May 2017

Comparing the Efficacy and Efficiency of Varying Task Interspersal Ratios

Sophie Claire Knutson
University of Wisconsin-Milwaukee

COMPARING THE EFFICACY AND EFFICENCY OF VARYING TASK INTERSPERSAL RATIOS

by

Sophie C. Knutson

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Psychology

at

The University of Wisconsin-Milwaukee

May 2017

ABSTRACT

COMPARING THE EFFICACY AND EFFICIENCY OF VARYING TASK INTERSPERSAL RATIOS

by

Sophie C. Knutson

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Tiffany Kodak

Task interspersal (TI) is a procedural variation of discrete trial training that has been implemented with children with autism spectrum disorder (ASD) to facilitate the acquisition of novel skills, and may reduce problem behavior during instruction. The literature shows equivocal results regarding the efficiency of TI, but there is limited literature indicating the effects on level of problem behavior. The current study extended the literature on TI by comparing the efficacy and efficiency of varying TI ratios implemented in early intervention practices with children with ASD and related disorders on acquisition and levels problem behavior. The four ratios of mastered to acquisition stimuli included 75% mastered to 25% acquisition, 50% mastered to 50% acquisition, 25% mastered to 75% acquisition, and 0% mastered to 100% acquisition. An adapted alternating treatments design was implemented to compare the number of stimuli mastered, and the level of problem behavior. A condition was considered efficacious if at least one stimulus was mastered and problem behavior was reduced by 50% of the pre-test level. The condition that resulted in the most stimuli mastered in the fewest trial presentations was considered the most efficient intervention procedure. The results showed that the 0%M/100%A condition was the most efficient intervention procedure for all four participants. Results were inconsistent on the efficacy of the procedures regarding levels of problem behavior.

TABLE OF CONTENTS

LIST OF FIGURES

In the field of behavior analysis, discrete trial training (DTT) is an effective approach to teach a variety of basic and advanced skills to children with autism spectrum disorder (ASD) and related disorders (Smith, 2001). Trials in DTT have a discrete beginning and ending that consist of discriminative stimuli, prompts, responses, consequences, and intertrial intervals (Smith, 2001). Training trials are brief, with many trials conducted in succession. Although there are basic guidelines for the implementation of DTT, there are variations and extensions to the procedures so that training can be individualized in order to efficiently and effectively facilitate the acquisition of novel skills.

Task interspersal (TI) is a commonly implemented variation of the DTT procedure in clinical practice with children with ASD and related disorders (Rapp & Gunby, 2016). According to Chong and Carr (2005), TI is typically 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 (e.g., known tasks) prior to presenting an acquisition task (e.g., unknown task; Dunlap, 1984). Ratios of mastered to acquisition tasks may include 3:1 (e.g., Henrickson, Rapp, & Ashbeck, 2015; Majdalany, Wilder, Greif, Mathisen, & Saini, 2014; Nicholson, 2013), 1:1 (e.g., Dunlap, 1984; Nicholson, 2013; Volkert, Lerman, Trosclair, Addison, & Kodak, 2008), and 1:3 (e.g., Nicholson, 2013). In addition, previously mastered tasks are interspersed using either (a) the substitutive method, which replaces acquisition trials with mastered task trials to maintain the number of trials per session, or (b) the additive method, which increases the number of trials per session with the addition of mastered task trials (Cates, 2005). In both methods, the interspersed items are previously mastered items of a response class functionally related to (Chong & Carr, 2005; Dunlap, 1984; Neef, Iwata, & Page, 1980; Volkert et al., 2008) or unrelated to (Charlop, Kurtz, & Milstein, 1992; Majdalany et al., 2014; Volkert et al., 2008) the acquisition targets. That is, interspersed items are previously mastered items that are similar or dissimilar to the acquisition targets. For example, when targeting tacts (i.e., a vocal response evoked by a nonverbal stimulus; Skinner, 1957) as the acquisition skill, interspersed items could be previously mastered tacts (similar) or previously mastered motor tasks (dissimilar).

The extant literature suggests improvement in correct responding when implementing interspersal techniques across populations (Benavides & Poulson, 2009; Dunlap & Koegel, 1980; Koegel & Koegel, 1986). For example, Koegel and Koegel (1986) investigated the effects of TI on the performance (e.g., correct responding) and motivation (e.g., subjective ratings of affect) of a childhood stroke victim. They compared an acquisition-only condition to an interspersal condition. In the acquisition-only condition, no previously mastered items were presented. In the interspersal condition, up to two acquisition items were presented with interspersed previously mastered items. The results showed that the participant had improved performance indicated by higher levels of correct responding in the interspersal condition. The participant also had an increase in motivation in the interspersal condition, indicated by higher ratings of interest, enthusiasm, and happiness.

Research also has examined the effects of TI on correct responding in children with ASD (e.g., Benavides & Poulson, 2009; Dunlap & Koegel, 1980). For example, Benavides and Poulson (2009) found higher levels of correct responding during an interspersal procedure for three children with ASD. Participants were first exposed to an acquisition-only baseline condition in which only unmastered match-to-sample tasks were presented during trials. Thereafter, the authors introduced two interspersal conditions in which mastered tasks were interspersed with acquisition tasks in a 1:1 ratio. In the first interspersal condition, reinforcement was available for correct responding in each trial. After correct responding increased, they introduced the second interspersal condition, reducing the density of reinforcement so that reinforcement was available for correct responding in half of the total trials (e.g., 12 out of 24 trials). Correct responding increased and remained stable with the staggered introduction of the interspersal conditions across stimulus sets.

It is hypothesized that the improvement in performance associated with TI may be the result of an increased density of reinforcement. Interspersing previously mastered items increases the number of opportunities to obtain reinforcement, thus increasing the probability of a higher density of reinforcement (Charlop et al., 1992). Neef et al. (1980) addressed this hypothesis by comparing the effects of TI to a high-density reinforcement condition in which social praise was delivered for task-related behavior to match reinforcement across conditions. They found that all participants acquired and maintained more words in TI than in the high-density reinforcement condition, although both conditions showed improvement over baseline. Thus, increased opportunities for reinforcement for responses to instructional tasks (e.g., mastered tasks) may improve performance on all instructional tasks (e.g., acquisition tasks) (Dunlap, 1984).

A second hypothesis is that interspersal 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 (Horner et. al, 1991). Difficult academic tasks (e.g., acquisition tasks) may require greater response effort to complete than engaging in problem behavior which may result in the removal of the demand. Individuals who engage in frequent problem behavior to avoid difficult academic tasks during instruction may rarely contact the contingencies of reinforcement in place for correct responding. Interspersing previously mastered tasks may allow an individual to contact reinforcement after correct responses. Contacting the reinforcement contingencies in place with potentially easier tasks may reduce the establishing operation for a break from the more difficult acquisition task (Rapp & Gunby, 2016). Thus, TI may reduce level of problem behavior during instruction.

Horner, Day, Sprague, O'Brien, and Heathfield (1991) investigated the effects of TI on aggression and self-injurious behavior in adolescents with intellectual disabilities. They compared levels of responsiveness, aggression, and self-injury across three conditions: easy tasks only, hard tasks only, and hard tasks interspersed with three to five easy previously mastered tasks. Results indicate low levels of aggression and self-injury in both the easy tasks only condition and the hard tasks with interspersed simple tasks condition. Levels of aggression and self-injury remained high in the hard tasks only condition. The establishing operation for the termination of demands was effectively reduced in the easy tasks only condition and hard tasks with interspersed simple tasks condition. The results suggest interspersal procedures may effectively reduce aggression and self-injury in children with intellectual disabilities, but do not indicate the efficacy in reducing problem behavior in children with ASD.

Only one study investigated the efficacy of TI in regards to the levels of problem behavior that occur in children with ASD during skill acquisition. Henrickson et al. (2015) examined the use of massed-trial teaching (MTT) versus 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 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. More research is needed in this area to come to more definitive conclusions of the efficacy of TI in reducing problem behavior in children with ASD.

The evidence on the efficiency of TI in facilitating acquisition has been inconsistent (Rapp & Gunby, 2016). Several studies indicate that TI is superior to other DTT variations (Dunlap, 1984; Neef et al., 1980). Dunlap (1984) investigated the efficiency of interspersal procedures with five children with ASD. They implemented three experimental conditions, including: (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 varied acquisition with maintenance tasks condition for all five participants. The results showed interspersing previously mastered or maintenance items was superior to massed-trial teaching.

Nevertheless, other studies indicate that TI is inferior to other DTT procedures, and may reduce the efficiency of instruction (Majdalany et al., 2014; Henrickson et al., 2015; Volkert et al., 2008). In a comparison of MTT, distributed-trial instruction (DTI), and TI, Majdalany et al. (2014) examined the efficacy and efficiency of the three procedures on the acquisition of tacts of countries in six children with ASD. In MTT, three countries were randomly presented with no interspersal of previously mastered items. In DTI, three countries were randomly presented with intertrial intervals (ITIs) of 10 s during which the children did not have access to social interaction or tangible items. In TI, three countries were randomly presented and three previously mastered tasks were presented during each 10-s ITI. The results showed that MTT resulted in a quicker rate of acquisition for five out of the six participants, while DTI was most efficient for one participant. The results showed that TI is inferior to other methods of instruction.

The literature on TI shows equivocal findings regarding efficiency, but TI may be effective in reducing problem behavior (Rapp & Gunby, 2016). From the limited number of studies that have investigated the effect of TI on problem behavior, it is unclear if TI is effective in reducing problem behavior during instruction, and if so, what ratio of TI will be most beneficial for reducing problem behavior. In addition, the specific ratio of acquisition to mastered tasks has varied across studies, and it remains unclear whether a specific ratio may be associated with higher levels of efficacy and efficiency. The purpose of the current study was to extend the literature on TI by comparing the efficacy and efficiency of varying TI ratios implemented in early intervention practices for children with ASD and related disorders on the rate of acquisition and level problem behavior.

METHOD

Participants

Four children diagnosed with ASD or who displayed ASD-like symptoms were recruited to participate in the study. Children with ASD or ASD-like symptoms were recruited because the research question evaluated in the present investigation related to the efficacy and efficiency of common early intervention practices implemented with individuals with ASD and related disorders. Participants with and without problem behavior were included in the study to evaluate whether the interspersal procedures produced differential outcomes based on the presence or absence of problem behavior.

Owen was a 5-year-old boy with a moderate cognitive impairment who displayed ASD-like symptoms and had a diagnosis of global developmental delay. He had a limited vocal-verbal repertoire, with a score of 12.5 on the early echoic skills assessment (EESA), a subtest of the VB-MAPP, conducted at the onset of his early intervention services. He received early intervention services for 4 months prior to inclusion in the study. He communicated using a picture exchange communication system (PECS) and with a limited number of phonemes (e.g., “pa” for iPad). Owen was included as a participant who engaged in problem behavior during instruction, identified through a pre-test and functional analysis (FA) (described below). Owen engaged in various topographies of problem behavior including aggression (e.g., hitting, kicking) disruption (e.g., swiping, mouthing, and spitting on materials), elopement, negative vocalizations (e.g., crying), and vocal noncompliance.

Finn was a 3-year-old boy with a mild cognitive impairment who was diagnosed with ASD by 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 early intervention services one month prior to the onset of the study. Finn was included as a participant who engaged in problem behavior during instruction, determined by a pre-test and FA (described below). Finn engaged in several topographies of problem behavior including aggression (e.g., kicking, raking fingers across skin), disruption (e.g., swiping and hitting materials, placing his foot or feet on or above the table surface), and negative vocalizations (e.g., growling).

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 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 (e.g., disruption, flopping, negative vocalizations), he did not meet the criterion for problem behavior during the pre-test (described below). Thus, he was included 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 pre-test (described below). Thus, Benny was included 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. A token board and tokens were included during Benny's sessions. This token economy was established within his clinical services and used during all of his skill acquisition programs, including the conditions in the study.

Response Measurements, Interobserver Agreement, and Procedural Fidelity

The dependent measures included independent correct responses, prompted correct responses, incorrect responses, and trials with the occurrence of problem behavior. All dependent measures were recorded for acquisition and mastered stimuli. An *independent correct response* was defined as the occurrence of a pre-defined response to the target stimulus within 5 s of its presentation. A *prompted correct response* was defined as the occurrence of a pre-defined response to the target stimulus within 5 s of a gestural, model, or physical prompt. An *incorrect response* was defined as an error or a 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). *Aggression* was defined as contact between the participant's body and another person's body, including hitting, kicking, slapping, pinching, or raking fingers across skin. *Disruption* was defined as 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* was defined as moving more than 3 feet from the therapist from a seated position in a chair. *Negative vocalizations* were defined as crying, whining, screaming, or growling. *Vocal noncompliance* was defined as statements indicating vocal refusal to comply. Data collectors recorded problem behavior as an occurrence or non-occurrence per trial, and recorded the frequency of problem behavior for Lucas.

The dependent variables were the cumulative number of stimuli mastered in each condition and the percentage of problem behavior during sessions in each condition. Mastery of a stimulus was defined as four consecutive independent correct responses to a stimulus. The cumulative number of stimuli mastered was calculated for each condition by adding the number of stimuli mastered in each session across sessions of each condition. The percentage of problem behavior was calculated by dividing the number of trials with an occurrence of problem behavior by the total number of trials per session, and multiplied by 100.

The efficacy of the procedures was defined as the training procedures producing mastery of acquisition stimuli. Thus, any condition in which at least one acquisition stimulus was mastered was identified as efficacious. The efficacy of the procedures was also defined as the extent to which the training procedures reduced problem behavior. Thus, a condition was also efficacious if problem behavior was reduced by at least 50% from pre-test levels.

Efficiency was measured in several ways. First, efficiency was determined by calculating the cumulative number of acquisition stimuli mastered per condition. The condition with the most stimuli mastered was considered the most efficient intervention procedure. The efficiency of the procedures also was defined as acquiring stimuli in the fewest trial presentations, determined by the average trials to mastery. The mean number of trials to mastery per stimulus in each condition was calculated by dividing the total number of acquisition trial presentations for all sessions within the condition by the number of acquisition stimuli acquired in the condition (e.g., 240 trials/10 stimuli = 24 trials to mastery per acquisition stimulus). The mean number of trials to mastery per stimulus was then compared across conditions. Finally, 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).

Two independent observers recorded data on all dependent measures during 33.9% to 67.3% of sessions across all experimental conditions for all participants. Trial-by-trial interobserver agreement (IOA) was obtained 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 measures during the trial. IOA was calculated for each dependent measure in each condition by dividing the trials in which an agreement was scored by the total number of trials in the session, and multiplying by 100. The average agreement was calculated across dependent measures and conditions for each participant. Mean IOA was 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 for Owen and Finn, and IOA was calculated for the combined topographies of problem behavior. Sessions were divided into 10-s intervals to calculate agreement. Proportional agreement was scored 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. Proportional agreement was averaged for each session by adding percentages of agreements for each interval and dividing by the number of intervals in a session. Mean IOA was 95.5% (range, 93.3% to 100%) for Owen, and 87.8% (range, 80.5% to 100%) for Finn.

Observers also collected treatment integrity data during 33.9% to 40.5% of sessions to determine the extent to which the experimenter implemented the procedures as intended. The measures of integrity included: (a) presenting the correct discriminative stimulus as indicated on the data sheet, (b) presenting prompts immediately after incorrect or no responses (if relevant), (c) delivering reinforcement, defined as the presentation of praise and a tangible item for independent correct responses to acquisition and mastered stimuli, and the presentation of praise only for prompted correct responses, and (d) ignoring problem behavior and continuing the task, as described in the protocol. Treatment integrity was measured for each trial and was scored as either a one for correct implementation of the entire trial or a zero for incorrect implementation of any aspect of the trial. The percentage of treatment integrity was calculated for each session by dividing the number of trials implemented correctly by the total number of trials in the session, multiplying by 100. Treatment integrity 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.

Pre-Test

One skill (e.g., tacts) was targeted for each participant. The targeted skill was selected based upon individual treatment goals related to each participant's skill deficits determined by assessments conducted prior to the onset of the study. For Owen, the targeted skill was auditory-visual conditional discriminations (AVCD; i.e., receptive identification). Finn's targeted skill was tacts of common items (i.e., expressive object labels). For Benny, the targeted skill was tacts of item features (e.g., an elephant's trunk). Lucas's targeted skill was adjective-noun tacts (e.g., brown bear).

Participants completed a pre-test to identify stimuli to include in each condition and to identify participants who were categorized as displaying problem behavior during instruction. Stimuli were grouped into sets of three during each pre-test session. During the first two pre-test sessions, sessions consisted of 15 trials with three stimuli presented five times. The experimenter presented each stimulus, allowed up to 5 s for a response, and removed the stimulus following a correct, incorrect, or no response within 5 s. No prompts or reinforcement were provided during the first two sessions of the pre-test. The experimenter collected data on correct responses and problem behavior. Participants who engaged in problem behavior in four or more trials during the first 30 pre-test trials were categorized as displaying problem behavior during instruction (Owen and Finn). Owen engaged in problem behavior during 25 out of 30 (83.3%) pre-test trials, and Finn engaged in problem behavior during 13 out of 30 (43.3%) pre-test trials. Lucas engaged in problem behavior during 3 of 30 (10%) pre-test trials, therefore, he was not categorized as displaying problem behavior during instruction. Benny engaged in problem behavior during 0 of 30 (0%) pre-test trials, therefore, he was not categorized as displaying problem behavior during instruction.

After conducting the first two sessions of the pre-test, the pre-test procedures were modified. Sessions consisted of 9 trials with each stimulus presented three times. Experimenters interspersed previously mastered items on a VR2 schedule. Prompts and reinforcement were provided for independent or prompted correct responses to mastered items only. No prompts were provided during the pre-test following incorrect or no responses to the pre-test targets. Reinforcement was provided for independent correct responses to the pre-test tact targets, but reinforcement was not provided for independent correct responses to the AVCD pre-test targets to avoid teaching these skills during the pre-test.

The pre-test procedures for Owen varied from the other three participants for half of the pre-test sessions. The first five sessions of the pre-test consisted of 15 trials with each stimulus presented five times. He engaged in high levels of disruption during the first five pre-tests. We were concerned that the disruption of the materials would result in the incorrect identification of target stimuli; thus, we trained nine stimuli to mastery using his typical intervention procedures (5-s prompt delay with non-differential reinforcement). Thereafter, we modified the pre-test procedures to include the interspersal of recently mastered items on a VR2 with praise and a tangible, edible, or token delivered following an independent or prompted correct response. After conducting six sessions with the modified procedures, we made a second modification to reduce the length of the pre-test. We terminated a session when he engaged in two correct responses to each target in the session. Sessions ranged from 6 to 15 trials. After conducting 30 sessions with the second modification, we modified the procedures to be consistent with the pre-test procedures implemented with the other three participants (described above).

Stimuli to which the participant engaged in a correct response during 100% of pre-test trials (i.e. 3 out of 3 correct responses) were designated as mastered stimuli. Stimuli to which the participant engaged in a correct response during 0% of pre-test trials for tacts and no more than 33% of pre-test trials for AVCDs were designated as acquisition stimuli. Mastered stimuli were comprised of functionally related skills (e.g., both the mastered and acquisition stimuli were tacts of common items) for all participants.

We assigned unique sets of three acquisition stimuli per experimental condition using a logistical analysis method (Gast, 2010). That is, stimuli were assigned based on (a) a similar number of syllables in responses across conditions, (b) overlapping visual or auditory stimuli separated across condition, and (c) similar levels of correct responding during the pre-test. A unique set of mastered stimuli also was assigned to each condition. Although each condition included a set of three acquisition stimuli that were trained simultaneously, additional acquisition tasks were identified and assigned to each condition so that any acquisition stimulus that met the mastery criterion during training was replaced by a new acquisition stimulus. The specific stimuli assigned to each condition for each participant can be found in Appendices A – H.

Echoic Assessment

We conducted an echoic assessment with Lucas to identify the highest form of the approximation of the target response for each stimulus. Sessions consisted of 12 trials with six vocal stimuli presented two times. The experimenter presented the relevant vocal stimulus and allowed up to 5 s for an echoic response. If the participant echoed or approximated the vocal stimulus, the experimenter provided praise and a tangible item for 20 s. If the participant engaged in a poor approximation of the vocal model, the experimenter provided praise only. Following a poor approximation, the experimenter re-presented the vocal stimulus one more time and allowed up to 5 s for an echoic response. If the participant echoed or approximated the vocal stimulus, the experimenter provided praise and a tangible item for 20 s. If the participant did not engage in an echoic or approximated response, the experimenter re-presented the vocal model one more time. If the participant echoed or approximated the vocal stimulus, the experimenter provided praise and a tangible item for 20 s.

Functional Analysis

We conducted a FA for Owen and Finn who were identified as participants who engaged in problem behavior during the first two sessions of the pre-test. We conducted an abbreviated FA based on a portion of the procedures described Iwata, Dorsey, Slifer, Bauman, and Richman, (1982/1994) to test if problem behavior was maintained by escape from demands. All sessions were conducted at the table and were 5 min.

Escape. The experimenter initiated sessions by saying, "Owen/Finn, we have some work to do". The experimenter presented instructions using three-step prompting consisting of vocal, model, and physical prompts. The experimenter 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 participant's problem behavior.

Toy Play. The experimenter sat next to the participant and provided continuous vocal attention and brief physical contact at least every 30 s. No instructions were presented, and no consequence was provided following the occurrence of any topography of problem behavior.

This condition was included as a control to provide noncontingent access to all potential socially mediated reinforcers.

Preference Assessment

Tangible items were identified based on parent report. Initial MSWO preference assessments were conducted with Owen, Finn, and Lucas based on the procedures described Carr, Nicholson, & Higbee (2000). The experimenter placed a linear array of six to eight tangible items in front of the participant, and said, “pick one”. The participant received the selected item for 20 s, and then the selected item was removed from the array. The remaining tangible items were repositioned on the table, and this sequence continued until all items were chosen or the participant did not respond to an item for 30 s. The most preferred item was used as a reinforcer for the sessions. However, participants consistently engaged in mands for preferred items. Thus, daily MSWO assessments were not conducted with participants.

Benny completed several one-trial MSWO assessments (similar to DeLeon, Fisher, Rodriguez-Catter, Maglieri, Herman, & Markhefka, 2001) prior to the start of the study. Benny consistently selected the same item in the first trial across all assessments. That item, plus three other items included in the MSWO, were available during all sessions. Once Benny completed his token economy (i.e., he earned three tokens), he selected an item from an array of four items placed on the table.

Procedure

An adapted alternating treatments design was implemented to examine the effects of TI ratios on the efficacy and efficiency of intervention and levels of problem behavior. Each participant was exposed 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 were implemented in a pseudo-random order for each participant, alternating the order of each condition within a session block of four sessions before re-ordering for the next session block to control for order effects. 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 provide 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. Thus, the experimenter presented the stimulus material(s) and allowed up to 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.

All mastered tasks were presented at a 5-s prompt delay. The experimenter presented the relevant stimulus material(s) and allowed up to 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 failed to respond within 5 s or responded incorrectly to the mastered stimulus, the experimenter prompted the correct response, and delivered praise only for 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 was considered mastered if the participant engaged in an independent correct response for four consecutive presentations of the stimulus. Once an acquisition stimulus was mastered, it was removed from treatment and replaced with another acquisition stimulus assigned to the condition. Mastered acquisition stimuli were not added to the pool of mastered stimuli presented during training in order to assess maintenance of recently mastered stimuli in the absence of continued practice. Training was completed when 21 acquisition stimuli were mastered in at least one condition, or when a total of 30 sessions per condition were conducted.

75% mastered stimuli to 25% acquisition (75%M/25%A). Nine of the trials were designated to mastered stimuli, and three of the trials were designated to 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. Each mastered stimulus was presented three times per session, and each acquisition stimulus was presented one time per session. Thus, an acquisition stimulus could have been mastered in a minimum of six consecutive sessions.

50% mastered to 50% acquisition stimuli (50%M/50%A). Six of the trials were designated to mastered stimuli, and six of the trials were designated to 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. Each mastered stimulus was presented two times per session, and each acquisition stimulus was presented two times per session. Thus, an acquisition stimulus could have been mastered in a minimum of two and a half consecutive sessions.

25% mastered stimuli to 75% acquisition (25%M/75%A). Three of the trials were designated to mastered stimuli, and nine of the trials were designated to 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. Each mastered stimulus was presented one time per session, and each acquisition stimulus was presented three times per session. Thus, an acquisition stimulus could have been mastered in a minimum of two sessions.

0% mastered stimuli to 100% acquisition (0%M/100%A). All trials were designated to the acquisition stimuli. There was no interspersal of mastered stimuli in this condition. Each acquisition stimulus was presented four times per session. Thus, an acquisition stimulus could have been mastered in one and a half sessions of this condition.

Maintenance

Maintenance probes were conducted 1 and 2 weeks following the mastery of an acquisition stimulus. Maintenance probes did not include prompts or reinforcement during trials, and mastered stimuli were not interspersed between trials. Maintenance probes were conducted in blocks of five trials, with the presentation of one mastered acquisition stimulus in five consecutive trials. A stimulus was maintained at 1 and 2 weeks if the participant responded correctly to at least 3/5 (60%) of presentations for an AVCD (Owen) or at least 1/5 (20%) of presentations for a tact (Finn, Lucas, Benny).

RESULTS

Figure 1 shows the results of the FA for Owen. As previously stated, Owen was identified as a participant who engaged in problem behavior during instruction, based on the results of his pre-test. The results of the FA indicated that Owen's problem behavior was maintained by escape from demands. In the toy play condition, Owen engaged in an average of 0.2 instances of problem behavior per min. In the escape condition, Owen engaged in an average of 1.8 instances of problem behavior per min.

Figure 2 depicts the cumulative number of stimuli mastered in each experimental condition for Owen. Owen simultaneously met both termination criteria to complete the study (i.e., 30 sessions per condition and mastery of 21 stimuli in a condition). He mastered one stimulus in the 75%M/25%A condition, 15 stimuli in the 50%M/50%A condition, 10 stimuli in the 25%M/75%A condition, and 21 stimuli in the 0%M/100%A condition. Thus, all conditions were efficacious in teaching at least one acquisition stimulus. Nevertheless, the 25%M/75%A condition resulted in fewer stimuli mastered than in the 50%M/50%A condition. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered. This indicates that interspersing mastered tasks decreased the number of stimuli mastered during instruction.

Figure 3 depicts the average trials to mastery per stimulus for Owen. A total of 90 (75%M/25%A), 180 (50%M/50%A), 270 (25%M/75%A), and 360 (0%M/100%A) acquisition trials were conducted across experimental conditions. The highest mean number of trials to mastery occurred in the 75%M/25%A condition, with an average of 90 trials to mastery per stimulus. The lowest average trials to mastery per stimulus occurred in the 50%M/50%A condition, with an average of 12 trials to mastery per stimulus. Thus, the most efficient condition in terms of trials to mastery was the 50%M/50%A condition.

Figure 4 depicts the average min to mastery per stimulus for Owen. The total training times per condition were 242 min (75%M/25%A), 250 min (50%M/50%A), 229 min (25%M/75%A), and 232 min (0%M/100%A). The highest average min to mastery per stimulus occurred in the 75%M/25%A condition, with an average of 242 min to mastery per stimulus. The lowest average min to mastery per stimulus occurred in the 0%M/100%A condition, with an average of 11 min to mastery per stimulus. Thus, the most efficient condition in terms of min to mastery was the 0%M/100%A condition.

Figure 5 depicts the percentage of trials with problem behavior during experimental conditions for Owen. Owen engaged in comparable levels of problem behavior across experimental sessions, with problem behavior occurring during an average of 1.1% of trials per session in the 75%M/25%A condition, 2.2% of trials per session in the 50%M/50%A condition, 0.6% of trials per session in the 25%M/75%A condition, and 1.7% of trials per session in the 0%M/100%A condition. This corresponds to a 99%, 96%, 99%, and 97% reduction from pre-test levels of problem behavior for 75%M/25%A, 50%M/50%A, 25%M/75%A, and 0%M/100%A conditions, respectively. Thus, all conditions were efficacious in reducing problem behavior.

Figure 6 depicts the percentage of stimuli maintained one and two weeks following the mastery of a stimulus for Owen. Owen maintained the stimulus mastered in the 75%M/25%A condition one week following mastery, but did not maintain the target two weeks following mastery. Owen maintained 73% (11 of 15) of stimuli mastered in the 50%M/50%A condition one week following mastery, and maintained 80% (12 of 15) of stimuli two weeks following mastery. He maintained all stimuli mastered in the 25%M/75%A condition one week following mastery (10 of 10 stimuli), but decreased to 70% (7 of 10 stimuli) maintenance of the mastered targets two weeks following mastery. He maintained 71% (15 of 21) of stimuli mastered in the 0%M/100%A condition one week following mastery, but decreased to 62% (13 of 21) maintenance of the mastered targets two weeks following mastery. Thus, there were small differences in maintenance observed across conditions with the highest overall levels of maintenance observed in the 25%M/75%A condition.

Overall, all experimental conditions were efficacious in teaching at least one acquisition stimulus and reducing problem behavior by at least 50% for Owen. The average percentage of trials with problem behavior was reduced to near zero levels across conditions; thus, all conditions were efficacious in reducing problem behavior. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli acquired and the lowest number of min to mastery per stimulus. The 50%M/50%A condition was the most efficient condition in terms of trials to mastery. This pattern of acquisition for Owen suggests that interspersal of mastered tasks was detrimental to his acquisition.

Figure 7 shows the results of the FA for Finn. Finn was identified as engaging in problem behavior during instruction. The results of Finn’s FA identified that his problem behavior was maintained by escape from demands. In the toy play condition, Finn engaged in an average of 0.07 instances of problem behavior per min. In the escape condition, he engaged in an average of 5.51 instances of problem behavior per min.

Figure 8 depicts the cumulative number of stimuli mastered in each experimental condition for Finn. Finn met the termination criterion of mastering 21 stimuli in a condition; thus, training was terminated following 19 sessions per condition. He mastered five stimuli in the 75%M/25%A condition, 11 stimuli in the 50%M/50%A condition, 14 stimuli in the

25%M/75%A condition, and 21 stimuli in the 0%M/100%A condition. Therefore, all conditions were efficacious in teaching at least one acquisition stimulus. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered.

Figure 9 depicts the average trials to mastery per stimulus for Finn. A total of 57 (75%M/25%A), 114 (50%M/50%A), 171 (25%M/75%A), and 228 (0%M/100%A) acquisition trials were conducted across experimental conditions. The number of trials to mastery per stimulus ranged from 10 to 12 across experimental conditions. The highest mean number of trials to mastery occurred in the 25%M/75%A condition, with an average of 12 trials to mastery per stimulus. The lowest average trials to mastery per stimulus occurred in the 50%M/50%A condition, with an average of 10 trials to mastery per stimulus. Thus, the most efficient condition in terms of trials to mastery was the 50%M/50%A condition.

Figure 10 depicts the average min to mastery per stimulus for Finn. The total training times per condition were 118 min (75%M/25%A), 114 min (50%M/50%A), 111 min (25%M/75%A), and 105 min (0%M/100%A). The average min to mastery per stimulus ranged from 5 to 23.6 across experimental conditions. The highest average min to mastery occurred in the 75%M/25%A condition, with an average of 23.6 min to mastery per stimulus. The lowest average min to mastery per stimulus occurred in the 0%M/100%A condition, with an average of 5 min to mastery per stimulus. Thus, the most efficient condition in terms of min to mastery was the 0%M/100%A condition.

Figure 11 depicts the percentage of trials with problem behavior during experimental conditions for Finn. Finn engaged in variable levels of problem behavior across experimental conditions, with problem behavior occurring during an average of 24.1% of trials per session in the 75%M/25%A condition, 26.8% of trials per session in the 50%M/50%A condition, 33.3% of trials per session in the 25%M/75%A condition, and 40.4% of trials per session in the 0%M/100%A condition. This corresponds to a 44.3%, 38.3%, 23.1%, and 6.8% reduction from pre-test levels of problem behavior for 75%M/25%A, 50%M/50%A, 25%M/75%A, and 0%M/100%A conditions respectively. Thus, although we observed reductions in problem behavior, none of the conditions were efficacious in reducing problem behavior according to our definition of efficacy.

Figure 12 depicts the percentage of stimuli maintained one and two weeks following the mastery of a stimulus for Finn. Finn maintained all stimuli mastered in the 75%M/25%A ,50%M/50%A condition, and 25%M/75%A condition one and two weeks following mastery. He maintained 100% of stimuli mastered in the 0%M/100%A condition one week following mastery, but decreased to 95% (20 of 21) maintenance of the mastered targets two weeks following mastery. Thus, all conditions resulted in similar levels of maintenance across weeks.

In summary, all experimental conditions were efficacious in teaching at least one acquisition stimulus for Finn. None of the experimental conditions were efficacious in reducing problem behavior to 50% of pre-test levels. The largest reduction in the average percentage of trials with problem behavior occurred in the 75%M/25%A condition. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered and the lowest number of min to mastery per stimulus. The 0%M/100%A condition and the 50%M/50%A condition resulted in comparable mean trials to mastery; thus, both conditions were efficient. The overall pattern of Finn's results suggests that interspersal decreased the efficiency of intervention and resulted in some reduction in problem behavior.

Although Lucas engaged in problem behavior during the first two sessions of the pre-test, it did not meet our criteria of inclusion as a participant who engaged in problem behavior during instruction. Therefore, we did not conduct a FA. Figure 13 depicts the cumulative number of stimuli mastered in each experimental condition for Lucas. Lucas met the termination criterion of mastering 21 stimuli in a condition; thus, training was terminated following 27 sessions per condition. He mastered zero stimuli in the 75%M/25%A condition, nine stimuli in the 50%M/50%A condition, 16 stimuli in the 25%M/75%A condition, and 22 stimuli in the 0%M/100%A condition. Therefore, all but the 75%M/25%A condition were efficacious in teaching at least one acquisition stimulus. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered.

Figure 14 depicts the average trials to mastery per stimulus for Lucas. A total of 81 (75%M/25%A), 162 (50%M/50%A), 243 (25%M/75%A), and 324 (0%M/100%A) acquisition trials were conducted across experimental conditions. The highest mean number of trials to mastery occurred in the 50%M/50%A condition, with an average of 18 trials to mastery per stimulus. The lowest average trials to mastery per stimulus occurred in the 25%M/75%A condition and the 0%M/100%A condition, with an average of 15 trials to mastery per stimulus. We were unable to calculate the average trials to mastery per stimulus in the 75%M/25%A condition because Lucas did not acquire any stimuli in this condition. Thus, the most efficient conditions in terms of trials to mastery were the 25%M/75%A and 0%M/100%A conditions.

Figure 15 depicts the average min to mastery per stimulus for Lucas. The total training times per condition were 160 min (75%M/25%A), 150 min (50%M/50%A), 158 min (25%M/75%A), and 155 min (0%M/100%A). The highest average min to mastery per stimulus occurred in the 50%M/50%A condition, with an average of 16.7 min to mastery per stimulus. The lowest average min to mastery per stimulus occurred in the 0%M/100%A condition, with an average of 7.1 min to mastery per stimulus. We were unable to calculate the average min to mastery in the 75%M/25%A condition because Lucas did not acquire any stimuli in this condition. Thus, the most efficient condition in terms of min to mastery was the 0%M/100%A condition.

Figure 16 depicts the percentage of stimuli maintained one and two weeks following the acquisition of a stimulus for Lucas. Lucas maintained 22% (2 of 9) of stimuli mastered in the 50%M/50%A condition one week following mastery, but decreased to 0% (0 of 9) maintenance of stimuli two weeks following mastery. He maintained 18.75% (3 of16) of stimuli mastered in the 25%M/75%A condition one week following mastery, but decreased to 0% (0 of 16) maintenance of the acquired targets two weeks following mastery. He maintained 31.8%% (7 of 22) of stimuli mastered in the 0%M/100%A condition one week following mastery, but decreased to 4.5% (1 of 22) maintenance of the mastered targets two weeks following mastery. Thus, none of the condition resulted in high levels of maintenance.

In summary, three of the four experimental conditions were efficacious in teaching at least one acquisition stimulus to Lucas. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered and the lowest number of min to mastery per stimulus. The 0%M/100%A and 25%M/75%A conditions resulted in comparable mean trials to mastery; thus, both conditions were efficient. The 75%M/25%A condition was the least efficacious and efficient. The overall pattern of Lucas's results suggests that interspersal decreased the efficiency of intervention.

Benny did not engage in problem behavior during the first two sessions of the pre-test, therefore we did not conduct a FA. Figure 17 depicts the cumulative number of stimuli mastered in each experimental condition for Benny. Benny met the 30 sessions per condition termination criteria to complete the study. He mastered one stimulus in the 75%M/25%A condition, one stimulus in the 50%M/50%A condition, three stimuli in the 25%M/75%A condition, and five stimuli in the 0%M/100%A condition. Thus, all conditions were efficacious in teaching at least one acquisition stimulus. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli mastered.

Figure 18 depicts the average trials to mastery per stimulus for Benny. A total of 90 (75%M/25%A), 180 (50%M/50%A), 270 (25%M/75%A), and 360 (0%M/100%A) acquisition trials were conducted across experimental conditions. The highest mean number of trials to mastery occurred in the 50%M/50%A condition, with an average of 180 trials to mastery per stimulus. The lowest average trials to mastery per stimulus occurred in the 0%M/100%A condition, with an average of 72 trials to mastery per stimulus. Thus, the most efficient condition in terms of trials to mastery was the 0%M/100%A condition.

Figure 19 depicts the average min to mastery per stimulus for Benny. The total training times per condition were 92 min (75%M/25%A), 85 min (50%M/50%A), 84 min (25%M/75%A), and 90 min (0%M/100%A). The highest average min to mastery per stimulus occurred in the 75%M/25%A condition, with an average of 92.0 min to mastery per stimulus. The lowest average min to mastery per stimulus occurred in the 0%M/100%A condition, with an average of 16.6 min to mastery per stimulus. Thus, the most efficient condition in terms of min to mastery was the 0%M/100%A condition.

Figure 20 depicts the percentage of stimuli maintained one and two weeks following the acquisition of a stimulus for Benny. Benny maintained the stimuli mastered in the 75%M/25%A and 50%M/50%A conditions one and two weeks following mastery. Benny maintained 33% (1 of 3) of stimuli mastered in the 25%M/75%A condition one and two weeks following mastery. He maintained 80% (4 of 5) of stimuli mastered in the 0%M/100%A condition one week following mastery, but decreased to 60% (3 of 5) maintenance of the acquired targets two weeks following mastery. Thus, there were differences in maintenance observed across conditions with the overall highest levels of maintenance occurring in the conditions with more interspersed mastered target trials.

Overall, Benny's results showed all experimental conditions were efficacious in teaching at least one acquisition stimulus. The 0%M/100%A condition was the most efficient condition, resulting in the largest number of stimuli acquired, the lowest number of trials to mastery per stimulus, and the lowest number of min to mastery per stimulus. Although Benny acquired a small number of stimuli across experimental conditions, the overall pattern of his results suggests that interspersal decreased the efficiency of intervention.

DISCUSSION

The current study extended the literature on TI by comparing the efficacy and efficiency of varying TI ratios on the rate of acquisition and level of problem behavior of children with ASD and related disorders. Only one known study has examined the efficiency of varying TI ratios (Nicholson, 2013). Our results were consistent with Nicholson (2013), suggesting all interspersal ratios were detrimental to the acquisition of the participants. The 0%M/100%A condition was the most efficient condition for all participants according to min to mastery and the cumulative number of stimuli mastered. This is consistent with studies in which interspersal procedures were found to be inferior to other instructional procedures (Majdalany et al., 2014; Henrickson et al., 2015; Volkert et al., 2008).

When examining the efficiency of instructional approaches, it is important to include multiple measures of efficiency. This is because one measure may provide a more accurate representation of efficiency, while the others may still provide relevant information. 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, 2013) as measures of efficiency. In the current study, we included three measures of efficiency: the cumulative number of stimuli mastered, the trials to mastery, and the time to mastery. Each measure provides valuable information, but they should be interpreted together to prevent false conclusions.

Although the cumulative number of stimuli mastered and trials to mastery provide valuable information, they may not be as sensitive a measure as the min to mastery per stimulus (Kodak et al., 2016; Yaw et al., 2014). For example, a condition with the lowest trials to mastery may not be synonymous to a condition in which the most stimuli were acquired in the least amount of time. In the case of substitutive interspersal, as in the current study, the number of acquisition trials per session decreased with the interspersal of previously mastered item trials. Therefore, comparable trials to mastery across conditions would indicate more time to mastery for conditions with interspersal. There were discrepancies amongst the measures of efficiencies that would have gone undetected if only one measure of efficiency was included. For Owen, 50%M/50%A resulted in fewer trials to mastery than 0%M/100%A, but 0%M/100%A resulted in less time to mastery than 50%M/50%A. If trials to mastery was the only measure of efficiency included, interspersal could be falsely concluded to be more efficient than no interspersal. Minutes to mastery may be a more accurate representation of efficiency to indicate that a condition resulted in the least amount of training time per stimulus when the number of acquisition trials is not held constant across conditions.

The conditions with TI decreased the efficiency of instruction by including mastered tasks within the instructional period. Interspersing mastered tasks decreased the number of acquisition trials per session, thus decreasing the proportion of session time spent teaching acquisition targets. We can see this affect with Finn (Figures 9 and 10). Finn had comparable average trials to mastery per stimulus across conditions, but the average min to mastery per stimulus was higher in the conditions with interspersal. If interspersed trials were removed from sessions and replaced with acquisition trials, he may have acquired stimuli in comparable amounts of time. The results of this study are consistent with Henrickson et al. (2015), who found that a 3:1 interspersal ratio resulted in more time to mastery than a condition without interspersal. The results of this study add to the literature suggesting that clinicians should carefully consider any perceived benefits of using TI because it has been shown to reduce the efficiency of instruction.

The study also extended the literature on TI by experimentally identifying escape from demands as the maintaining variable for problem behavior with a FA and measuring level of problem behavior across experimental conditions. Although interspersal resulted in a reduction of problem behavior during instruction for both participants who engaged in problem behavior, the results do not allow for a definitive conclusion on the efficacy of interspersal procedures on reducing problem behavior. Owen engaged in similar levels of problem behavior across conditions, including the condition with no interspersed mastered tasks. The results for Owen were similar to the participants in Henrickson et al. (2015), in which all participants engaged in low and comparable levels of problem behavior during instruction with and without interspersed mastered tasks.

In comparison, Finn engaged in variable levels of problem behavior across conditions, but his results showed the lowest levels of problem behavior in the 75%M/25%A condition. Finn’s reduction in problem behavior coincided with the interspersal of mastered task trials, suggesting interspersal may be efficacious in reducing problem behavior. Although the 75%M/25%A condition was efficacious in reducing problem behavior, it was the least efficient condition for acquisition. When implementing an instructional format, it is important to maximize efficiency of acquisition while reducing problem behavior to near-zero levels. None of the conditions in the current study achieved this goal with Finn. In addition, Finn continued to engage in moderate levels of problem behavior (i.e., 24.1% of trials) during intervention despite a high proportion of interspersed mastered tasks in the 75%M/25%A condition. It is possible that problem behavior was not significantly reduced for Finn because all functional reinforcers for problem behavior were not identified with our modified FA. Problem behavior could have been multiply maintained, or different topographies could have been maintained by different functional reinforcers. This was not identified since we conducted a pairwise assessment with escape and toy play conditions only, and excluded conditions that tested for problem behavior maintained by social positive reinforcement. Future researchers seeking to examine the relation between problem behavior and the efficacy of TI procedures could conduct a full FA of problem behavior.

It also is possible that Finn’s problem behavior was not reduced because it was adventitiously reinforced. Finn may have engaged in problem behavior in close temporal proximity to a correct response resulting in access to a brief break from demands that included a tangible reinforcer. Therefore, the contingent relation between problem behavior and escape may not have been disrupted. This is an important consideration when combining putative interventions for problem behavior with skill acquisition procedures. Including TI procedures during acquisition programs as a strategy to reduce problem behavior may not sufficiently accomplish this goal if problem behavior continues to be reinforced on an intermittent schedule due to the timing of problem behavior and correct responses. Clinicians using TI procedures for this purpose could omit reinforcement (and provide additional prompts) if problem behavior and correct responses occur simultaneously or in close temporal proximity to prevent adventitious reinforcement of problem behavior during training trials. Nevertheless, the addition of prompts will likely extend time in instruction and may lead to further reductions in instructional efficiency.

Two previous studies examining TI implemented escape extinction (Horner et al., 1991; Volkert et al., 2008). Horner et al., (1991) found that TI paired with escape extinction reduced problem behavior, but the authors did not examine the efficiency of instruction. Although Volkert et al. (2008) implemented TI with escape extinction, they did not report data on problem behavior. Therefore, it is unknown what affect TI with escape extinction had on problem behavior. From the limited literature including escape extinction with TI, it is unclear if it alters the efficiency of the intervention and/or the efficacy of reducing problem behavior. Thus, the decision to modify TI procedures to include escape extinction should be made in consideration of the importance of the efficiency of instruction.

Previous research on TI procedures found differences in maintenance of targeted skills across conditions. For example, Henrickson et al. (2015) found that targets acquired in ITT were maintained, whereas targets acquired in MTT were not maintained. We did not observe differences in maintenance across conditions for all participants. Two out of the four participants maintained a high percentage of acquired targets across all conditions at one- and two-week maintenance. One participant, Benny, maintained a higher percentage of stimuli mastered in the interspersal condition. The discrepancy between our results and those of Henrickson et al. may be a result of differences in maintenance criteria. Henrickson et al. included a maintenance criterion of mastery level responding, whereas we included a minimum level of responding for each stimulus (i.e., at least 3/5 correct responses for AVCD's, and at least 1/5 correct response for tacts). If we had included a more stringent maintenance criterion (e.g., mastery level responding), participants would have had lower levels of maintenance across all conditions.

Maintenance trials in the present study did not include feedback for correct responses and included repeated presentation of the same stimulus across all trials. Thus, it is possible that there were discriminable changes in the trial arrangement and reinforcement contingencies from experimental to maintenance sessions. One participant's maintenance results are consistent with this account. Lucas did not maintain many of the stimuli acquired across conditions. He typically did not respond to any of the trials in a session. When he engaged in a correct response, it occurred during the first maintenance trial and he did not respond to the remaining trials in the session. This may indicate noncompliance due to absence of reinforcement contingencies for correct responses.

It is important to note that even a single correct response during a session is indicative of maintenance of the target, because Lucas's targets required a vocal response to a picture, and the session was comprised of five presentations of only one stimulus. Previous studies included maintenance trials embedded within skill acquisition sessions to reduce the discriminability of reinforcement contingencies across maintenance and training trials (e.g., Allan, Vladescu, Kisamore, Reeve, & Sidener, 2015). Although our TI procedures prevented the inclusion of maintenance trials within sessions, including recently mastered skills within the programmed TI trials during clinical services could present natural opportunities to assess and promote maintenance of these stimuli. It is possible that periodic practice of mastered skills within TI could lead to better maintenance for clients who may not maintain skills without extended practice of these skills over time. 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 25%M/75%A 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. Similarly, Benny only acquired 10 stimuli across 120 total experimental sessions. This may have been the result of faulty stimulus control resulting from a limited number of initial 0-s prompt delay trials followed by differential reinforcement. An increase in the number of initial 0-s prompt delay trials could allow the response to come under control of the prompt. Thereafter, stimulus control could be transferred from the prompt to the discriminative stimulus with differential reinforcement. Including a criterion to re-conduct trials with an immediate prompt could have facilitated the transfer of stimulus control, resulting in fewer trials to mastery and allowing more targets to be introduced into training.

The lack of efficiency of the interspersal conditions may be attributed to the limited number of trials per acquisition stimulus per session. There was an unequal number of exposures to targets in sessions across conditions. It is unclear if interspersal of mastered target trials was detrimental to acquisition, or if the number of exposures in a session to an acquisition target was detrimental. This in an inherent flaw in the procedures of substitutive interspersal, with fewer acquisition target trials as more mastered target trials are added to sessions. To address this limitation, future research should examine the efficacy and efficiency of varying ratios of TI using the additive method of interspersal, in which the number of exposures to each stimulus per session is held constant.

The limited number of participants who engaged in problem behavior in the present study, and the mixed results amongst these participants, limits the conclusions that can be drawn regarding the efficacy of TI in reducing escape-maintained problem behavior during instruction. Future research should include more participants who engage in escape-maintained problem behavior to determine if TI results in concomitant reductions in problem behavior while teaching novel skills.

In conclusion, the current study suggests that interspersal is detrimental to acquisition but may reduce problem behavior during instruction for some individuals with escape-maintained problem behavior. When implementing an instructional format, it is important to maximize the efficiency of instruction as well as reduce problem behavior to near-zero levels. Although TI may lead to a reduction in problem behavior, the reduction in the efficiency of instruction based on the inclusion of mastered tasks should be carefully considered when selecting intervention procedures.

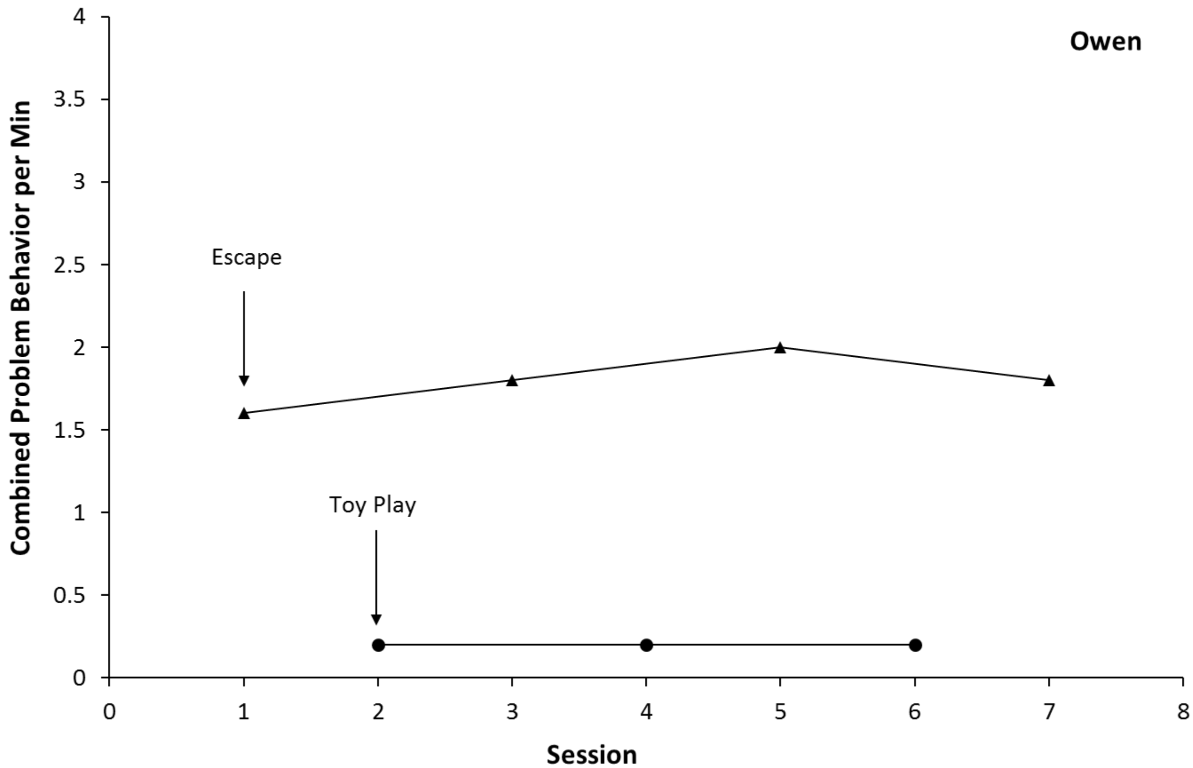

Figure 1. Combined problem behavior per min in escape (closed triangles) and toy play (closed circles) conditions of the functional analysis for Owen.

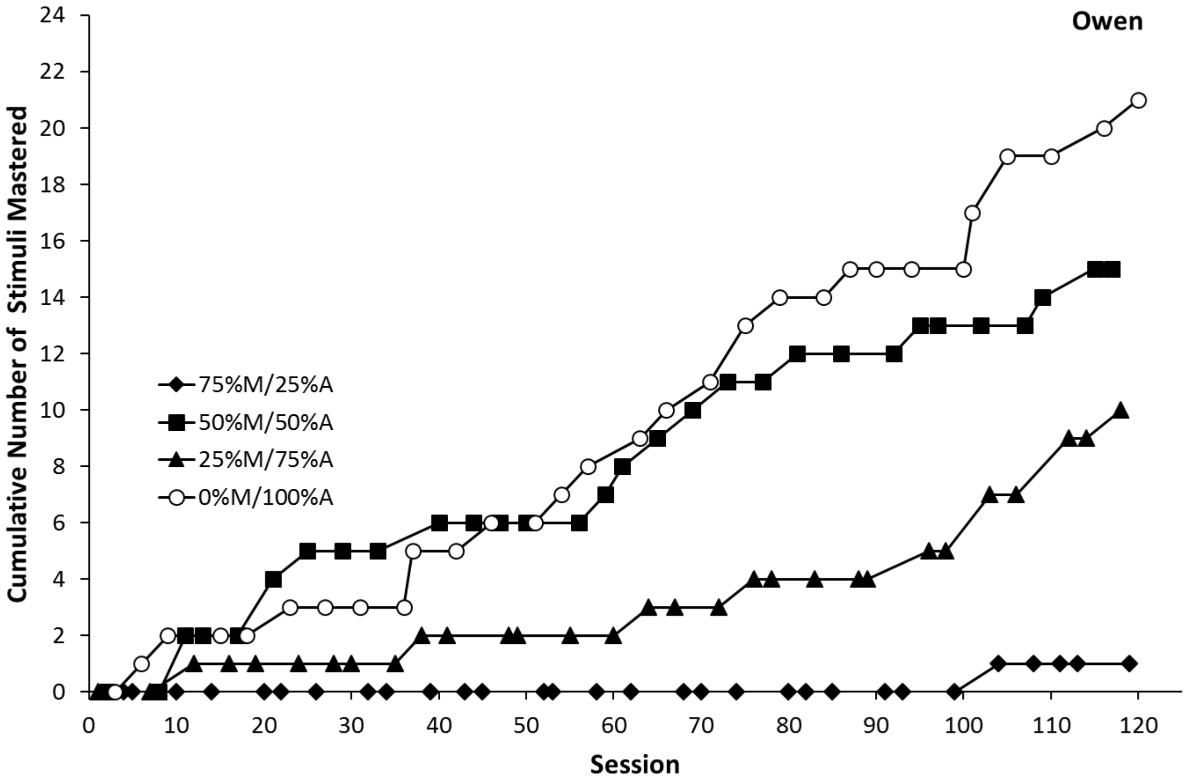

Figure 2. The cumulative number of stimuli mastered across conditions for Owen.

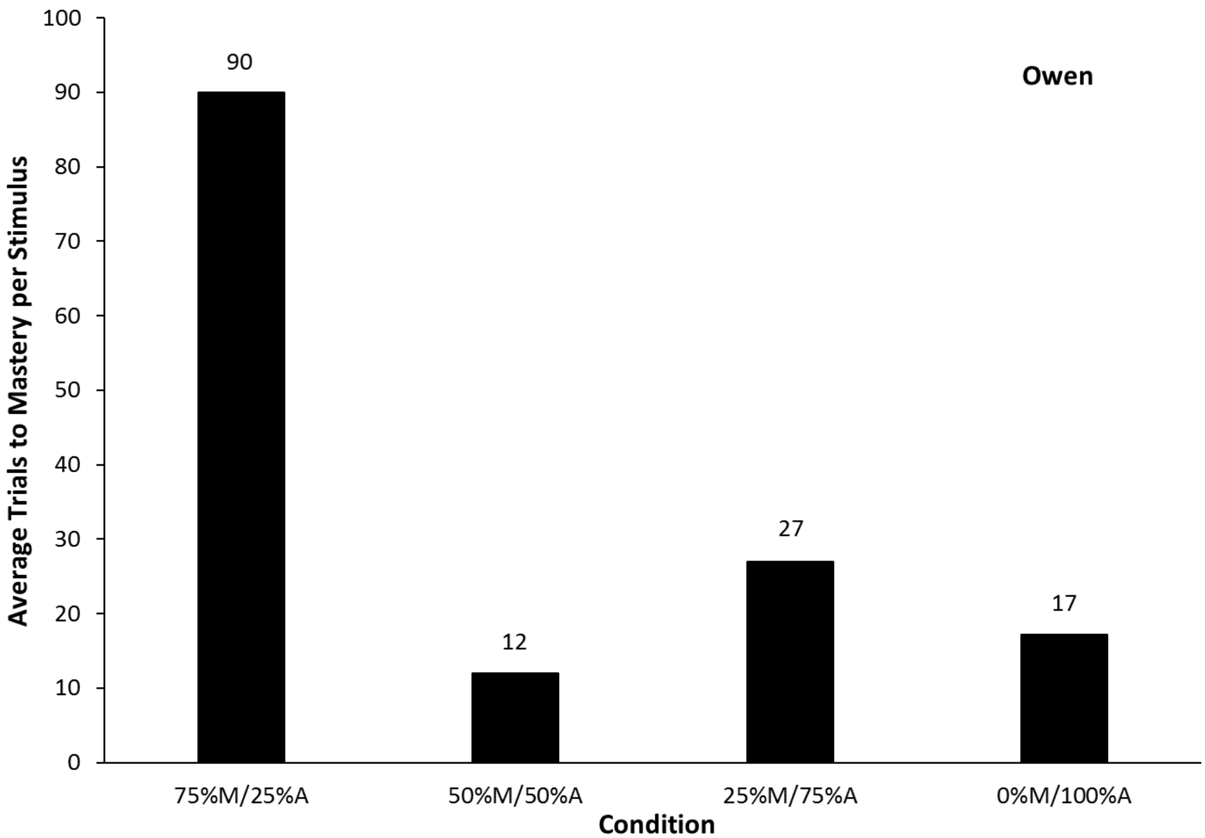

Figure 3. The average trials to mastery per stimulus for Owen.

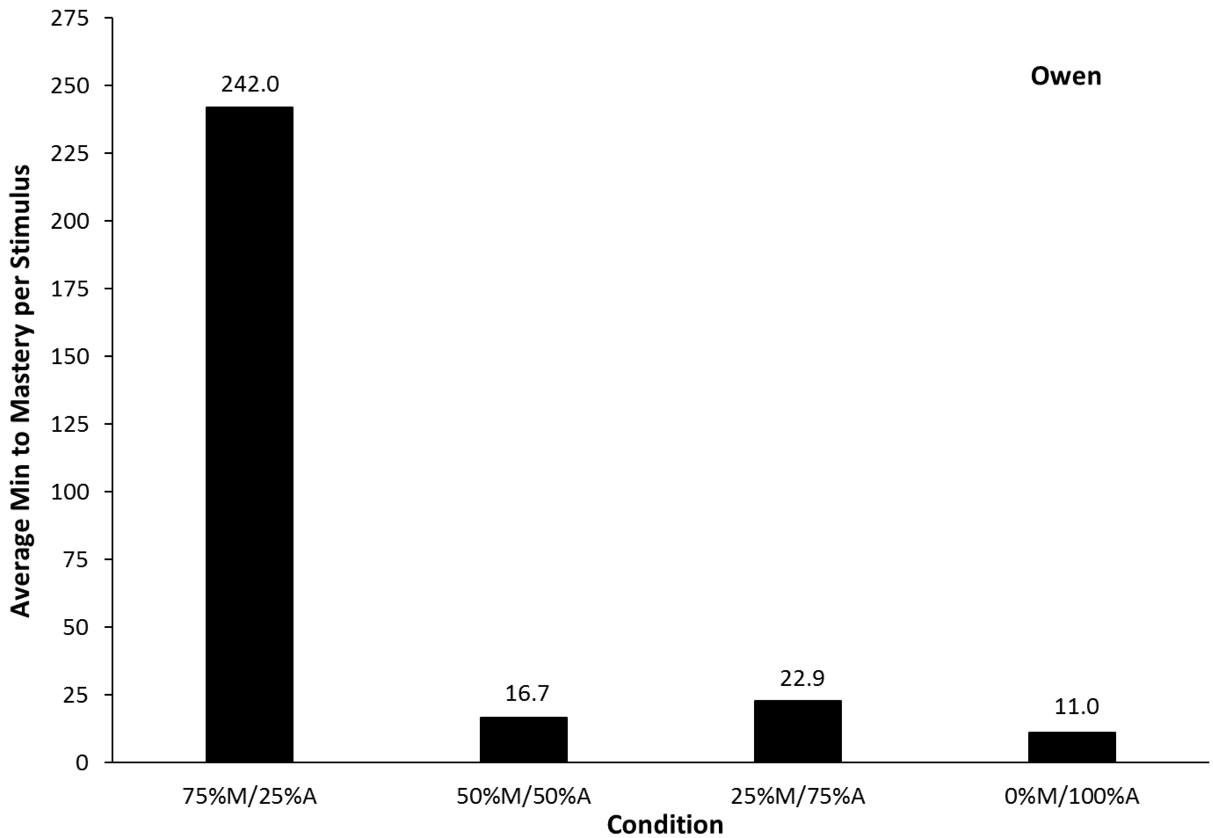

Figure 4. The average min to mastery per stimulus for Owen.

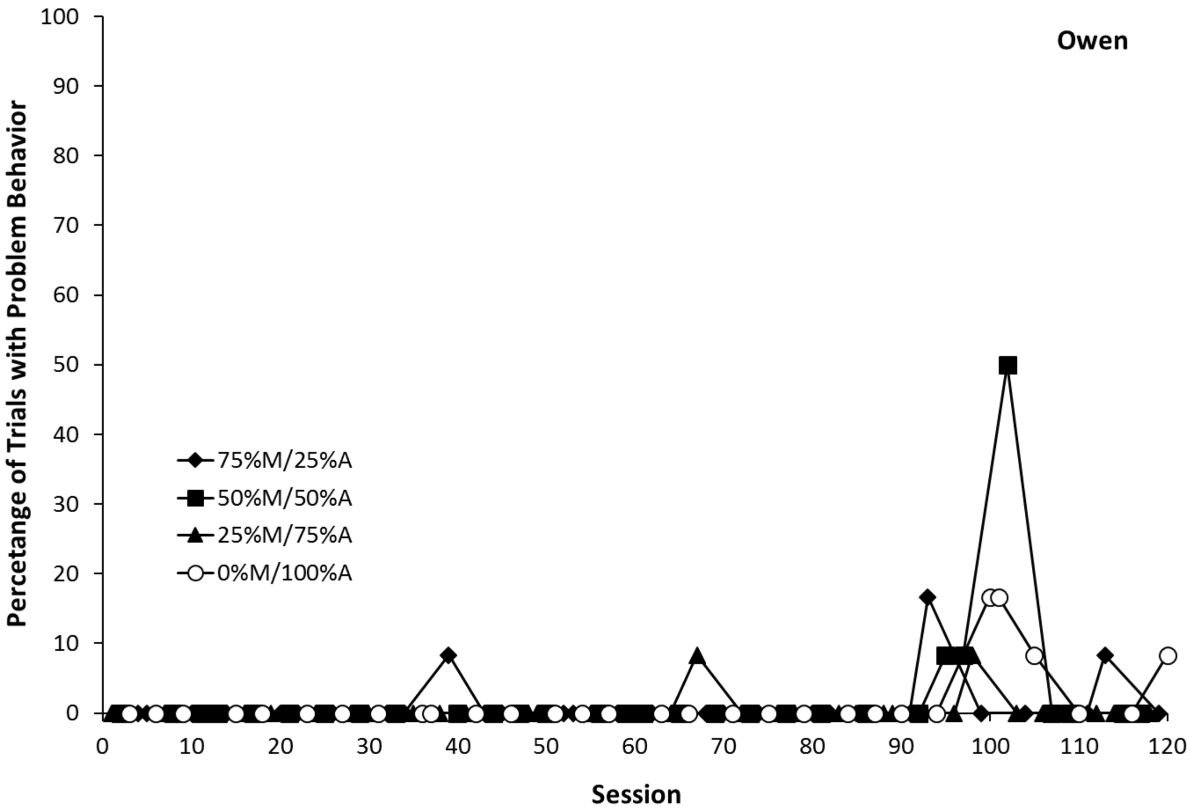

Figure 5. The percentage of trials with problem behavior across conditions for Owen.

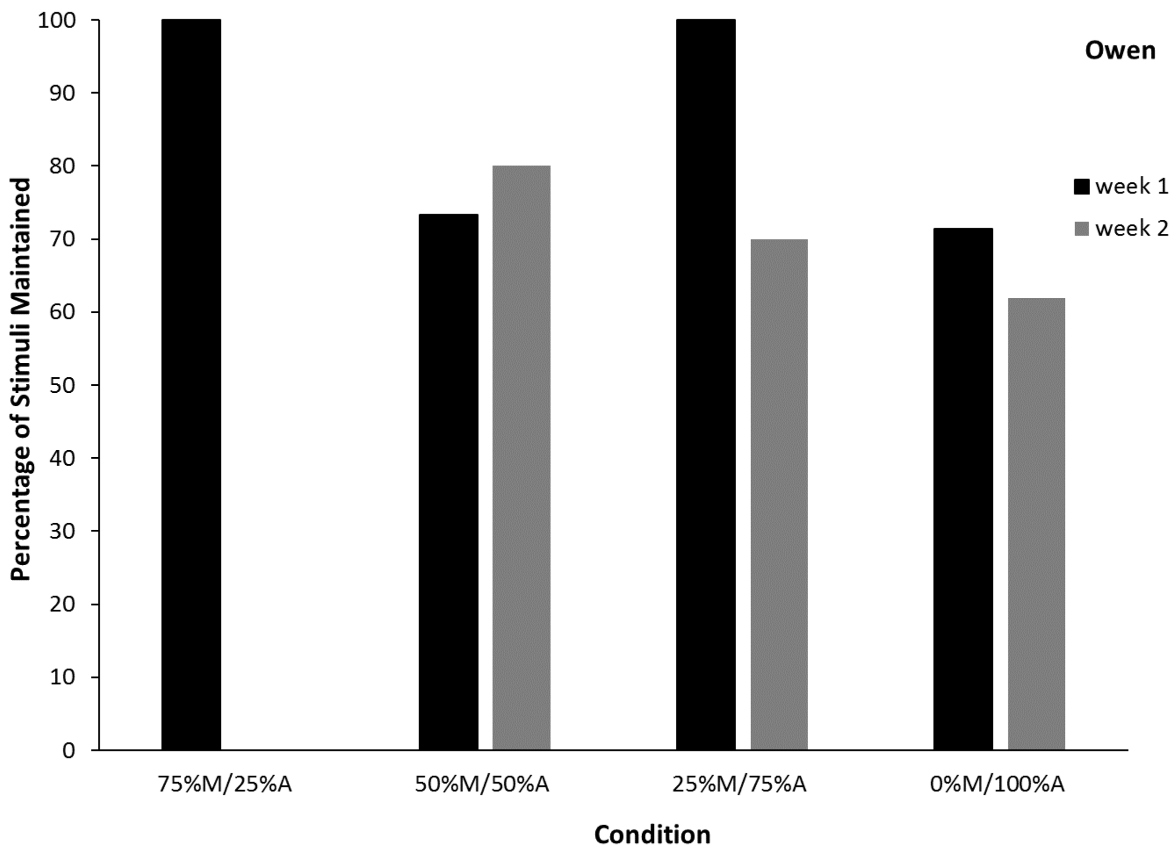

Figure 6. The percentage of stimuli maintained during one- and two-week maintenance probes across conditions for Owen.

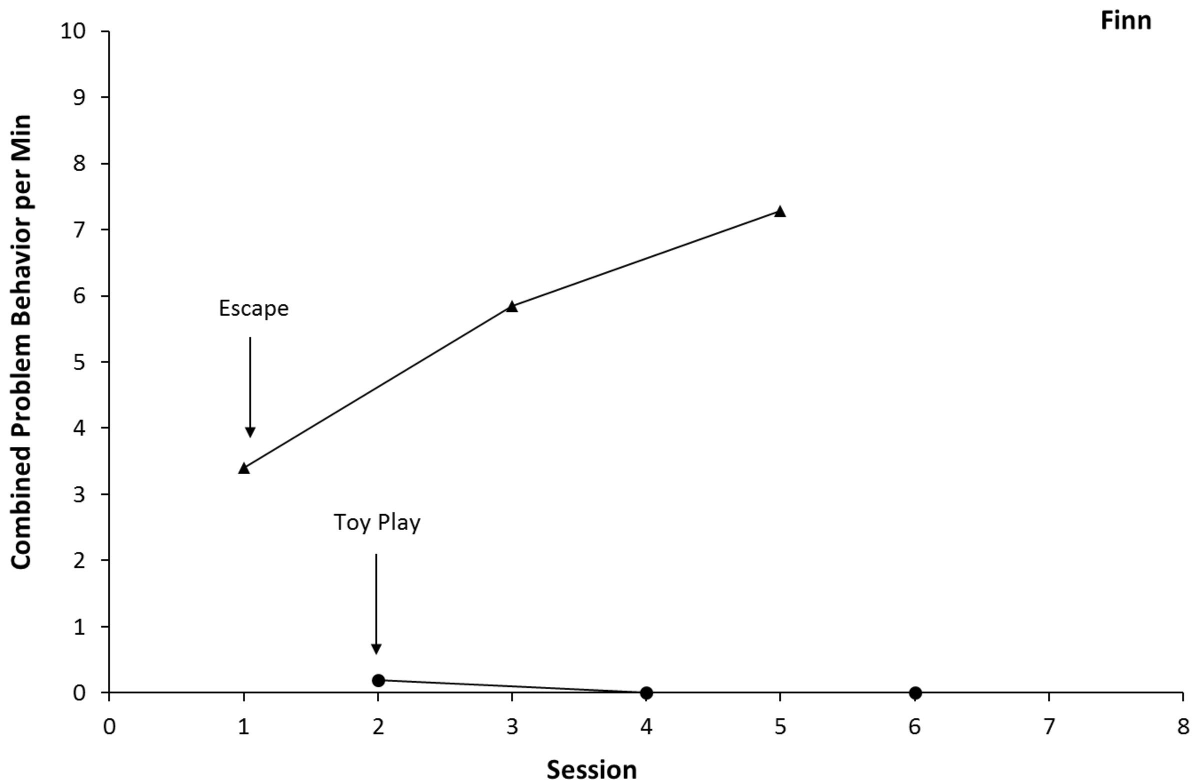

Figure 7. Combined problem behavior per min in escape (closed triangles) and toy play (closed circles) conditions of the functional analysis for Finn.

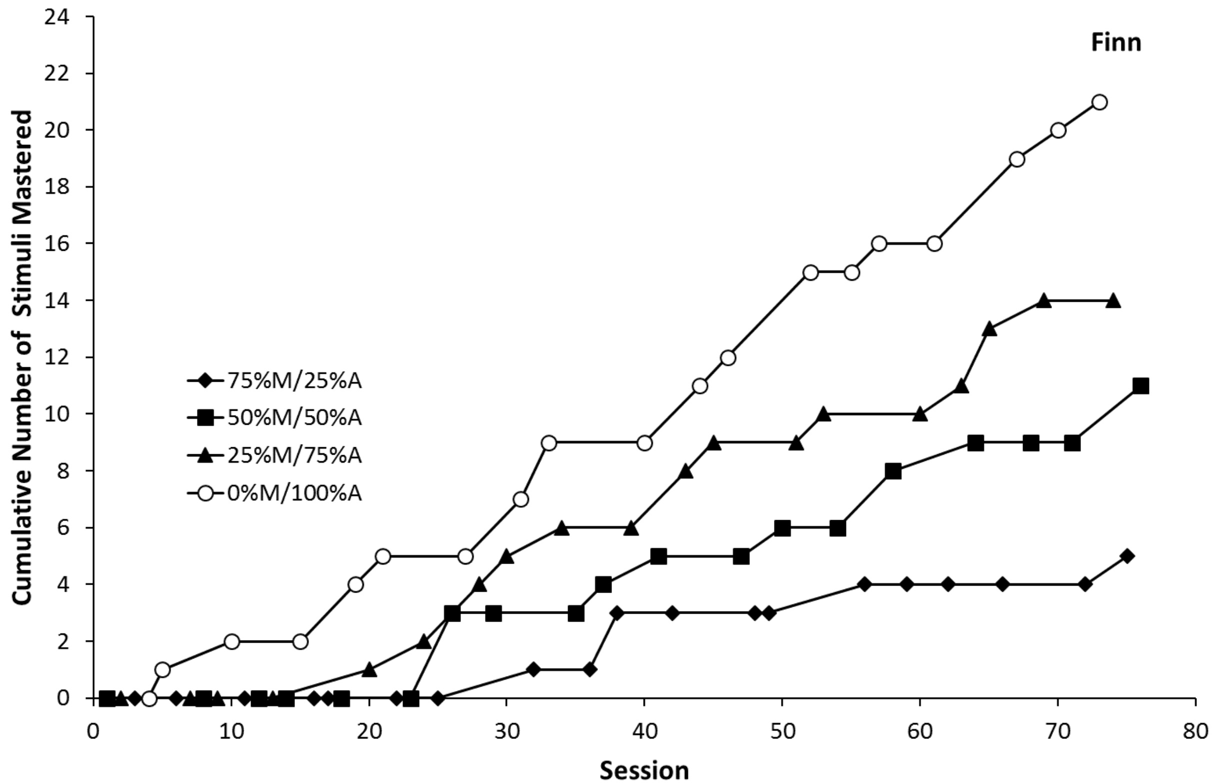

Figure 8. The cumulative number of stimuli mastered across conditions for Finn.

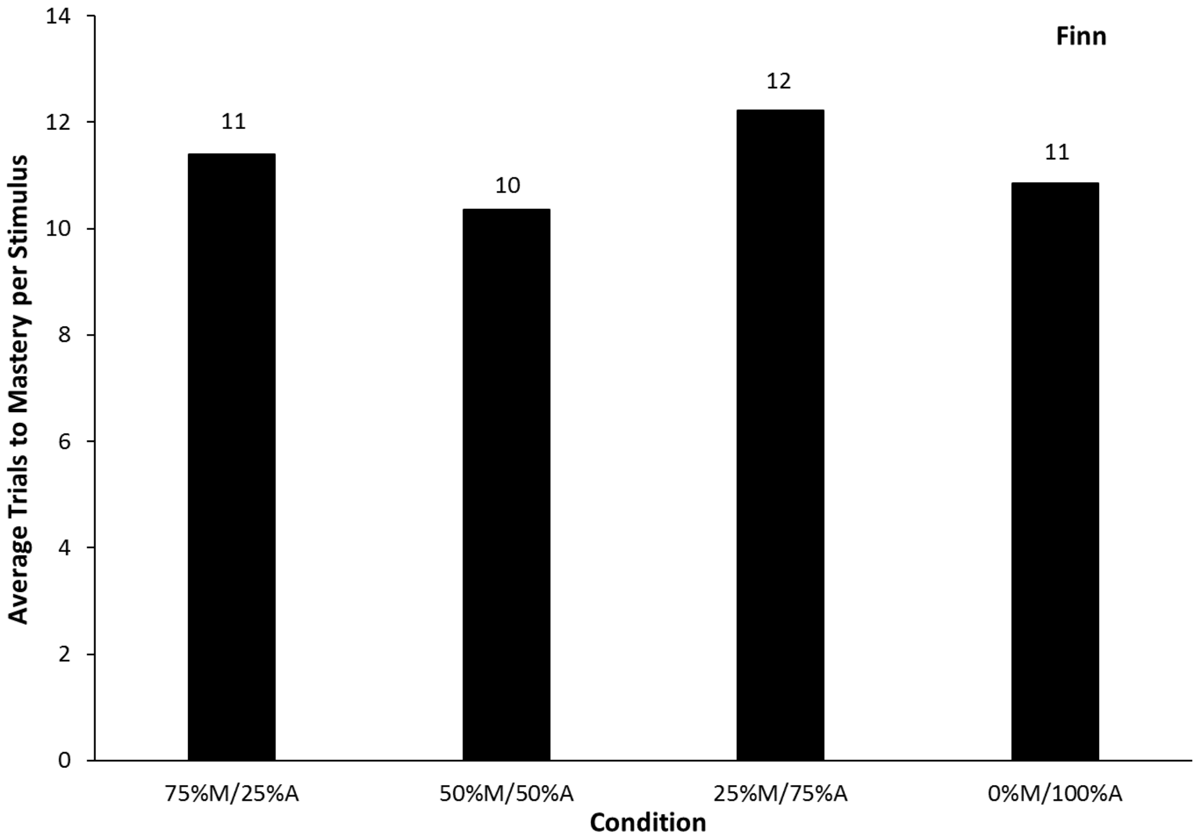

Figure 9. The average trials to mastery per stimulus across conditions for Finn.

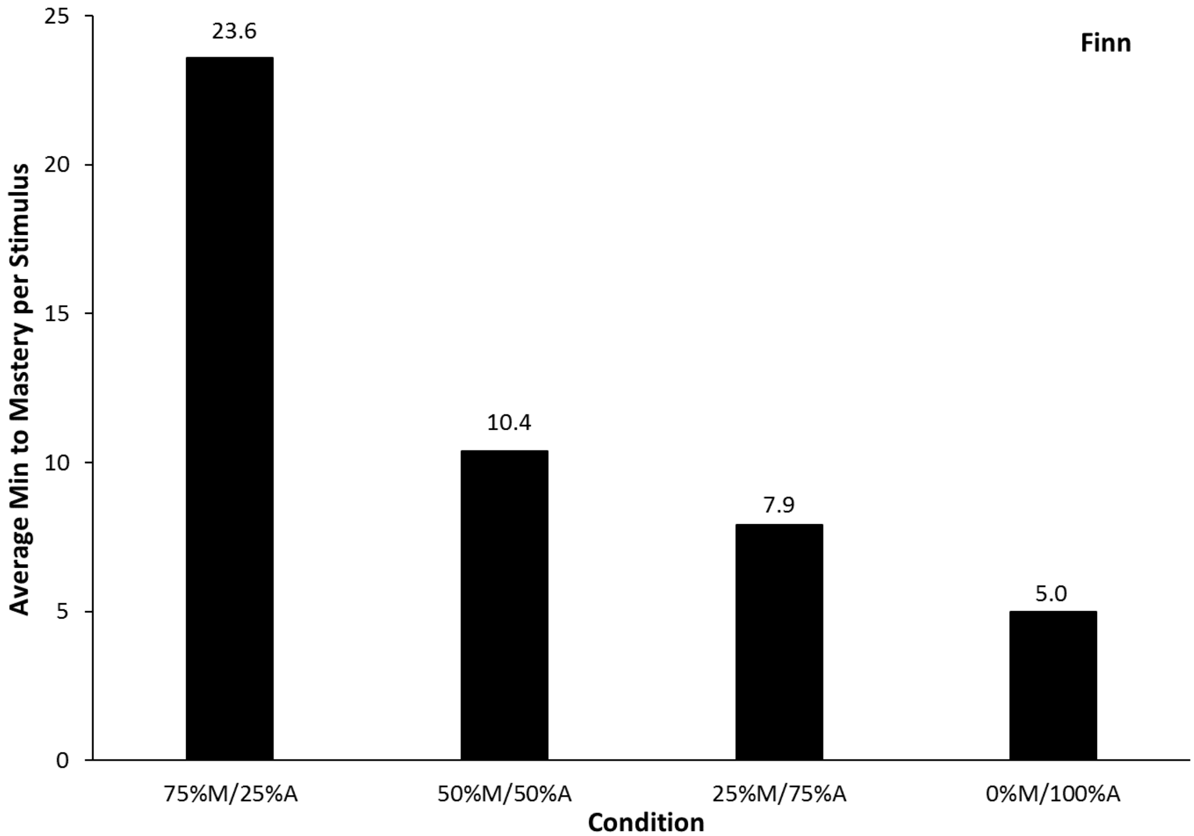

Figure 10. The average min to mastery per stimulus across conditions for Finn.

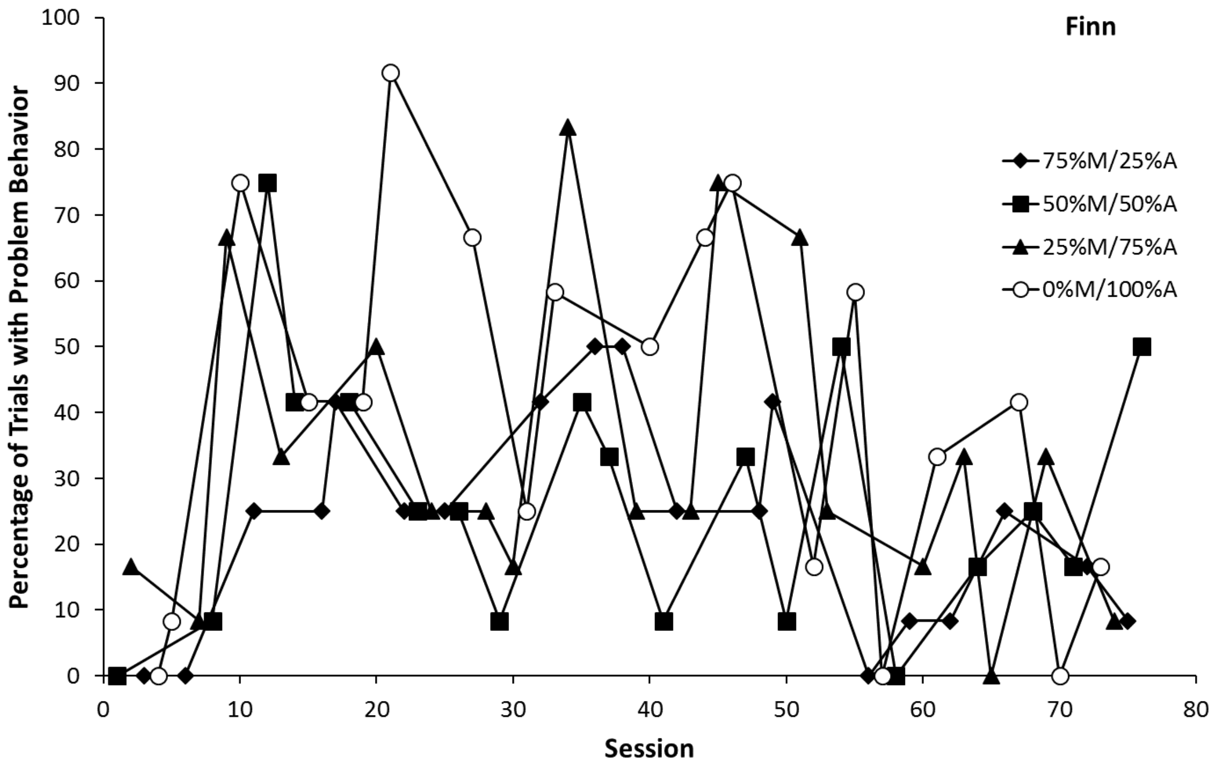

Figure 11. The percentage of trials per session with problem behavior across conditions for Finn.

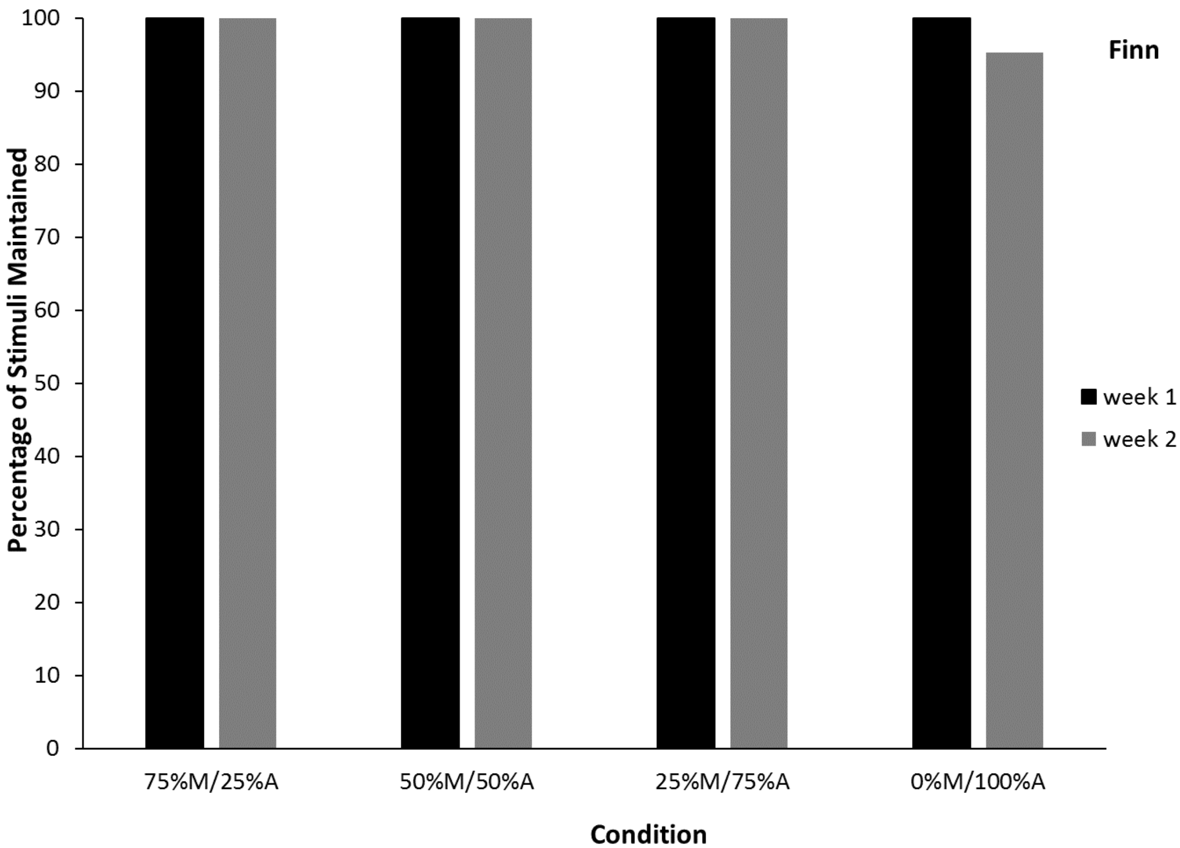

Figure 12. The percentage of stimuli maintained during one- and two-week maintenance probes across conditions for Finn.

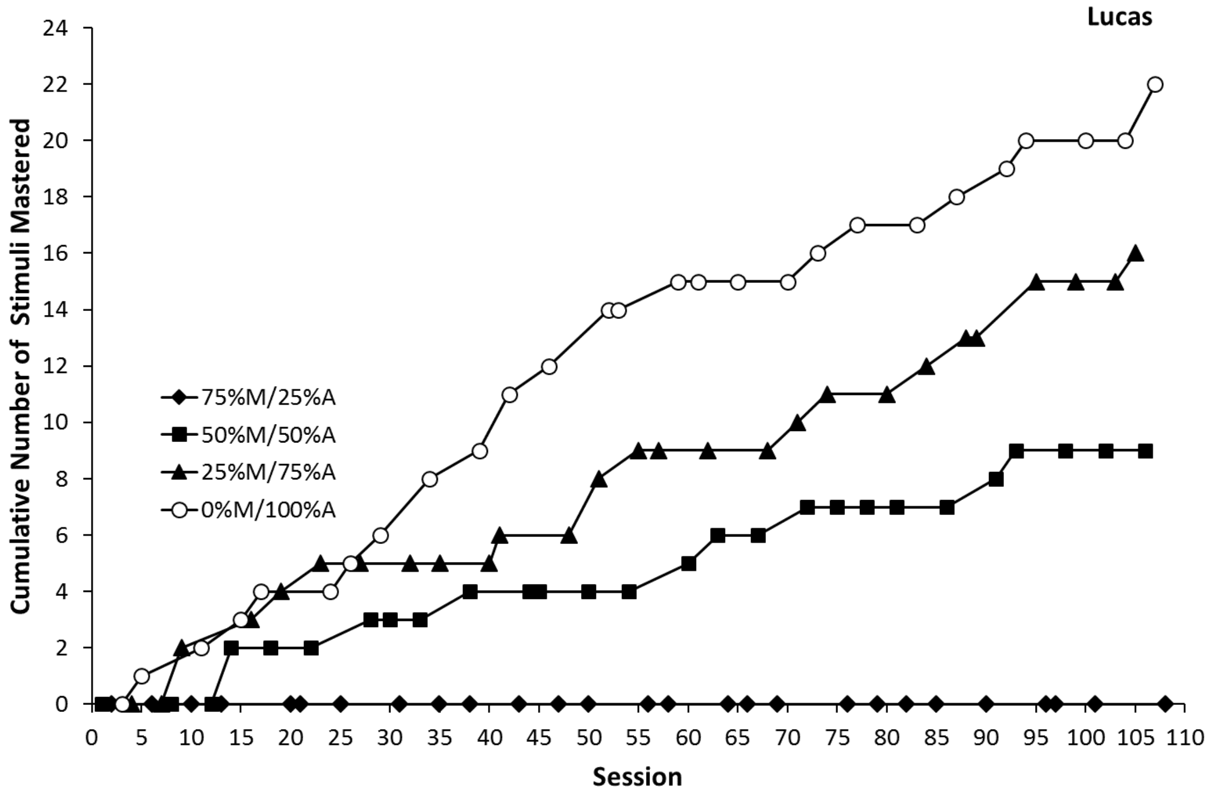

Figure 13. The cumulative number of stimuli mastered across conditions for Lucas.

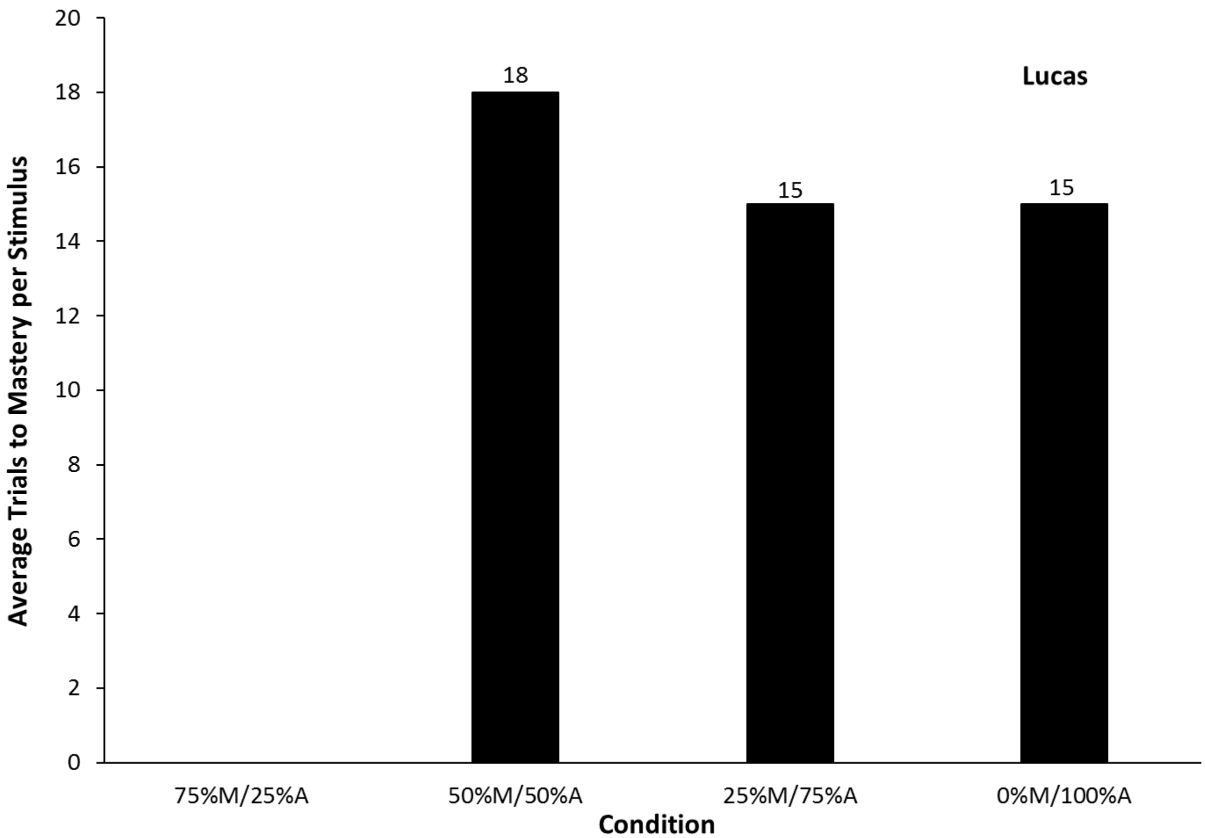

Figure 14. The average number of trials to mastery per stimulus for Lucas.

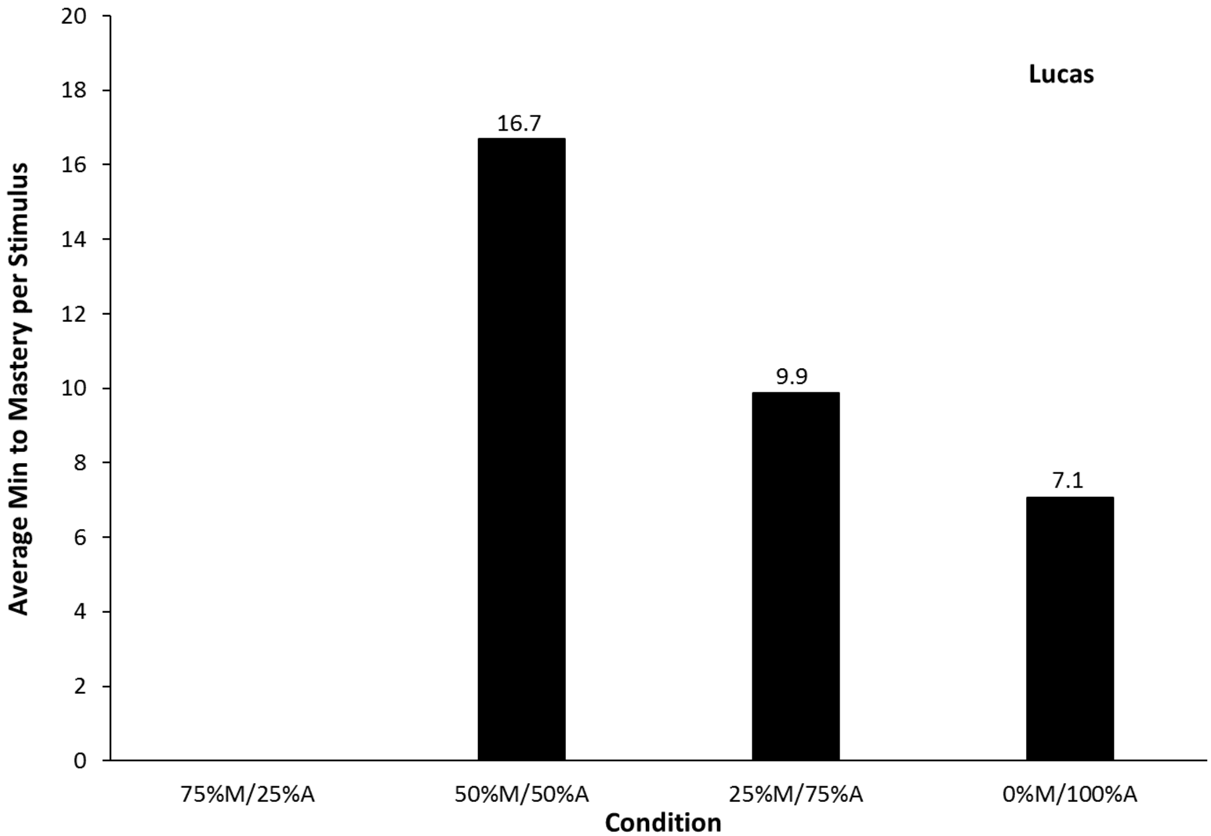

Figure 15. The average min to mastery per stimulus for Lucas.

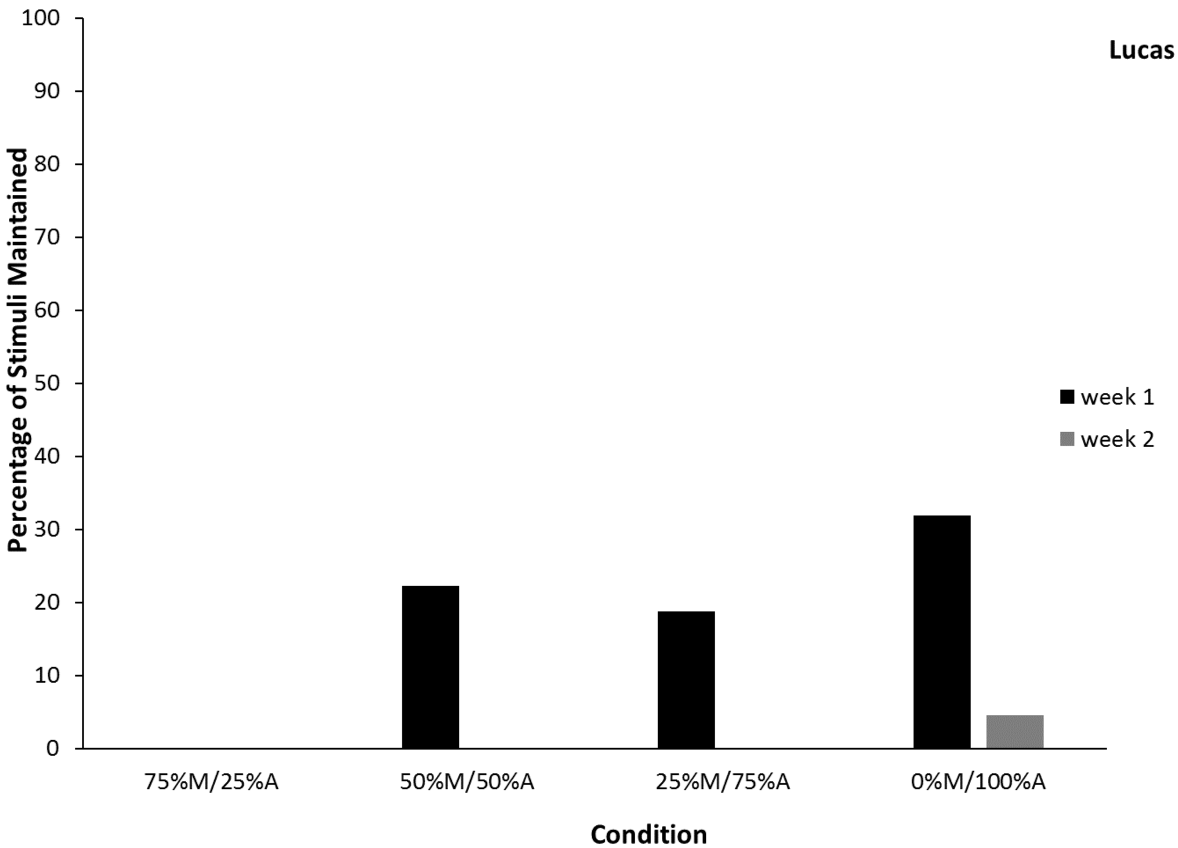

Figure 16. The percentage of stimuli maintained during one- and two-week maintenance probes across conditions for Lucas.

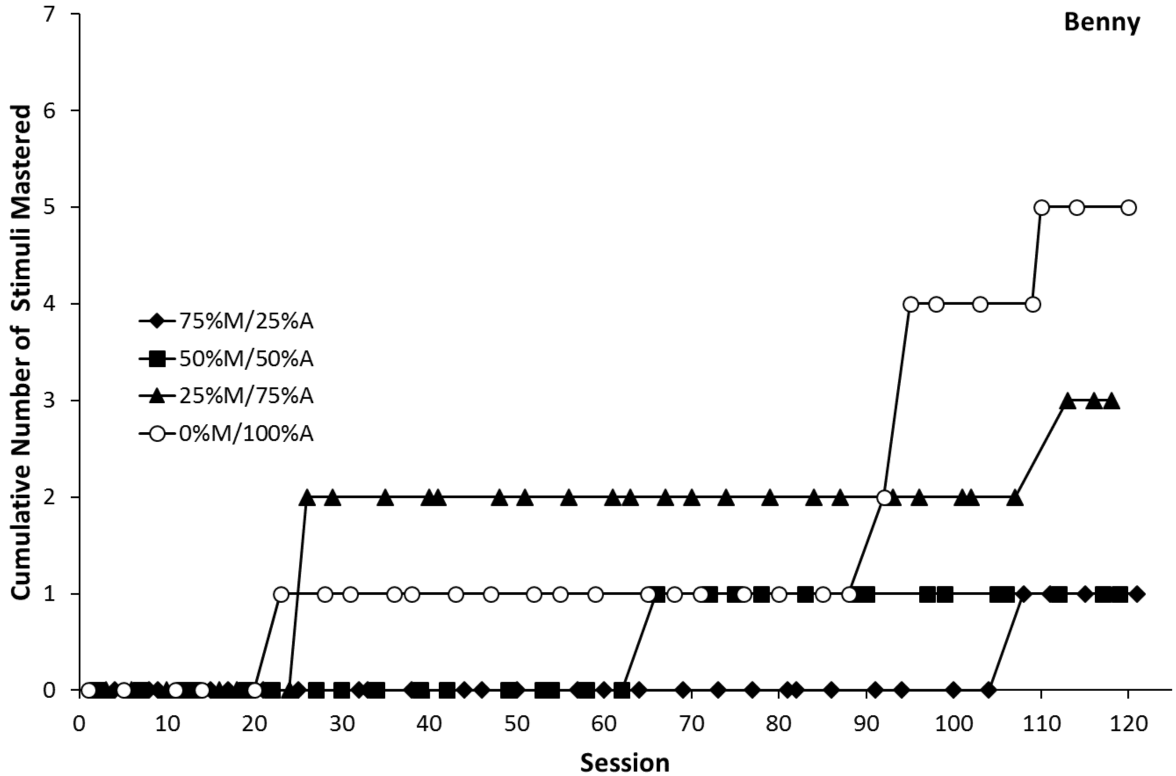

Figure 17. The cumulative number of stimuli mastered across conditions for Benny.

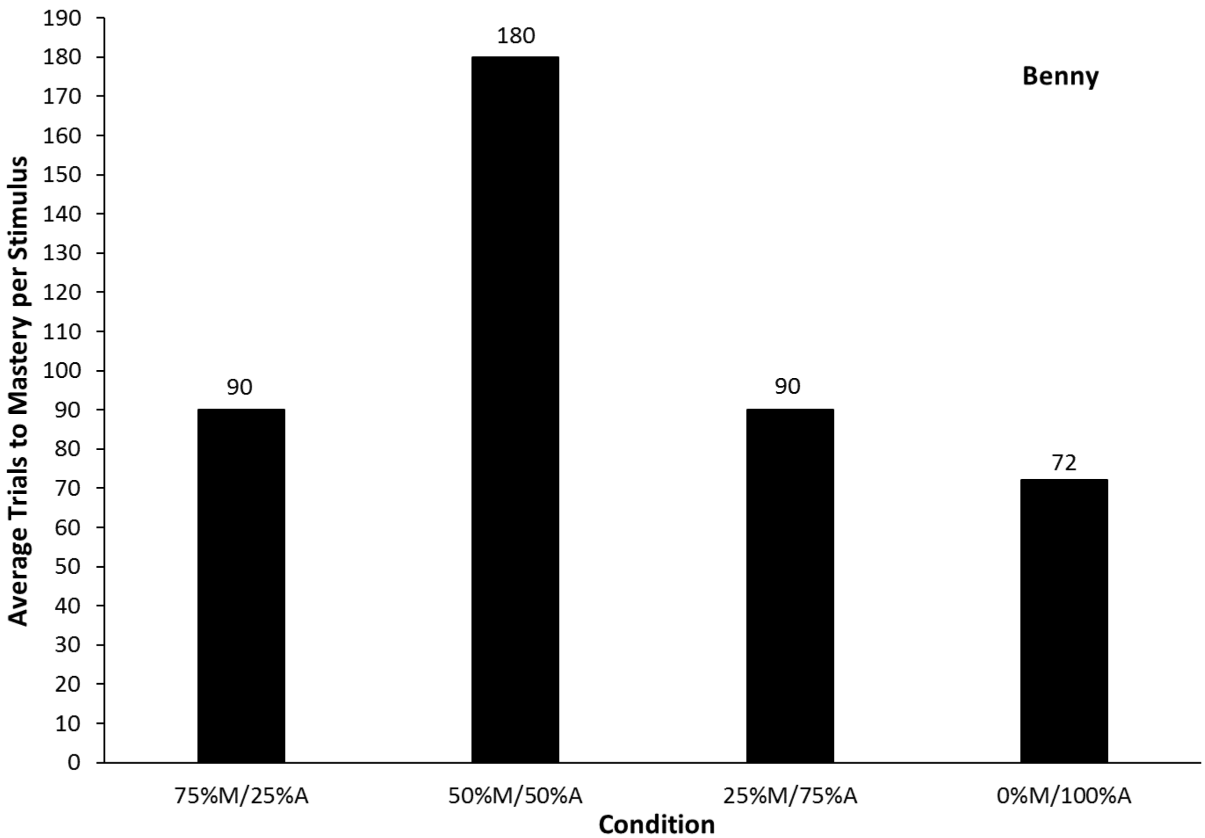

Figure 18. The average trials to mastery per stimulus for Benny.

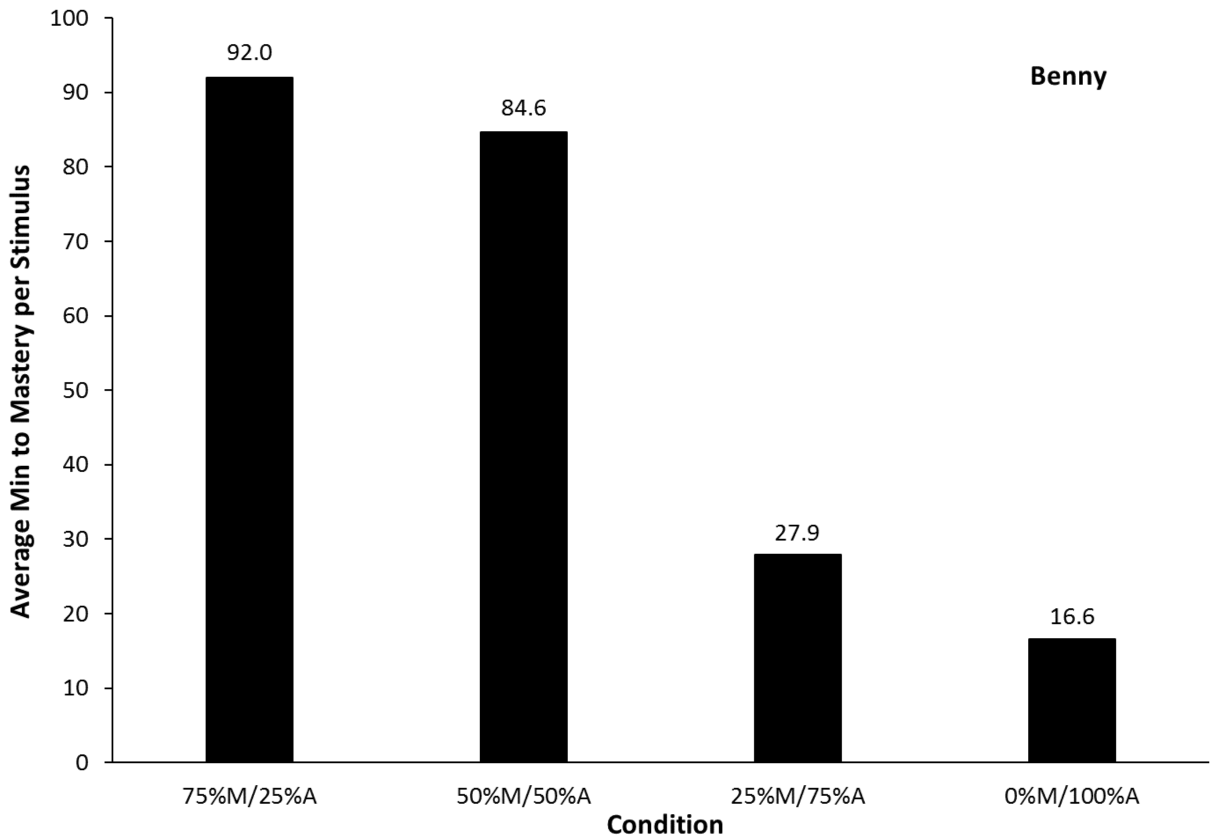

Figure 19. The average min to mastery per stimulus for Benny.

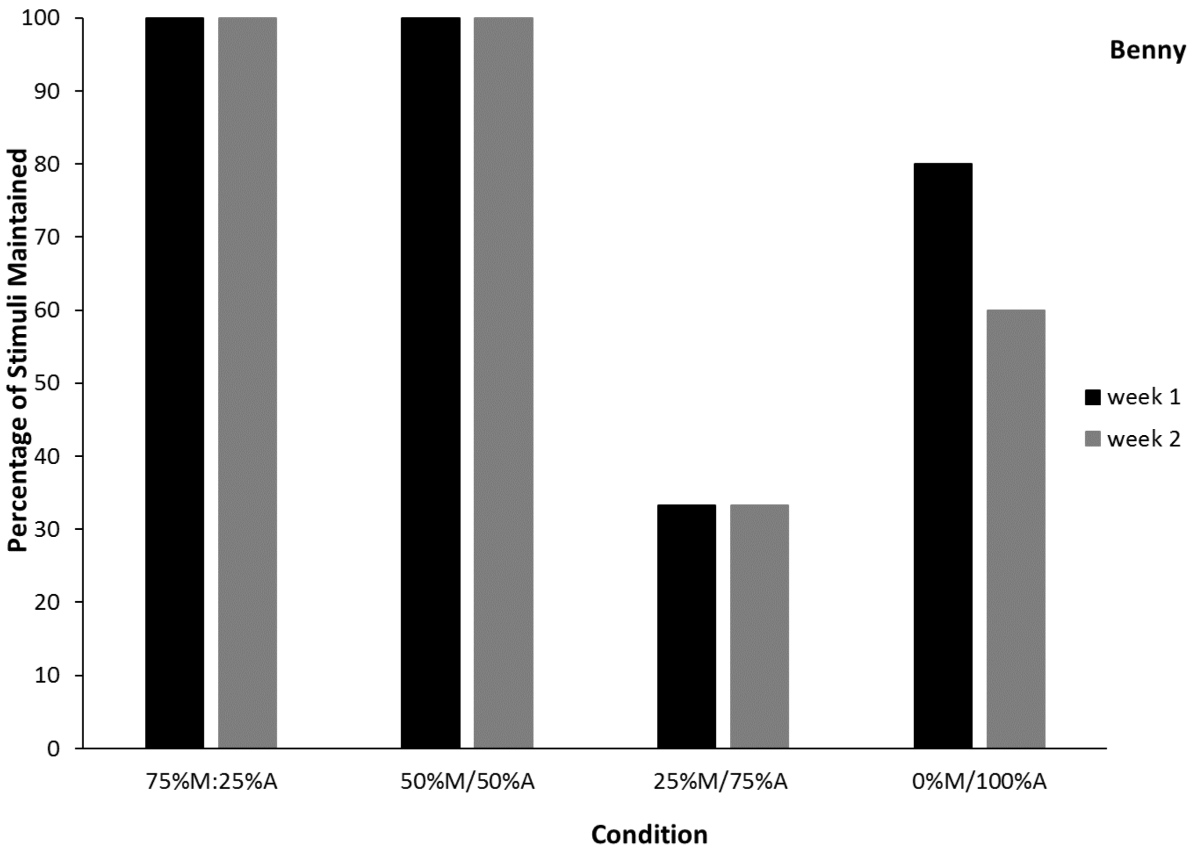

Figure 20. The percentage of stimuli maintained during one- and two-week maintenance probes across conditions for Benny.

REFERENCES

Allan, A. C., Vladescu, J. C., Kisamore, A. N., Reeve, S. A., & Sidener, T. M. (2015). Evaluating the emergence of reverse intraverbals in children with autism. The Analysis of Verbal Behavior, 31, 59-75. doi: 10.1007/s40616-014-0025-8

Benavides, C.A., & Poulson, C. L. (2009). Task interspersal and performance of matching tasks by preschoolers with autism. *Research in Autism Spectrum Disorders, 3*(3), 619-629. doi: 10.1016/j.rasd.2008.12.001

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*(3), 353-357. doi: 10.1901/jaba.2000.33-353

Cates, G. L. (2005). A review of the effects of interspersing procedures on the stages of academic skill development. *Journal of Behavioral Education, 14,* 305-325. doi: 10.1007/s10864-005-8652-8

Charlop, M. H., Kurtz, P. F., & Milstein, J. P. (1992). Too much reinforcement, too little behavior: assessing task interspersal procedures in conjunction with different reinforcement schedules with autistic children. *Journal of Applied Behavior Analysis, 25,* 4, 795-808.

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

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis*, *13*(4), 619-627.

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*(1), 36-50. doi: 10.1002/bin.1396

Horner, R. H., Day, H. M., Sprague, J. R., O'Brien, M., & Tuesday Heathfield, L. (1991). Interspersed requests: a nonaversive procedure for reducing aggression and self-injury during instruction. *Journal of Applied Behavior Analysis, 24*(2), 265-278. doi: 10.1901/jaba.1991.24-265

Iwata, B. A., Dorsey, M., Slifer, K., Bauman, K., & Richman, G. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27,* 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities*, *2*, 3-20, 1982).

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 preposition to children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, *49,* 532-547.

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*(4), 425-430.

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*(3), 657-662. doi: 10.1002/jaba.149

Nicholson, C. (2013). An analysis of variables affecting task interspersal among children with autism. *ProQuest Dissertations and Theses*.

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*(1), 153-158.

Rapp, J. T., & Gunby, K. (2016). Task interspersal for individuals with autism and other neurodevelopmental disorders. *Journal of Applied Behavior Analysis, 49*(3), 1-5. doi: 10.1002/jaba.319

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*(2), 86-92. doi: 10.1177/108835760101600204

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*(3), 335-350. doi: 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:10.1002/jaba.126

Appendix A

ACVD Acquisition Targets for Owen

75%M/25%A	**50%M/50%A**	**25%M/75%A**	**0%M/100%A**
Flute *	Clock *	Brown *	Hand *
Red	Hose *	Mop	Pink *
Blender	Pants *	Ring *	Robe *
Ambulance	Tape *	Saw *	Blanket *
	Canoe *	Backpack *	Circle *
	Eagle *	Ferret *	Iron *
	Ostrich *	Life Vest *	Magnet *
	Rain Coat *	Saxophone *	Parrot *
	Clarinet *	Triangle *	Grasshopper *
	Pentagon *	Skillet	Piano *
	Rolling Pin *	Drill *	Scorpion *
	Spatula *	Remote *	Teddy Bear *
	Purple *	Window *	Mosquito *
	Sting Ray		Rectangle *
	Van *		Wheel Barrel *
	Wrench		Thimble *
	Shopping Cart		Pot
	Clip Board *		Motorcycle *
			Dump Truck *
			Screwdriver *
			Light Switch *
			Square *
			Mug

* targets mastered by Owen

Appendix B

AVCD Mastered Targets for Owen

75%M/25%A	**50%M/50%A**	**25%M/75%A**
Apple	Cat	Dog
Grapes	Cookies	Banana
Horse	Pig	Snake

Appendix C

Tact Acquisition Targets for Finn

75%M/25%A	**50%M/50%A**	**25%M/75%A**	**0%M/100%A**
Cup *	Bee *	Blimp *	Armadillo *
Buffalo	Clown *	Boots *	Bread *
Frog *	Desk *	Clip Board	Brush *
Iron	Dust Pan *	Crown *	Couch *
Moose *	Grill	Deer *	Eggplant *
Panda *	Ladybug *	Dress	Fridge *
Parrot	Mixer *	Feather *	Gloves *
Toilet Paper	Mouse *	Gorilla *	Hamster *
Turtle *	Owl *	Grapes *	Harp *
	Ring *	Necklace *	Jet Ski *
	Scooter *	Pear *	Kangaroo *
	Screwdriver	Penguin *	Lamp *
	Worm *	Raccoon *	Mailbox *
		Roller Blades *	Ostrich
		Sink *	Pen *
		Tie *	Pliers *
		Watermelon	Rhino *
			Rooster
			Seal *
			Stove *
			Sunglasses *
			Swan *
			Tomato *

* targets mastered by Finn

Appendix D

Tact Mastered Targets for Finn

75%M/25%A	**50%M/50%A**	**25%M/75%A**
Banana	Cookie	Fork
Chicken	Lion	Lemon
Shirt	Spoon	Trumpet

Appendix E

Adjective-Noun Tact Acquisition Targets for Lucas

75%M/25%A	**50%M/50%A**	**25%M/75%A**	**0%M/100%A**
Blue Sky	Big Whale *	Brown Bear *	Black Shoes *
Long Neck	Cold Ice *	Checkered Flag	Bumpy Ball *
Smelly Skunk	Dirty Pants *	Crispy Chicken *	Chewy Candy
Straight Line	Flat Tire	Curly Hair *	Crunchy Chips *
	Green Grass *	Dry Desert	Dark Night *
	Pink Pig *	Empty Cup *	Fast Cheetah
	Pretty Flower	Fat Hippo *	Flaky Biscuit *
	Round Tomato	Full Moon *	Fluffy Dog *
	Salty Pretzels *	Knotted Rope *	Gray Shark *
	Scaly Fish	Polished Nails *	Happy Boy *
	Soft Pillow *	Red Marker *	Hard Hat *
	Spiky Cactus	Rotten Apple *	New Toy
	Striped Zebra *	Sharp Knife *	Old Woman *
	Thick Book *	Shiny Penny	Orange Pumpkin *
	Yummy Cookies	Sleepy Lion *	Purple Grapes *
		Small Ant *	Slimy Worm *
		Ugly Witch *	Slow Turtle *
		White Shirt *	Soapy Hands *
		Yellow Lemon *	Sparkly Dress *
		Young Baby	Spicy Pepper *
			Spotted Cow *
			Sticky Gum *
			Tall Giraffe *
			Wet Sponge *
			Wrinkly Clothes *

* targets mastered by Lucas

Appendix F

Tact Mastered Targets for Lucas

75%M/25%A	**50%M/50%A**	**25%M/75%A**
Bee	Ear	Bottle
Dresser	Fridge	Glue
Heart	Sunglasses	Straw

Appendix G

Tact Feature Acquisition Targets for Benny

75%M/25%A	**50%M/50%A**	**25%M/75%A**	**0%M/100%A**
Collar (dog)	Branch (tree)	Beard (face)	Address (envelope)
Cord (charger)	Driveway (house)	Chimney (house)	Antenna (car)
Crust (pizza)	Ear (elephant)	Cover (book)	Back (chair)
Eyebrow (face)	Eyelid (eye)	Drain (sink)	Bristles (toothbrush) *
Garage Door (house)	Fingernail (hand)	Drawstring (sweatshirt)	Faucet (sink)
Headlight (car)	Handle (toothbrush)	Eraser (pencil)	Handlebars (bike)
Lashes (eye)	Hood (sweatshirt)	Hoof (horse) *	Keyboard (computer)
Leaf (flower)	Laces (shoe)	Knuckle (hand) *	Leash (dog)
Paw (elephant)	Lead (pencil) *	Leg (chair)	Mane (horse) *
Pedal (bike)	Petals (flower)	Monitor (computer) *	Pages (book) *
Seed (apple)	Seat (chair)	Pocket (sweatshirt)	Pupil (eye) *
Spokes (bike)	Spine (book)	Prongs (charger)	Roof (house)
Tongue (shoe)	Stamp (envelope)	Sole (shoe)	Sleeve (sweatshirt)
Zipper (sweatshirt) *	Steering Wheel (car)	Tire (car)	Stem (apple) *
	Toppings (pizza)	Trunk (tree)	Tusk (elephant)

* targets mastered by Benny

Appendix H

Tact Feature Mastered Targets for Benny

75%M/25%A	**50%M/50%A**	**25%M/75%A**
Mouse (computer)	Bed (bedroom)	Door Knob (door)
Tub (bathroom)	Mustache (face)	Table (kitchen)
Window (house)	Tail (horse)	TV (entertainment center)