Spring 2019

Comparison of task interspersal ratios on efficiency of learning and problem behavior for children with autism spectrum disorder

Sophie C. Knutson

Tiffany Kodak

Dayna R. Costell

Terra Cliett

Psychology Faculty Research and Publications/College of Arts and Sciences

Journal of Applied Behavior Analysis, Vol. 52, No. 2 (Spring 2019) : 355-369. DOI.

Comparison of task interspersal ratios on efficiency of learning and problem behavior for children with autism spectrum disorder

Sophie C Knutson: University of Wisconsin-Milwaukee

Tiffany Kodak: University of Wisconsin-Milwaukee

Dayna R. Costello: University of Wisconsin-Milwaukee

Terra Cliett: University of North Texas

Abstract

The current study extends the literature on task interspersal (TI) by comparing the effects of four different TI ratios on the efficiency of skill acquisition and on levels of problem behavior in children with autism spectrum disorder and related disorders. The four ratios of TI were 3:1, 1:1, 1:3, and 0:1 mastered-to-acquisition tasks. An adapted alternating treatments design was implemented to compare the cumulative number of stimuli mastered, mean training time to mastery, rate of acquisition, and the level of problem behavior. The results

showed that the 0:1 condition was the most efficient intervention procedure for all four participants. In addition, TI did not lead to a greater reduction in levels of problem behavior.

Task interspersal (TI) is a commonly implemented variation of discrete trial training (DTT) in clinical practice with children with autism spectrum disorder (ASD) and related disorders (Rapp & Gunby, **2016**). According to Chong and Carr (**2005**), TI is usually implemented to facilitate the acquisition of novel skills by providing an increased rate of reinforcement. Although there are currently no specific guidelines to direct the implementation of TI, it typically consists of the presentation of a specified ratio of previously mastered tasks (i.e., known tasks) prior to presenting acquisition tasks (i.e., unknown tasks; Dunlap, **1984**).

Evidence regarding the efficiency (e.g., rate of acquisition) of learning during TI has been inconsistent (Rapp & Gunby, **2016**). Some studies indicate that TI is superior to other DTT variations (e.g., Dunlap, **1984**; Neef, Iwata, & Page, **1980**). For example, Dunlap (**1984**) compared (1) a constant task condition in which one acquisition task was repeatedly presented, (2) a varied acquisition tasks condition in which five acquisition tasks were randomly presented, and (3) a varied acquisition with maintenance tasks condition in which five acquisition tasks and five maintenance tasks were randomly presented. The results showed that acquisition was similar in the constant task condition and varied acquisition tasks condition, while learning was most efficient in the TI condition for all five participants. Nevertheless, other studies indicate that TI is inferior to other DTT procedures, and may reduce the efficiency of instruction (e.g., Henrickson, Rapp, & Ashbeck, **2015**; Majdalany, Wilder, Greif, Mathisen, & Saini, **2014**; Volkert, Lerman, Trosclair, Addison, & Kodak, **2008**). Majdalany et al. (**2014**) compared massed-trial training (i.e., several acquisition targets, brief intertrial interval), distributed-trial training (i.e., several acquisition targets, 10-s intertrial interval), and TI (i.e., several acquisition targets, mastered tasks interspersed during the 10-s intertrial interval) for six children with ASD. The results showed that massed-trial training resulted in a quicker rate of acquisition for five out of the six participants by requiring 60% less instructional time than other conditions, and TI was inferior to other methods of instruction.

It is possible that the inconsistent findings regarding the efficiency of TI relate to differences in the ratio of TI across studies. Ratios of mastered-to-acquisition tasks evaluated in prior studies include 3:1 (e.g., Henrickson et al., **2015**; Majdalany et al., **2014**; Nicholson, **2014**), 1:1 (e.g., Dunlap, **1984**; Nicholson, **2014**; Volkert et al., **2008**), and 1:3 (e.g., Nicholson, **2014**). Higher ratios of TI are not consistently associated with increased instructional efficiency.

The method in which different ratios are incorporated into TI also may influence the efficiency of instruction. Mastered tasks can be interspersed during instruction using two methods. In the additive method, the number of trials per session increases with the addition of mastered tasks. For example, Volkert et al. (**2008**) arranged additive TI training for children with ASD. They conducted 10 training trials of acquisition stimuli during a no-interspersal condition and increased the number of trials to 20 (10 trials of acquisition stimuli and 10 trials of mastered stimuli) during the additive TI conditions. The second method of arranging TI involves use of a substitutive method that replaces acquisition trials with mastered task trials to maintain the same number of trials or length of training time per session (e.g., Dunlap & Koegel, **1980**; Forbes et al., **2013**; Koegel & Koegel, **1986**). For example, Forbes et al. (**2013**) compared the rate of learning for students with disabilities during two conditions that contained 15 trials. In the no-interspersal condition, 15 trials were composed of 15 acquisition words presented on flashcards. In the substitutive TI condition, 15 trials were composed of 3 acquisition and 12 mastered words. Although both methods of TI have been shown to improve learning outcomes in some cases, each method of TI has advantages and disadvantages. One disadvantage of the substitutive method is that participants have fewer exposures to each acquisition stimulus. In comparison, the additive method allows for an equal number of exposures to each acquisition stimulus because the interspersed mastered trials are added to the overall number of trials. Nevertheless, by adding more trials, a longer interval of time may be necessary to allocate to instruction, which could reduce the overall efficiency of learning (i.e., more time spent responding to mastered tasks takes instructional time away from instruction on acquisition stimuli). In addition, if children engage in problem behavior during academic instruction, increasing the overall number of demands in the additive TI method may be an establishing operation for escape from those tasks. Therefore, the selection of additive versus substitutive TI methods may depend on the goal(s) of instruction.

In a review of the literature on TI, Rapp and Gunby (**2016**) suggest additional lines of research to evaluate the benefits of TI on several dependent variables during instruction. These authors note that one variable that has not received much attention in the TI literature is the effects of TI on problem behavior. It is hypothesized that TI procedures may function as an abolishing operation for problem behavior that typically results in escape from demands (Rapp & Gunby, **2016**). Some individuals engage in problem behavior, such as aggression, during instruction to escape from or avoid aversive stimuli such as difficult academic tasks (Carr & Durand, **1985**; Horner, Day, Sprague, O'Brien, & Heathfield, **1991**). Difficult academic tasks (e.g., acquisition tasks) may require greater response effort to complete than engaging in problem behavior. Individuals who engage in frequent problem behavior to escape difficult academic tasks may rarely contact the contingencies of reinforcement in place for correct responding. Interspersing previously mastered tasks may allow an individual to contact positive reinforcement after correct responses, which may reduce the establishing operation for a break during subsequent acquisition tasks (Mevers, Fisher, Kelley, & Fredrick, **2014**; Rapp & Gunby, **2016**). Thus, TI may reduce levels of problem behavior during instruction.

Only one study has investigated the effects of TI on levels of problem behavior during skill acquisition for children with ASD. Henrickson et al. (**2015**) compared massed-trial training (MTT) to TI to teach children with ASD, and they recorded data on the percentage of trials with problem behavior per session. The experimenter implemented a 3:1 ratio of mastered-to-acquisition tasks in the TI condition and equated reinforcement in the MTT condition by providing social praise for behavior such as sitting and listening. The results showed that interspersing previously mastered tasks was inferior to MTT in rate of acquisition, and participants engaged in similar levels of problem behavior across the two conditions. Nevertheless, the authors did not conduct a functional analysis to determine the function of the participants' problem behavior. Therefore, it remains unclear whether TI may be effective for reducing escape-maintained problem behavior. Further, Rapp and Gunby (**2016**) describe the necessity of additional research that evaluates the effects of different ratios of TI on escape-maintained problem behavior.

The purpose of the current study was to extend the literature on TI and directly evaluate recommendations of Rapp and Gunby (**2016**). We compared the effects of four commonly used ratios of TI on the efficiency of acquisition for children with ASD or a related disorder. To extend prior studies and align our measures of efficiency with current trends in skill-acquisition research, we included four measures of efficiency. In addition, participants with and without problem behavior were included in the study to evaluate whether the interspersal procedures produced differential outcomes on the presence or absence of problem behavior. We conducted a functional analysis for the two participants who engaged in problem behavior to examine the effects of TI on escape-maintained problem behavior.

METHOD

Participants

We recruited four children diagnosed with ASD or a related disorder to participate in the study. Owen was a 5-year-old boy with a moderate cognitive impairment who was diagnosed with global developmental delay by an independent psychologist not affiliated with the study. He had a limited vocal-verbal repertoire, with a score of 12.5 on the early echoic skills assessment (EESA), a subtest of the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP; Sundberg, **2008**). He received behavior-analytic early intervention services for

4 months prior to inclusion in the study. He communicated using a picture exchange communication system and with a limited number of phonemes (e.g., "pa" for iPad). Owen engaged in several topographies of problem behavior during instruction, as demonstrated through a pretest and functional analysis (FA; described below).

Finn was a 3-year-old boy with a mild cognitive impairment who was diagnosed with ASD by a psychologist at a psychology clinic specializing in the assessment of neurodevelopmental disorders. He communicated using one-word responses at the start of the study. Finn began receiving behavior-analytic early intervention services 1 month prior to the onset of the study. Finn engaged in several topographies of problem behavior during instruction (described below).

Lucas was a 5-year-old boy with a moderate cognitive impairment and a diagnosis of ASD provided by an independent psychologist. Lucas communicated using short phrases or sentences. He received behavior-analytic early intervention services for 15 months prior to the onset of the study. Although Lucas had a history of engaging in a low rate of problem behavior during instruction, he did not meet the criterion for "problem behavior" during the pretest (described below). Thus, we included him as a participant who did not engage in problem behavior during instruction.

Benny was a 15-year-old boy with a moderate cognitive impairment and a diagnosis of ASD provided by an independent psychologist. He communicated using short phrases. He received behavior-analytic intervention services for 1.5 years prior to the onset of the study. Benny did not have a history of problem behavior during instruction, which was confirmed by a pretest (described below). Thus, we included Benny as a participant who did not engage in problem behavior during instruction.

Setting and Materials

Sessions took place at the family kitchen table for Owen, Finn, and Benny. Lucas's sessions took place at a child-sized table in his bedroom. Participants sat at a table next to or across from the experimenter during all sessions.

Session materials included data sheets, pens, timers to record session duration, preferred items typically delivered as reinforcers during clinical service for each participant, a video camera to record sessions, and instructional stimuli. We included a token board and tokens during Benny's sessions. We had previously established a token economy within his clinical services and used it during all of his skill acquisition programs, including the conditions in the study.

Response Measurement, Interobserver Agreement, and Procedural Fidelity

Observers collected data on independent correct responses, prompted correct responses, incorrect responses, and problem behavior. All measures were recorded for acquisition and mastered stimuli. An *independent correct response* was defined as the occurrence of a predefined response to the target stimulus within 5 s of its presentation. A *prompted correct response* was defined as the occurrence of a predefined response to the target stimulus within 5 s of a gestural, model, or physical prompt. An *incorrect response* was defined as an error or no response within 5 s of the initial presentation of a stimulus or within 5 s of a gestural, model, or physical prompt. Problem behavior included aggression (Owen and Finn), disruption (Owen and Finn), elopement (Owen), negative vocalizations (Owen and Finn), and vocal noncompliance (Owen). We defined aggression as contact between the participant's body and another person's body, including hitting, kicking, slapping, pinching, or raking fingers across skin. Disruption included swiping, hitting, ripping, or piling materials, placing materials in the mouth, spitting on materials, standing on the chair/table, pushing the table or chairs over, or placing a foot or both feet on or above the table surface. Elopement included moving more than 3 feet from the therapist from a seated position in a chair. We defined negative vocalizations as crying, whining, screaming, or growling. Vocal noncompliance included statements indicating vocal refusal to comply with a demand (e.g., "no"). Data collectors recorded problem behavior as an occurrence or nonoccurrence per trial, and calculated the percentage of trials with problem behavior by dividing the number of trials in which problem behavior occurred by the total number of trials, and multiplying by 100.

The primary dependent variables were the *efficiency of learning* and *level of problem behavior* across conditions. The efficiency of learning was calculated in several ways. First, efficiency was determined by examining the cumulative number of acquisition stimuli mastered per condition, which was calculated by adding the number of stimuli mastered across sessions of each condition. Second, the mean training time per stimulus was calculated as a measure of efficiency by dividing the total duration of all training sessions within a condition by the total number of acquisition stimuli mastered within the condition (e.g., 110 min/10 stimuli = mean of 11 min to mastery per acquisition stimulus). Last, the rate of acquisition per condition was calculated by dividing the number of acquisition stimuli mastered within a condition by the total duration of all training sessions within a condition, multiplied by 60 min to convert the data to hours (e.g., 21 stimuli mastered/105 min = 0.2 stimuli acquired per min X 60 min = 12 stimuli mastered per hour). Although this measure includes the same data as mean training time per stimulus, calculations of rate show trends in the speed of learning based on the amount of time in instruction. The condition with the lowest mean training time per stimulus and highest rate of acquisition was considered the most efficient format of instruction. Levels of problem behavior were evaluated by comparing the percentage of problem behavior across sessions in each condition.

Two independent observers recorded data on all dependent measures during 33.9% to 67.3% of sessions across all experimental conditions for all participants. We obtained trial-by-trial interobserver agreement (IOA) by comparing the data collected by these two observers during each trial in a session. An agreement was scored for each dependent measure if observers recorded the exact same dependent measure during the trial. We calculated IOA for each dependent measure in each condition by dividing the trials with an agreement by the total number of trials in the session, and multiplying by 100. The average agreement was calculated across all dependent measures (excluding problem behavior) and conditions for each participant. Mean IOA scores for all dependent measures (except problem behavior) were 99.1% (range, 58.3% to 100%) for Owen, 95.8% (range, 40% to 100%) for Finn, 96.8% (range, 60% to 100%) for Lucas, and 95.7% (range, 77.8% to 100%) for Benny.

A second observer recorded data on problem behavior during 43% to 50% of FA sessions and 46% to 67% of baseline and training sessions for Owen and Finn, and we calculated IOA for the combined topographies of problem behavior. Sessions were divided into 10-s intervals to calculate agreement for the FA. We calculated proportional agreement in an interval by dividing the lowest number of instances of problem behavior scored by an observer by the highest number of instances of problem behavior scored by the other observer, and multiplying by 100 (Mudford, Martin, Hui, & Taylor, **2009**). We averaged proportional agreement for each session by adding percentages of agreements for each interval and dividing by the number of intervals in a session. Mean IOA scores for the FA were 95.5% (range, 93.3% to 100%) for Owen and 87.8% (range, 80.5% to 100%) for Finn. We calculated trial-by-trial agreement for problem behavior during baseline and training conditions. Mean IOA for problem behavior during baseline and training sessions was 99.4% (range, 75% to 100%) for Owen and 91% (range, 58% to 100%) for Finn.

Observers also collected procedural fidelity data during 34% to 41% of sessions to determine the extent to which the experimenter implemented the procedures as intended. Fidelity was assessed based on whether the experimenter: (a) presented the correct discriminative stimulus as indicated on the data sheet, (b) presented prompts immediately after incorrect or no responses (if relevant), (c) delivered reinforcement, defined as the presentation of praise and a tangible item for independent correct responses to acquisition and mastered stimuli for the specified reinforcement interval, and the presentation of praise only for prompted correct responses, and (d) ignored problem behavior and continued the task, as described in the protocol (if relevant). Procedural fidelity was measured for each trial and was scored as either a “one” for correct implementation of all relevant components for the entire trial or a “zero” for incorrect implementation of any aspect of the trial.

We calculated procedural fidelity as a percentage for each session by dividing the number of trials implemented correctly by the total number of trials in the session, and multiplying by 100. Procedural fidelity averaged 98.3% (range, 75% to 100%) for Owen, 99.7% (range, 91% to 100%) for Finn, 93.9% (range, 44.4% to 100%) for Lucas, and 94.4% (range, 20% to 100%) for Benny. In the sessions with low integrity scores for Lucas and Benny, the video camera failed to capture necessary components of fidelity, which we scored as a nonoccurrence because the data collector could not observe the behavior.

Pre-experimental Procedures

Pretest

One skill (e.g., tacting) was targeted for each participant. We selected the targeted skill based upon individual treatment goals related to each participant's skill deficits as determined by assessments conducted prior to the start of the study. For Owen, the targeted skill was auditory–visual conditional discrimination (AVCD) presented in a three-comparison array. Finn's targeted skill was tacting common items. Benny's targeted skill was tacting item features (e.g., an elephant's trunk). Lucas's targeted skill was adjective–noun tacting (e.g., brown bear).

We conducted a pretest to identify participants who displayed problem behavior during instruction. Stimuli were grouped into sets of three during each pretest session. The first two pretest sessions consisted of 15 trials with three stimuli presented five times each. The experimenter presented each stimulus, allowed up to 5 s for a response, and removed the stimulus. No prompts or reinforcement were provided during the first two sessions of the pretest. The experimenter collected data on correct responses and problem behavior. Participants (Owen and Finn) who engaged in problem behavior on six or more trials during the first 30 pretest trials were considered to “display problem behavior during instruction” for the purposes of this study. Owen and Finn engaged in problem behavior during 25 out of 30 (83.3%) and 13 out of 30 (43.3%) pretest trials, respectively. Lucas and Benny engaged in problem behavior during 3 of 30 (10%) and 0 of 30 (0%) pretest trials, respectively. Therefore, they were not considered to display problem behavior during instruction.

After conducting the first two sessions of the pretest to identify individuals who displayed problem behavior during instruction, we modified the pretest procedures to identify stimuli for inclusion in the study. Sessions consisted of 12 trials (nine pretest trials and three interspersed mastered task trials) with each pretest stimulus presented three times. Experimenters interspersed previously mastered stimuli approximately every three trials. The experimenter provided prompts, as well as reinforcement for independent or prompted correct responses, for mastered stimuli only. The experimenter provided no prompts following incorrect or no responses to the pretest targets. Independent correct responses to the pretest tact targets resulted in the delivery of reinforcers, but the experimenter did not provide reinforcers for independent correct responses to the AVCD pretest targets.

Stimuli to which the participant engaged in a correct response during all three pretest trials were designated as mastered stimuli. We assigned specific mastered stimuli to each condition. Stimuli to which the participant engaged in a correct response during zero pretest trials for tacts and no more than one of three pretest trials for AVCD were designated as acquisition stimuli. All of Owen's mastered and acquisition targets were AVCDs; all targets for the other participants were tacts.

We assigned unique sets of three acquisition stimuli per experimental condition using a logistical analysis method (Gast, **2010**). That is, we assigned stimuli such that (a) there was a similar number of syllables in responses across conditions, (b) overlapping visual or auditory stimuli were separated across conditions, and (c) similar levels of correct responding were observed during the pretest. We also assigned a unique set of mastered stimuli to each condition. Although each condition included a set of three acquisition stimuli that were trained simultaneously, we identified additional acquisition tasks and assigned them to each condition so that any acquisition stimulus that met the mastery criterion during training was replaced by a new acquisition stimulus. See Supporting Information for stimuli assigned to each condition for each participant.

Functional analysis

We conducted FAs for Owen and Finn. For both participants, we conducted an abbreviated FA in a test–control pairwise design (Iwata & Dozier, **2008**) to test if problem behavior was maintained by escape from demands. All sessions occurred at the table and were 5 min.

During the escape condition, the experimenter presented instructions similar to those included in the pretest (but with different stimuli), using least-to-most prompting consisting of vocal, model, and physical prompts, and provided praise following independent and prompted correct responses. Following the occurrence of any topography of problem behavior, the experimenter removed the instructional materials and turned away from the participant for 30 s. This condition was included to determine if social negative reinforcement (i.e., escape from demands) functioned as a reinforcer for the participants' problem behavior. During the toy play condition, the experimenter sat next to the participant and provided continuous vocal attention and brief physical contact at least every 30 s. The experimenter did not present demands nor deliver any consequences following the occurrence of any topography of problem behavior. This condition served as a control to provide noncontingent access to all potential socially mediated reinforcers.

Preference assessment

We identified tangible items based on parent report. We conducted initial multiple stimulus without replacement (MSWO) preference assessments with Owen, Finn, and Lucas based on the procedures described by Carr, Nicholson, and Higbee (**2000**). The experimenter used the most preferred item as a reinforcer. Participants consistently engaged in mands for preferred items during sessions; therefore, we did not conduct daily MSWO assessments.

Benny completed several one-trial MSWO assessments (similar to DeLeon et al., **2001**) prior to the start of the study. Benny consistently selected the same item in the first trial across all assessments. That item, plus two other items included in the MSWO, was available during all sessions. Once Benny earned three tokens, he selected an item from an array of three items placed on the table.

Procedure

We implemented an adapted alternating-treatments design to examine the effects of TI ratios on the efficiency of skill acquisition and levels of problem behavior. We exposed each participant to four conditions consisting of different ratios of mastered-to-acquisition stimuli. The experimenter conducted one or two sessions of each condition per day, with an equal number of sessions conducted across conditions each day. The four experimental conditions occurred in a random order for each participant to control for order effects; the order of each condition within a session block of four sessions alternated before reordering for the next session block.

Each session consisted of 12 trials, with three acquisition stimuli presented in each session. The number of presentations of each acquisition stimulus in a session and the sequence of mastered-to-acquisition stimuli depended on the condition ratio of mastered-to-acquisition stimuli.

Training in each condition with each acquisition stimulus began with a 0-s prompt delay until the participant engaged in two consecutive correct prompted responses to the acquisition stimulus. That is, the experimenter presented the relevant stimulus material(s) and immediately provided a prompt (e.g., vocal model prompt, physical prompt). Correct prompted responses produced praise and a token (Benny) or tangible item (Owen, Finn, and Lucas) for 20 s. Following two consecutive correct prompted responses to each acquisition stimulus, the experimenter implemented a 5-s prompt delay for all stimuli (acquisition and mastered). Thus, the experimenter presented the stimulus material(s) and allowed 5 s for a response. If the participant engaged in an independent correct response, the experimenter provided praise and a token or tangible item for 20 s. If the participant engaged in an error or did not respond within 5 s, the experimenter provided a prompt, and delivered praise only following a correct prompted response. If the participant did not engage in a correct prompted response within 5 s of the prompt, the experimenter implemented the next trial.

An acquisition stimulus met the mastery criterion if the participant engaged in an independent correct response for four consecutive presentations of the stimulus. Once a participant mastered an acquisition stimulus, the experimenter removed the stimulus from treatment and replaced it with another acquisition stimulus assigned to the condition. We did not add mastered acquisition stimuli to the pool of mastered stimuli presented during training in order to assess maintenance of recently mastered stimuli in the absence of continued practice. Participants completed training when they mastered 21 acquisition stimuli in at least one condition, or when they completed a total of 30 sessions per condition.

Three to one

Nine of the trials consisted of mastered stimuli, and three of the trials consisted of acquisition stimuli. The first three trials of each session included the presentation of three randomly ordered mastered stimuli, followed by the presentation of one acquisition stimulus. Thereafter, the experimenter presented three more randomly ordered mastered stimuli followed by one acquisition stimulus. This sequence continued across the 12-trial session. The experimenter presented each mastered stimulus three times per session and each acquisition stimulus one time per session.

One to one

Six of the trials consisted of mastered stimuli, and six of the trials consisted of acquisition stimuli. The first trial of each session included the presentation of one mastered stimulus, followed by the presentation of one acquisition stimulus. This sequence continued across the 12-trial session. The experimenter presented each mastered stimulus two times per session and each acquisition stimulus two times per session.

One to three

Three of the trials consisted of mastered stimuli, and nine of the trials consisted of acquisition stimuli. The first trial of each session included the presentation of one mastered stimulus, followed by the presentation of three randomly ordered acquisition stimuli. This sequence continued across the 12-trial session. The experimenter presented each mastered stimulus one time per session and each acquisition stimulus three times per session.

Zero to one

All trials consisted of acquisition stimuli; the experimenter did not present interspersed mastered stimuli in this condition. Sessions included four presentations of each acquisition stimulus presented in random order.

Maintenance

Following mastery of one stimulus, the experimenter conducted maintenance probes for that stimulus after 1 and 2 weeks elapsed. Maintenance probes did not include prompts or reinforcement, and the experimenter did not intersperse mastered stimuli between trials. Maintenance probes occurred in a massed-trial format, with the presentation of one mastered acquisition stimulus in five consecutive trials.

RESULTS

Figure **1** shows the results of Owen's and Finn's FAs. Elevated rates of problem behavior occurred during the escape condition, and near-zero rates of problem behavior were observed during the toy play condition. Thus, we concluded that both Owen and Finn engaged in problem behavior maintained by escape from demands.

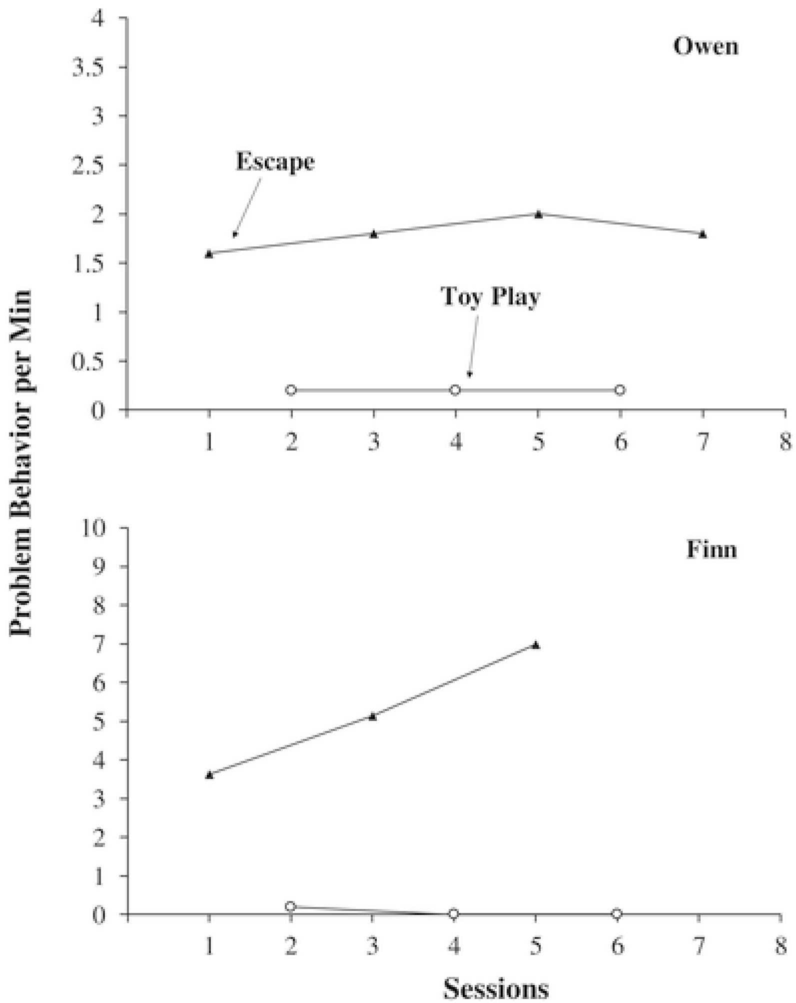

Figure 1

Problem behavior per minute in escape (closed triangles) and toy play (closed circles) conditions of the functional analysis for Owen (top) and Finn (bottom).

Figure 2 depicts the cumulative number of stimuli mastered in each experimental condition for Owen, Finn, Lucas, and Benny. The 0:1 mastered-to-acquisition stimuli condition was the most efficient condition, resulting in the largest number of stimuli mastered for all four participants. The 3:1 condition resulted in the fewest number of stimuli mastered for Owen and Finn, produced mastery of zero stimuli for Lucas, and resulted in the same number of stimuli mastered as the 1:1 condition for Benny.

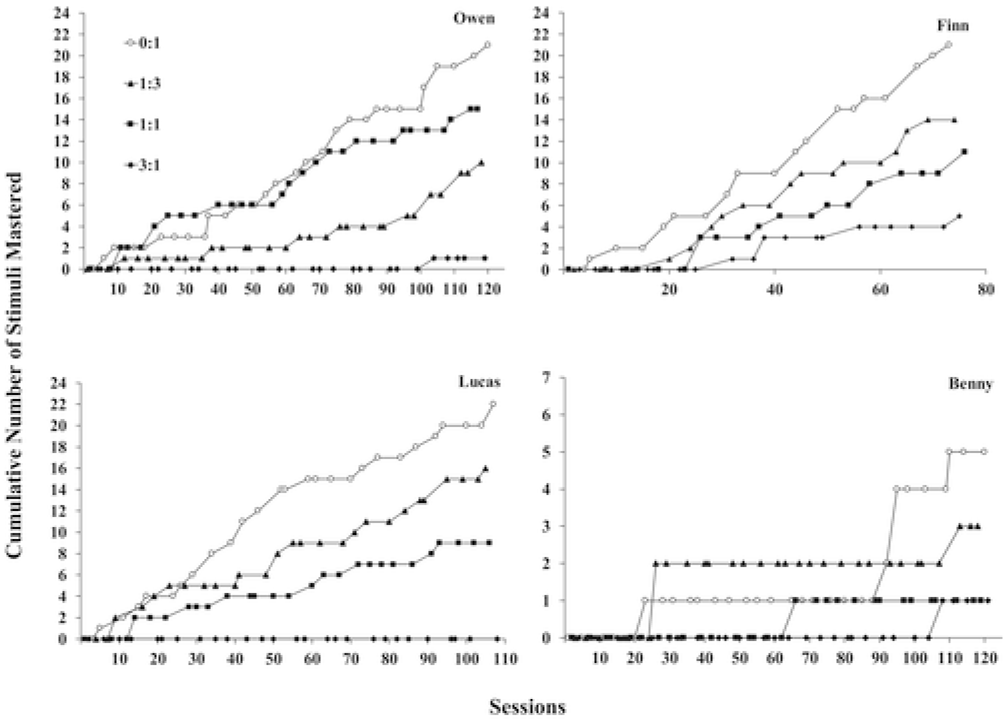

Figure 2

The cumulative number of stimuli mastered across conditions for Owen, Finn, Lucas, and Benny.

Figure 3 shows the mean minutes to mastery per stimulus for Owen, Finn, Lucas, and Benny. For three of the four participants (Finn, Lucas, and Benny), the mean minutes to mastery per stimulus increased across conditions as the number of interspersed mastered stimuli increased. Said another way, allocating increasing amounts of instructional time to already mastered stimuli decreased the amount of time spent training acquisition stimuli, which resulted in longer durations of instruction and lower levels of mastery of acquisition stimuli. Owen's data are the exception; his mean minutes to mastery per stimulus for the 1:1 condition was lower than the mean in the 1:3 condition. Nevertheless, the 0:1 condition had the lowest mean minutes to mastery per stimulus for Owen and the other three participants.

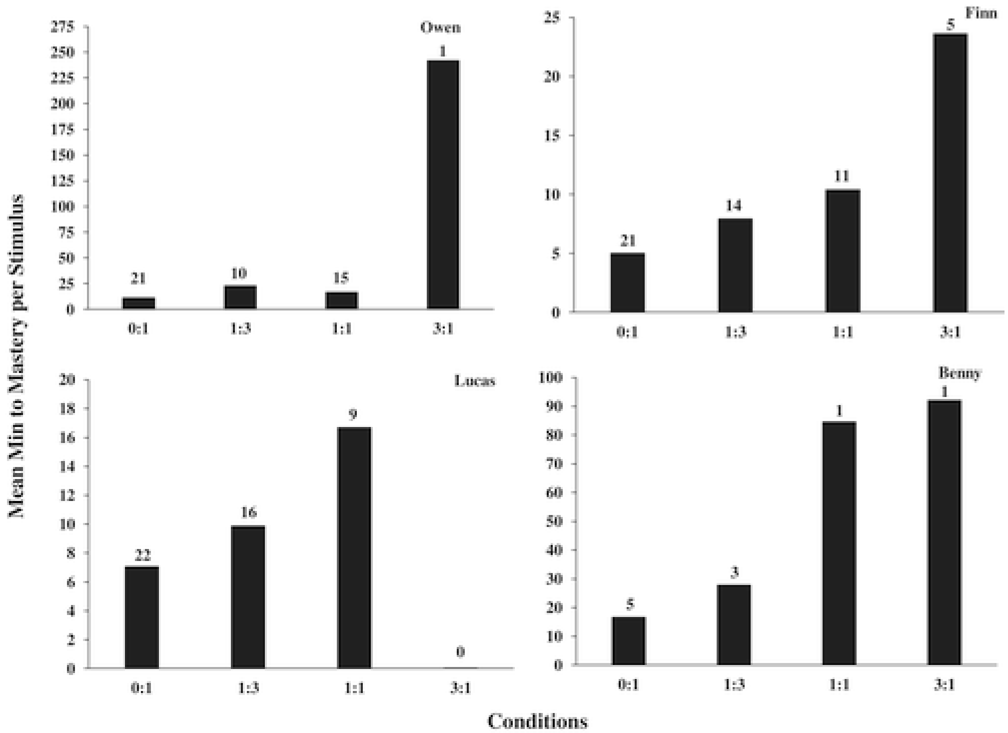

Figure 3

The mean minutes to mastery per stimulus for Owen, Finn, Lucas, and Benny. The numbers above each bar in represent the number of stimuli mastered in the condition.

We also calculated the rate of acquisition per condition to provide a measure of efficiency that shows how rapidly participants acquired stimuli per hour in each condition. The rate of acquisition per condition is shown in Figure 4. Overall, we observed an inverse relation between ratio and acquisition; as the ratio of interspersed mastered tasks decreased, the number of stimuli mastered per hour increased. Owen's data are the exception, because he had a lower rate of acquisition in the 1:3 condition in comparison to the 1:1 condition. However, the condition with no task interspersal produced the highest rate of acquisition for all participants.

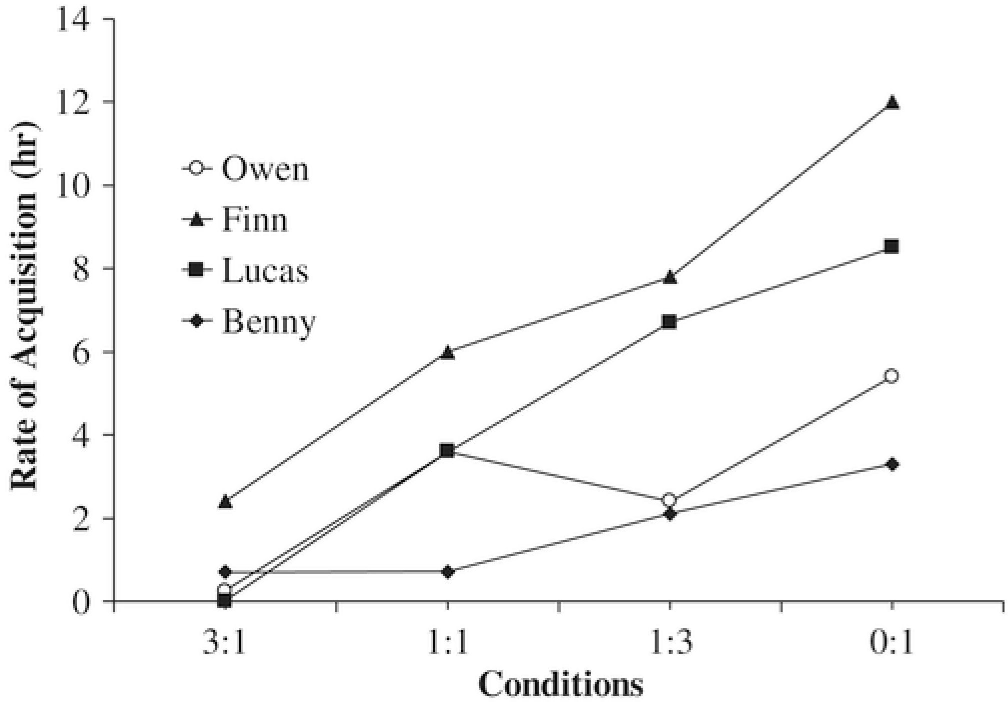

Figure 4

Acquisition of stimuli per hour across conditions for all participants.

Figure 5 depicts the percentage of trials with problem behavior during experimental conditions for Owen and Finn. Owen engaged in low or zero levels of problem behavior in most sessions across all conditions, including the condition with no interspersed mastered stimuli (top panel). Finn's levels of problem behavior were highly variable, and problem behavior occurred in all conditions (bottom panel). Therefore, interspersing mastered tasks did not appear to produce reductions in problem behavior for either participant, in comparison to the condition with no task interspersal (i.e., 0:1 condition).

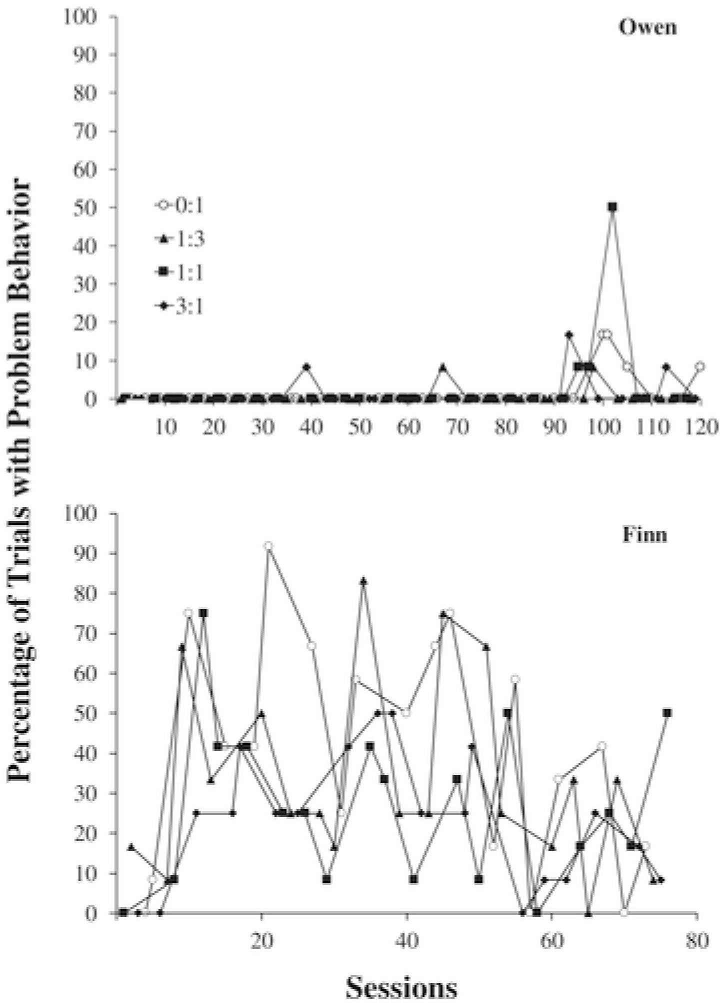

Figure 5

The percentage of trials per session with problem behavior across conditions for Owen (top) and Finn (bottom).

Maintenance probes for mastered tasks showed inconsistent outcomes across conditions and participants with no clear advantage of any interspersal ratio on response maintenance. Owen's maintenance probes showed an average of 70%, 100%, 75%, and 100% correct responses to stimuli in the 0:1, 1:3, 1:1, and 3:1 conditions, respectively. Finn had 100% correct responses to all stimuli across all four conditions. Lucas rarely responded during maintenance probes, and had an average of 30%, 19%, and 22% correct responses for the 0:1, 1:3, and 1:1 conditions, respectively (he did not master any stimuli in the 3:1 condition). Benny had an average of 80%, 33%, 100%, and 100% correct responses to stimuli in the 0:1, 1:3, 1:1, and 3:1 conditions, respectively. Thus, Owen and Finn had generally high levels of correct responses in maintenance probes across conditions. Benny had high levels of correct responses in all conditions with the exception of the 1:3 condition. In contrast, Lucas had overall low levels of correct responses due to nonresponding during most probes.

DISCUSSION

The current study extends the literature on TI by comparing TI ratios on the efficiency of learning and level of problem behavior of children with ASD and related disorders. Our results were consistent with those of Nicholson (**2014**), suggesting that none of the TI ratios lead to more efficient learning regardless of whether substitutive or additive TI is arranged during instruction. The 0:1 condition (i.e., no TI) was the most efficient condition for all participants according to all three measures of efficiency. In addition, TI did not reduce problem behavior in comparison to a condition with no TI. These findings are consistent with the outcomes of other studies in which TI procedures did not lead to superior outcomes in comparison to other instructional procedures (e.g., Henrickson et al., **2015**; Majdalany et al., **2014**; Volkert et al., **2008**). The current study suggests that substitutive TI does not result in improved efficiency of instruction or reductions in escape-maintained problem behavior.

When examining the efficiency of instructional approaches, it is beneficial to include multiple measures of efficiency. One measure may provide a more accurate representation of efficiency, while the others may allow for the comparison of results across studies. Previous studies have included trials to mastery (Dunlap, **1984**; Nicholson, 2013), sessions to mastery (Henrickson et al., **2015**; Majdalany et al., **2014**), time to mastery (Dunlap, **1984**; Henrickson et al., **2015**), and rate of mastery (Nicholson, **2014**) as measures of efficiency. In the current study, we included three measures of efficiency: the cumulative number of stimuli mastered, the time to mastery (i.e., mean minutes to mastery per stimulus), and the rate of acquisition. Each measure provides valuable information, but they should be interpreted together.

The cumulative number of stimuli mastered per condition may not be as sensitive a measure as the minutes to mastery per stimulus or rate of acquisition (Kodak et al., **2016**; Yaw et al., **2014**). For example, due to our use of the substitutive TI method, participants had fewer exposures to acquisition stimuli in conditions that included TI. Thus, it is reasonable to assume that given the same number of instructional sessions, fewer stimuli should be mastered in conditions with TI. However, the amount of training time remains an important variable to consider, despite differences in exposures to acquisition trials. Regardless of the TI method (i.e., additive or substitutive), more instructional time will be allocated to mastered rather than acquisition stimuli as the ratio of TI increases. Therefore, calculations of the mean duration of instructional time required to master stimuli and the rate of acquisition can provide unbiased measures of efficiency. Our results showed that interspersing mastered stimuli into instruction lead to longer mean minutes to mastery per stimulus and reduced the rate of acquisition. This outcome is also consistent with prior studies conducted with children with ASD during DTT that found reduced efficiency in instruction with the inclusion of additive TI (e.g., Henrickson et al., **2015**; Nicholson, **2014**) or no benefits to learning (e.g., Volkert et al. **2008**).

The study also extended the literature on TI by using functional analysis to identify escape from demands as the maintaining variable for problem behavior, and by measuring levels of problem behavior across experimental conditions. Rapp and Gunby (**2016**) noted the paucity of research on the effects of TI on problem behavior. In the limited studies that have investigated TI with children with ASD who engage in problem behavior, either results have indicated low rates of problem behavior across all conditions (similar to Owen's data in the current investigation; Henrickson et al., **2015**), or the authors did not report the effects of TI on problem behavior (Volkert et al., **2008**). Further, previous studies did not include FAs to identify the function of participants' problem behavior. The current investigation included FAs that showed Owen and Finn engaged in problem behavior maintained by escape from demands. During the TI evaluation, Finn displayed elevated levels of problem behavior across all conditions, regardless of the inclusion of TI. In comparison, Owen engaged in low levels of problem behavior across all conditions, which is likely due to the introduction of differential reinforcement of correct responding plus escape extinction during sessions across all conditions. TI did not produce further reductions in problem behavior.

Previous research on TI has found inconsistent outcomes regarding maintenance of acquisition targets. For example, Henrickson et al.'s (**2015**) results showed that participants had zero or relatively low (0% to 60%) correct responses across 2-, 4-, and 6-week maintenance probes in both the TI and massed-trial instruction conditions. However, Henrickson et al. provided reinforcement for correct responses during maintenance probes. In the current study, maintenance probes did not include reinforcement or prompts. Nevertheless, three participants showed relatively high levels of correct responses during unreinforced maintenance probes across conditions. The other participant (Lucas) had low levels of correct responses across conditions. It is possible that differences in maintenance probe procedures (e.g., the inclusion of reinforcement and/or prompts) and the length of time between mastery and maintenance probes (e.g., 1 week versus 6 weeks) may account for discrepancies in maintenance outcomes between Henrickson et al.'s study and the current study. The long-term maintenance of skills that are embedded within TI procedures was not examined within this study and may be a worthy topic of additional research.

There were several limitations of the current study. First, Owen, Lucas, and Benny required many instructional trials to acquire certain stimuli or did not acquire some targets across conditions. For example, Owen acquired "saw" in 119 trials, but he never acquired "mop" in the 1:3 condition. These targets were introduced near the beginning of training. Because only three stimuli were targeted at the same time, delayed acquisition of these two stimuli affected the number of stimuli that Owen acquired in that condition, which can be seen in his pattern of acquisition in Figure **2**.

The lack of efficiency of the interspersal conditions may be attributed to the unequal number of exposures to acquisition stimuli across conditions. This in an inherent flaw in the procedures of substitutive TI, with acquisition target trials decreasing as more mastered target trials are added to sessions. Nevertheless, it was necessary to use the substitutive method of TI in the current study due to the dual focus on efficiency of learning and level of problem behavior. Use of the additive method would have required different numbers of demands across conditions (i.e., 12 trials in the 0:1 condition, 15 trials in the 1:3 condition, 18 trials in the 1:1 condition, and 20 trials in the 3:1 condition), which could have altered the value of escape from these tasks and increased the likelihood of problem behavior during TI conditions. Thus, we chose the substitutive method to maintain an identical number of demands across conditions and ensure a fair comparison of each TI ratio on levels of problem behavior.

To address the limitation of the substitutive method of TI on differences in exposure to acquisition stimuli, we calculated the rate of acquisition across conditions. Rate measures permit a comparison of learning across conditions based on time spent in instruction rather than the specific number of exposures to acquisition stimuli. Had we conducted the evaluation in the 3:1 condition for three times the number of sessions that we conducted in the 0:1 condition (which would have permitted exposure to an identical number of acquisition trials in both conditions), the duration of training in the TI conditions would have been considerably longer. Thus, any benefit of additional stimuli mastered would be offset by the added duration of training to provide an identical number of exposures to stimuli. As such, use of the additive method of TI is unlikely to improve instructional efficiency, although future studies could directly compare similarities in outcomes across additive and substitutive methods of TI.

The mastery criterion in the current study may also be a limitation. Rapp and Gunby (**2016**) noted that previous studies on TI used different mastery criteria, which may hinder comparisons of outcomes across studies. We selected a mastery criterion of four consecutive correct responses per stimulus, which required at least two consecutive sessions with correct responses, based on prior research that has used similar or less stringent mastery criteria for skill acquisition (Forbes et al., **2013**). Our selection of a more stringent mastery criterion could have resulted in limited or no acquisition in certain TI conditions within the comparison. Previous studies on TI that used more stringent mastery criteria arranged training so that *sets of stimuli* were exposed to the mastery criterion rather than establishing a mastery criterion *per stimulus* (e.g., Henrickson et al., **2015**; Majdalany et al., **2014**). Nevertheless, three of the four participants in the current study showed maintenance of the stimuli mastered across conditions, suggesting that the mastery criterion was sufficient to produce sustained levels of correct responding in the absence of reinforcement and prompts during maintenance probes. These outcomes are consistent with those of Fuller and Fienup (**2017**), who showed that more stringent mastery criteria led to higher levels of maintenance.

References

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis,* **18**, 111- 126. https://doi.org/10.1901/jaba.1985.18-111.

Carr, J. E., Nicholson, 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. doi:https://doi.org/10.1901/jaba.2000.33-353

Chong, I. M., & Carr, J. E. (2005). An investigation of the potentially adverse effects of task interspersal. *Behavioral Interventions,* **20**, 285- 300. doi:https://doi.org/10.1002/bin.202

DeLeon, I. G., Fisher, W. W., Rodriguez-Catter, V., Maglieri, K., Herman, K., & Markhefka, J. M. (2001). Examination of relative reinforcement effects of stimuli identified through pretreatment and daily brief preference assessments. *Journal of Applied Behavior Analysis,* **34**, 463- 473. doi:https://doi.org/10.1901/jaba.2001.34-463

Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology,* **37**, 41- 64. doi:https://doi.org/10.1901/jaba.1980.13-619

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis,* **13**, 619- 627. doi:https://doi.org/10.1901/jaba.1980.13-619

Forbes, B. E., Skinner, C. H., Black, M. P., Yaw, J., Booher, J., & Delisle, J. (2013). Learning rates and known-to-unknown flash-card ratios: Comparing effectiveness while holding instructional time constant. *Journal of Applied Behavior Analysis,* **46**, 832- 837. doi:https://doi.org/10.1002/jaba.74

Fuller, J. M., & Fienup, D. M. (2017). A preliminary analysis of mastery criterion level: Effects on response maintenance. *Behavior Analysis in Practice,* **10**, 1- 8. https://doi.org/10.1007/s40617-017-0201-0

Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York: Routledge.

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. doi:https://doi.org/10.1002/bin.1396

Horner, R. H., Day, H. M., Sprague, J. R., O'Brien, M., & Heathfield, L. T. (1991). Interspersed requests: A nonaversive procedure for reducing aggression and self-injury during instruction. *Journal of Applied Behavior Analysis,* **24**, 265- 278. doi:https://doi.org/10.1901/jaba.1991.24-265

Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, **1**, 3- 9.

Kodak, T., Campbell, V., Bergmann, S., LeBlanc, B., Kurtz-Nelson, E., Cariveau, T., . . . Mahon, J. (2016). Examination of efficacious, efficient, and socially valid error-correction procedures to teach sight words and prepositions to children with autism spectrum disorder. *Journal of Applied Behavior Analysis,* **49**, 532- 547. doi:https://doi.org/10.1002/jaba.310

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

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

Mevers, J., Fisher, W., Kelley, M., & Fredrick, L. (2014). The effects of variable-time versus contingent reinforcement delivery on problem behavior maintained by escape. *Journal of Applied Behavior Analysis,* **47**, 277- 292. doi: https://doi.org/10.1002/jaba.110

Mudford, O. C., Martin, N. T., Hui, J. K. Y., & Taylor, S. (2009). Assessing observer accuracy in continuous recording of rate and duration: Three algorithms compared. *Journal of Applied Behavior Analysis,* **42**, 527- 539. doi: https://doi.org/10.1901/jaba.2009.42-527

Nicholson, C. A. (2014). An analysis of variables affecting task interspersal among children with autism. Dissertation Abstracts International: Section B: The Sciences and Engineering. ProQuest Information & Learning. Retrieved from http://0-search.ebscohost.com.libus.csd.mu.edu/login.aspx?direct=true&db=psyh&AN=2014-99180-356&site=eds-live

Neef, N. A., Iwata, B. A., & Page, T. J. (1980). The effects of interspersal training versus high-density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis,* **13**, 153- 158. doi: https://doi.org/10.1901/jaba.2003.36-21

Rapp, J. T., & Gunby, K. (2016). Task interspersal for individuals with autism and other neurodevelopmental disorders. *Journal of Applied Behavior Analysis,* **49**, 730- 734. doi: https://doi.org/10.1002/jaba.319

Sundberg, M. L. (2008). *Verbal behavior milestones assessment and placement program: The VB-MAPP*. Concord, CA: AVB Press.

Volkert, V. M., Lerman, D. C., Trosclair, N., Addison, L., & Kodak, T. (2008). An exploratory analysis of task-interspersal procedures while teaching object labels to children with autism. *Journal of Applied Behavior Analysis,* **41**, 335- 350. doi: https://doi.org/10.1901/jaba.2008.41-335

Yaw, J., Skinner, C. H., Skinner, A. L., Maurer, K., Cihak, D., Wilhoit, B., . . . Booher, J. (2014). Measurement scale influences in the evaluation of sight-word reading interventions. *Journal of Applied Behavior Analysis*, **47**, 360- 379. doi:https://doi.org/10.1002/jaba.126