

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications

Psychology, Department of

3-2019

A Comparison of Stimulus Set Size on Tact Training for Children with Autism Spectrum Disorder

Tiffany Kodak

M. Nerissa Halbur

Samantha Bergmann

Dayna R. Costello

Brittany Benitez

See next page for additional authors

Follow this and additional works at: https://epublications.marquette.edu/psych_fac

 Part of the [Psychology Commons](#)

Authors

Tiffany Kodak, M. Nerissa Halbur, Samantha Bergmann, Dayna R. Costello, Brittany Benitez, Miranda Olsen, Ella Gorgan, and Terra Cliett

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications/College of Arts and Sciences

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Journal of Applied Behavior Analysis, (March 2019). [DOI](#). This article is © Wiley and permission has been granted for this version to appear in [e-Publications@Marquette](#). Wiley does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Wiley.

A Comparison of Stimulus Set Size on Tact Training for Children with Autism Spectrum Disorder

Tiffany Kodak

Marquette University

Mary Halbur

Marquette University

Samantha Bergmann

University of North Texas

Dayna R. Costello

University of Wisconsin-Milwaukee

Brittany Benitez

University of Wisconsin-Milwaukee

Miranda Olsen

University of Wisconsin-Milwaukee

Ella Gorgan

University of Wisconsin-Milwaukee

Terra Cliett

University of North Texas

Abstract

Previous studies on skill acquisition have taught targets in stimulus sets composed of different numbers of stimuli. Although the rationale for selection of a stimulus set size is not clear, the number of target stimuli trained within a set is a treatment decision for which there is limited empirical support. The current investigation compared the efficiency of tact training in 4 stimulus set sizes, each of which included 12 stimuli grouped into (a) 4 sets of 3 stimuli, (b) 3 sets of 4 stimuli, (c) 2 sets of 6 stimuli, and (d) 1 set of 12 stimuli. Results of all 4 participants with autism spectrum disorder show tact training with larger (i.e., 6 and 12) stimulus set sizes was more efficient than training with smaller (i.e., 3 and 4) stimulus set sizes.

Skill-acquisition programs for learners with autism spectrum disorder (ASD) require careful arrangement to maximize learning outcomes and meet intervention goals. One type of instruction that is frequently used in a portion of a learner's intervention program involves trial-based instruction (i.e., discrete-trial instruction). These trials include the presentation of one or more discriminative stimuli, a prompt, a response opportunity, and a consequence (Smith, [2001](#)). Trials often occur in rapid succession and may include instruction on stimulus sets composed of one or more stimuli.

Stimulus set size is the number of stimuli that are simultaneously targeted for instruction. The selection of a stimulus set during trial-based instruction should be based on the targeted skill. Several frequently-used early intervention curricular manuals recommend training small sets of stimuli, particularly during early instructional programming for receptive identification training (e.g., Leaf & McEachin, [1999](#); Lovaas, [2003](#); Maurice, Green, & Foxx, [2001](#)). For example, Lovaas ([2003](#)) recommended a method of receptive identification training that involves presenting trials of one stimulus (i.e., massed-trial instruction) in isolation during sessions. Once responses to three stimuli are trained with massed-trial instruction, several stimuli are rotated within a session. That is, it is first recommended to include one stimulus per set and increase to a total of three stimuli per set. In comparison, other published recommendations suggest including stimulus sets of at least three stimuli throughout receptive identification training (e.g., Grow & LeBlanc, [2013](#); LaMarca & LaMarca, [2018](#); MacDonald & Langer, [2018](#)). Training and including a minimum of three stimuli in an array is recommended to teach the learner to engage in behavior necessary for acquisition of conditional discriminations (e.g., comparing stimuli within an array, scanning the array) as well as to minimize the development of faculty stimulus control (e.g., side bias, matching by exclusion) during training.

The selection of a stimulus set for instruction on other repertoires also should be based on arrangements that facilitate stimulus control. For example, early tact training (i.e., teaching a learner to emit a vocal response in the presence of a nonverbal stimulus; Skinner, [1957](#)) may include stimuli that are visually discrepant in a set (e.g., a red ball, a brown dog, and a yellow banana), whereas subsequent tact training may include the gradual introduction of visually similar stimuli within a set (e.g., a long-haired dog, a long-haired cat, a llama, and a goat). In early tact training, the learner's response may be under the control of one component of a stimulus (e.g., shape), whereas more advanced tact training establishes responding under multiple sources of control (e.g., shape, features, and size). Throughout tact training, stimulus sets that are relatively small (e.g., massed-trial instruction of one stimulus; two or three stimuli rotated per session) may contain an insufficient number of nonexemplars to which learners can compare the features that are unique to each stimulus and will aid in the discrimination. Thus, the stimulus set size is an important consideration, although it is a treatment decision for which there is limited empirical support.

The arrangement of smaller stimulus sets during tact training allows for multiple presentations of each stimulus. For example, researchers who arranged stimulus sets of three or four stimuli presented each stimulus at least three times during a training session (e.g., Carroll, Owsiany, & Cheatham, **2018**; Marchese, Carr, LeBlanc, Rosati, & Conroy, **2012**). In comparison, arranging larger stimulus sets may prevent repeated exposures to each stimulus during training. Clinicians may arrange smaller stimulus sets to allow for repeated presentations of each stimulus in an effort to produce rapid learning. Rapid mastery of responses to stimuli may permit expansion of learner repertoires and attainment of skills not directly targeted during intervention, and learners may acquire skills at a rate that helps to bridge the gap between their skills and those of their typically developing peers. Furthermore, practices that increase the efficiency of instruction by producing rapid mastery of stimuli may have beneficial outcomes for the field of Applied Behavior Analysis and the community such as decreasing the cost of intervention by reducing the overall length of time spent in treatment. Nevertheless, it remains unclear whether smaller stimulus sets during tact training increase the efficiency of instruction.

Studies on skill acquisition vary widely with respect to the stimulus set size selected for training (e.g., Dass, Kisamore, Vladescu, Reeve, Reeve, & Taylor-Santa, 2018; Henrickson, Rapp, & Ashbeck, **2015**; Marchese et al., **2012**; Schnell, Vladescu, Kodak, & Nottingham, **2018**). For example, Schnell et al. (**2018**) and Dass et al. (**2018**) conducted tact training with stimulus sets of three and four, respectively, for participants diagnosed with ASD. In comparison, other studies arranged larger stimulus set sizes. For example, Haq et al. (**2015**) taught tacts and textual responses with sets of 10 stimuli to two participants who received special education services in their local school district. Similarly, Yaw et al. (**2014**) conducted training with sets of 15 stimuli for students with ASD or specific learning disabilities. None of these studies provided a rationale for their selection of the number of stimuli included in a stimulus set.

The experimenters were unable to identify any skill-acquisition studies conducted with individuals with ASD that evaluated the effects of stimulus set size on learning. Empirical evaluations of common clinical practices (such as the selection of stimulus set sizes) can help to ensure that behavior-analytic methods are leading to better treatment outcomes for individuals who receive behavioral interventions targeting skill acquisition. The purpose of the present investigation was to compare the efficiency of tact training with four stimulus set sizes. Each condition included 12 stimuli trained in groups of (a) four sets of three stimuli, (b) three sets of four stimuli, (c) two sets of six stimuli, and (d) one set of twelve stimuli. The efficiency of training can be evaluated in several ways. Training methods that require fewer sessions to mastery or less instructional time (e.g., lower average minutes-to-mastery per stimulus) are considered more efficient than methods requiring more sessions and time (e.g., Carroll, Joachim, St. Peter, & Robinson, **2015**; Haq et al., **2015**). However, the number of exposures to each stimulus that are required to produce acquisition is an important variable to consider when stimuli are presented repeatedly within the same session (e.g., Cariveau, Kodak, & Campbell, **2016**). Sessions to mastery, which is a common measure of efficiency in comparison studies, should not be used to determine efficiency if conditions are composed of differing exposures to each stimulus per session (e.g., one versus multiple exposures per stimulus). The experimenters calculated several measures of efficiency to examine the data in a manner that permitted comparisons across efficiency measures and to prior skill acquisition studies.

METHOD

Participants, Setting, and Materials

Participants included three children and one adolescent who were diagnosed with ASD by a psychologist not affiliated with this study. All participants received behavioral intervention services and had intervention goals related to skill acquisition and verbal behavior. Tavi was a 3-year-old boy who communicated with one-to-three-word phrases. His mand repertoire was within Level 1 (0–18 months), and his tact repertoire was within Level 2 (18–30 months) on the Verbal Behavior-Milestones Assessment and Placement Program (VB-MAPP);

Sundberg, [2008](#)) at the onset of the study. Tavi had prior exposure to skill-acquisition programs in which three to six targets were trained simultaneously in 9- to 12-trial sessions. Silas was a 5-year-old boy who communicated with one-to-three-word phrases. Silas' mand and tact repertoires were consistent with Level 2 (18–30 months) on the VB-MAPP. Silas had prior exposure to skill-acquisition programs in which 3 to 10 targets were trained simultaneously in 9- to 12-trial sessions. Raul was a 6-year-old boy who communicated in short sentences. He scored a Level 2 on the mand and tact domains of the VB-MAPP. Raul had prior exposure to skill-acquisition programs in which 3 to 10 targets were trained simultaneously in 6- to 12-trial sessions. Duke was a 15-year-old young man who communicated with one-to-four-word phrases and short sentences. The experimenters did not conduct the VB-MAPP with Duke. However, he had acquired at least 20 mands, 300 tacts, 20 intraverbals, and generalized motor and vocal imitation repertoires. Duke had prior exposure to skill-acquisition programs in which 2 to 12 targets were trained simultaneously in 6- to 12-trial sessions. Duke had a lengthy history of receiving tokens following correct responses during skill-acquisition programs; thus, the experimenters used Duke's tokens and token board during the study.

The experimenters conducted sessions where the participant typically received behavioral skill-acquisition instruction (e.g., kitchen or bedroom of the participant's home). Data collectors, a table, chairs, relevant session materials (e.g., data sheets, picture cards), preferred items, and a video camera on a tripod for data collection were present during sessions.

Response Measurement, Interobserver Agreement, and Procedural Integrity

The primary dependent variable was an *independent correct response*, which was defined as providing a predetermined vocalization to the target stimulus within the 5-s response interval. The experimenters calculated the percentage of independent correct responses for each session by dividing the number of trials with an independent correct response by the total number of trials in the session, multiplied by 100. Observers also collected data on secondary dependent variables including *prompted responses* and *errors* during trials (data available from first author). A *prompted response* was defined as engaging in the predetermined vocalization within 5 s of a vocal model prompt. An *error* was defined as emitting any vocalization other than the predetermined response or not responding within 5 s of the initiation of the trial.

Observers also collected data that permitted an analysis of the efficiency of treatment across conditions. The efficiency measures included *exposures-to-mastery per stimulus* and the average *minutes-to-mastery per stimulus*. The experimenters calculated exposures-to-mastery per stimulus by summing the total number of trials (excluding prompts) for each stimulus in each session until it was mastered. Observers collected data on the duration of each session to calculate the average minutes-to-mastery per stimulus. The experimenter started the timer immediately before the onset of the first trial and stopped the timer after the reinforcement interval for the last trial to obtain session duration. Average minutes-to-mastery per stimulus was calculated by dividing the duration of all sessions within each condition by the number of targets mastered in the condition (e.g., 84 min divided by 12 stimuli equals 7 min per stimulus).

Two observers independently recorded data on independent correct responses, prompted responses, and session duration during 25% to 58% of participants' baseline and treatment sessions across conditions. An agreement was scored when both data collectors recorded the occurrence or nonoccurrence of each dependent variable during a trial and when the recorded session duration differed by no more than 2 s. The experimenters calculated trial-by-trial interobserver agreement (IOA) by dividing the number of trials with an agreement by the total number of trials in a session and multiplying by 100. The experimenters calculated total agreement IOA for session duration by dividing the number of sessions for which observers agreed on duration by the total number of sessions and multiplying by 100. Mean IOA across all dependent variables was 99% (range, 83% to 100%) for Tavi's first comparison, 98% (range, 75% to 100%) for Tavi's second comparison, 97% (range, 83% to 100%) for

Silas's first comparison, 97% (range, 75% to 100%) for Silas's second comparison, 99% (range, 92% to 100%) for Raul, and 99% (range, 83% to 100%) for Duke.

An observer also collected data on the experimenter's procedural integrity during 25% to 58% of participants' baseline and treatment sessions across conditions. Procedural integrity focused on the extent to which the experimenter implemented all components of the trial as described in the protocol (e.g., presenting the correct target stimulus assigned to the trial, securing attending, providing the programmed response interval, delivering the correct consequences following independent correct and prompted responses or errors). Trials were scored as either a 1 (all components of the trial were implemented correctly) or 0 (one or more of the trial components were not implemented correctly). The experimenters calculated procedural integrity by dividing the number of trials implemented correctly by the total number of trials per session and converted the ratio to a percentage. Mean procedural integrity was 99% (range, 92% to 100%) for Tavi's first comparison, 97% (range, 62% to 100%) for Tavi's second comparison, 98% (range, 58% to 100%) for Silas's first comparison, 99% (range, 92% to 100%) for Silas's second comparison, 99% (range, 92% to 100%) for Raul, and 99% (range, 83% to 100%) for Duke.

Preference Assessment

The experimenters conducted three-array modified multiple stimulus without replacement (MSWO; Carr, Nicolson, & Higbee, [2000](#)) preference assessment (Tavi and Duke) or used existing participant mands (Silas, Raul) to select tangible and edible stimuli to provide following independent correct and prompted responses. A one-trial momentary selection procedure with the top two items occurred prior to sessions (Tavi) or following the accumulation of three tokens (Duke). The experimenter placed two tangibles (Tavi) or edibles (Duke) frequently used as reinforcers during other skill-acquisition programs in a horizontal array and prompted the participant to select one. The participant received the selected item for a brief interval. That item was provided during treatment. The experimenter did not use an MSWO with Silas and Raul because they did not consistently indicate their preference for stimuli within an array. However, all participants consistently engaged in mands for preferred items that were used in skill-acquisition programs and were successful in producing mastery of their targeted skills. Thus, experimenters provided items during reinforcement intervals following participants' mands.

Identification of Stimuli

The experimenter conducted a pretest to identify stimuli for inclusion in stimulus sets. The stimuli included in the pretest were selected based on the participant's intervention goals (e.g., adjective–noun tacts). Pretest sessions consisted of 12 trials, and the experimenter presented each stimulus a total of three (Tavi, Silas, and Duke) or four times (Raul only). During each trial, the experimenter presented the stimulus and asked, “what is it?” Participants had 5 s to engage in a response, and no differential consequences were provided for independent correct or incorrect responses. The experimenter presented a mastered task trial after approximately every two pretest trials, and independent correct responses to mastered tasks produced praise and 20-s access to a tangible (Tavi, Silas, and Raul) or a token (Duke). The experimenters selected 48 stimuli (12 stimuli per condition) to include in stimulus sets, for which the participant did not engage in any independent correct responses. We assigned stimuli to conditions and sets based on dissimilar visual characteristics (e.g., stimuli contained different colors and shapes) and auditory characteristics (e.g., stimuli did not rhyme or start with the same sound), the absence of independent correct responses during the pretest, and similarity in the number of syllables (Gast, [2010](#)). Tavi's, Raul's, and Duke's target stimuli included adjective–noun tacts. Silas's target stimuli were one-word tacts of common items (see the online supporting information target lists).

Stimulus Set Size

All conditions included 12 target stimuli. However, the number of target stimuli included in each stimulus set varied across conditions. All conditions had 12-trial sessions; the number of exposures per stimulus per session varied based on the size of the stimulus set.

Three stimuli

Each stimulus set included a group of three target stimuli. The experimenter presented three target stimuli in a semirandom order (i.e., stimuli could not be presented more than twice in a row) four times per session. Four sets of three stimuli were taught during treatment.

Four stimuli

Each stimulus set included a group of four target stimuli. The experimenter presented four target stimuli in a semirandom order three times per session. Three sets of four stimuli were taught during treatment.

Six stimuli

Each stimulus set included a group of six target stimuli. Sessions included six target stimuli presented in a random order twice per session. Two sets of six stimuli were taught during treatment.

Twelve stimuli

The stimulus set included a group of 12 target stimuli. The experimenter presented all 12 stimuli in a random order once per session. One set of 12 stimuli were taught during treatment.

Procedure

One to two sessions of each condition were conducted per day, 2 to 5 days per week. An adapted alternating treatments design was used to evaluate the effects of stimulus set size on independent correct responses. In addition, a multiple-probe design across stimulus sets was used in three of the conditions (i.e., three, four, and six stimuli). The multiple-probe design included baseline probes conducted prior to training of each set. For example, in the three-stimuli condition, the experimenter conducted a baseline probe of Sets 2, 3 and 4 during the baseline for Set 1. Once responding met mastery in Set 1, the experimenter initiated the baseline phase for Set 2 and included a baseline probe for Sets 3 and 4. Refer to online supporting information for a flow chart that details the sequence of baseline probes in each condition. Sessions of each condition occurred in session blocks in which the experimenter conducted one session of each condition in a random order before reordering the sessions for the next session block. The mastery criterion for each set of stimuli was two consecutive sessions with 100% independent correct responses. A replication of the treatment comparison was conducted with two participants (Tavi and Silas).

Baseline

The experimenter presented the target stimulus, ensured attending to the stimulus (prompted the participant to "look," if necessary), asked, "what is it?" and waited 5 s for a response. If the participant engaged in an independent correct response or an error, the experimenter removed the stimulus and initiated the next trial. Mastered tasks (e.g., nontarget tacts, motor imitation) were interspersed approximately every other trial to maintain participant responding. The experimenter provided praise and 20-s access to a tangible item (Tavi, Silas, and Raul) or token (Duke) for independent or prompted correct responses to mastered tasks. After conducting one session of all future stimulus sets (in the first baseline) and at least three sessions of the target stimulus set with stable responding, we moved to treatment.

Treatment

The trial arrangement was similar to baseline except that the experimenter did not intersperse mastered tasks and provided prompts and reinforcement. Training of each stimulus set began with a 0-s prompt delay. The

experimenter immediately provided a vocal model prompt after presenting the target stimulus and asking, "What is it?" Following a prompted response, the experimenter delivered praise and a tangible for 20 s or a token. Errors resulted in removal of the stimulus and the initiation of the next trial. Once the participant responded correctly in at least 90% prompted responses, the prompt delay increased to 5 s. During the 5-s prompt delay, the participant had 5 s to engage in an independent correct response, which produced praise and a tangible for 20 s or a token. Following an error, the experimenter provided a vocal model prompt. A prompted response resulted in praise and 20-s access to a tangible or token until the participant engaged in independent correct responses during at least 50% of trials for one session. Thereafter, the experimenter provided praise only following prompted responses. An error following a prompt resulted in the initiation of the next trial. Treatment continued until the participant's independent correct responses met the mastery criterion (i.e., two consecutive sessions with 100% correct independent responses). Thereafter, the experimenter conducted baseline and treatment for additional stimulus sets in a condition until all stimulus sets reached the mastery criterion (refer to the online supporting information for a flow chart of procedures).

Maintenance

At 1 and 2 weeks following mastery of a stimulus set, the experimenter conducted a session of maintenance. The procedures were identical to baseline except that sessions included two presentations of each stimulus. For example, the three-stimuli condition included six trials, and the six-stimuli condition included twelve trials. Thus, the number of trials per session varied across conditions. Participants did not practice nor were they exposed to mastered sets of stimuli prior to the completion of maintenance probes.

RESULTS

Figures [1-6](#) depict the results of each participant's comparison analyses. Although the stimulus set size conditions were compared within an adapted alternating treatments design, data for each condition are shown across tiers due to the timing of baseline sessions for additional stimulus sets following mastery of one stimulus set (i.e., the multiple-probe design for sets within a condition). Thus, each tier in the figure shows the multiple-probe design across sets for each condition, and the adapted alternating treatments design is shown across tiers. Experimental control is demonstrated when independent correct responding only increased when each stimulus set was exposed to treatment and not during repeated baseline measures. In addition, experimental control is demonstrated in the embedded adapted alternating treatments design when there were fewer total sessions to mastery across conditions for all participants.

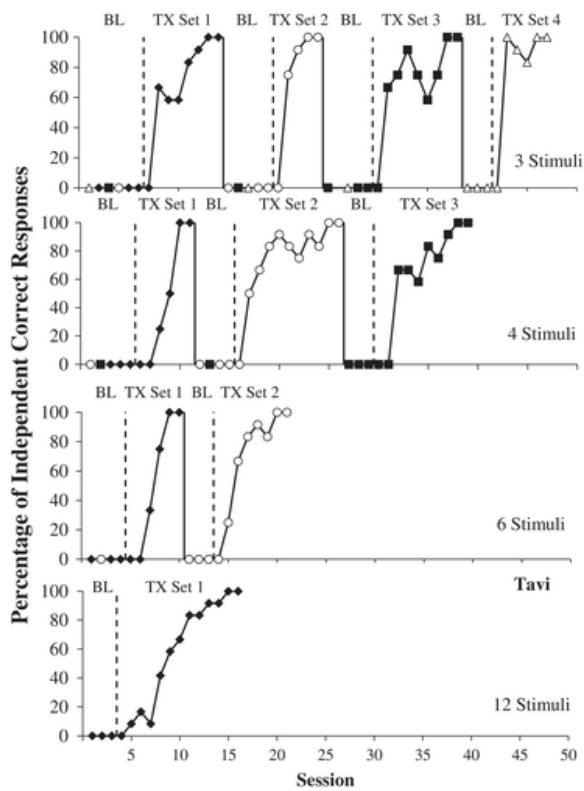


Figure 1 Tavi's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

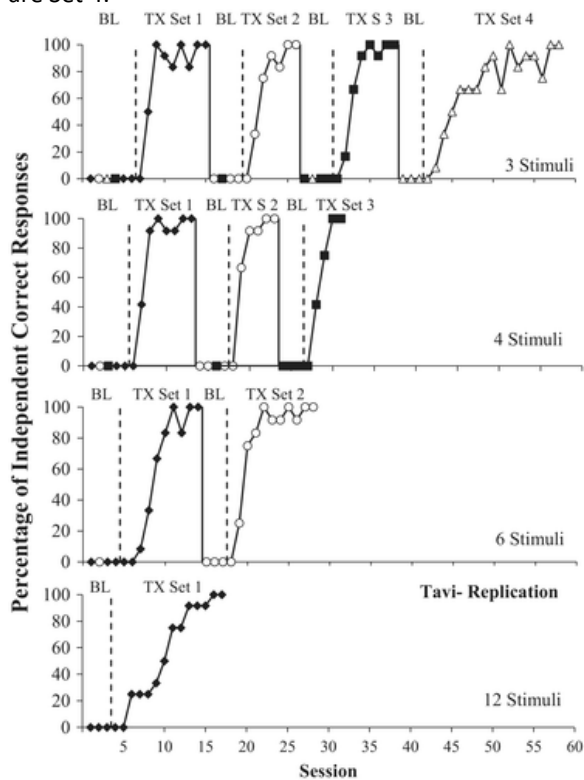


Figure 2 Tavi's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions of his replication. S = Set. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

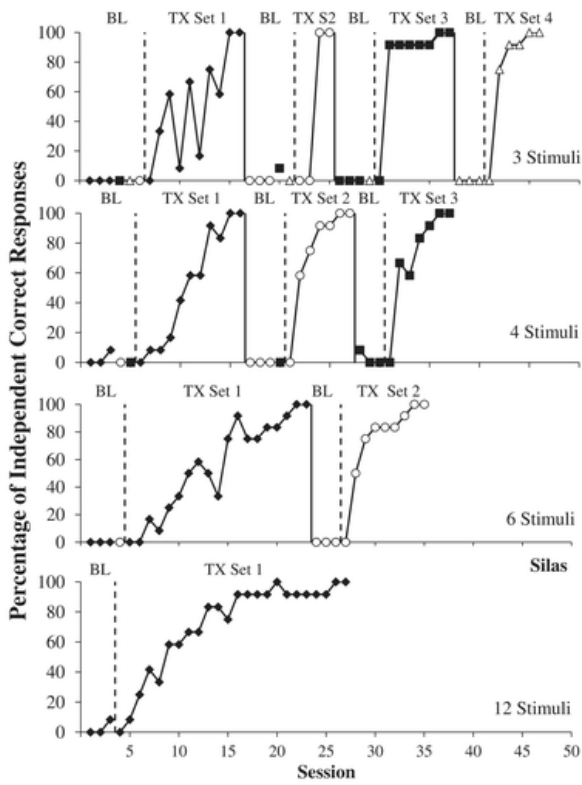


Figure 3 Silas's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions. S = Set. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

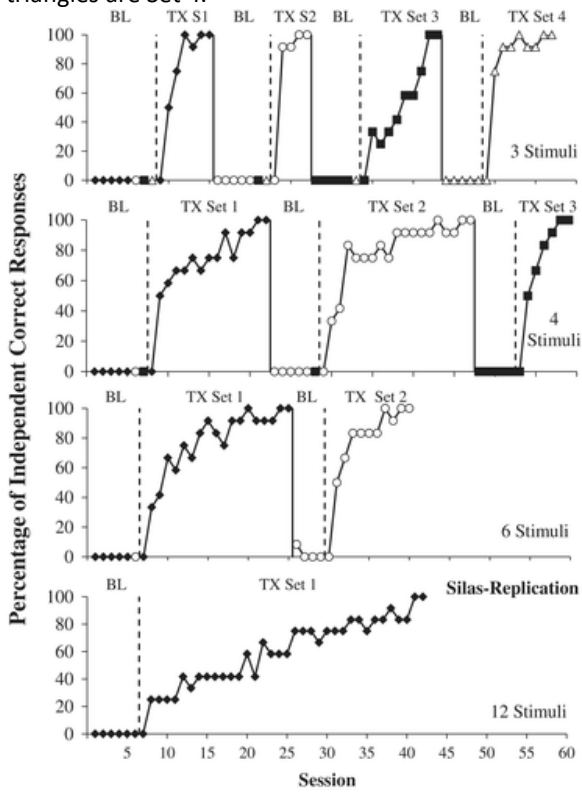


Figure 4 Silas's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions of his replication. S = Set. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

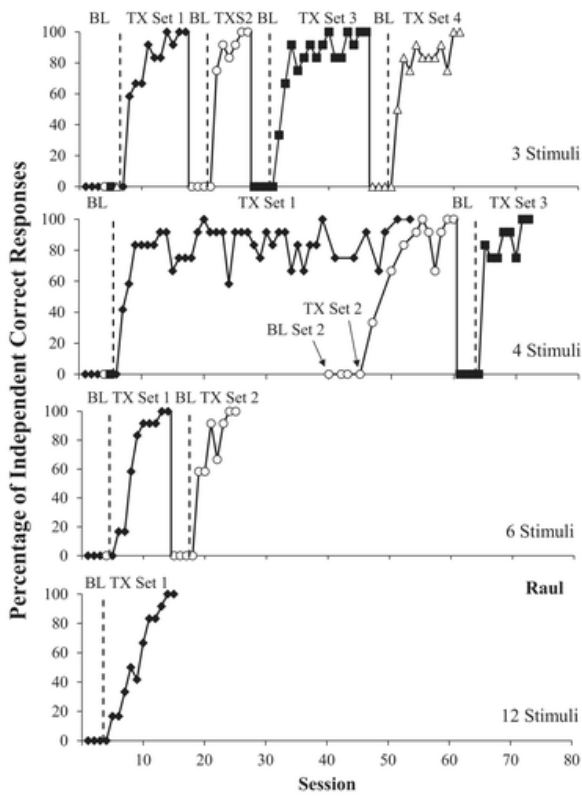


Figure 5 Raul's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions. BL and TX for Set 2 overlap with TX of Set 1 in the four-stimuli condition. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

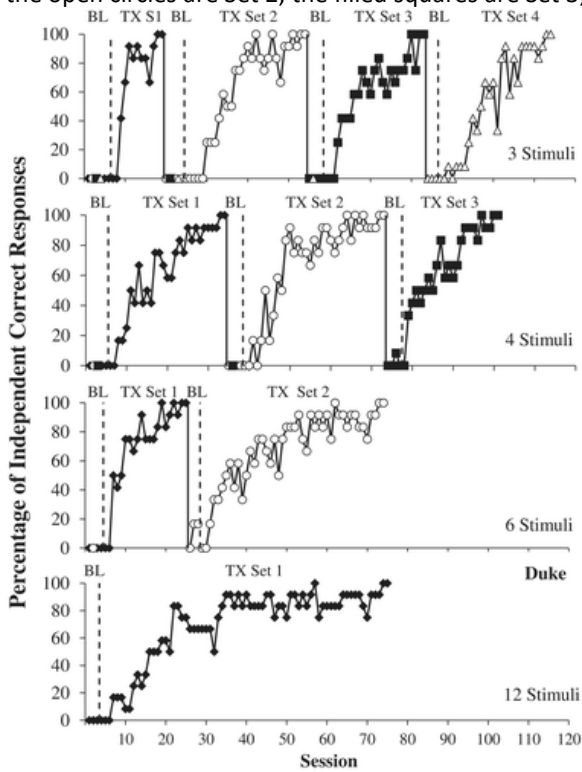


Figure 6 Duke's percentage of independent correct responses in baseline (BL) and treatment (TX) for each set of stimuli across conditions. S = Set. The filled diamonds are Set 1; the open circles are Set 2; the filled squares are Set 3; the open triangles are Set 4.

Tavi did not engage in an independent correct response in any of the baseline sessions across stimulus sets and conditions during his first (Figure 1) or second (Figure 2) comparisons. During both comparisons, Tavi mastered all stimuli in each set. No acquisition was evident in baseline for any skill set, which provides a demonstration of experimental control of the teaching procedures over the acquired responses. Tavi mastered all stimulus sets in the three-, four-, six-, and twelve-stimuli conditions in 28 (total exposures-to-mastery per stimulus: 112 exposures; average minutes-to-mastery per stimulus: 14.3 min), 27 (81 exposures; 13.5 min), 14 (28 exposures; 7.5 min), and 13 (13 exposures; 7 min) training sessions, respectively. During his replication, we observed similar outcomes. Tavi mastered all stimulus sets in the three-, four-, six-, and twelve-stimuli conditions in 39 (156 exposures; 23 min), 20 (60 exposures; 10.5 min), 21 (42 exposures; 11.7 min), and 14 (14 exposures; 8 min) training sessions, respectively. Overall, findings from both comparisons indicate that Tavi required the fewest exposures-to-mastery per stimulus and the lowest average minutes-to-mastery per stimulus in the twelve-stimuli condition (Figure 7). The three-stimuli condition was associated with the most exposures-to-mastery per stimulus and highest average minutes-to-mastery per stimulus.

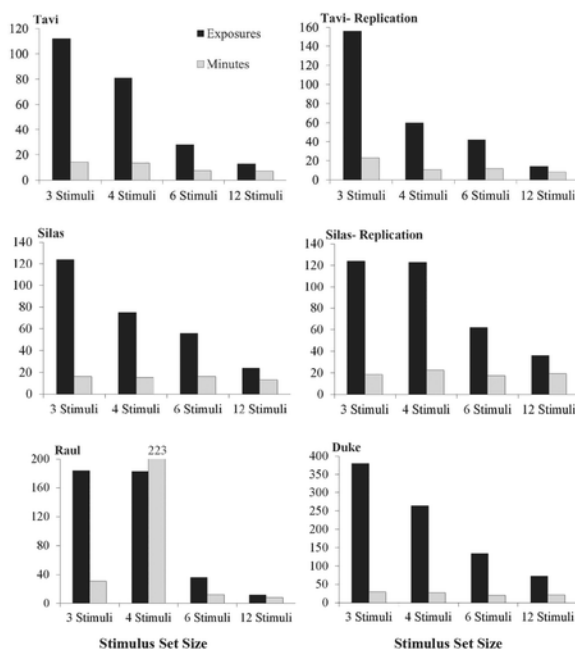


Figure 7 Exposures- and minutes-to-mastery per stimulus across stimulus set size conditions for Tavi, Silas, Raul, and Duke. Number above Raul's 4 Stimuli bar = total minutes to mastery.

Silas had near-zero levels of independent correct responses in baseline across conditions during his first (Figure 3) and second (Figure 4) comparisons. In both comparisons, Silas mastered all stimuli assigned to each set only after treatment was introduced. During his first comparison, Silas mastered all stimulus sets in the three-, four-, six-, and twelve-stimuli conditions in 31 (124 exposures; 16.3 min), 25 (75 exposures; 15 min), 28 (56 exposures; 16.1 min), and 24 (24 exposures; 13.1 min) training sessions, respectively. During his replication, Silas mastered all stimulus sets in the three-, four-, six-, and twelve-stimuli conditions in 31 (124 exposures; 18.5 min), 41 (123 exposures; 22.5 min), 31 (62 exposures; 17.4 min), and 36 (36 exposures; 19.3 min) training sessions, respectively. Overall, findings from both comparisons indicate that Silas required the fewest exposures-to-mastery per stimulus in the twelve-stimuli condition (Figure 7). Silas required the lowest average minutes-to-mastery per stimulus in the six- or twelve-stimuli conditions. The three-stimuli condition was associated with the highest exposures-to-mastery per stimulus.

Raul had no independent correct responses in baseline sessions across conditions, and he mastered all stimuli in each set (Figure 5). Raul mastered all stimulus sets in the three-, four-, six-, and twelve-stimuli conditions in 46

(184 exposures; 31 min), 61 (183 exposures; 233.2 min), 18 (36 exposures; 12.3 min), and 12 (12 exposures; 8.2 min) training sessions, respectively. Raul had an unusual pattern of independent correct responding in Set 1 for the four-stimuli condition. After a rapid increase in independent correct responding in the first four treatment sessions, he had inconsistent patterns of errors to two stimuli and varied levels of independent correct responding thereafter. The experimenter initiated treatment with Set 2 in the four-stimuli condition prior to mastery of Set 1 to prevent a delay in the completion of training in this condition. Even if data from Set 1 in the 4-stimuli condition were excluded from the analysis, Raul still would have required more exposures and average minutes-to-mastery to acquire the remaining eight stimuli in the four-stimuli condition (i.e., 90 exposures; 160 min) than was necessary to teach all 12 stimuli in both the 6- and 12-stimuli conditions. Overall, Raul's results show he required the fewest exposures-to-mastery per stimulus and the lowest average minutes-to-mastery per stimulus in the 12-stimuli condition. The 4-stimuli condition was associated with the highest average minutes-to-mastery per stimulus.

Duke had near-zero levels of independent correct responses in baseline sessions across conditions (Figure 6). Duke mastered all stimuli in each set only after treatment was introduced in the three-, four-, six-, and twelve-stimuli conditions in 95 (380 exposures; 29.8 min), 88 (264 exposures; 26.9 min), 67 (134 exposures; 19.9 min), and 72 (72 exposures; 21.8 min) training sessions, respectively. Overall, findings indicate that Duke required the fewest exposures-to-mastery per stimulus in the 12-stimuli condition. Duke required the lowest average minutes-to-mastery per stimulus in the 6-stimuli condition. The 3-stimuli condition was associated with the most exposures-to-mastery per stimulus and highest average minutes-to-mastery per stimulus.

Maintenance sessions conducted 1 and 2 weeks following mastery of each stimulus set show overall high levels of independent correct responding across sets and conditions for Tavi, Raul, and Duke (data are available from first author). Their mean independent correct responses per condition during maintenance sessions was at or above 90% (range, 70% to 100%). In contrast, Silas emitted low levels of independent correct responses across all sets and conditions during maintenance sessions. Silas displayed a pattern of responding during maintenance sessions that involved engaging in a correct independent response on the first trial which contacted extinction, and then he did not respond for the rest of the session. Following the study, Silas was exposed to a reinforcement thinning protocol to maintain high levels of independent correct responses across sets and conditions under more naturalistic reinforcement contingencies (data available from first author).

DISCUSSION

No previous research has examined how set size may influence acquisition for learners with ASD. The current study evaluated the effectiveness and efficiency of tact instruction across four different set sizes. Across participants, all set-size conditions were efficacious, but there were differences in efficiency across conditions. More specifically, in all six comparisons the six- or twelve-stimuli conditions required fewer exposures-to-mastery and a lower average minutes-to-mastery per stimulus relative to the three- and four-stimuli conditions. Thus, tact training with stimuli presented in larger stimulus sets (six or twelve stimuli) was more efficient than training with smaller stimulus sets for all participants.

The results are contrary to recommendations in some early intervention curricular manuals that recommend tact training (referred to as expressive labeling) of small stimulus sets (e.g., Leaf & McEachin, [1999](#); Lovaas, [2003](#)). Although some of the stimulus sets that consisted of just three stimuli resulted in rapid acquisition in the current investigation (e.g., Tavi's Set 2 in the three-stimuli condition in his first comparison, Raul's Set 2 in the three-stimuli condition, Duke's Set 1 in the three-stimuli condition), other sets within the same condition required a similar or greater number of sessions to reach mastery compared to somewhat larger stimulus set sizes (i.e., four or six stimuli per set). In addition, training of larger sets of stimuli (i.e., six or twelve stimuli) required the fewest exposures and average minutes-to-mastery per stimulus for all six treatment

comparisons for our four participants. Nevertheless, the experimenters did not evaluate training of even smaller sets of stimuli (e.g., massed-trial instruction for one stimulus), which are recommended practices in some curricular manuals (e.g., Lovaas, **2003**). Massed-trial instruction was omitted from the current investigation because prior studies that compared massed-trial instruction to varied-trial instruction (i.e., treatment for a small set of stimuli) show that varied-trial instruction is more effective and efficient (e.g., Cariveau et al., **2016**; Dunlap & Koegel, **1980**). In addition, none of our participants had prior exposure to massed-trial instruction. However, researchers could conduct additional evaluations to identify the ideal stimulus set size for training based on the sources of stimulus control being targeted during instruction.

Although the current study arranged sets of stimuli that were visually discrepant, the results suggest that larger stimulus sets facilitated rapid acquisition. It is possible that arranging stimuli in sets of six or twelve provided a range of nonexemplars of each stimulus, and smaller stimulus set sizes did not include sufficient variation in stimuli to facilitate the rapid development of stimulus control. Nevertheless, the ideal stimulus set size may vary based on the targeted skill and complexity of the stimuli. Additional research is needed to investigate beneficial arrangements of exemplars and nonexemplars of stimuli that lead to rapid acquisition.

The number of sessions to mastery provided a broad comparison of efficiency across conditions, and the results show that larger stimulus sets required fewer overall sessions to mastery. However, the number of exposures per stimulus in each session varied across conditions; training of smaller stimulus set sizes provided more exposures per stimulus in each session (e.g., four exposures to each stimulus per session in the three-stimuli condition). Thus, the inclusion of exposures-to-mastery per stimulus was an important efficiency measure, particularly because it allowed us to evaluate the assumption that more exposures to a stimulus may increase the rate of learning. The results of this efficiency measure for all six comparisons show the more exposures a participant received to each stimulus in the session, the less efficient the training outcomes (Figure 7). This outcome is in line with studies that have compared massed- versus varied-trial instruction. For example, Cariveau et al. (**2016**) compared massed-trial instruction in which one target was presented in all nine trials to varied-trial instruction in which three targets were presented three times each per nine-trial session. Both participants required more trials- and average minutes-to-mastery per target during massed-trial instruction. Thus, presenting the same stimulus on all nine trials did not increase the efficiency of training.

Similarly, previous research on massed and distributed practice (Haq & Kodak, **2015**; Haq et al., **2015**) compared the effects of grouping many practice opportunities into one session (referred to as massed practice) versus spacing practice opportunities across days (referred to as distributed practice). The results of those studies showed participants required fewer trials- and minutes-to-mastery during distributed practice opportunities. The format of instruction in sessions in the three-, four-, and six-stimuli conditions could be described as a type of massed practice based on repeated practice opportunities for each stimulus per session. The collective results of the present study and those on massed-trial instruction, varied-trial instruction, massed practice, and distributed practice suggest that arranging more exposures to the same stimulus within a session may not increase the efficiency of instruction. Thus, the effects of timing and distribution of practice opportunities on acquisition and instructional efficiency remains an area in need of further investigation.

Exposures to each stimulus within a session were not equated across conditions. That is, conditions with fewer targets in each stimulus set included more exposures (i.e., trials) to each stimulus within a session. The experimenters elected to arrange the training conditions and stimulus set sizes to match procedures in previous studies. For example, previous investigations provided repeated presentations of a stimulus within sessions when conducting training with smaller stimulus set sizes (e.g., Carroll, Owsiany, & Cheatham, **2018**; Marchese et al., **2012**; McGhan & Lerman, **2013**; Richardson et al., **2017**). Nevertheless, it is possible that conducting sessions in which participants receive one exposure per stimulus (e.g., a session composed of three trials with three

stimuli presented one time each) could produce different outcomes. Researchers could extend our findings by conducting this type of comparison across stimulus set sizes.

An error occurred in the timing of the introduction of treatment in two conditions of Silas's first treatment comparison. Treatment began after an increase in independent correct responses occurred in baseline for Set 1 in the four-stimuli and twelve-stimuli conditions. Silas may have mastered those stimulus sets more quickly if one of the target stimuli in those sets had already been acquired. However, the stimulus set in the four-stimuli condition to which he had an independent correct response in baseline required the most treatment sessions to reach the mastery criterion in comparison to the other stimulus sets in that same condition. In addition, Duke received an extra 0-s prompt delay session for all stimulus sets except two (Set 4 in the three-stimuli condition and Set 3 in the four-stimuli condition) due to experimenter error. Although the additional exposure to immediate prompts could result in more rapid acquisition for these stimulus sets, Duke required a similar or greater number of treatment sessions to reach mastery for these sets in comparison to other stimulus sets within the same condition or across conditions.

Implementation of a consistent mastery criterion across conditions resulted in an increasingly stringent requirement for independent correct responding to stimuli as the stimulus set size decreased. For example, participants had to engage in six consecutive independent correct responses to each stimulus to meet the mastery criterion in the four-stimuli condition (e.g., three presentations of each stimulus per session multiplied by two consecutive sessions equal six consecutive independent correct responses per stimulus). In comparison, participants only had to engage in two consecutive independent correct responses per stimulus to meet the mastery criterion in the twelve-stimuli condition. Although the mastery criterion and stimulus arrangements were designed to be consistent with clinical practice and previous studies on skill acquisition, the more stringent criterion for consecutive correct responses per stimulus may have reduced the efficiency of training for smaller stimulus set sizes. Alternative procedural arrangements in which each stimulus is presented only once per session could address this limitation in future studies.

The participants in the study had prior exposure to several of the stimulus sets sizes included in the investigation, though the extent of their exposure to larger stimulus sets varied. Although some prior programs included sets of 6, 10, or 12 stimuli, most of the participants' skill-acquisition programs included stimuli taught in sets of three or four presented in sessions with nine or twelve trials (i.e., three to four exposures per session). Unlike the results of Coon and Miguel (2012) that show prior exposure to prompt-fading procedures increased the efficiency of those procedures, the current findings show that the most efficient conditions were those to which the participants had limited exposure. All four participants required the fewest exposures-to-mastery per stimulus in the 12-stimuli condition, yet only Duke had previously received training on a set of 12 stimuli (once in a 30-month period prior to the study). If prior exposure to stimulus sets of three and four stimuli enhanced the efficiency of those procedures in the current investigation, then it is possible that training stimuli in larger set sizes is even more efficient than our results suggest. Further investigation of stimulus set sizes could include participants who have either (a) no prior exposure to trial-based instruction nor training in stimulus sets, or (b) similar exposure to all stimulus set sizes.

Although the experimenters used common methods for equating stimuli across conditions (Gast, 2010), it is possible that certain responses to stimuli within a stimulus set were more difficult to master. This appeared to be the case for Set 1 in Raul's four-stimuli condition, which required 41 treatment sessions to reach the mastery criterion. Despite these differences, the outcomes of all six comparisons across our four participants suggest that targeting smaller stimulus sets during tact training did not result in more efficient intervention. Such consistent results across participants is relatively uncommon in skill-acquisition studies, and additional research is necessary to replicate these results with more participants and skills.

The current investigation is one of the first to examine the effects of stimulus set size on the efficiency of skill acquisition for children with ASD. The results of six comparisons across four participants suggest larger stimulus set sizes during tact training lead to more efficient acquisition. Early intervention programming requires consideration of variables that may alter a learner's performance during intervention. Stimulus set size appears to be one variable that requires further investigation.

REFERENCES

- Cariveau, T., Kodak, T., & Campbell, V. (2016). The effects of intertrial interval and instructional format on skill acquisition and maintenance for children with autism. *Journal of Applied Behavior Analysis, 49*, 809–825. <https://doi.org/10.1002/jaba.322>
- Carr, J. E., Nicolson, A. C., & Higbee, T. S. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33*, 353–357. <https://doi.org/10.1901/jaba.2000.33-353>
- Carroll, R. A., Joachim, B. T., St. Peter, C. C., & Robinson, N. (2015). A comparison of error-correction procedures on skill acquisition during discrete-trial instruction. *Journal of Applied Behavior Analysis, 48*, 257–273. <https://doi.org/10.1002/jaba.205>
- Carroll, R. A., Owsiany, J., & Cheatham, J. M. (2018). Using an abbreviated assessment to identify effective error-correction procedures for individual learners during discrete-trial instruction. *Journal of Applied Behavior Analysis, 51*, 482–501. <https://doi.org/10.1002/jaba.460>
- Coon, J. T., & Miguel, C. F. (2012). The role of increased exposure to transfer-of-stimulus-control procedures on the acquisition of intraverbal behavior. *Journal of Applied Behavior Analysis, 45*, 657–666. <https://doi.org/10.1901/jaba.2012.45-657>
- Dass, T. K., Kisamore, A. N., Vladescu, J. C., Reeve, K. F., Reeve, S. A., & Taylor, S. C. (2018). Teaching children with autism spectrum disorder to tact olfactory stimuli. *Journal of Applied Behavior Analysis, 51*, 538–552. <https://doi.org/10.1002/jaba.470>
- Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis, 13*, 619–627. <https://doi.org/10.1901/jaba.1980.13-619>
- Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York: Routledge.
- Grow, L., & LeBlanc, L. (2013). Teaching receptive language skills: Recommendations for instructors. *Behavior Analysis in Practice, 6*, 56–75. <https://doi.org/10.1007/BF03391791>
- Haq, S. S., & Kodak, T. (2015). Evaluating the effects of massed and distributed practice on acquisition and maintenance of tacts and textual behavior with typically developing children. *Journal of Applied Behavior Analysis, 48*, 85–95. <https://doi.org/10.1002/jaba.178>
- Haq, S. S., Kodak, T., Kurtz-Nelson, E., Porritt, M., Rush, K., & Cariveau, T. (2015). Comparing the effects of massed and distributed practice on skill acquisition for children with autism. *Journal of Applied Behavior Analysis, 48*, 454–459. <https://doi.org/10.1002/jaba.213>
- Henrickson, M. L., Rapp, J. T., & Ashbeck, H. A. (2015). Teaching with massed versus interspersed trials: Effects on acquisition, maintenance, and problem behavior. *Behavioral Interventions, 30*, 36–50. <https://doi.org/10.1002/bin.1396>
- LaMarca, V., & LaMarca, J. (2018). Designing receptive language programs: Pushing the boundaries of research and practice. *Behavior Analysis in Practice*. <https://doi.org/10.1007/s40617-018-0208-1>
- Leaf, R., & McEachin, J. A., (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York: DRL Books
- Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: PRO- ED.

- MacDonald, R., & Langer, S. (2018). *Teaching essential discrimination skills to children with autism: A practical guide for parents and educators*. Bethesda, MD: Woodbine House.
- Marchese, M. V., Carr, J. E., LeBlanc, L. A., Rosati, T. C., & Conroy, S. A. (2012). The effects of the question “What is this?” on tact-training outcomes of children with autism. *Journal of Applied Behavior Analysis*, **45**, 539– 547. <https://doi.org/10.1901/jaba.2012.45-539>
- Maurice, C., Green, G., & Foxx, R. M. (2001). *Making a difference: Behavioral intervention for autism*. Austin, TX: PRO- ED.
- McGhan, A. C., & Lerman, D. C. (2013). An assessment of error-correction procedures for learners with autism. *Journal of Applied Behavior Analysis*, **46**, 626– 639. <https://doi.org/10.1002/jaba.65>
- Richardson, A. R., Lerman, D. C., Nissen, M. A., Luck, K. M., Neal, A. E., Bao, S., & Tsami, L. (2017). Can pictures promote the acquisition of sight-word reading? An evaluation of two potential instructional strategies. *Journal of Applied Behavior Analysis*, **50**, 67– 86. <https://doi.org/10.1002/jaba.354>
- Schnell, L. K., Vladescu, J. C., Kodak, T., & Nottingham, C. L. (2018). Comparing procedures on the acquisition and generalization of tacts for children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, **51**, 769– 783. <https://doi.org/10.1002/jaba.480>
- Skinner, B. F. (1957). *Verbal behavior*. Action, MA: Copley Publishing Group.
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities*, **16**, 86– 92. <https://doi.org/10.1177/108835760101600204>
- Sundberg, M. L. (2008) *Verbal behavior milestones assessment and placement program: The VB-MAPP*. Concord, CA: AVB Press.
- Yaw, J., Skinner, C. H., Delisle, J., Skinner, A. L., Maurer, K., Cihak, D., ... Booher, J. (2014). Measurement scale influences in the evaluation of sight-word reading interventions. *Journal of Applied Behavior Analysis*, **47**, 360– 379. <https://doi.org/10.1002/jaba.126>