorders*, *Behavior Modification*, *Behavior Analysis in Practice*, *Behavioral Interventions*, and the *Journal of Behavioral Education*.

Other Basic and translational studies (i.e., other) were published in seven journals, listed here from most to least: *The Journal of Applied Behavior Analysis*, *Developmental Neurorehabilitation*, *Research in Autism Spectrum Disorders*, *Creativity Research Journal*, *Psychonomic Bulletin and Review*, the *International Journal of Creativity and Problem Solving*, and the *Journal of Experimental Psychology: Human Perception and Performance*.

Participant Characteristics

Applied studies included 77 participants, and other studies included 398 participants, for a total of 475 people who participated in these studies. The percentages of participants in each age group for applied and other studies are displayed in Fig. 2. Of the 475 people who served as participants in these studies, 60% ($n = 285$) were adults and 40% ($n = 190$) were children. Only 15 adults (all were in applied studies) were reported to have diagnoses of ASD, but 70 children (nearly all were in applied studies, some in translational) were

Fig. 1 Cumulative publications per domain of behavior analysis



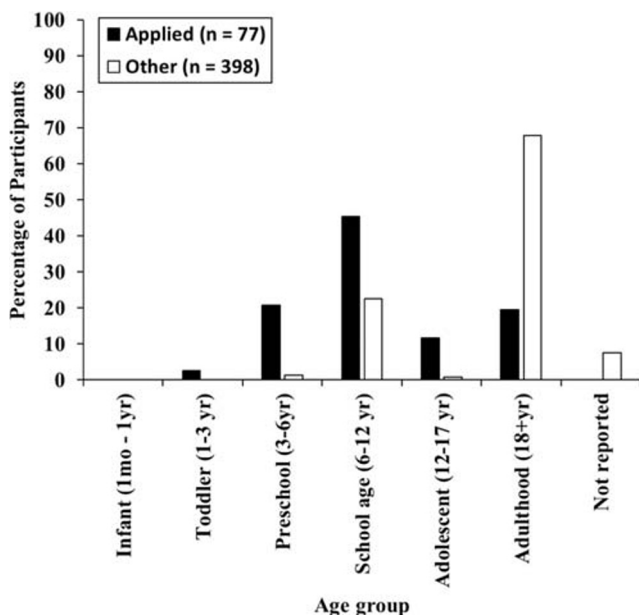


Fig. 2 The percentage of participants in each age group from applied and other (i.e., basic and translational) studies. The large majority of participants under 18 years of age were in applied studies and had a diagnosis of ASD

reported to have diagnoses of ASD. Thus, basic studies mostly included adults without ASD, and applied and translational studies mostly included children with ASD.

Applied No participants were infants ($n = 0$; age 0–1 year), 3% of participants were toddlers ($n = 2$; age 1–3 years), 21% were preschoolers ($n = 16$; age 3–6 years), 45% were school age ($n = 35$; age 6–12 years), 12% were adolescent ($n = 9$; age 12–17 years), and 19% were adults ($n = 15$; age 18+ years). Participants were mostly male (male = 55; 71%, female = 20; 26%), and gender was not reported for two participants (3%). Twenty-three percent of experiments ($n = 7$) reported ethnicity (Caucasian, African American, Hispanic). Fifty percent ($n = 15$) of experiments reported information about participants' treatment or educational history. All experiments reported their participants' diagnoses or conditions, most participants (79%) had diagnoses of ASD, and other diagnoses or conditions included intellectual disability and learning disability.

Other No participants were infants or toddlers, 1% were preschoolers ($n = 5$), 23% were school age ($n = 90$), 1% were adolescent ($n = 3$), 68% were adults ($n = 270$), and for 8% of participants, we could not classify them by age group. Participants were mostly female (male = 76; 19%, female = 312; 78%), and gender was not reported for 10 (3%) participants. No experiments reported ethnicity. Forty-seven percent ($n = 7$) of experiments (i.e., some studies contained multiple experiments) reported information about participants' education or treatment history. Thirty-three percent ($n = 5$) of experiments reported diagnoses or conditions, and of those studies,

participants were either typically developing or had diagnoses of ASD.

Study Characteristics

We classified twenty-nine studies as applied research which consisted of 30 experiments, four studies as translational research which consisted of five experiments, and five studies as basic research which consisted of 10 experiments. We concluded that Camilleri and Hanley (2005) did not use a lag schedule because reinforcement was contingent on responses that differed from all prior responses within a session (i.e., differential reinforcement of novelty, not variability). Therefore, we excluded the coding guide data from this study when calculating dependent variables pertaining to lag schedules (e.g., lag schedule values or whether lag schedules were evaluated in isolation from other treatment components). Study characteristics for applied and translational studies are summarized in Tables 1, 2, 3, and 4.

Applied

Settings, Procedural Fidelity, and Interobserver Agreement

Some studies collected data in more than one setting. The settings in which experiments were conducted were mostly schools (48%), various other settings (24%; e.g., university-based classroom or vocational rehabilitation center), homes (15%), and clinics (14%). Eighteen experiments (60%) assessed procedural fidelity. Of those experiments, fidelity was perfect or nearly perfect (i.e., 90% or greater) for 83%, acceptable (i.e., 80–89%) for 11%, and poor (i.e., < 80%) for 6%; 39% selected all sessions or selected sessions randomly, and 6% assessed interobserver agreement for sessions implemented by non-research persons. All applied experiments reported interobserver agreement. Of those experiments, agreement was perfect or nearly perfect for 73% ($n = 22$) and acceptable for 27% ($n = 8$), and 27% ($n = 8$) of experiments sampled all sessions or sampled sessions randomly (see Table 3).

Methodology, Research Questions, and Manualized Interventions

All experiments used single-subject design experimental methodology, and researchers conducted those experiments using multiple baseline or multiple probe designs ($n = 12$; 35%), combined designs ($n = 11$; 32%), withdrawal or reversal designs ($n = 8$; 24%), multielement designs ($n = 2$; 6%), and changing criterion designs ($n = 1$; 3%). When we combined applied and translational studies that used single-subject design experimental methodology, it was found that researchers used these methods to answer different research questions (Kennedy 2005), which we classified as demonstration ($n = 24$; 70%), sequential analysis ($n = 7$; 21%), parametric ($n = 2$; 6%), comparative ($n = 2$; 6%), or component

Table 1 Summary of study characteristics for determining the best available evidence in the evidence-based practice of ABA for applied and translational experiments which evaluated lag schedules in isolation

Authors/ experiment	Evidential certainty	Lag values	Dependent variables	Age range	Setting	Potential unwanted effects
Esch and Esch et al. (2009)	3	1	Phonemes/vocal response variability	Toddler and school age	School	Unknown
Heldt and Schlinger (2012)	3	3	Tacts/different responses	Preschool and adolescent	Home	Unknown
Silbaugh and Falcomata (2019)	3	1	Sign manding/variant, invariant, novel, and different	Preschool	School	None
Silbaugh et al. (2017)	3	1, 2	Food consumption/variant and different	Preschool	Home	None
Silbaugh et al. (2018)	3	1	Vocal manding/variant, invariant, novel, different	Preschool	Home	None
Koehler-Platten et al. (2013)	2	1	Phonemes/variant, novel	Toddler, preschool, school age	Therapy center	Mild escape behavior and perseveration on a single phoneme
Lee and Sturmey (2014)	2	1	Intraverbals/variant and appropriate	School age	School	Unknown
Baruni et al. (2014)	1	1, 2	Toy play and engagement/novel	School age, adolescent	School	Decreased toy engagement
Lee et al. (2002)	1	1	Intraverbals/variant and appropriate	School age, adulthood	School, day habilitation program	Higher-order stereotypy
Lee and Sturmey (2006)	1	1	Intraverbals/variant and appropriate, novel, different	Adolescent, adulthood	School	Higher-order stereotypy
Lepper et al. (2017)	1	1, 2	Verbal statements/related or unre- lated to circumscribed interests	School age, adolescent	School	Contextually inappropriate response forms
Wiskow and Donaldson (2016)	1	1	Intraverbals/variant, novel, different	Preschool, school age	University-based lab classroom	None

For evidential certainty, level 3 is conclusive, level 2 is preponderant, and level 1 is suggestive

analysis ($n = 1$; 3%) (see Table 3). Lag schedules were embedded into a manualized intervention for seven (23%) experiments (e.g., O'Neill and Rehfeldt 2014; Radley et al. 2017a).

Dependent Variables, Data Collection, Stability Criteria, and Condition Signaling Of those experiments that used single-subject design methodology, to evaluate the effects of the intervention, the researchers measured variant responses or sequences in 21 experiments (62%), novel response or sequences in 16 experiments (47%), different (i.e., when “different” referred to an aspect of responding other than variability such as the number of different responses per session; e.g., Contreras and Betz 2016) responses or sequences in 13 experiments (35%), repetitive responses or sequences in two experiments (6%), and collateral changes in behavior factored into data collection for 18 experiments (53%). Examples of collateral changes include improvements in toy engagement, activity engagement, challenging behavior, appropriate play, stereotypy, appropriateness of verbal behavior, circumscribed interest-related and unrelated verbal behavior, skill accuracy, disruptive mealtime behavior, and relevance of behavior.

Researchers collected data manually in all experiments. Eight experiments (27%) used stability criteria to make phase change decisions. Three studies (10%) programmed stimuli to signal condition changes.

Invariance Assessments, Reinforcer Identification, and Academic and Functional Areas Three studies (10%) conducted invariance assessments prior to the treatment evaluation. Seventeen studies (57%) conducted assessments to identify reinforcers prior to the treatment evaluation. Of those studies, 14 (72%) used an empirical stimulus preference assessment, four (22%) used indirect assessment methods, three (17%) used functional analysis of challenging behavior, one (6%) used a reinforcer assessment, and no studies used functional behavior assessment (see Table 3). Using a classification scheme for academic and functional skills (Scheuermann et al. 2019), we classified 24 studies (75%) as targeting functional skills, three studies (9%) as targeting academic skills, three studies (9%) as targeting vocational skills, and two studies (6%) as targeting personal care skills. No studies targeted tool skills or fine arts skills.

Table 2 Summary of study characteristics for determining the best available evidence in the evidence-based practice of ABA for applied and translational experiments which evaluated lag schedules in combination with other components

Authors/experiment	Evidential certainty	Lag values	Dependent variables	Age range	Setting	Potential unwanted effects
Brodhead et al. (2016)	3	2, 3	Mand frames/different	Preschool	University-based program	None
Contreras and Betz (2016)	3	1, 3	Intraverbals/variant, novel, different	Preschool, school age	Home/clinic	Higher-order stereotypy
Falcomata et al. (2018)	3	1, 2, 3, 4, 5	Manding, challenging behavior/variant	School age	School	None
Napolitano et al. (2010)	3	1	Block play/variant	School age	School	None
O'Neill and Rehfeldt (2014)	3	1	Intraverbals/accuracy	Adulthood	Vocational training center	None
Radley et al. (2018a)	3	1, 2, 3, 4, 5	Mands, tacts/variant, novel, appropriate	School age, adolescent	School	Unknown
Radley et al. (2018b)	3	2	Mands, tacts, intraverbals novel, appropriate	School age, adolescent	School	None
Silbaugh et al. (2017)	3	1, 2	Food consumption, disruptive mealtime behavior/variant, different	Preschool	Home	None
Adami et al. (2017)	2	1	Manding, challenging behavior/variant	School age, adulthood	Home/school	Unknown
Ivy et al. (2019)	2	2	Leisure activity selection/variant, activity engagement	School age, adolescent	School	Higher-order stereotypy
Lang et al. (2014)	2	1, 2	Play, stereotypy/appropriateness	Preschool	School	None
Radley et al. (2017b)	2	2, 4	Unable to classify	School age	Clinic	Unknown
O'Neill et al. (2015)	1	1, 2	Intraverbals/accuracy	Adulthood	Vocational training center	Higher-order stereotypy
O'Neill and Rehfeldt (2016)	1	1	Intraverbals/novel, accuracy	Adulthood	Vocational training center	None
O'Neill and Rehfeldt (2017)	1	1, 2, 3	Intraverbals/accuracy	Adulthood	Vocational training center	None
Radley et al. (2017a)	1	2	Mands, intraverbals/novel, appropriateness	Preschool, school age	Clinic	Prompt dependence
Susa and Schlinger (2012)	1	1, 2, 3	Intraverbals/variant, relevance	School age	Home	Unknown
Wiskow et al. (2018), Exp. 1	1	1, 2, 4, 8, 9	Intraverbals/variant, novel, different	School age	University-based classroom	Vocal verbal protests about difficulty
Wiskow et al. (2018), Exp. 2	1	1, 2, 9	Intraverbals, tacts/variant, novel, different	School age	University-based classroom	Higher-order stereotypy

For evidential certainty, level 3 is conclusive, level 2 is preponderant, and level 1 is suggestive

Table 3 Summary of additional study characteristics

Authors/experiment	Fidelity	IOA	Research question	Invariance assessment	Treatment components	Reinforcement type
Adami et al. (2017)	N	Y (> 90%)	Demo	N	Lag + other	Pos, Neg
Baruni et al. (2014)	N	Y (> 90%)	Demo	N	Lag	Positive
Brodhead et al. (2016)	Y (> 90%)	Y (> 90%)	Demo	N	Lag + other	Positive
Camilleri and Hanley (2005)	N	Y (80–89%)	Demo	N	Lag	Positive
Contreras and Betz (2016)	Y (> 90%)	Y (> 90%)	Demo	Y	Lag + other	Positive
Dracolby et al. (2017), Exp. 2	Y (> 90%)	Y (> 90%)	Comp	N	NA	Positive
Dracolby et al. (2017), Exp. 3	Y (> 90%)	Y (> 90%)	Demo	N	NA	Positive
Esch and Esch et al. (2009)	Y (> 90%)	Y (> 90%)	Demo	N	Lag	Positive
Falcomata et al. (2018)	Y (> 90%)	Y (> 90%)	Para	N	Lag + other	Pos, Neg
Heldt and Schlinger (2012)	Y (> 90%)	Y (> 90%)	Demo	N	Lag	Positive
Ivy et al. (2019)	Y (> 90%)	Y (> 90%)	Demo	N	Lag + other	Positive
Koehler-Platten et al. (2013)	Y (> 90%)	Y (80–89%)	Demo	N	Lag	Positive
Lang et al. (2014)	Y (> 90%)	Y (80–89%)	Sequ	N	Lag + other	Positive
Lee et al. (2002)	N	Y (> 90%)	Demo	Y	Lag	Positive
Lee and Sturmey (2006)	N	Y (> 90%)	Demo	N	Lag	Positive
Lee and Sturmey (2014)	N	Y (> 90%)	Demo	N	Lag	Positive
Lepper et al. (2017)	N	Y (> 90%)	Demo	N	Lag	Positive
Napolitano et al. (2010)	Y (> 90%)	Y (80–89%)	Demo	N	Lag + other	Positive
O'Neill and Rehfeldt (2014)	Y (> 90%)	Y (> 90%)	Demo	N	Lag + other	Positive
O'Neill et al. (2015)	Y (< 80%)	Y (> 90%)	Sequ	N	Lag + other	Positive
O'Neill and Rehfeldt (2016)	N	Y (80–89%)	Demo	N	Lag + other	Positive
O'Neill and Rehfeldt (2017)	N	Y (80–89%)	Demo/Para	N	Lag + other	Positive
Radley et al. (2017a)	Y (> 90%)	Y (> 90%)	Sequ	N	Lag + other	Positive
Radley et al. (2017b)	Y (> 90%)	Y (> 90%)	Sequ	N	Lag + other	Positive
Radley et al. (2018a)	Y (> 90%)	Y (> 90%)	Sequ	N	Lag + other	Positive
Radley et al. 2018b)	Y (> 90%)	Y (> 90%)	Sequ	N	Lag + other	Positive
Silbaugh and Falcomata (2019)	Y (> 90%)	Y (> 90%)	Demo	N	Lag	Positive
Silbaugh et al. (2017)	Y (> 90%)	Y (> 90%)	Demo/Comp	N	Lag + other	Pos + Neg
Silbaugh and Falcomata (2017)	Y (> 90%)	Y (> 90%)	Demo	N	Lag	Positive
Silbaugh et al. (2018)	Y (80–89%)	Y (> 90%)	Demo	Y	Lag	Positive
Susa and Schlinger (2012)	N	Y (> 90%)	Demo	N	Lag + other	Positive
Wiskow and Donaldson (2016)	Y (> 90%)	Y (> 90%)	Demo	Y	Lag	Positive
Wiskow et al. (2018), Exp. 1	N	Y (> 90%)	Para	N	Lag + other	Positive
Wiskow et al. (2018), Exp. 2	N	Y (> 90%)	Sequ	N	Lag + other	Positive

Comp comparative analysis, Demo demonstration, N no, Neg negative reinforcement, Para parametric analysis, Pos positive reinforcement, Sequ sequential analysis, Y yes

Treatment Components, Type of Contingency, Lag Value, and Reinforcement Type For all experiments that evaluated treatments (i.e., applied and translational), most experiments ($n = 15$; 47%) evaluated lag schedules combined with other treatment components, 13 (41%) evaluated the effects of lag schedules alone, three (9%) evaluated the effects of lag schedules alone and combined with other treatment components, but no studies compared the effects of lag schedules in isolation to alternative treatments (see Table 3). All applied experiments evaluated the effects of individual lag schedule contingencies, and no studies evaluated group contingencies with lag

schedules. Twenty-seven (90%) applied experiments evaluated the effects of lag schedules on responses (e.g., bites of food, instances of manding, tacts). No applied experiments evaluated the effects of lag schedules on response sequences. Three (10%) applied experiments evaluated the effects of lag schedules on aspects of behavior we could not classify as specific responses or sequences such as activity selections (e.g., Camilleri and Hanley 2005; Ivy et al. 2019) and topical conversation content (e.g., Lepper et al. 2017). During treatment evaluations in applied studies, programmed contingencies targeted topographical response variability in 17 (57%)

Table 4 Summary of additional study characteristics

Authors/experiment	Prompts	Rules and instructions	Generalization assessment	Maintenance assessment	Social validity assessment
Adami et al. (2017)	N	N	N	N	N
Baruni et al. (2014)	N	N	N	N	Y (+/-)
Brodhead et al. (2016)	Ve, Ph, Sc	Y, Inc	Y, Set	Y (0–2 months)	N
Camilleri and Hanley (2005)	N	Y, Inc	N	N	N
Contreras and Betz (2016)	Pic, TD, Dtf	N	N	N	N
Dracolby et al. (2017), Exp. 2	N	N	N	N	N
Dracolby et al. (2017), Exp. 3	N	Y, Inc	N	N	N
Esch and Esch et al. (2009)	N	N	N	N	N
Falcomata et al. (2018)	N	N	N	Y (0–2 months)	N
Heldt and Schlinger (2012)	Ge	N	N	Y (0–2 months)	N
Ivy et al. (2019)	N	N	N	Y (0–2 months)	N
Koehler-Platten et al. (2013)	N	N	N	N	N
Lang et al. (2014)	Ve	N	Y, Comb	Y (0–2 months)	Y
Lee et al. (2002)	Ve, Ec, Dtf	N	Y, Comb	N	N
Lee and Sturmey (2006)	N	N	N	N	N
Lee and Sturmey (2014)	N	N	Y, Comb	N	Y
Lepper et al. (2017)	Ve	N	Y, Per	N	N
Napolitano et al. (2010)	Ve, Dtf	Y, Inc	Y, Mat	Y, 0–2 months, 3–6 months	N
O'Neill and Rehfeldt (2014)	N	N	Y, Comb	N	N
O'Neill et al. (2015)	N	N	Y, Comb	Y (0–2 months)	N
O'Neill and Rehfeldt (2016)	N	N	N	Y (0–2 months)	Y
O'Neill and Rehfeldt (2017)	N	N	N	Y (0–2 months)	N
Radley et al. (2017a)	Ve	N	N	Y (0–2 months)	N
Radley et al. (2017b)	N	Y, Com	N	Y (0–2 months)	Y
Radley et al. (2018a)	N	N	N	Y (0–2 months)	N
Radley et al. (2018b)	N	Y, Inc, Com	N	Y (0–2 months)	N
Silbaugh and Falcomata (2019)	Ec, TD, LTM	N	N	N	N
Silbaugh et al. (2017)	N, Dtf	N	N	N	N
Silbaugh and Falcomata (2017)	N	N	N	N	N
Silbaugh et al. (2018)	Ec, TD	N	N	N	N
Susa and Schlinger (2012)	Ec,	N	N	N	N
Wiskow and Donaldson (2016)	N	Y, Inc, Com	Y, Set	N	N
Wiskow et al. (2018), Exp. 1	N	Y, Inc	N	N	N
Wiskow et al. (2018), Exp. 2	Vi	Y, Inc	N	N	N

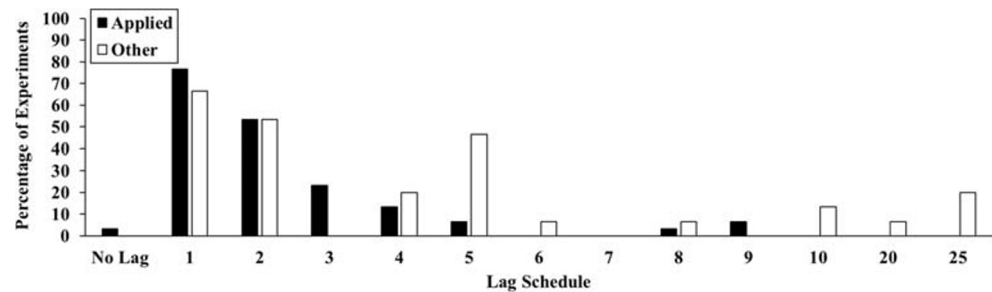
Com complete rule or instruction (i.e., contingency stated), *Comb* combination of setting, persons, or materials, *Dtf* due to lag schedule failure, *Ec* echoic prompt, *Ge* gesture prompt, *Inc* incomplete rule or instruction (i.e., contingency not stated), *LTM* least to most prompting, *Mat* generalization across materials, *N* no, *Per* generalization across persons, *Ph* physical prompt, *Sc* script, *Set* generalization across settings, *TD* time delay prompt fading, *Ve* verbal prompt, *Vi* visual prompt, *Y* yes, (+/-) positive and negative responses

experiments, selection-based response variability in six (20%) experiments, and a combination of topographical and selection-based response variability in nine (30%) experiments.

The range of lag schedule values evaluated in the literature is displayed in Fig. 3. The most frequently evaluated lag schedule values were Lag 1 ($n = 23$; 77%), followed by Lag 2, Lag 3, and Lag 4. Lag 9 was the highest lag schedule value

evaluated by an applied experiment. Of the 77 participants in applied experiments, lag schedules of positive reinforcement were evaluated in 74 (96%) participants, lag schedules of negative reinforcement were evaluated in two (3%) experiments, and for one experiment, we could not determine whether the lag schedule is consisted of positive or negative reinforcement.

Fig. 3 The percentage of applied and other (i.e., basic and translational) studies that evaluated each lag schedule value



Prompts, Self-Monitoring, Rules and Instructions, and Verbal Operants Thirteen experiments (43%) included prompts in lag schedule conditions, and 17 experiments (57%) did not include prompts in lag schedule conditions. Of those experiments that included prompts, four (31%) included them as a remediation strategy after lag schedules failed to produce the desired increases in variability. Examples of prompts include vocal prompts, model prompts, gestural prompts, physical prompts, scripts, pictorial prompts, time delay, and prompts embedded in error correction procedures. Some prompts were faded, eliminated, or both, and others were not (see Table 4). Anecdotally it appears that all prompts were response prompts. One applied study used self-monitoring (Radley et al. 2018b). Seven experiments (23%) included rules, instructions, or both in treatment evaluations, and 23 experiments (77%) did not (see Table 4). Twenty-one applied experiments (70%) targeted verbal operants in the treatment evaluation, and six experiments (20%) targeted nonverbal operants in the treatment evaluation. Of those experiments that targeted verbal behavior, seven (33%) targeted manding, four (19%) targeted tacts, and 14 (67%) targeted intraverbals (and potentially autoclitics).

Generalization, Maintenance, Social Validity, and Unwanted Effects Nine (27%) experiments (applied and translational) assessed generalization (i.e., measuring the dependent variable in a nontreatment context such as a novel setting, novel materials, or with novel people). Of those experiments, none assessed generalization across behaviors. However, generalization was assessed in two (22%) experiments across settings, in one (11%) experiment across persons, in one (11%) experiment across materials, and five (56%) experiments across a combination of settings, persons, or materials (see Table 4). Also, of those experiments that assessed generalization, full generalization was obtained in six (67%), some generalization (i.e., at least some participants) was obtained in two (22%), and no generalization was observed in one (11%). Generalization was assessed for fewer than three data points in five (56%) experiments and more than three data points in five (56%) experiments. Thirteen (43%) applied experiments assessed maintenance (i.e., the continuation of treatment effects following the removal of one or more treatment

components). Of those experiments, maintenance was assessed at 0–2 months ($n = 11$; 85%) and 3–6 months ($n = 1$; 8%). Seven experiments (54%) demonstrated full maintenance (i.e., maintenance of treatment effects) for all participants, and six (46%) demonstrated some maintenance (i.e., observed for at least some participants). Three (23%) experiments assessed maintenance for less than three data points, and 10 (77%) assessed maintenance for three or more data points (see Table 4).

Five (17%) applied experiments assessed social validity (see Table 4). All of those experiments used an indirect assessment instrument, and no experiments used objective direct measures of social validity (e.g., giving participants a choice of treatment in a concurrent operants arrangement). Eighty percent ($n = 4$) of those experiments reported only positive social validity scores, and 20% reported some negative scores. Twelve (40%) applied experiments and one (7%) other experiment (i.e., translational) reported unwanted effects such as higher-order stereotypy (Schwartz 1982) associated with a lag schedule condition (see Tables 1 and 2 for unwanted effects).

Evidential Certainty When all applied and translational experiments were examined, evidential certainty was rated conclusive for 13 (41%) experiments, preponderant for six (19%) experiments, and suggestive for 12 (42%) experiments. The 12 experiments that evaluated lag schedules in isolation are summarized in Table 1. Of these experiments, five (38%) provided a conclusive level of evidential certainty, and the dependent variables were manding, food consumption, tacts, and vocalizations.

Other

Settings, Procedural Fidelity, and Interobserver Agreement

Translational studies were conducted in the spare room of a child development center (Dracolby et al. 2017), a university-based lab classroom (Wiskow and Donaldson 2016), a lab (Murray and Healy 2013), and a home setting (Silbaugh et al. 2017). Due to the use of automated procedures, procedural fidelity was not applicable for 10 (67%) experiments. Four translational experiments (27%) assessed procedural fidelity, which was perfect or nearly perfect for those studies. Four translational experiments (27%) assessed interobserver

agreement and reported perfect or nearly perfect agreement, which was not applicable due to automated procedures for seven (47%) experiments.

Methodology and Manualized Interventions Four experiments (27%; translational studies) used single-subject experimental methodology, and 11 experiments (73%) used group design methodology. For 13 (87%) experiments, treatment manualization was not applicable, and no studies embedded lag schedules in a manualized treatment even when potentially applicable (i.e., translational experiments).

Dependent Variables, Data Collection, and Condition Signaling All basic studies measured sequence variability to evaluate the effects of lag schedules. Ten (67%) experiments used automated data collection methods (i.e., for computer-based tasks), and four translational experiments used manual data collection methods (i.e., observation and recording). One experiment (7%; Dracolby et al. 2017, Exp. 3) used condition signaling, and all other experiments used did not use condition signaling, or it was not applicable to the study.

Invariance Assessments, Reinforcer Identification, and Academic and Functional Areas One translational experiment (7%) used an invariance assessment, and invariance assessment was not applicable for 13 (87%) experiments. One translational experiment (7%) used empirical stimulus preference assessments to identify potential reinforcers. Reinforcer identification methods were not applicable for 13 experiments. One (7%) translational experiment targeted personal care skills (i.e., feeding), and one (7%) targeted academic skills (i.e., naming members of categories).

Type of Contingency, Lag Value, and Reinforcement Type Fourteen (93%) experiments programmed individual contingencies of lag schedules, and one (7%) experiment programmed a group contingency lag schedule. Four (27%) experiments targeted response variability, and 11 (73%) targeted sequence variability. One (7%) experiment targeted topography-based variability, and 14 (93%) experiments targeted selection-based variability, but no experiments targeted a combination of topography-based and selection-based variability. Experiments evaluated lag schedule values ranging from 1 to 25 (see Fig. 3). All experiments evaluated lag schedules of positive reinforcement.

Prompts, Self-Monitoring, Rules and Instructions, and Verbal Operants Prompts were either not embedded in lag schedule conditions or not applicable in all experiments. No experiments used self-monitoring. Three (20%) experiments used rules or instructions, six (40%) did not, and rules or instructions were not applicable in the remaining experiments. Verbal operants were targeted in one (7%) translational experiment,

and nonverbal operants were targeted in one (7%) translational experiment.

Discussion

We used a multistep search strategy to identify all experimental studies of lag schedules in humans published in peer-reviewed journals. Thirty-eight studies met inclusion criteria, and we summarized study and participant characteristics and assessed evidential certainty. The results suggest that lag schedules are emerging as a promising applied behavioral technology for increasing operant variability in individuals with ASD but that the research should be considered fairly limited for any given skill domain (e.g., verbal behavior, play, feeding, activity selection) given that the methods vary considerably across studies and many of the different applications of lag schedules await further replication. Additionally, more lag schedule research in adults with ASD generally, typically developing adults using single-subject design methodology, and typically developing children and adults is needed to better understand the effects of lag schedules on verbal and nonverbal behavior in typically and atypically developing individuals. We discuss the results in relation to the purposes of the current study: (a) determine what is and is not known about the effects of lag schedules in individuals with ASD and other populations, (b) determine for whom and when lag schedule-based interventions are appropriate, (c) propose current practice guidelines based on the evidential certainty from the studies, and (d) identify future avenues of research.

What We Know and Do Not Know About Lag Schedules Basic and applied researchers have proposed a wide range of potential clinical applications of lag schedules such as the mitigation of resurgence of challenging behavior following functional communication training (e.g., Adami et al. 2017), the treatment of food selectivity (e.g., Silbaugh and Falcomata 2017), enhancing the adaptive use of play skills (e.g., Baruni et al. 2014), verbal skills (e.g., Lee et al. 2002), generating novel and creative behavior (Stokes 1999), strengthening problem-solving skills or accelerating learning (e.g., Stokes et al. 2008b), and in general replacing repetitive and stereotyped behavior with increased operant variability (e.g., Silbaugh et al. 2017). Overall, the data suggest that lag schedules increase variability and novel behavior in individuals across age groups from preschool to adulthood, including typically developing individuals and those with developmental disorders or intellectual disability.

Affected Behaviors and Unwanted Effects Increased variant and novel responding were demonstrated across a variety of verbal and nonverbal behavior, mostly in individuals with ASD, such as phonemes, mands (mand frames, nonvocal

modalities, sign, vocal), tacts, intraverbals, activity selections, feeding, toy play, and conversational content. Depending on multiple variables such as the context, the participant characteristics, the programmed lag schedule contingency, and the domain of operant behavior targeted, when lag schedules are arranged to increase operant variability, practitioners can expect collateral improvements in toy engagement, activity engagement, challenging behavior, appropriate play, stereotypy, appropriateness of verbal behavior, verbal behavior unrelated to circumscribed interests, skill accuracy, disruptive mealtime behavior, or the relevance of behavior. However, we caution that lag schedules may produce unwanted effects such as higher-order stereotypy (outlined in Tables 1 and 2) under some conditions, typically with lower lag schedule values (e.g., Lag 1, Lag 2) which suggests that it may be important for practitioners to anticipate potential unwanted effects and plan for rapid data-based adjustments.

Outcome Generality and Maintenance Most experiments did not assess generalization. Of the experiments that did assess generalization, most assessed generalization across a combination of variables such as both persons and settings. Most demonstrated full generalization (i.e., for all participants), but only half of those experiments provided three or more data points to allow for visual analysis and a determination of whether generalization was transient. Perhaps more importantly, of the experiments that assessed generalization (Brodhead et al. 2016; Lang et al. 2014; Lee et al. 2002; Lee and Sturmey 2014; Lepper et al. 2017; Napolitano et al. 2010; O'Neill et al. 2014, 2015; Wiskow and Donaldson 2016), only four (Lee et al. 2002; Lee and Sturmey 2014; Lepper et al. 2017; Wiskow and Donaldson 2016) evaluated the effects of lag schedules alone (i.e., the remaining five combined lag schedules with other intervention components), and all targeted intraverbals or verbal statements directed at a conversational partner in mostly school age children using Lag 1 or Lag 2 schedules of positive reinforcement. Therefore, the evidence for generality (i.e., across settings, people, stimuli) of increased variability or novel responding attributable to lag schedule interventions alone is limited to a handful of experiments that targeted a relatively narrow range of verbal operant behavior in a narrow range of ages of children with ASD with low lag schedule values and positive reinforcement. No experiments evaluated the extent to which increased variability generalized to untrained behavior as would be expected if variability was a pivotal response. Less than half of experiments assessed maintenance, and almost all experiments that assessed maintenance did not exceed 2 months. Of those experiments, most provided more than three data points in the maintenance phase but only roughly half demonstrated full maintenance. The findings converge to indicate that the extent to which lag schedule interventions can produce generalized behavior change such as widespread lasting meaningful

replacement of repetitive or stereotyped behavior with adaptive variability in individuals with ASD remains a largely unanswered empirical question.

Basic Research Findings What we know from basic research about the effects of lag schedules in humans is limited to sequence variability in college students (e.g., pressing a sequence of keys on a keyboard to move across a pyramid in a computer-based task). In one experiment (Stokes 1999), response sequence variability under a continuous schedule of reinforcement was higher following a history of reinforcement for varying than a history of reinforcement not selective for response variability. The author suggested that early challenges (i.e., contingencies with large variability requirements) may adjust a “set point” or “level of variability” that is more resistant to change across time and schedules with increasingly lower variability requirements. That is, organisms may learn to vary more persistently when the early learning history of the behavior requires high levels of variability to produce reinforcement.

Basic studies in the current synthesis may inform our understanding of the effects of lag schedules on sequence variability in adults, such as the impact of lag schedule values and reinforcement history on the adaptive use of reinforced variability. Based on the results of two experiments, Stokes and Balsam (2001) proposed that to obtain higher sustained levels of variability, an organism needs to contact “(1) an initial period of reinforcement, followed by (2) an early change in the criterion that results in lower reinforcement and (3) an increase in variability that helps satisfy the new criterion” (Stokes and Balsam 2001, p. 181). Further, they suggested that the opportunity to establish a set point of variability may be localized to a particular optimal stage in an organism’s training in which if variability is targeted, subsequently we can expect relatively more maintenance or persistence than if variability was targeted in less optimal stages of training. Stokes et al. (2008a) demonstrated that baseline variability levels were higher in fifth grade students than students in lower grades on a computer maze game, and as sequence variability increased across lag schedule values, the group differences disappeared. When sequence variability was reinforced, older children made fewer errors than younger children, suggesting that older children on average may exhibit more sequence variability than younger children during problem-solving and make fewer errors when sequence variability is reinforced. In 2008, Stokes and Lai conducted five experiments, and in all, an early optimal period for acquiring a variability level was identified. In experiments 1, 2, and 3, they showed that early constraints that require high variability sustain higher variability levels than early constraints that can be mastered without high variability. Experiments 3, 4, and 5 showed that high levels of variability facilitate greater transfer (i.e., stimulus generalization). Early mastery of a specific task

did not facilitate transfer. Rather, high variability appeared to facilitate transfer to the novel task because the lag group had become sensitive to changes in contingency, not because it had already mastered an earlier version of the task. In summary, basic experiments have informed our understanding of the reinforcement of sequence variability in adults. The findings from these experiments suggest that high lag schedule values engender more variability and an early history of reinforcement for relatively high levels of variability early in a learner's history during a critical timepoint may improve generalized persistence of sequence variability by helping the learner more effectively adapt to changes in the environment.

Types of Response Variability Further examination of basic, translational, and applied experiments suggests a disconnect between what we know about the effects of lag schedules from these different research domains. Basic experiments exclusively targeted sequence variability, only a handful of experiments were conducted by a small number of researchers, and experiments did not use within-subjects designs necessary for demonstrating functional relations between lag schedules and participants' performance. Therefore, confident conclusions about lag schedule effects on sequence variability in humans require considerably more research and replication. Moreover, reinforcement of sequence variability seems to have little relevance in clinical contexts. Applied and translational experiments examined variability or novelty in topography-based responding (e.g., signs used to mand for a reinforcer), selection-based responding (e.g., activity selection or nonvocal functional communication modalities), or a combination (e.g., selecting and saying responses to interview questions). However, most of these experiments too have yet to be replicated, and whether lag schedule effects on selection-based sequence variability are similar to their effects on topography-based or selection-based variability in individual responses in humans is another empirical question requiring further research.

Where, when, and for whom are lag schedule interventions appropriate? The majority of participants in applied studies were children with diagnoses of ASD suggesting that lag schedule interventions may be useful for producing clinically indicated increased operant variability in this population. Generally, when deciding whether a given behavioral intervention is appropriate for a client, practitioners conduct a prescriptive pretreatment assessment, which includes considering the function(s) of existing problem behavior and what is expected of the typical learner. For example, in the early stages of developing an intervention for food selectivity, a practitioner would be expected to develop feeding goals and select relevant interventions based on the consideration of typical feeding skill and related social skill development. The results of formal indirect (e.g., feeding problem questionnaires) and

direct (i.e., observation) assessments would be analyzed to rule out medical factors and determine the following: (a) if behavioral intervention is appropriate; (b) the function of inappropriate mealtime behavior (e.g., using functional behavior assessment); and (c) appropriate feeding skills, mealtime behavior, or foods to target during intervention (e.g., Piazza et al. 2015).

Similarly, a prescriptive pretreatment assessment for lag schedule-based interventions for problems with "invariance" (Rodriguez and Thompson 2015) might include (a) administering a standardized repetitive behavior questionnaire (e.g., Bodfish et al. 2000) or comparing the level of adaptive variability exhibited by age-matched, independent, typically developing peers, in the same or similar contexts, to the level(s) exhibited by one's client; (b) determining the consequence(s) controlling invariant responding; and (c) determining whether repetitive responding is due to a limited repertoire. None of the studies in the current synthesis took this approach to intervene on an invariance behavioral problem. When we examined how applied and translational experiments assessed invariance problems prior to intervention, we identified only three (Contreras and Betz 2016; Lee et al. 2002; Silbaugh et al. 2018) applied and one (Wiskow and Donaldson 2016) translational experiment which reported using what we could consider a pretreatment "invariance assessment."

Pretreatment Invariance Assessments Contreras and Betz (2016) assessed participants' responses to feature/function/class questions such as "What can you find in the kitchen?" by repeating the question and prompting the participant to answer differently by saying "what else" or "tell me more." The assessment was completed when the participant stopped emitting new responses after three to five prompts for more responses. This assessment was used not to determine if repetitive responding was problematic for the participants, but rather to aid in the interpretation of the effects of the lag schedule after intervention. Similarly, Wiskow and Donaldson (2016) conducted listener identification, tact, and category pretests to rule out the possibility that repetitive responding was due to a deficient verbal repertoire. Lee et al. (2002) interviewed staff familiar with the participants to identify questions to which the participants always gave the same answer. Silbaugh et al. (2018) described a novel "mand topography invariance assessment" in which the experimenter provided the participant with multiple toys and snacks and repeatedly took turns with the participant until a toy or snack was shown to engage the participant when they had continuous access and temporary removal or withholding of the toy or snack evoked repetitive mand topographies.

In summary, no studies included in the current synthesis compared clients' levels of variability to levels observed in similar peers in similar contexts or assessed difficulties exhibiting appropriate behavior to maximize reinforcement in the natural

environment, as a basis for determining if there was an invariance problem. That is, the literature does not offer practitioners a prescriptive systematic approach to determining when a client exhibits an invariance problem (Rodriguez and Thompson 2015) or how to determine when the problem has been solved (i.e., the client now exhibits the appropriate level of adaptive variability). Other important questions not addressed by the current literature are the extent to which naturally occurring schedules of reinforcement selective for variability exist, how they might be detected through systematic assessment, and how such schedules influence skill development over time. The literature in the current review exclusively focused on interventions to increase variability. However, it may be important for researchers to develop training programs that teach individuals not only to increase variability but also to vary levels of variability across changes in contingencies which include increasing variability, decreasing variability, or maintaining current levels of variability as needed to maximize reinforcement (i.e., adaptive variability; e.g., Sidman 1960).

Social Validity The positive results of social validity assessments in five experiments suggest that lag schedules and their effects may sometimes be considered socially valid by stakeholders; however, the evidence is limited, and it is unclear whether lag schedule interventions are socially valid from the perspective of the intervention recipient. Some basic non-human research has shown that preference for conditions selective for variability depends on the level of variability required to produce reinforcement (Abreu-Rodrigues et al. 2005). Future research examining objective measures of social validity such as allowing clients opportunities to maximize reinforcement and choose their intervention (i.e., assess choice of schedules selective for repetition or variability in concurrent chained schedule arrangements; Hanley 2010) may help inform our understanding of when lag schedules are adaptive and appropriate.

Proposed Practice Guidelines for Using Lag Schedules in the Evidence-Based Practice of ABA Wolf et al. (2014) synthesized research on behavioral interventions that promote variability in individuals with ASD, which included lag schedule studies. Based on their results, they proposed a general preliminary approach to targeting variability in practice. Specifically, they suggested, “(a) evaluating the number and form of different responses needed to produce a desirable amount of variability given the behavior and context, (b) assessing the responses currently in the individual’s repertoire, (c) teaching new responses if needed, (d) implementing a differential reinforcement procedure, and (e) incorporating prompts, if needed, to help the learner contact reinforcement for varying” (Wolf et al. 2014, p. 9). The results of the current synthesis suggest that these guidelines still hold up in light of recent advancements

in the literature, but we would add that practitioners should also consider observing typically developing children in similar contexts to verify if there is a problem with invariance.

Updated Practice Guidelines With the tentative nature of our findings, we also recommend a conservative approach to deciding whether a lag schedule intervention is warranted, feasible, and appropriate. If the decision is to proceed with a lag schedule intervention, it should be implemented within the evidence-based practice of ABA (Slocum et al. 2014) by also considering client characteristics, characteristics of the invariance problem and associated context, and the evidential certainty and relevance of existing lag schedule interventions. Towards this aim, practitioners can refer to Tables 1 and 2.

Using the Tables in Practice Table 1 summarizes applied and translational experiments that evaluated the effects of lag schedules alone. Implementing lag schedules with high independent and dependent variable integrity can be difficult, especially when increasing the value of the lag schedule because it can be very difficult to determine in real time whether a response is eligible for reinforcement and deliver the reinforcer immediately contingent on the target response while simultaneously collecting accurate and reliable data (i.e., multiple therapists may need to be present). Therefore, we recommend that practitioners keep things simple at first by using a lag schedule without other treatment components. However, prompting and prompt fading should be applied when new responses must be added to the learner’s restricted repertoire for selection by the lag schedule. For example, first the practitioner should verify that the client exhibits an invariance problem (e.g., repetitive tacts in the presence of multiple eligible nonverbal discriminative stimuli), which warrants intervention and is not otherwise due to another skill deficit (e.g., deficient tact repertoire), and confirms the availability of adequate resources (e.g., video recording equipment and/or multiple therapists). Next, the practitioner should inspect Table 1 for studies with conclusive evidential certainty (e.g., Heldt and Schlinger 2012), with a similar dependent variable (e.g., tacting), in children of a similar age group (e.g., preschool), identify the initial lag schedule value (e.g., Lag 3), and note any potential unwanted side effects reported by researchers. To the extent that client characteristics differ from the study (e.g., adolescent), practitioners should modify the intervention as needed (with as little as possible deviation from the validated procedure) to suit the client and intensify progress monitoring. If sufficient improvements in the invariance problem are not obtained, after verifying high independent and dependent variable integrity, the practitioner should make adjustments to the intervention based on similar studies beginning again with those with high evidential certainty. If the practitioner still struggles to obtain sufficient improvements in the invariance problem, they should consider adding additional

treatment components sequentially by referring to studies in Table 2, moving from conclusive (e.g., Radley et al. 2018b) to suggestive (e.g., Wiskow et al. 2018, Exp. 2) evidential certainty studies targeting similar dependent variables in similar participants, again being careful to consider potential unwanted side effects (e.g., Wiskow et al. 2018, Exp. 2).

Compliance with Ethical Standards

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

References

References marked with an asterisk indicate studies included in the review

- Abreu-Rodrigues, J., Lattal, K. A., dos Santos, C. V., & Matos, R. A. (2005). Variation, repetition, and choice. *Journal of the Experimental Analysis of Behavior*, *83*, 147–168. <https://doi.org/10.1901/jeab.2005.33-03>.
- *Adami, S., Falcomata, T. S., Muething, C. S., & Hoffman, K. (2017). An evaluation of lag schedules of reinforcement during functional communication training: Effects on varied mand responding and challenging behavior. *Behavior Analysis in Practice*, *10*, 209–213. <https://doi.org/10.1007/s40617-017-0179-7>.
- *Baruni, R. R., Rapp, J. T., Lipe, S. L., & Novotny, M. A. (2014). Using lag schedules to increase toy play variability for children with intellectual disabilities. *Behavioral Interventions*, *29*, 21–35. <https://doi.org/10.1002/bin.1377>.
- Bijou, S. W. (1993). *Behavior analysis of child development* (p. 12). Reno: Context Press.
- Blough, D. S. (1966). The reinforcement of least-frequent interresponse times. *Journal of the Experimental Analysis of Behavior*, *9*, 581–591. <https://doi.org/10.1901/jeab.1966.9-581>.
- Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, *30*, 237–243. <https://doi.org/10.1023/A:1005596502855>.
- *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*, 34–48. <https://doi.org/10.1002/jaba.280>.
- *Camilleri, A. P., & Hanley, G. P. (2005). Use of a lag differential reinforcement contingency to increase varied selections of classroom activities. *Journal of Applied Behavior Analysis*, *38*, 111–115. <https://doi.org/10.1901/jaba.2005.34-04>.
- *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*, 3–16. <https://doi.org/10.1002/jaba.271>.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks: SAGE Publications.
- Doughty, A. H., & Galizio, A. (2015). Reinforced behavioral variability: Working towards an understanding of its behavioral mechanisms. *Journal of the Experimental Analysis of Behavior*, *104*, 252–273. <https://doi.org/10.1002/jeab.171>.
- *Dracolby, 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*, 600–621. <https://doi.org/10.1002/jaba.392>.
- *Esch, J. W., Esch, B. E., & Love, J. R. (2009). Increasing vocal variability in children with autism using a lag schedule of reinforcement. *The Analysis of Verbal Behavior*, *25*, 73–78. <https://doi.org/10.1007/BF03393071>.
- *Falcomata, T. S., Muething, C. S., Silbaugh, B. C., Adami, S., Hoffman, K., Shpall, C., & Ringdahl, J. E. (2018). Lag schedules and functional communication training: Persistence of mands and relapse of problem behavior. *Behavior Modification*, *42*, 314–334. <https://doi.org/10.1177/0145445517741475>.
- Goetz, E. M., & Baer, D. M. (1973). Social control of form diversity and the emergence of new forms in children's block building. *Journal of Applied Behavior Analysis*, *6*, 209–217. <https://doi.org/10.1901/jaba.1973.6-209>.
- Hanley, G. P. (2010). Toward effective and preferred programming: A case for the objective measurement of social validity with recipients of behavior-change programs. *Behavior Analysis in Practice*, *3*, 13–21. <https://doi.org/10.1007/BF03391754>.
- *Heldt, J., & Schlinger, H. D. (2012). Increased variability in tacting under a lag 3 schedule of reinforcement. *The Analysis of Verbal Behavior*, *28*, 131–136. <https://doi.org/10.1007/BF03393114>.
- Holth, P. (2012a). Variability as an operant? *The Behavior Analyst*, *35*, 243–248. <https://doi.org/10.1007/BF03392283>.
- Holth, P. (2012b). The creative porpoise revisited. *European Journal of Behavior Analysis*, *13*(1), 87–89. <https://doi.org/10.1080/15021149.2012.11434408>.
- *Ivy, J. W., Payne, J., & Neef, N. A. (2019). Increasing across-session variability of leisure activity selection for children with autism. *Behavior Analysis: Research and Practice*, *19*(3), 261–272. <https://doi.org/10.1037/bar0000132>.
- Kennedy, C. H. (2005). *Single-case designs for educational research* (pp. 63–75). Pearson Education, Inc..
- *Koehler-Platten, K., Grow, L. L., Schulze, K. A., & Bertone, T. (2013). Using a lag reinforcement schedule to increase phonemic variability in children with autism spectrum disorders. *The Analysis of Verbal Behavior*, *29*, 71–83. <https://doi.org/10.1007/BF03393125>.
- *Lang, R., Machalicek, W., Rispoli, M., O'Reilly, M., Sigafoos, J., Lancioni, G., et al. (2014). Play skills taught via behavioral intervention generalize, maintain, and persist in the absence of socially mediated reinforcement in children with autism. *Research in Autism Spectrum Disorders*, *8*, 860–872. <https://doi.org/10.1016/j.rasd.2014.04.007>.
- *Lee, R., & Sturmey, P. (2006). The effects of lag schedules and preferred materials on variable responding in students with autism. *Journal of Autism and Developmental Disorders*, *36*, 421–428. <https://doi.org/10.1007/s10803-006-0080-7>.
- *Lee, R., & Sturmey, P. (2014). The effects of script-fading and a Lag-1 schedule on varied social responding in children with autism. *Research in Autism Spectrum Disorders*, *8*, 440–448. <https://doi.org/10.1016/j.rasd.2014.01.003>.
- *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*, 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*, 313–316. <https://doi.org/10.13109/17518423.2016.1152612>.
- Miller, N., & Neuringer, A. (2000). Reinforcing variability in adolescents with autism. *Journal of Applied Behavior Analysis*, *33*, 151–165. <https://doi.org/10.1901/jaba.2000.33-151>.
- *Murray, C., & Healy, O. (2013). Increasing response variability in children with autism spectrum disorder using lag schedules of reinforcement. *Research in Autism Spectrum Disorders*, *7*, 1481–1488. <https://doi.org/10.1016/j.rasd.2013.08.004>.
- *Napolitano, D. A., Smith, T., Zarcone, J. R., Goodkin, K., & McAdam, D. B. (2010). Increasing response diversity in children with autism.

- Journal of Applied Behavior Analysis*, 43, 265–271. <https://doi.org/10.1901/jaba.2010.43-265>.
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, 9, 672–705. <https://doi.org/10.3758/BF03196324>.
- *O'Neill, J., & Rehfeldt, R. A. (2014). Selection-based instruction and emergence of topography-based responses to interview questions. *The Analysis of Verbal Behavior*, 30, 178–183. <https://doi.org/10.1007/s40616-014-0013-z>.
- *O'Neill, J., & Rehfeldt, R. A. (2016). Selection-based instruction with touch-screen video and the emergence of exact, recombinative, and novel topography-based responses to interview questions. *The Analysis of Verbal Behavior*, 32(2), 194–204. <https://doi.org/10.1007/s40616-016-0068-0>.
- *O'Neill, J., & Rehfeldt, R. A. (2017). Computerized behavioral skills training with selection-based instruction and lag reinforcement schedules for responses to interview questions. *Behavior Analysis: Research and Practice*, 17, 42–54. <https://doi.org/10.1037/bar0000043>.
- *O'Neill, J., Blowers, A. P., Henson, L., & Rehfeldt, R. A. (2015). Further analysis of selection-based instruction, lag reinforcement schedules, and the emergence of topography-based responses to interview questions. *The Analysis of Verbal Behavior*, 31, 126–136. <https://doi.org/10.1007/s40616-015-0031-5>.
- Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology*, 11, 429–452. <https://doi.org/10.1037/0097-7403.11.3.429>.
- Palmer, D. C. (2012). Is variability an operant? Introductory remarks. *The Behavior Analyst*, 35, 209–211. <https://doi.org/10.1007/BF03392279>.
- Piazza, C. C., Milnes, S. M., & Shalev, R. A. (2015). A behavior-analytic approach to the assessment and treatment of pediatric feeding disorders. In H. S. Roane, J. L. Ringdahl, & T. S. Falcomata (Eds.), *Clinical and Organizational Applications of Applied Behavior Analysis*. Elsevier.
- Pryor, L. W., Haag, R., & O'Reilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior*, 12, 653–661. <https://doi.org/10.1901/jeab.1969.12-653>.
- *Radley, K. C., Dart, E. H., Moore, J. W., Lum, J. D. K., & Pasqua, J. (2017a). Enhancing appropriate and variable responding in young children with autism spectrum disorder. *Developmental Neurorehabilitation*, 20, 538–548. <https://doi.org/10.1080/17518423.2017.1323973>.
- *Radley, K. C., Dart, E. H., Moore, J. W., Battaglia, A. A., & LaBrot, Z. C. (2017b). Promoting accurate variability of social skills in children with autism spectrum disorder. *Behavior Modification*, 41, 84–112. <https://doi.org/10.1177/0145445516655428>.
- *Radley, K. C., Battaglia, A. A., Dadakhodjaeva, K., Ford, W. B., & Robbins, K. (2018a). Increasing behavioral variability and social skill accuracy in children with autism spectrum disorder. *Journal of Behavioral Education*, 27(3), 395–418. <https://doi.org/10.1007/s10864-018-9294-y>.
- *Radley, K. C., Dart, E. H., Helbig, K. A., & Schrieber, S. R. (2018b). 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>.
- Rodriguez, N. M., & Thompson, R. H. (2015). Behavioral variability and autism spectrum disorders. *Journal of Applied Behavior Analysis*, 48, 167–187. <https://doi.org/10.1002/jaba.164>.
- Scheuermann, B., Webber, J., & Lang, R. (2019). *Autism: Teaching makes a difference* (2nd ed.). Boston: Cengage.
- Schwartz, B. (1982). Failure to produce response variability with reinforcement. *Journal of the Experimental Analysis of Behavior*, 37, 171–181. <https://doi.org/10.1901/jeab.1982.37-171>.
- Sidman, M. (1960). *Tactics of Scientific Research* (pp. 199–206). New York: Basic Books, Inc.
- *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, 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, 124–132. <https://doi.org/10.1007/s40617-018-00273-x>.
- Silbaugh, B. C., Penrod, B., Whelan, C. M., Hernandez, D. A., Wingate, H. V., Falcomata, T. S., & Lang, R. (2016). A systematic synthesis of behavioral interventions for food selectivity in children with autism spectrum disorders. *Review Journal of Autism and Developmental Disorders*, 3, 345–347. <https://doi.org/10.1007/s40489-016-0087-8>.
- *Silbaugh, B. C., Wingate, H. V., & Falcomata, T. S. (2017). Effects of lag schedules and response blocking on variant food consumption by a girl with autism. *Behavioral Interventions*, 32, 21–34. <https://doi.org/10.1002/bin.1453>.
- *Silbaugh, B. C., Falcomata, T. S., & Ferguson, R. H. (2018). Effects of a lag schedule of reinforcement with progressive time delay on topographical mand variability in children with autism. *Developmental Neurorehabilitation*, 21, 166–177. <https://doi.org/10.1080/17518423.2017.1369190>.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Slocum, T. A., Detrich, R., Wilczynski, M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. *Behavior Analyst*, 37, 41–56. <https://doi.org/10.1007/s40614-014-0005-2>.
- *Stokes, P. D. (1999). Learned variability levels: Implications for creativity. *Creativity Research Journal*, 12, 37–45. https://doi.org/10.1207/s15326934crj1201_5.
- *Stokes, P. D., & Balsam, P. (2001). An optimal period for setting sustained variability levels. *Psychonomic Bulletin & Review*, 8, 177–184. <https://doi.org/10.3758/BF03196155>.
- *Stokes, P. D., Lai, B., Holtz, D., Rigsbee, E., & Cherrick, D. (2008a). Effects of practice on variability, effects of variability on transfer. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 640–659. <https://doi.org/10.1037/0096-1523.34.3.640>.
- *Stokes, P. D., Holtz, D., Massel, T., Carlis, A., & Eisenberg, J. (2008b). Sources of variability in children's problem solving. *The International Journal of Creativity & Problem Solving*, 18, 49–67.
- *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, 125–130. <https://doi.org/10.1007/BF03393113>.
- Verschuur, R., Didden, R., Lang, R., Sigafos, J., & Huskens, B. (2014). Pivotal response treatment for children with autism spectrum disorders: A systematic review. *Review Journal of Autism and Developmental Disorders*, 1, 34–61. <https://doi.org/10.1007/s40489-013-0008-z>.
- *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, 472–484. <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>.
- Wolf, K., Slocum, T. A., & Kunnatana, S. S. (2014). Promoting behavioral variability in individuals with autism spectrum disorders: A literature review. *Focus on Autism and Developmental Disabilities*, 29, 1–11. <https://doi.org/10.1177/1088357614525661>.