AN EVALUATION OF SIGHT WORD INSTRUCTION PROCEDURES FOR CHILDREN WITH AUTISM: INTERSPERSED TRIAL TRAINING

By

Melissa J. Gard

Submitted to the graduate degree program in Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson: Glen White

_____

Claudia Dozier

_____

Jomella Watson-Thompson

_____

Debra Kamps

_____

Michael Wehmeyer

Date Defended: March 8th, 2012

The Dissertation Committee for Melissa J. Gard

certifies that this is the approved version of the following dissertation:

AN EVALUATION OF SIGHT WORD INSTRUCTION PROCEDURES FOR CHILDREN WITH AUTISM: INTERSPERSED TRIAL TRAINING

_____

Chairperson Glen White

Date approved: March 8, 2012

Abstract

Interspersal procedures have been used to decrease challenging behavior and facilitate skill acquisition. However, little is known about the effect of interspersing mastered tasks that are similar versus dissimilar to the target task. This study examined the effects of different teaching conditions on the acquisition, maintenance and generalization of novel sight words by four children with autism. Rates of challenging behavior and participant preference data are also reported. Phase 1 compared teaching conditions with and without interspersed trials, and the results showed that all four participants acquired more sight words when mastered tasks were interspersed during teaching. Phase 2 compared the interspersal of mastered sight words (similar) and mastered motor imitation (dissimilar). For two participants interspersal of similar tasks resulted in the acquisition of more words, for one participant results were mixed, and one participant acquired slightly more words when dissimilar tasks were interspersed. Rates of challenging behavior throughout the study were low but variable across all conditions. Higher rates of challenging behavior for some participants were observed during baseline. Participants preferred interspersal to no interspersal and interspersal of similar tasks to dissimilar tasks. Implications for the motivation and learning of children with autism are discussed.

Acknowledgements

I am deeply grateful to my advisor, Glen White, who not only provided the support and guidance necessary to complete this dissertation, but also taught and inspired me throughout my graduate experience.

I am grateful to Claudia Dozier, for her insight both on this dissertation research topic and conducting applied research in general. Her expertise in and passion for the field of ABA are remarkable, and I learned much from each interaction with her.

To the other members of my committee, Debra Kamps, Mike Wehmeyer, and Jomella Watson-Thompson, thank you for your participation and direction from the initial research proposal to the completion of this dissertation.

I would also like to thank the many others who aided me during this process, including Carol Sella, Rachel Adamek, Jeff Gordon, Eric Larsson and the staff at the Lovaas Institute Midwest.

Finally, I want to express my gratitude to my family and friends. You have been a constant source of love and support, and I am grateful for your belief in me.

Contents

List of Tables

List of Figures

An Evaluation of Sight Word Instruction Procedures for Children with Autism: Interspersed Trial Training

Autism Spectrum Disorders (ASD) is an increasingly prevalent group of developmental disorders. The prevalence of ASD as reported by the Centers for Disease Control and Prevention (CDC) is estimated to be 1 in every 110 children (this was estimated from a multi-state sample in 2006). This is a 57% increase in prevalence from the last report in 2002. Although the exact etiology is not known, there are several possible explanations for the increasing prevalence of autism, including an increase in societal awareness, greater sensitivity of diagnostic measurements, improved training of medical professionals who are providing an increasing number of diagnoses, and not least of all, a possible increase in the incidence of autism due to genetic factors, environmental factors, or a combination of the these factors (see Committee on Children with Disabilities, 2001 as an example of medical professionals' education and a discussion of factors contributing to increased prevalence). Children diagnosed with ASD (including autism, pervasive developmental disorder not otherwise specified, and Asperger's Disorder) demonstrate delayed or atypical development in language and social skills, as well as the presence of restricted, stereotyped or repetitive behavior or interests (American Psychiatric Association, 2000). These symptoms can have a significant impact on an individual's ability to learn the skills necessary to live independently and to be able to form and maintain mutual and meaningful social relationships.

The symptoms of autism may also contribute to a pattern of difficulty learning from instructional situations to such a degree that the individual will actively avoid those opportunities. Therefore, whether the goal of treatment is the remediation of the symptoms of autism or to enhance a specific skill set (e.g., provide functional life skills), the effective and efficient teaching of new skills is imperative. Interspersal procedures represent one strategy that may both decrease the avoidance of and increase the efficacy of instructional situations. The

field of applied behavior analysis has a rich repository of other procedures using behavioral principles such as reinforcement (Baer & Sherman, 1964; Baer, Peterson, & Sherman, 1967, Durand & Carr, 1991; Glover, Roane, Kadey, & Grow, 2008; Ingvarsson, Kahng, & Hausman, 2008) and extinction (Richman, Wacker, Asmus, & Casey, 1998), which have been used to increase specific skill sets or decrease targeted challenging behaviors of children with autism (for review see Matson, Benavidez, Compton, Paclawskyi & Baglio, 1996).

This paper will review the literature on interspersal procedures across populations and dependent variables, evaluate how interspersal procedures can impact learning in general, and more specifically the learning of children with behavioral, intellectual, or developmental disabilities. First, this paper will provide a brief historical overview of, with a brief exploration of the behavioral studies that informed the development of interspersal procedures, followed by a review of early seminal studies on interspersal procedures. Next, the studies that implemented interspersal procedures with general education students will be reviewed. This will be followed by the studies conducted with atypical populations, first those that explored the use of interspersal procedures to decrease challenging behavior, then those that evaluate the procedure to facilitate skill acquisition. Finally, articles that evaluate other key variables related to interspersal will be examined. Subsequently, the methodology and results of the investigation on the effects of interspersal in sight word acquisition by four children with autism will be presented, ending with a discussion of the findings as they relate to future research and application.

In order to identify the relevant literature, a search was conducted using PubMed with the search for ‘in all fields’ using the keywords ‘behavior analysis’ and ‘interspersal,’ which yielded no results. Another search was conducted, again selecting the ‘in all fields’ option but with only the keyword ‘interspersal,’ which yielded 13 results. Five of these were relevant to the

behavioral procedure of interspersal, met the additional criteria of being an empirical study that contributes to the understanding of interspersal procedures, and was conducted with a population of children or adolescents of any ability or individuals of any age with disabilities. A study was considered to contribute to the understanding of interspersal procedures if interspersal was a defined independent variable, and the study investigated the impact of interspersal by either comparing this procedure to another, or by conducting a systematic manipulation of variables within the interspersal procedure (such as drill ratio or rates of reinforcement).

Behavioral momentum, or the high probability to low probability sequence, is occasionally grouped with interspersal procedures, and may function due to the same behavioral process; however, these conclusions are tentative and warrant a closer examination of each procedure individually. Therefore, this paper will draw parallels to the high p to low p sequence, but these were not included in the interspersal literature review. The impact of interspersal procedures with college students and other typically developing adults is outside of the scope of the current study (conducted with children with autism) and were therefore excluded.

There are various teaching packages derived from interspersal procedures in the literature including incremental rehearsal (Burns, 2005; Volpe, Mulé, Briesch, Joseph, & Burns, 2011) and a drill sandwich (MacQuarrie et al. 2002). Because the influence of the other components of the teaching packages may obscure the effect of interspersal, for the purposes of this review, studies that demonstrated the effect of the above teaching packages, but that did NOT examine other variables relevant to interspersal were not included.

Due to the limited results found through the PubMed database, another search was conducted using PsychInfo. Using the advanced search function, the option ‘anywhere’ was selected to search using the keywords ‘applied behavior analysis,’ and ‘interspersal’ or ‘interspersed.’ Results were limited to English language and only peer-reviewed journals. This

yielded 185 results. Thirty-six of these met the additional inclusion criteria listed above. Finally, the references of the studies identified through the database literature searches above were inspected until all relevant studies that met the identified parameters were included.

Interspersal procedures (which will also be referred to as "interspersal" in this study) refers to the instructional practice of presenting easy (mastered) tasks within an instructional sequence that includes more challenging (acquisition) tasks. In contrast, mass-trial or constant-task conditions target one task during the course of an instructional sequence (often prompting and fading prompts for that task during the instructional sequence but only presenting one task). Interspersal has also been referred to as distributed practice, task variation, varied-task condition, and interspersed tasks. 'Interspersal' will refer to the inclusion of maintenance tasks within an instructional sequence that targets one or more acquisition tasks. This paper will use the term 'no interspersal' when one or more acquisition task, but no maintenance tasks, are presented during an instructional sequence. Massed trial training is another common term, but this implies only one acquisition task. The studies included in this review used a no-interspersal condition that included either one acquisition task or multiple acquisition tasks. Incremental rehearsal is a derivation of interspersal procedures and uses a systematic increase in the number of maintenance trials between trials of acquisition trials. For example, when M indicates a maintenance task and T indicates a target or acquisition task, the sequence of trials would be as follows: T1 M1 T1 M1 M2 T1 M1 M2 M3 T1 M1 M2 M3 T1 M1…..M9 T1. Incremental rehearsal most commonly uses a ratio of 90% maintenance tasks to 10% acquisition tasks and it also often includes a procedural embedding of the acquisition task into the sequence as a maintenance trial once it has been mastered. The term "drill sandwich" refers to an interspersal procedure that uses a ratio of 50% maintenance tasks to 50% acquisition tasks, presented in a

sequence of three maintenance tasks alternating with three acquisition tasks. It also commonly embeds the newly mastered skill as a maintenance task in subsequent trials.

Task Variation

Previous studies have compared a constant-trial condition, in which one acquisition skill was targeted, with a varied-task condition, in which more than one acquisition skill was targeted. These studies demonstrated improved responding in the varied-task condition (Dunlap & Koegel 1980; Weber & Thorpe, 1989). Dunlap and Koegel (1980) compared a constant-task condition, in which one acquisition skill was targeted, and a varied-task condition, in which multiple acquisition skills were targeted. They measured the percentage of correct responses and the affect of two young children with autism across the two conditions. Reinforcement and prompting procedures were used consistently across both conditions, and the types of tasks were not systematically different across conditions. All tasks were taken from the child's curriculum and spanned a variety of skill areas such as counting, color identification, vocal imitation, and matching. The data for the percentage of correct responses to the tasks indicated a declining trend during the constant-task condition and an immediate increase followed by a continued increasing trend upon the introduction of the varied-task condition. Ratings of child affect, which were scored via video by naïve observers for one participant, closely mimicked the percentage of correct responding data. The scores indicated declining ratings for enthusiasm, interest, happiness, general behavior, during the constant-task condition and increases to positive and stable ratings on the same measures during the varied-task condition. The authors note as a limitation that high levels of correct responding during the early trials of the study may suggest that the tasks were not completely unknown to the participants. They further suggest that if some responses were already known, the findings of the study may be more relevant to the impact of task variation on the maintenance of behaviors and a child's motivation to respond to tasks.

Inclusion of a formal assessment to empirically identify unknown tasks and the inclusion of acquisition data in addition to data on the percentage of correct responses would have helped to clarify the role of the varied-task condition in the process of response acquisition. Although some experimental control was demonstrated through the use of the multiple baseline design, this design did not control for potential order effect, as the constant-task condition was always presented first and followed by the varied task condition. Nonetheless, Dunlap and Koegel (1980) did demonstrate that task variation increased the percentage of correct responses and yielded more positive ratings of child affect, which combined, may be measures of child motivation.

Early Interspersal Studies

Expanding the concept of task variation, researchers began to investigate the effects of varying the tasks and interspersing maintenance tasks during teaching. These studies were among the first to evaluate the effects of interspersal procedures. Some key investigations include Neef, Iwata, and Page (1977, 1980), Dunlap (1984), Rowan and Pear (1985), Koegel and Koegel (1986), and Winterling, Dunlap, and O'Neill (1987) (see Appendix A for a table summarizing these studies).

Neef, Iwata, and Page (1980) demonstrated that for three men with developmental disabilities, a teaching condition that included an interspersal procedure was more effective in teaching spelling words than two other conditions that did not include interspersal, even when one of the non-interspersal conditions included a richer rate of reinforcement. The dependent variables were the number of spelling words mastered in each session and the number of spelling words retained during a maintenance probe 10 days after the spelling words were mastered and 10 to 14 days after the study completion. This study extended the literature in several ways: use of the alternating treatments design with a baseline enhanced experimental control by allowing a

comparison of acquisition and maintenance rates from no interspersal training with praise for correct responding to both the no interspersal with a higher density of reinforcement condition and the interspersal condition. Additionally skill acquisition was the primary dependent variable, a pretest provided empirical verification of both known and unknown words before their inclusion in the study, and a measure to assess participant preference was included. Preference was assessed by allowing the participant to choose the word list, which was either associated with the interspersal condition or the high-density reinforcement condition. One participant selected the interspersal list exclusively, and the other two participants selected the interspersal list on three of four opportunities, suggesting that all participants preferred the interspersal condition to the high-density reinforcement condition. Although the choice of a treatment is a valid measure of preference, conclusions here are limited by the infrequent use of the preference assessment (participants were given four opportunities to choose, but 35-70 sessions were conducted during the study). Valid preference assessments require that the participant be able to easily distinguish between the conditions, and can be facilitated by assigning distinct characteristics to each condition (i.e., dissimilar materials, color code, different assigned instructors). More information regarding if and how participants in this study were likely discriminating between the two conditions would increase the validity of the preference assessment.

The results suggest that the interspersal condition facilitated maintenance of the words learned. However, because the words mastered in the study were then included as the interspersed items, it is unclear if the higher percentage of correct responses during the retention tests were due to the use of the interspersal procedure during teaching, or due to the additional systematic practice of those words as they were interspersed after mastery.

Dunlap (1984) investigated the role of interspersal on the rate of acquisition of new skills for five children with autism, ages four to seven years old. The number of trials to a pre-determined mastery criterion was measured across three conditions: constant-task (mass-trial), varied-tasks with only acquisition tasks (no interspersal), and interspersing acquisition and maintenance tasks (interspersal). Similar to the tasks taught in Dunlap and Koegel (1980), the acquisition tasks were varied unknown tasks appropriate to the child's general level of function, and included spelling, sequencing story cards, receptive labeling, and matching, among others. The acquisition tasks and mastered tasks were verified as unknown or known through a pretest. To ensure that tasks were equivalent across conditions, acquisition tasks were matched according to stimulus characteristic (skill set) and each item of the matched set was randomly assigned to a condition. Prompting and fading was used across all conditions as a proactive teaching strategy to teach the targeted acquisition task and was re-implemented if three consecutive incorrect responses occurred. The results indicate no significant difference in number of trials to criterion between the two acquisition-only conditions (constant and varied-task), but the interspersal condition yielded the fastest acquisition. The average number of trials to criterion in the interspersal condition ranged from 21.2 to 119.6, but most children required fewer than 50 trials to meet the mastery criterion. In the constant-task condition, the average number of trials to criterion ranged from 45.6 to 177.6, but most children required more than 90 trials to meet criterion. In the varied-task condition (all acquisition tasks), the average number of trials to criterion ranged from 47.8 to 500. In addition to the data on number of trials to criterion, Dunlap (1984) also provided data on the time needed to teach the tasks, which is an important variable when evaluating the efficiency of a given procedure. The average duration in minutes to teach a task was shortest for the interspersal condition across all participants. Three of the five

participants needed the most time to learn a task in the constant-task condition and two of the five participants required the most time to learn a task in the varied-acquisition condition.

The ratings of child affect both replicated and extended the findings of Dunlap and Koegel (1980). Similar to the previous study, the varied acquisition condition had higher overall ratings of enthusiasm, interest, happiness and general behavior than did the constant-task condition. The inclusion of the interspersal condition, in additional to the constant and varied-task conditions, allows for a further analysis and this condition produced more positive scores for child affect and on-task responding than the other two conditions, which supports the use of interspersal to enhance child motivation.

This study extended the literature on interspersal to another population, children with intellectual disabilities, and by measuring acquisition using the number of trials to criterion, comparing three treatment conditions, including measures of child affect with interspersal procedures, and including measures of relative treatment efficiency. There were also some limitations of this investigation. Although the patterns of responding were consistent across participants, the amount of data per condition was relatively sparse, as it was based on the acquisition of only five tasks per condition. Additionally, no maintenance or generalization data were reported. Finally, the mastered tasks that were interspersed with each acquisition task were not reported, so the reader is unable to determine if these skills were or were not significantly similar to the acquisition task nor is it known if this would have impacted the results.

The use of interspersal procedures to facilitate acquisition is also supported by the results of Rowan and Pear (1985). It demonstrated the relative efficacy of interspersal when compared to a concurrent task (no interspersal) condition in teaching picture names to children with intellectual disabilities. An ABA with counterbalance design was used to compare a concurrent training (no interspersal) condition with an interspersal training condition. The concurrent

training procedure alternated the target unknown picture with three other unknown pictures. The interspersal training procedure alternated the target picture with three known pictures. In each condition the target was first alternated with one picture (known or unknown) for a number of trials, then the same target was alternated with the second picture (known or unknown) for a number of trials, then the sequence was repeated with the third known or unknown picture. Both conditions included prompts for correct responding and a set schedule of reinforcement. Pretests empirically confirmed that pictures were known and unknown, and an additional probe of three trials was conducted immediately before the start of teaching to ensure that the target was still unknown. During the pretest, the researchers also ensured that the child could vocally imitate the object label names, thus ensuring that the vocal model prompts that were later used in teaching were within the child's abilities. The results indicate that more words were learned in the interspersal condition for all participants. Also, there were fewer words discontinued in the interspersal condition. A word was discontinued due to lack of progress at a predetermined number of successive errors. Finally, the words taught with interspersal required significantly fewer trials to meet the mastery criterion. There was little correlation between the teaching condition and the rates of maintenance at a one week probe, or generalization across teachers or to object labeling. The total number of words learned varied per condition so maintenance data reflects the percentage of words maintained or generalized of the total words learned in that condition. Although slight, the concurrent condition resulted in a higher percentage of words maintained.

Koegel and Koegel (1986) further expanded the application of interspersal procedures to facilitating the skill acquisition of an eight year-old boy who had previously had a stroke. Before this child's stroke he had an average IQ. A post stroke IQ assessment score was 69. The skills targeted in the study included spelling, reading, word-finds, and memory tasks. These were skills

that had been in this child's repertoire before the stroke but not present post-stroke. Both known and unknown tasks were identified during a pre-assessment requiring consistent responding on two separate days. The effects of the interspersal procedure were measured using a multiple baseline across behaviors design, first implementing an acquisition-only (no interspersal) condition, followed by an acquisition with interspersed maintenance trials (interspersal) condition. During the interspersal condition no more than two acquisition tasks were presented consecutively, and an equal number of maintenance and acquisition tasks were presented during the session, although the sequence of trials was not specified. The dependent variables were percentage of correct, unprompted responses and subjective ratings of affect. The duration of the sessions was held constant at 15 minutes across each of the four subject areas for both conditions. Consistent with the previous research, the implementation of the interspersal condition resulted in a higher percentage of correct responses across all subject areas. The authors noted that the number of target tasks varied per session based on the participant's rate of responding, but that the number of target tasks completed did not differ consistently across the experimental conditions. Because the duration of the sessions was constant, this may indicate that the overall rate of responding was typically higher in the interspersal condition, if all response acquisition and maintenance are considered. Ratings of child affect were also similar to previous results, and were scored with the same 6-point Likert-type rating scale that was used by Dunlap (1984). During the acquisition only condition, affect was scored as neutral (range of 2-3 on a 6 point Likert scale where 6 is the most positive rating), and the acquisition with interspersed maintenance trials had more positive ratings (4). The authors included an additional measure of the validity of the behavior change through the use of a standardized assessment implemented pre and post intervention. This demonstrated skill development in the areas that were targeted in the intervention and no skill development in the areas that were not targeted

(control). The gains in reading and reading comprehension ranged from 0.4 to 2 years and the gains in spelling ranged from 0.1 to 0.4 years. Math, the control subtest, demonstrated either no change or a decrement of 0.3 years.

Winterling, Dunlap, and O'Neill (1987) evaluated the effect of interspersal on rates of aberrant behavior in two separate studies. In the first study, a constant-task (no interspersal) and a varied-task (interspersal) condition were implemented with two children with autism who engaged in high rates of aberrant behavior which included whining, crying, aggression, and leaving his or her seat. Although labeled as the varied-task condition, the interspersed skills were "four different discrimination tasks that had been previously acquired" (p.108), which indicates that this was an interspersal condition. The procedures for both conditions were based on those outlined in Dunlap (1984), but the dependent variable was the percentage of 15-s intervals in which the participant engaged in aberrant behavior. A reversal design revealed consistently higher rates of aberrant behavior during the constant-task condition, reaching 100% in two of the three constant-task phases. Upon introduction of the varied-task condition the rate of aberrant behavior consistently decreased to zero or near zero levels. The second participant demonstrated similar patterns of responding, although her rates of challenging behavior only reached 100% in one condition-- the final constant-task condition. The percentage of correct responding was provided for the second participant, replicating previous findings that higher percentages of correct responses were consistently observed in the varied-task (interspersal) condition.

The second study conducted by Winterling et al. (1987) evaluated both the frequency of challenging behavior and the number of trials to mastery for acquisition tasks. The participant was a 20 year-old woman with autism whose aberrant behavior included crying, stereotypy, property destruction, and aggression. This study used an alternating treatment design with a constant-task (no interspersal) and a varied-task (interspersal) condition. The resultant distinct

data paths representing the cumulative frequency of aberrant behavior in these two alternating conditions demonstrated experimental control. As in the first study, the higher frequency of aberrant behavior was in the constant-task condition. There were seven instances of aberrant behavior during the constant-task and only one in the varied-task condition. Acquisition data, again, reflects fewer trials to achieve a mastery criterion in the varied-task (interspersal) condition.

Collectively, the above studies provide the initial indications that the interspersal of mastered tasks with acquisition tasks can increase percentages of correct responses and acquisition rates as well as reduce the rates of challenging behavior. The interspersal conditions were also associated with more favorable ratings of child affect, and when provided the opportunity, most participants chose the interspersal condition. Although not directly measured, these data suggested that interspersal conditions were also associated with an overall higher rate of responding. The maintenance and generalization data revealed mixed findings. Since these early investigations, research has been conducted on the effects of interspersal across a variety of populations such as general education students (Schmidgall & Joseph, 2007), students with learning disabilities (Wildmon, Skinner, Watson, & Garrett, 2004), individuals with intellectual disabilities (Burns, 2007), and individuals with autism (Koegel, Singh, & Koegel, 2010). In addition, studies have used interspersal to teach various skills such as math (Wildmon, Skinner, McCurdy, and Sims, 1999), spelling (Cates et al., 2003), and sight words (Browder & Shear, 1996). Research has also been done on the effect of interspersal procedures across response classes such as self-injurious behavior (Horner et al., 1991) and on-task behavior (McCurdy, Skinner, Grantham, Watson, & Hindman, 2001). The following section will provide a summary and critique of the interspersal research conducted with general education students, presumed to be typically developing (see Appendix B for a table summarizing these studies).

Interspersal in General Education

Mathematics.

Studies conducted with typically developing children and youth who are enrolled in general education settings frequently occurred in schools and commonly used brief interventions. Notably the studies comparing interspersal to other teaching conditions demonstrated minimal differential effect, although there were some exceptions. The skill areas targeted included math, reading, and sight word acquisition.

A series of brief investigations, ranging from one to three sessions, were conducted to determine student perceptions and preferences as related to interspersal procedures. Billington, Skinner, and Cruchon (2004) found that a group of 44 students from five different sixth grade classes preferred the multiplication math assignment with easier multiplication problems interspersed, even though the interspersal assignment had more total problems (24 as compared to 18 on the control assignment). A majority of the students ranked the interspersal assignment more favorably in terms of perceived difficulty, effort needed to complete, and time needed to complete. Additionally, when given a choice between the two assignments as homework, 86% of students selected the interspersal assignment. Logan and Skinner (1998) replicated these results with 39 sixth grade students using addition assignments. Seventy-eight percent of these students chose the assignment with the interspersed problems, and 73% of students perceived the control assignment to be more effortful. There was no difference between conditions on the student perception of time needed to complete the assignments. Regarding student performance on the math skills, the mean number of problems completed was higher in the interspersal assignment, but there were no significant differences between assignments on the measures of total number of target problems completed or the accuracy of the target problems.

Similar student perceptions and preferences were demonstrated by Rhymer and Morgan (2005), with 45 third grade students and with a choice between a subtraction assignment with interspersal or a subtraction assignment using explicit timing, in which each minute the student would be notified and there was a time limit for the completion of the assignment. In this comparison, students again scored interspersal more favorably regarding difficulty, effort to complete, and time to complete, and 76% of the students indicated an overall preference for the interspersal procedure. On the measures of work completion, significantly more problems were completed in the interspersal assignment, but more target problems were completed during the explicit timing. The percentage of target problems completed accurately was similar across both conditions. Rhymer and Cates (2006) replicated the above findings with 187 second grade students using a within groups design comparing an interspersal assignment and an explicit timing assignment, both with addition problems. Consistent with the previous studies, student perceptions were more favorable for interspersal regarding difficulty, time and effort to complete the assignment. However, there was no difference between conditions on student preference (measured by which type of assignment they chose for homework). Unlike the previous study there was no significant difference on the number of target problems completed correctly nor was there a significant difference on the percentage of target problems completed correctly, or on problem completion rates. There was a significant difference on the amount of time necessary to complete the assignments, with the explicit timing requiring 134.42 seconds versus 169.80 seconds for the interspersal. However, even this difference was not reflected in the students' perception of time needed to complete assignments, which rated explicit timing as requiring more time.

The above studies extended the literature on interspersal through implementation of interspersal on worksheet skills that were completed independently by the student. The

investigations were brief, with the longest assessment requiring approximately 20 minutes for the students to complete. Additionally, the results were based on limited exposure to the experience with the conditions; students experienced each condition between one to four times. However, the findings were consistent across grade levels and types of computation. These students demonstrated both a preference for, and had more favorable perceptions of the interspersal assignments, despite the fact that this required them to complete additional problems. Although student preference was a primary interest in the above studies, it is notable that the use of interspersal in these cases did not result in significant differences on the rate or accuracy of the work completed. Comparable results also with mathematics skills can be found in Hawkins, Skinner, and Oliver (2005); Montarello and Martens (2005); and Robinson and Skinner (2002).

Two studies that demonstrated positive effects of interspersal with math worksheets were conducted by McCurdy, Skinner, Grantham, Watson, and Hindman (2001) and Belfiore, Lee, Vargas, and Skinner (1997). McCurdy et al. (2001) increased the percentage of on-task intervals for a fourth grade girl during scheduled independent math time. An alternating treatment design was used comparing control assignments, consisting of the typical classroom assignments, to assignments with an easier problem interspersed for every three target problems. Momentary time sampling with five-second intervals was used to measure on-task behavior. Clear increases in on-task behavior were evident during the interspersal assignments. On-task behavior was demonstrated during 55.5% of the intervals with the control assignments, and 72.5% of the intervals during the interspersal assignments. Although the on-task behavior of this student did not meet the percentage of intervals that a comparison peer demonstrated (85%), the use of interspersal did result in improved responding. This study was notable for its application in the classroom, during naturally occurring activities, and for its inclusion of the social validity measures of the peer behavior.

Belfiore et al. (1997) used a reversal design to demonstrate a reduction in response latency to begin math problems through interspersal of easier math problems. Effects were replicated across both participants, with the largest effect demonstrated in the first condition change and diminishing in subsequent reversals. The diminishing effects were due to response latency decreasing across both conditions. Practice with the math skills during the experiment may have been sufficient to improve the participants' ability to complete the problems such that the original difficult problems were now easy. This pattern of responding, if replicated in other research, may indicate that interspersal procedures will produce a stronger effect when the acquisition skill approaches a certain threshold of difficulty for the individual.

Reading and spelling.

Roberts, Turco, and Shapiro (1991) used a group design to compare the effect of different drill ratios of known to unknown words on the acquisition of sight words by 42 fifth graders. Sessions occurred three times per week for eight weeks. Students were assigned to one of the four groups for the duration of the study. Results indicated that the highest number of words were learned by the group with the 50% known to 50% unknown ratio. The mean number of words learned across drill ratios was 62.66 for the 50:50 group, 48.65 for the 60:40 group, 24.44 for the 80:20 group, and 12.50 for the 90:10 group. However, the number of new words introduced in a session varied across the groups as a function of their ratio. In this case, a total of ten words were presented in each session, which means that the fewest new words a student was exposed to in a session was one new word and the most would be five. So, the most words were learned in the 50:50 group, but the most unknown words were also presented to the 50:50 group. When the percent of words learned from the total words possible in each condition was further evaluated and tested, no significant difference across conditions was found, with students learning 50-54% of the possible words introduced across all conditions. These results may

indicate that acquisition is a function of the students' opportunity to respond. No student preference data were reported.

An extension of the Roberts et al. (1991) study was conducted by Roberts and Shapiro (1996) with a group of 42-second grade students. In this study, assessments empirically identified known and unknown words for each participant before the start of the study, the number of words drilled per session was increased (from 10 in the previous study to 20), and an assessment-only condition was included as a control. The drill ratios in this study expanded on previous research by investigating the effect of more unknown words to known words. The conditions included 80% known to 20% unknown, 50% known to 50% unknown, and 20% known to 80% unknown. Again, overall more words were learned in the condition with the highest percentage of unknown words (80% unknown: 20% known). However, this ratio of unknown to known words was also correlated with higher number of opportunities to respond to new words (i.e., a greater number of new words were presented to the student). When the data were evaluated as the percentage of words learned proportional to the number of words that could be presented per condition, this proportion was significantly greater for the students in the 80% known group than the other two conditions. And the proportion of words learned in the 50% known group was significantly greater than that of the 20% known group. Of the unknown words that were presented, the students learned 67%, 50%, and 35%, respectively. These results provide some direction for application based on if the goal of teaching is increasing the overall total words learned (by providing exposure to more words using an 80% known ratio) or teaching more words of a specific group (by using a 20% known ratio). However, the effects of the intervention overall was weak, as indicated by the comparable gains made by the assessment only group. Student preference of instructional condition was not measured.

MacQuarrie et al. (2002) used a within-groups design to compare the effects of three types of flashcard drill instructions with a group of 25 third grade and 26 seventh grade students. The dependent variable was number of Esperanto words pronounced and translated correctly. The Esperanto International Language is an artificial language based on European languages and first used in 1892 (Mirriam-Webster, 2012). It was selected for use in this study to increase the internal validity of the findings, as it was unlikely that students would have a prior history or other uncontrolled exposure to Esperanto words. The three types of instruction included traditional flashcard instruction (no interspersal), drill sandwich (interspersal), and incremental rehearsal (interspersal). Traditional drill sessions continued until the student responded with three correct responses to each of the nine Esperanto flashcards, providing the correct pronunciation for the Esperanto word and the English translation. During the traditional instruction, when the presented flashcard was mastered it was removed from the stack. In the drill sandwich approach, words were presented in sets of three unknown words with six known words. Three known words were presented before the Esperanto flashcard, which was followed by three more known words, another Esperanto flashcard, three known words and the final Esperanto flashcard. Each set was practiced three times, and then the three previous unknown words became the 'known' words for the next set of flashcards. Incremental rehearsal practiced the first Esperanto word with an incrementally increasing number of known words between trials until the first Esperanto word had been practiced with all nine known words, then a known word was removed and the previously unknown Esperanto word was included as a 'known' word for the next sequence with the second Esperanto word. This continued until all nine Esperanto words had completed the same sequence. Each student completed each condition and maintenance was probed the following day, plus two, three, seven, and thirty days after teaching had occurred. For the third grade students, the mean number of words maintained was significantly higher for

words taught using incremental rehearsal versus the other two conditions. The mean number of words that were maintained was slightly greater for those that were taught in the traditional drill than for those learned in the drill sandwich. The seventh grade students demonstrated the same pattern of maintenance across conditions, with the most words maintained from the incremental rehearsal condition followed by traditional drill and the drill sandwich. Although this study provides a comparison of total number of words learned across three conditions, there are many variables that differed between the conditions that could impact skill acquisition and maintenance. Further analysis is needed to determine how different durations of instructional sessions (20-40 minutes for incremental rehearsal, 20-30 minutes for traditional drill, and 10-15 minutes for drill sandwich) and different number of opportunities to respond affect skill acquisition. More information on these variables for this and other populations would clarify what the controlling variables are for each of the various instructional methods.

When the time, or the number of opportunities to respond is controlled, many studies conducted with general education students on reading or spelling skills result in undifferentiated results across conditions. Cates et al. (2003) found comparable results for the cumulative number of spelling words mastered across a no interspersal condition, a condition that interspersed one known word for every three target words, and a condition that interspersed three known words before each target word. Additionally, maintenance was comparable across all conditions. When learning rates were considered, five of the six participants learned more words per instructional minute in the no interspersal condition, followed by the condition that interspersed one known words for every three targets. Student preference regarding instructional condition was not collected.

Like Cates et al. (2003), Joseph and Nist (2006) compared the effectiveness of a traditional drill (no interspersal), a high-probability sequence (interspersing three known words

before each unknown word), and an interspersal condition with three known words following three unknown words, on the acquisition of sight words by two fifth grade boys and one sixth grade boy. An alternating treatments design was used, and each condition was implemented daily across eight consecutive school days. Known and unknown words were empirically tested in a single probe pretest. Six unknown words were assigned to each condition, and as each word was mastered it was removed and an additional unknown word was added. At the start of each session, and following any errors, the instructor modeled each unknown word. Verbal praise was provided for all correct responses. The total number of words mastered per condition was similar across all conditions for two of the three participants. The third participant had the greatest difference between conditions on the number of words learned, with 20 words learned in the high-p sequence condition, and 15 in the other conditions. As a group, the participants learned the most words in the high-p sequence (58), followed by the traditional drill (54), and least words in the interspersing condition (48). However, all participants demonstrated the fastest learning rate in the traditional drill condition, followed by the interspersal condition, and the lowest number of words mastered per instructional minute in the high-p sequence. The percentage of words learned in each session that were maintained was highest in the traditional drill condition (85%) followed by the interspersal condition (77%) and lowest in the high-p sequence condition (74%). However, the maintenance probe was conducted one day after the last instructional session. The teaching package that was implemented at the start of each session and following any errors may have been sufficiently powerful to mask the effects of the conditions. However, considering the small difference in total number of words mastered, and the difference in the learning rates and percentage of words maintained, this investigation supports the use of traditional drill and practice over either of the interspersal conditions when teaching sight words

to fifth grade general education students. Student preference however, was not reported in this study.

Volpe, Mule, Briesch, Joseph, and Burns (2011) also evaluated instructional effectiveness (cumulative number of words mastered) and instructional efficiency (number of words learned per instructional minute). This study used a series of multielement designs to compare the effects of an incremental rehearsal drill (interspersal) to a traditional drill and practice format (no interspersal) on sight word learning for four first graders. Again, the cumulative number of words learned was undifferentiated across the two conditions when the number of opportunities to respond was controlled. Across all participants, the most words learned per instructional minute was in the traditional drill and practice condition. Unlike the student preferences in many of the math studies, students in this study did not demonstrate a preference for the interspersal condition (two of the four participants preferred each condition).

Nist and Joseph (2008) also found comparable results in their assessment of an incremental rehearsal condition, a traditional drill and practice condition (no interspersal), and an interspersal drill. Incremental rehearsal resulted in the highest number of cumulative words learned, followed by interspersal. Additionally the words learned in the incremental rehearsal condition had a higher level of generalization. However, when the duration of instructional time was considered, and the learning rate was evaluated, the traditional drill had the highest learning rate followed by interspersal, and incremental rehearsal had the lowest cumulative learning rate. In this study, each of the six unknown words assigned to each condition was practiced nine times during the session. The interspersal condition was comprised of six unknown and three known words. The sequence of presentation was one known word followed by three unknown words. In the incremental rehearsal condition each of six unknown words was presented nine times incrementally with nine known words. As is common to incremental rehearsal, once an unknown

word reached the end of the sequence it was then considered a known word and replaced one of the known words for the next sequence. This resulted in significantly higher opportunity to respond to the target than was provided in the other conditions, which may skew the acquisition and generalization data.

Procedurally, this study had several strengths, including matching the words assigned to each condition such that each had an equal distribution of various consonant – vowel patterns. This study also included generalization measures from reading on flashcards to reading the target words in sentences. However, the sentences, created by the researcher were written in a single sentence format typed on a white paper. Measures of generalization to a natural reading activity and media format would have enhanced the study. The authors also included social validity measures that recruited feedback from both the teachers and students regarding the appropriateness, efficacy, efficiency, and preference of the various conditions. Following the conclusion of data collection and solely for the purposes of the social validity assessment, the conditions were color-coded and the students were asked which of the conditions they preferred to complete. This was conducted on three occasions. All participants chose the traditional drill and practice.

Many of the studies reviewed here conducted with general education students implemented interspersal using the incremental rehearsal procedure. Several of these studies have evaluated the effect of different drill ratios or have compared incremental rehearsal to other interspersal procedures. However, the different drill ratios and procedures consist of differing numbers of opportunities for the student to respond to the acquisition task. Szadokierski and Burns (2008) held the opportunity to respond constant and manipulated the drill ratios, comparing 10% unknown to 50% unknown. The results confirmed that the number of opportunities to respond using the incremental rehearsal procedure (including the embedding of

the unknown word from the first sequence as a known word for the second and subsequent sequences) method had a stronger effect than the percent of known or unknown tasks included in the training.

Martin, Skinner, & Neddenriep (2001) conducted an investigation with 48 seventh grade students regarding the impact of interspersal procedures on reading fluency when the required response of the student was a passage (contrasted with the previous sight word studies). A within-subjects repeated measures design was used to compare a control condition, which consisted of a three paragraph passage, and an experimental condition that consisted of a three paragraph passage but interspersed with two additional easier paragraphs (with a first grade reading level). Perception, preference, and rate of correct words per minute were measured. Significantly more students rated the control passages as less time consuming. No differences were found regarding perceived effort, preference, or number of words read correctly per minute between conditions. The authors suggested that the effects of interspersal might be more significant if the task is more discrete (e.g., sight word or math problem) versus continuous (reading).

The research conducted with general education students and interspersal procedures revealed mixed results. Generally students demonstrated a preference for and favorable perceptions of math worksheets when easier problems were interspersed. Less data is available about the student preferences and perception of the various reading and spelling interventions implemented, but the data suggest a preference for the alternative, less effortful interventions, which in the above studies was not the interspersal intervention.

Of note, few of the above studies demonstrated strong effects of the interspersal procedure when implemented with general education students, especially when controlling for number of opportunities to respond or the length of instruction across conditions. Other potential

uncontrolled variables might include the type or complexity of the material being taught, the teaching packages that were included in many of the procedures, or the general behavioral characteristics of the students participating. Possibly, the general education students are representative of the group of students who have strong learning histories and relative high success rates, such that the learning (demand) contexts are not aversive conditions. It is also possible that the interspersal procedures produce more of an effect for those students who have experienced academic or learning difficulties, and for whom the learning contexts are aversive conditions.

The following section will review the literature on the use of interspersal procedures with atypical populations. These include students identified as having a specified or unspecified learning disability (Burns & Dean, 2005; Calderhead, Filter, & Albin, 2006; Cooke, Gazaukas, Pressley & Kerr, 1993; Wildmon, Skinner, Watson, & Garrett, 2004), students diagnosed with specific disorders, such as emotional and behavioral disorders (Cooke, Gazaukas, Pressley & Kerr, 1993; Skinner, Hurst, Teeple, and Meadows, 2002), intellectual disabilities (Burns, 2007; Burns & Boice, 2009; Horner, Day, Sprague, O'Brien& Heathfield, 1991; Koegel & Koegel, 1986; Rowan & Pear, 1985; Neef et al., 1980) and autism or developmental disabilities (Adcock & Cuvo, 2009; Benavides & Poulson, 2009; Browder & Shear, 1996; Charlop, Kurtz, & Milstein, 1992; Chong & Carr, 2005; Dunlap, 1984; Koegel, Singh, & Koegel, 2010; Neef et al, 1977; Reed, Luiselli, Moizio, & Child, 2010; Volkert, Lerman, Trosclair, Addison, & Kodak, 2008, Winterling, Dunlap, & O'Neil, 1987) (see Appendix C for a table summarizing these studies).

Interspersal with Atypical Populations

Many individuals with atypical development will demonstrate both challenging behavior and specific or generalized skill deficits. Research had been conducted on the effects of

interspersal in both areas. First research using interspersal procedures focused on reducing challenging behaviors and increasing on-task behavior will be discussed. Then those studies that are focused on increasing abilities through skill acquisition will be reviewed.

Reducing challenging behavior and increasing on-task behavior.

Horner et al. (1991) used a reversal design replicated across three participants to evaluate the effect of task difficulty and interspersal procedures on rates of aggression and self-injury. These conditions included easy, difficult, and difficult + interspersed requests. Easy and difficult tasks were recruited from the staff and defined as a task in which 70% or more or 33% or fewer of the trials in a session were likely to be performed correctly. One of each task was identified for each participant and implemented for the duration of the study. In the difficult + interspersed requests condition, three to five simple responses that the participant had a high probability of completing correctly were provided at the beginning of each session, after approximately every three training (difficult) trials, and following any resistant behavior. Initial reversals between easy and difficult task conditions were conducted as an assessment to determine the function of problem behavior, and the data verified that the challenging behavior varied as a function of task difficulty. Subsequent reversals demonstrated the same patterns of responding for two of the three participants with zero to near zero levels of problem behavior during the difficult + interspersed condition and immediate high rates of challenging behaviors in the difficult conditions. For the third participant the difficult + interspersed requests condition did result in significantly lower rates of challenging behavior than the difficult condition, but did not return to the same zero levels as demonstrated in the easy condition. For all participants, the behavior change was durable across a new trainer implementing the difficult + interspersed requests procedures. For two of the participants the behavior change was also generalized to a new trainer implementing a *new* task. The authors also reported the percentage of trials that the participants

attempted to respond to the task. One participant attempted all tasks regardless of condition. However, for the other two participants their attempts to complete the task did vary by condition, and the patterns of reversal mimicked those of the problem behavior, with higher attempts in the easy and difficult + interspersed requests and lowest percentage of attempts in the difficult task condition. These results may be interpreted as a measure of participant motivation, and taken with earlier studies on interspersal support the conclusion that interspersal may increase motivation, or decrease the aversiveness of the demand situation.

Although this study demonstrates a functional relationship between the use of interspersal and both challenging behavior and attempts to respond, there are some limitations. First, the use of the interspersed requests as an antecedent procedure (at the start of every session and after approximately every three target trials) and the use of the interspersed requests as a consequence for resistant behavior make conclusions unclear about how interspersal functions. Additionally, no data are provided on the rate of resistant behavior, or the frequency of the use of interspersed requests as a consequence. Another limitation is the difficult + interspersed request condition was only alternated with the difficult task condition, so one cannot rule out that the behavioral changes might be due to the relative difference of task difficulty between the two conditions, which may have differed had the easy condition been alternated with the difficult + interspersed condition.

Horner et al. (1991) conducted a second study to determine if the interspersal procedure would be effective across longer training sessions and in a school setting with a 14 year old boy diagnosed with a mild intellectual disability whose challenging behavior included aggression. This study replicated the first seven phases of the initial study, which included a reversal between the easy and difficult task conditions, followed by a reversal between the difficult and difficult + interspersed requests conditions. It also included several easy and difficult tasks to be

targeted throughout the study (the initial study only identified one easy and one difficult task for each participant). Finally, the difficult + interspersed requests condition did not use the interspersed requests as a consequence for challenging behavior. The results replicated those of the initial study with highest rates of aggression in the difficult task condition and evident reversals to low or zero rates in the easy and difficult + interspersed conditions. The percentages of intervals with attempts to respond to the task indicate some moderate reversals between the easy task and difficult task phases with higher rates of attempt in the easy task condition. However, attempts to respond across the difficult task and the difficult + interspersed request remained at a high rate across both conditions, possibly due to practice effects with the skills targeted.

Both studies in Horner et al. (1991) provide support for the use of interspersal to decrease challenging behaviors for individuals with intellectual disabilities, and study two specifically supports the use of interspersal in a school setting. The results of McCurdy et al. (2001) support the use of interspersal on a math worksheet to increase on-task behavior, but for a general education student. The results of Skinner et al. (2002), however, only partially replicate these findings. Alternating treatments designs were used to evaluate the effect of interspersal on the on-task behavior of four students, aged nine to eleven years old. Participants had a diagnosis of an emotional and behavioral disorder, and the study took place in their assigned classroom. Four mathematics assignment pairs were created, each with a control (no interspersal) and an experimental (interspersal) assignment. Each control assignment contained 30 target problems, and each experimental assignment contained 30 similar target problems plus an easier problem after every third target. Therefore the experimental assignments had a total of 40 problems. Momentary time sampling with a 5s interval was used to measure the percentage of on-task behavior. Although in some cases slight, the overall mean on-task behavior for each participant

favored the experimental assignments. Mean on-task levels for each participant were 88%, 88%, 95%, and 81% during the experimental assignments and 69%, 76%, 92%, and 83% for the control conditions.

A possible rationale for the weak effects is that students had limited exposure to the conditions, and only completed seven assignments in each of the two conditions. Although the data are variable, and a visual inspection of the graphs reveal no clear difference between the conditions for most participants, three of the four participants demonstrated trends toward differential responding across the conditions. An extended evaluation would be helpful to determine if a pattern was developing.

Skinner et al. (2002) took place in a self-contained classroom for nine children with emotional and behavioral disorders. The McCurdy et al. (2001) study took place in a classroom with 27 other general education students, and the comparison data indicated that these peers demonstrated the desired on-task behaviors at a high level. This peer model of the desired behavior may have had impact on the increases of on-task behavior demonstrated by the study participant. Although her peers' behavior did not control the on-task behavior of the study participant (otherwise her behavior would have come under control of peer models without the additional intervention), it may have supported her behavior change. More information regarding the behavior of the other students from the Skinner et al. (2002) study is necessary to determine if the peers' behavior was an important difference between the studies, and if so, how it might have impacted the failure to fully replicate.

The results of Burns and Dean (2005) indicate that the rate of on-task behavior for five fourth grade students receiving special education services may be affected by the ratio of known to unknown skills used during teaching using an Incremental Rehearsal procedure. Both the unknown words (from the Esperanto International Language) and the known words (from the

fourth grade Fry reading list) were confirmed as such in pretests. The conditions evaluated were 0% known (no interspersal), 50% known, 83% known, and 90% known. The number of new words in each session was controlled, therefore, to meet the different drill ratios for known to unknown words, the sessions differed in duration across the conditions. The mean number of minutes per condition was 6.1 for the 0% known condition, 7.1 for the 50% known, 13.9 for the 83% condition and 27.5 for the 90% known. Interestingly, for all participants, the highest rate of on-task behaviors occurred in the longest sessions, with 90% known words. No data points for this condition overlapped with any other condition. The rates of on-task behaviors were fairly comparable across the other drill ratio conditions. The 90% known condition also produced the highest number of words mastered and maintained one week after teaching occurred, no data point for this condition overlapped with any other. The number of words mastered or maintained in the other conditions were comparable with little difference in the range of scores between the no interspersal and the 50% known or 83% known conditions, suggesting that for these participants the 90% known might have produced the optimal level of on-task behavior for learning. These results provide evidence that attention can be enhanced through the use of interspersal procedures, but high ratios of known tasks may be necessary to produce an effect.

However, each participant was exposed to each condition only one time, so it is unclear if this pattern of on-task behavior would persist across time. None of the studies comparing the relative effects of drill ratios with general education students included a measure of on-task behavior, so comparisons to this population cannot be made.

A comparable study, conducted by Calderhead, Filter, and Albin (2006) evaluated the effects of different ratios of known to unknown tasks as related to the on-task behavior of a 12 year-old student receiving special education services. Pretests guided the development of easy and hard math problems, although specific data verifying them as such were not provided.

During the study, she responded with a high percentage of accuracy (92%) to the easy problems, suggesting that the selection was valid. An alternating treatment design was used which included a 0% interspersal (no interspersal), a 33% interspersal, in which every third item was easy, and a 67% interspersal, in which two easy items followed each hard item. On-task behavior was measured using whole-interval recording. Visual inspection of the data reveals that the behavior was most variable in the no interspersal condition, and although the overall rates of no interspersal were lower than either interspersal condition, there were several overlapping data points. In this study, participants worked on three worksheet packets per day, and each worksheet had 18 problems. Although the worksheet order was counterbalanced to control for order effects, participant fatigue may have contributed to the overall variability of the data (i.e., perhaps the last session of the day resulted in suppressed responding, regardless of condition). The mean percentage of on-task behavior was higher in both interspersal conditions (83.64 and 83.40%) relative to the no interspersal condition (66.67%). These results replicate the finding of McCurdy et al. (2001) and Skinner et al. (2002), in that interspersal facilitated on-task behavior of a student during independent mathematics tasks. The similarity of the rates of responding also support the findings of Burns and Dean (2005) in which different ratios of interspersal produced similar results until the percentage of known tasks reached 90%. However, the current study did demonstrate a difference between the no interspersal condition and the interspersal conditions, and the percentage of on-task behavior was high across both interspersal conditions, so a ceiling effect may be another explanation for the lack of differentiation between the two interspersal ratios.

Koegel et al. (2010) demonstrated decreases in response latency to begin an academic task and reductions in the rate of disruptive behavior of four children with autism through the implementation of a treatment package that included interspersal of easier tasks. The treatment

package also included child choice (of materials or setting), and the inclusion of natural reinforcers in the task. The tasks targeted were writing or math, and the level of difficulty and number of items per task was constant across sessions. A multiple baseline across participants design was used. All participants in all areas demonstrated an increasing trend during baseline in the response latency, as measured by the number of minutes to begin the task after instruction. Upon implementation of the intervention, response latency across all participants and both skill areas immediately decreased to zero or near zero levels, and remained low for the duration of the intervention. Additionally, the intervention level response latency were also demonstrated during follow up probes that were conducted at least two week post intervention and by an adult who had not participated in the study. Similar trends were observed in the rate of disruptive behavior (e.g., aggression, tantrum, bolting, crying), which was recorded using a continuous 30-second interval recording system. High or increasing rates of disruptive behavior, reaching 100% of intervals, occurred for all participants and both tasks during baseline. Implementation of the intervention resulted in an immediate reduction of at least 50% and in some cases 100% in the rate of disruptive behavior, which continued at low rates for the duration of intervention and at the follow up probe. Concomitant with reductions in response latency and challenging behavior were increases in the rate of problem completion for all participants and across both tasks. Child interest in the task, as scored by a observers using a 5 point Likert scale adapted from previous research, also increased during intervention and maintained at high ratings during follow up across all participants and both tasks. Although the effectiveness of the treatment package was demonstrated, the relative effect of each of the components was not. A component analysis, or reversals with different independent variables that comprised the treatment package would clarify the relative effect of each. And, although interspersal of shorter or easier tasks from the same skill area was identified as part of the treatment package, insufficient detail was provided

regarding the sequence and ratio of easy tasks to hard tasks or if the inclusion of interspersal added to the total number of tasks to be completed, or replaced some of the hard ones.

Using an ABCDE sequential modification design with embedded alternating treatments design, Reed et al. (2010) demonstrated reductions in the rate of the self-injurious behavior of a nine year-old girl with autism. Treatment was alternated across the following four conditions: low difficulty/high demand rate, 1:1 interspersal during which an easy task was presented before each difficult task, high difficulty/high demand rate, and a 3:1 interspersal during which three easy tasks were presented before each difficult task. Initially both interspersal conditions resulted in low rates of SIB, but the 3:1 interspersal had a continued decreasing trend while the 1:1 interspersal resulted in variable and increasing rates of SIB across sessions. The 3:1 interspersal condition was therefore evaluated in the context of other treatments during the next phase of the study. In addition to the 3:1 interspersal, the other conditions were a two-sets randomized design consisting of a high and low effort set that were presented in a randomized sequence, a three-sets randomized design, and a high difficulty/high demand rate condition. In this phase, the overall rate of SIB in the 3:1 interspersal condition was comparable to the prior phase, however the rates were significantly more variable. The increase in variability may have been a result of the interactions between the conditions; specifically that two of the other conditions also had an aspect of novelty in the randomization of the easy or difficult instructions.

The studies reviewed above provide support for the use of interspersal procedures as an antecedent intervention to decrease a variety of challenging behaviors and to increase on-task behaviors. Interspersal procedures have been used to facilitate the acquisition of a variety of skills, including expressive labeling (Rowan and Pear, 1985; Volkert et al., 2008), sight words (Burns & Boice, 2009; Browder et al., 1996; Knight, 2003), matching to sample tasks (Benavides & Poulson, 2009) and varied academic tasks (Adcock & Cuvo, 2009).

Skill acquisition.

In a series of three experiments using alternating treatment designs, Cooke et al. (1993) compared the effects of an interspersal procedure to a procedure without interspersal on the acquisition of spelling words, math facts, and reading. Each experiment included two conditions. The first interspersed skills at a 30% known with 70% unknown ratio, and the sequence of trials was three known prior to each unknown task. The second condition was a traditional drill with 100% new items (no interspersal). At the end of each experiment, student preference regarding instructional condition was recruited.

Experiment one was conducted with four male students with emotional and behavioral disorders, ranging in age from 14 to 17 years. A pretest of one trial for each word identified both known and unknown spelling words, all at their placement levels. During teaching, each of the 10 flashcards with the word printed on it was presented to each participant and then removed while the participant spelled the word. A five-step error correction occurred if the word was spelled incorrectly. The set of 10 words was practiced three times. Words mastered in the interspersal condition were then used as known words in subsequent sessions. Accuracy on daily probes, which immediately followed teaching, and on maintenance after one week was reported. Data indicate similar percentages of words spelled correctly during the daily probes. The percentage of words spelled correctly during the maintenance probes was also similar in both conditions, although slightly greater for three of the four students in the interspersal condition. This may be a result of the additional practice by using words mastered in the interspersal condition as the known words during subsequent sessions. Although the percentage of correct responses was comparable across conditions, the actual number of words mastered would therefore differ greatly, as 100% correct responding in the no interspersal condition indicates 10 words mastered, but only three in the interspersal. However, data were not reported on the total

number of words mastered per condition across participants. When asked, all students indicated that they preferred the interspersal condition.

Experiment two was conducted with three students with learning disabilities. Two boys, ages nine and ten, and one 11 year-old girl participated. Procedures were similar to experiment one, except the behavior of interest was fluency on math facts, as measured by the number of correctly written digits per minute on a multiplication test following teaching. The ratio of known to unknown facts during teaching differed according to the condition, but both conditions included 10 minutes of flashcard drill with the teacher which included instructive error correction and as many repetitions of the flashcards as the time allowed. Then, the student completed three written problems that had been targeted in that session; these responses were reviewed and corrected as necessary. Finally, the teacher presented each of the ten facts again, three times each. Mastered facts from the interspersal were then used as known in subsequent trials. Finally, the student completed as many problems on a worksheet with all of the facts taught in that session as possible during one minute. Fluency data were collected on the last test only. Improvements were observed in both conditions, but were more significant for the interspersal condition. Maintenance was comparable at a high percentage across both conditions. Two of the three participants indicated a preference for the interspersal condition. Although greater fluency was achieved using interspersal, the conclusions from this experiment are somewhat ambiguous because the difficulty of the problems was not controlled during the fluency tests. It is possible that the improved fluency was a result of the inclusion of easier problems during the test. A more compelling assessment of progress would have been testing the fluency of the target items only, but even this would be skewed by the duration of the task (three problems targeted in interspersal and 10 in no interspersal). Longer duration tasks may adversely affect fluency due to participant fatigue.

Experiment three measured the acquisition of sight words and reading fluency as measured by the number of correct words per minute, for six students accessing special education. Three students received the intervention, and three were assigned peer tutors who implemented the target sight words according to the condition they were assigned. More words were mastered in the no interspersal condition for all participants. Improvements in reading fluency were observed in both conditions, with most gain in the interspersal condition for two of the three participants. However responding was variable in both conditions, and does not clearly demonstrate experimental control of either of the conditions. Maintenance of words mastered was slightly higher for the interspersal condition (96-100% vs. 88-100% of no interspersal). All participants indicated a preference for the interspersal condition.

Although the experiments conducted by Cooke et al. (1993) all demonstrated learning across both conditions, the limited differentiation between the two conditions, especially in experiments one and three, restrict conclusions about the relative strength of the conditions. The lack of differentiation might be due in part to the intensity of the teaching package, which may have created a ceiling effect and masked the effect of the conditions. Second, the use of the alternating treatments design is not appropriate when fluency is the dependent variable; as it is susceptible to multiple treatment interference and there is a risk of carry over between the two conditions, in this case, fluency with multiplication facts and fluent reading. Finally, the process of embedding the recently mastered words in the teaching sequence of the new words may skew the maintenance of the responses.

In a brief study to evaluate the preferences of 39 seventh and 17 eighth grade students with learning disabilities, Wildmon, Skinner, Watson, and Garrett (2004) presented math assignments with and without interspersal. A within groups design was used and the order of assignments was counterbalance to control for sequence effects. Results indicated that more

problems were completed on the interspersal assignment, although number of target problems completed was comparable across conditions. Significantly more students perceived the interspersal as less difficult and less time consuming, and 87% of the group chose the interspersal assignment as their homework (an indication of preference), despite the fact that it had more total assignments than the control. These findings on student preference are consistent with those observed by Cooke et al. (1993) as well as the findings on preference in the general education population.

Browder and Shear (1996) demonstrated success in teaching three special education students novel sight words when those words were interspersed with other known sight words. A multiple probe design was used to evaluate the effect of the intervention on the number of words read correctly on a sight word test, the percentage of words read correctly, the percentage of words read incorrectly, and responding during weekly maintenance probes. Results of the sight words probes indicate a stable and low number of words read (zero to one) correctly during baseline across all participants. Within two sessions of the introduction of the intervention the number of words read correctly increased, and there was an increasing trend throughout the intervention. All participants also demonstrated high levels of correct responding during the weekly maintenance probes (reading at least nine of the ten words correctly on each probe and for most probes all ten were read correctly). The data on the reading probes were variable, but reflects an increasing trend for the number of words read correctly for two participants and a decreasing trend in the number of words read incorrectly across all three participants.

Although strong effects were demonstrated, there are some limitations of this study. The treatment package, which used incremental rehearsal, was not systematically compared to other treatments options, so conclusions cannot be made regarding relative effectiveness of different treatments. Nor can conclusions be made regarding the relative effectiveness of this intervention

to the participants' typical instruction because the baseline condition was a sight word test, and did not measure participants' acquisition or percentage of correct responses during typical instruction. There were several additional teaching conditions that may have also contributed to, or affected the results. For example, following the interspersal drill but before the generalization probe, the teacher developed and the participant read a passage that included the words that had been targeted in that day's session. During this activity, an instructive error correction was also used. No data were reported regarding performance during this reading activity. After this reading activity the generalization probe to the newspaper was conducted. Therefore the relative impact of the interspersal drill on the generalization to the newspaper cannot be determined. Despite these limitations, the data reflects an immediate and strong effect upon introduction of the treatment package and does support the use of an interspersal (or more specifically incremental rehearsal) to teach sight words to students in special education. As an additional indication of the effect of the intervention, the authors note that the three participants learned the 10 new words that were targeted in the study more quickly than they had previously learned other words through general reading instruction (10 words in 31 days versus 30 words after several years). These data indicate that more research on the treatment package and its components, which included interspersal, use of a systematic increase in number of maintenance trials between trials of acquisition words, and a five-step error correction sequence, may be beneficial in improving sight word instruction with special education students.

Knight et al. (2003) compared the effects of constant time delay and interspersal on sight word acquisition for two students with learning disabilities and two students with intellectual disabilities. Known and unknown words were empirically validated during a pretest that required three correct or incorrect responses. Words were assigned to each condition and mastered as word triads. In the constant time delay procedure the words were presented without interspersal

of other skills. Each word was presented 10 times and each session ended after the 30 trials were conducted. Praise was provided continuously for correct independent responses and no response for four seconds resulted in a model of the word. No correction was provided for other errors. During the interspersal condition three known words were interspersed with the three unknown words using the sequence common to incremental rehearsal. No praise was provided for correct responses. A five-step error correction procedure was implemented for any errors. Praise was provided for a correct imitation within the error correction procedure. The sequence was completed twice, which resulted in 58 opportunities to respond. The multiple differences between conditions in reinforcement contingencies, opportunities to respond, and procedures for incorrect responses limit the conclusions that can be made regarding the impact of the interspersal procedure. However, it is interesting that there were stronger differential effects of the conditions for the two students with intellectual disabilities than for the two students with learning disabilities. For the two students with learning disabilities the percentage of words read correctly, the number of word mastered, and the number of session to master words were largely undifferentiated between the two conditions throughout the study. Both conditions resulted in increases to differing degrees for those two students. However for the students with intellectual disabilities, the data paths separated early (at session three and five) and became more differentiated as the study continued. Both of these students had higher percentages of correct responses during the constant-task condition, and for both the percentage of correct responses to interspersal remained at a low level (20%) throughout the study. Both student mastered more words in the constant time delay condition, and only one word triad was mastered by one participant in the interspersal condition (contrasted with the four word triads that were mastered for the same participant during constant time delay). These results may suggest that different interventions may have different impacts across different populations.

In a replication of MacQuarrie et al. (2002), Burns and Boice (2009) compared three instructional strategies and evaluated their differential effects on the number of Esperanto words learned by 20 students in seventh and eighth grade. Also of interest was how the instructional condition interacted with IQ. All participants in the study had IQ scores that ranged between 61 and 85. An ABC design was used and each condition consisted of one session one week apart. Each participant was exposed to each condition. Conditions were presented in a counterbalance order across participants. Known words were empirically verified with a single probe pretest. The conditions were consistent with those of MacQuarrie et al. (2002), described above. Results were also consistent, but with a smaller effect size possibly due to the learning disability or intellectual disability. As in the MacQuarrie et al. (2002) study, the incremental rehearsal resulted in the most words maintained, but this was also the condition with the most opportunities to respond during teaching, and although duration was not explicitly measured it was estimated to be the longest of the three conditions.

The mixed findings of the effects of interspersal may be a result of the differing procedures (drill ratios, sequence of trials, opportunities to respond, rate of reinforcement, type of interspersed tasks) used across studies. The following section will review the studies that examined variables within the interspersal procedure, including the schedule of reinforcement and the type of task that is interspersed with the acquisition task.

Schedules of reinforcement.

Benavides and Poulson (2009) measured the percentage of correct responses to non-mastered matching to sample tasks across three conditions: baseline, interspersal, and low-density reinforcement. A 12 token reinforcement system was implemented in all sessions and across conditions. Similarly, there were 12 opportunities to respond to the acquisition matching task in each session across conditions. During baseline a token was provided for each correct

response to the acquisition task, or if the session ended without sufficient correct responses, the child could earn the remainder of the token through compliance with simple receptive instructions. During the interspersal condition the number of total tasks increased from 12 to 24. All correct responses were eligible for reinforcement, so the availability of reinforcement also doubled from baseline. The low-density reinforcement condition was implemented as a control that would equate the reinforcement density of baseline to determine if responding was controlled by reinforcement density or the interspersal procedure. In all other ways the low-density reinforcement condition was identical to the interspersal condition. The interspersed tasks were similar in response mode to the acquisition task (matching), but they were categorically different (interspersed tasks were pictures of preferred items like cartoon characters and the acquisition tasks were animals, shapes, numbers, and letters). The baseline condition did not intersperse any mastered tasks, but did teach across multiple exemplars simultaneously (no interspersal). Each of the three conditions taught across multiple exemplars simultaneously, and each of the three conditions included prompting in the form of physical guidance for incorrect responses. During baseline the percentage of correct responses was variable and ranged from 0% to 59% correct. Correct responding increased across all participants with the introduction of interspersal, although the data for the second participant indicates an increasing trend just before the condition change. Responding remained stable (within the range of variability established during the interspersal condition) for all participants with the introduction of the low-density reinforcement condition.

Limitations of the study include potential order effects of the experimental design, and practice effects of the task, which was constant throughout the study. The introduction of novel tasks and counterbalancing the order of the conditions would address these limitations. Although all participants demonstrated improved responding from baseline to interspersal, two of the three

participants demonstrated significant variability in the interspersal and low-density reinforcement conditions. This may have been addressed by extending the study until more consistent behavior was observed, or by conducting some reversals between baseline interspersal to better determine the effect of the independent variables.

Charlop, Kurtz, and Milstein (1992) examined the role of different schedules and types of reinforcement, used in conjunction with interspersal. They measured the impact of these variables on the percentage of correct responses to acquisition tasks. The acquisition tasks were chosen from each child's school curriculum, but all were receptive instructions of various types. Additionally, the maintenance tasks were also receptive instructions but were mostly related to gaining attention (e.g., hands down, look at me). Throughout the study correct responses to acquisition tasks were reinforced on a continuous reinforcement schedule (CRF) using food and praise. Acquisition tasks were prompted on the first two trials of each session, and for the two trials following five consecutive errors. Prompted trials were not scored in the data, but were reinforced. The schedule of reinforcement for the maintenance tasks was the primary independent variable. During baseline, the schedule was CRF (praise) and VR 3 (food); during the no-reinforcement condition, neither food nor praise was provided; during the praise-only condition praise was provided on a CRF schedule. The relative schedules of reinforcement between the acquisition tasks and the maintenance tasks were manipulated and the effect of these changes on the percentage of correct responses and rate of inappropriate behavior was measured. The goal of this study was to determine if the schedule of reinforcement needed to favor the acquisition tasks to produce optimal responding.

Across all participants and tasks, the percentage of correct responses to acquisition tasks increased during the treatment condition following baseline, whether it was the no-reinforcement condition or the praise only condition. With one exception (task two for Paul), the introduction

of the first treatment condition following baseline also resulted in a decrease in the rate of inappropriate behavior. However, this initial decrease was followed by an increase and continued variable rates for the duration of the study. Once mastered, the responding was stable at a high level of success across the remaining conditions. This speaks to the durability of the responses, but does not demonstrate differential effects of the different conditions of reinforcement.

Although the initial behavior change in this study suggested an interaction between the relative schedules of reinforcement within an interspersal procedure and its effectiveness, experimental control was not demonstrated and similar responding was observed across both the no-reinforcement and the praise only conditions. Because the same instructions were targeted from the beginning through the end of the study, it is possible that the more consistent responses at the end of the study could have been due to the cumulative effects of learning. To control for this, a more constant cycling of unknown skills could have been implemented (e.g., as soon as a skill is mastered then it is no longer implemented and instead a new, unknown skill is targeted).

Chong and Carr (2005) conducted a systematic replication of the earlier Charlop et al (1992) study and found that all participants met the mastery criterion for the unknown responses during baseline conditions, and they therefore did not have the opportunity to implement the differential schedule of reinforcement. Upon conducting a direct replication, they reported similar results and again were not able to implement the different schedules of reinforcement.

The findings of Benavides and Poulson (2009) were also distinct from the results from Charlop et al (1992). The participants in Charlop et al (1992) required differential reinforcement of acquisition tasks in order to achieve an increase in the percentage of correct responses to those acquisition tasks. In contrast, Benavides and Poulson (2009) reported increases in child responding from baseline to the interspersal condition even when equal reinforcement was provided for all correct responses, whether they were for maintenance or acquisition tasks.

Similarity of interspersed tasks to acquisition tasks.

Research evaluating the impact of how similar the interspersed tasks are to the target task is sparse. The following two studies will provide first a serial comparison of the interspersal of similar (IS) versus dissimilar (ID) tasks, and second a direct comparison.

Carr and Chong (2005) conducted two experiments, both targeting the acquisition of previously unknown skills and both implemented teaching in an interspersal context. However, in one experiment the maintenance tasks were functionally and topographically similar to the acquisition response, in the other experiment the maintenance tasks were functionally and topographically dissimilar to the acquisition response. Although both experiments resulted in the acquisition of the previously unknown response, when the maintenance tasks were from the same curricula area as the acquisition response, responding had less variability and the new responses were acquired with fewer tasks. However, without a direct comparison of the two conditions, these results are not robust enough to make definitive conclusions about the role of similarity in interspersal procedures but they do indicate a need for future research in this area.

Volkert et al. (2008) conducted the only systematic exploration of the differential effects of interspersal using similar or dissimilar tasks. Experiment one combined a multielement and non-concurrent multiple baseline design to measured the percentage of correct, independent object labels across four conditions: baseline, interspersal with similar tasks (other expressive object labels), interspersal with different tasks (receptive instructions to complete motor tasks), and a varied-task condition (no interspersal) (10 acquisition tasks were presented but no maintenance). All intervention conditions resulted in improvement from baseline levels. The percentage of correct responses to unknown object labels increased from zero in baseline to similar high rates across all conditions. However, little differential responding was shown between conditions. It may be possible that the effectiveness of the teaching package that was

implemented across conditions (proactive prompting plus instructive error correction procedure plus reinforcement) resulted in the increase in the percentage of correct responding following baseline and masked the effect of the interspersal conditions. The authors note several possible limitations, including the limited number of acquisition items per instructional session, which may have generated a ceiling effect on participant performance. In addition, the rapidly alternated conditions may have also produced carryover effects, and the number of trials per session, session length, and inter-trial intervals differed between conditions. Additional research that addresses the limitations of this study is warranted.

This study will evaluate the differential impact of various interspersal procedures in the learning of four children with autism. Specifically, it will measure the differential effects of a teaching procedure that does include the interspersal of mastered skills during teaching, as compared to the same procedure, but without interspersal. It will also compare two teaching procedures that both include interspersal of mastered tasks during teaching, but one that intersperses tasks that are similar to the target task, and one that intersperses tasks that are dissimilar to the target task. The overall aim of this study is to present an empirical analysis of the acquisition rate of new skills, and the rate of challenging behaviors across each of the conditions, as well as the child's preference for teaching procedures.

Method

Participants and Setting

The researcher contacted supervisors at a local organization providing Intensive Early Intervention Behavior Therapy services to recruit children with autism to participate this study. These supervisors disseminated information regarding the study to families they served. Four families contacted the researcher via email, who then discussed the study details with each family, including the consent form approved by the *Human Subject Committee of Lawrence*

(HSCL #18989). Interested parents signed the consent and a schedule was created for the pretests and the initial sessions.

Four children with autism participated who had sight word reading included in their current or upcoming school or home based curriculum. All children received intensive early intervention behavior therapy through the same local provider, averaging 20-40 hours per week. Two of the children also attended school (see Table 1 for demographic information). Before inclusion in the study, pretests were conducted to ensure that participants were able to match words with up to four letters, and to identify 30 already mastered sight words to be used in the interspersal condition.

Sessions were conducted in the child's home, or in community locations in which therapy was already regularly conducted for that child (e.g., library, community centers). Parents, siblings, or other therapists were often present in the home during sessions, and interactions with them between teaching sessions was also common. Each session was conducted individually and efforts were made to minimize distraction.

Table 1
Participant Demographic Information

Participant	Age	Gender	IQ	Diagnosis	Attending school	Grade	Grade
1	6 years, 4 mo	M	61	Autism	Y	1st	SpEd with full support in the classroom
2	7 years, 9 mo	F	74	Autism Spectrum Disorder	Y	3rd	Private school with minimal support
3	8 years, 2 mo	M	52	Autism Spectrum Disorder	N	-	-
4	9 years, 2 mo	M	57	Autism	N	-	-

Materials

Materials included a laptop computer, prepared data sheets, a timer, potential reinforcers, flashcards, and books. Data were collected using pencil and prepared data sheets that indicated the sequence of trials to be conducted. Data were also collected using a laptop computer, and a percentage of sessions was videotaped for the purposes of reliability. The webcam on the researcher's laptop and the PhotoBooth application was used to this end. A timer was used to record the duration of each session and each break. Preferred activities and edibles that were likely to function as reinforcers for the child were both gathered from the child's home and brought to sessions by the researcher. These items were specific to the child's interests and preferences as indicated by parent and therapist report, observation of the child's choices during no-demand contexts, and through a formal preference assessment. The items included books, movies, toys, games, chips, candy, and fruit.

All flashcards were specific to the study and created by the researcher. All were 3 by 5 inches with 36 to 72-point font for the word. The largest font that would allow the word to fit on one line was used. The word was always black on a white background, but the border at the edge of the card differed in color depending on the condition (see Table 2 for more specific descriptions of the experimental stimuli).

Table 2
Description of Flashcards Used in Each Condition.

Condition	Flashcard size	Word	Background	Set on color cardstock
Baseline	3 x 5 inches	Black, 36-72 point font	White	None
No Interspersal	3 x 5 inches	Black, 36-72 point font	White	Green
I- Similar	3 x 5 inches	Black, 36-72 point font	White	Blue
I- Dissimilar	3 x 5 inches	Black, 36-72 point font	White	Yellow

Children's easy reader books that included the sight words learned in the study were used to test generalization to a different reading task. (See Appendix D for a list of books used). Because each participant was taught different words, reading materials had to be developed by the researcher to test generalization when no books were found that contained a sufficient number of the target words. The experimenter-developed materials were tested using the Flesh-Kincaid grade level assessment to ensure that the grade level matched the level of the words targeted. The font and number of words on a page of the developed materials was consistent with the published easy readers.

General Procedures

Sessions were conducted 2-8 times per day, with a minimum of a 10-min break between each session. Although the number of sessions varied across days, the number of sessions was

counterbalanced to be equal for all experimental conditions in one given day. For example, if three no interspersal sessions were conducted, three sessions of interspersal similar were also conducted that day. The only exception was if at any point in the session the child did not provide assent for three successive invitations. At that point, sessions for that day were concluded, whether the number of sessions per condition were equal or not. Sessions were scheduled for two to seven days per week, depending on the family's availability and consent.

After 30 unknown sight words were identified through the pretest, these words were matched according to number of letters and to grade level of each word, and were then randomly assigned to one of two teaching conditions. When the words were not identified from the Dolch or Fry list, and did not have a grade level associated with them, they were matched according to number of letters and the selected source (IEIBT or other curriculum). Once a word was mastered, that word was removed from the target list and a new unknown word was added to the condition. This ensured that there were always ten unknown words in each condition.

The conditions were implemented on a quasi-random basis, such that any one condition would not be presented more than twice consecutively, and that the order of the sessions, including which session started the day varied. At the start of each session, a preference assessment was conducted to identify items that would then be provided to the participant for correct responses to the target words. The preference assessment was structured as a multiple stimulus presentation, and selections were made without replacement (MSWO) procedure as outlined by DeLeon and Iwata (1996).

Child Assent Procedures

Before each session, the researcher asked for the child's assent to participate in the study, using language and communicative gestures (e.g., pointing to the table, patting the chair to invite the child to sit) that were familiar and understandable to each child. The child demonstrated

assent verbally (saying yes) or nonverbally. Nonverbal assent included reaching for the materials, nodding his or her head yes, or accompanying the researcher to the work or play area.

In addition to a verbal decline to participate, leaving the work area was also viewed as a decline to participate. In order to avoid shaping escape-motivated inappropriate behavior, if the child provided assent to participate in a session but then engaged in challenging behavior during the teaching or probe session, the researcher ended the experimental session but implemented any relevant behavior control procedures that were recommended as a part of his or her clinical regular treatment plan. Challenging behavior was child specific but included leaving the work area, aggression, stereotypy, throwing materials or property destruction and the individualized procedures to address these behaviors included extinction, use of physical prompting to complete the response, and non-exclusionary or exclusionary timeout. If at any point the child did not assent, the researcher waited at least 10 min to initiate another session and request assent. However, the session would commence as soon as the child initiated to the researcher, even if 10 minutes had not passed. The challenging behavior that occurred during experimental sessions was not frequent or severe, and did not differ significantly from typical levels for that child.

Pretesting

Because reading has at least two response components, the visual discrimination of the written word and the verbal pronunciation of the word, the first pretest was conducted to ensure that each participant was capable of a visual discrimination task with word cards. For the sake of the visual discrimination both real and invented words were included. Five to seven flashcards were set out on the table in front of the child. An example of a set of words is: late, date, dale, dele, mele, dali, lake, leak. The researcher modeled the first response by setting the identical word on top of its match. For the subsequent responses the word card to be matched was handed to the child and he or she placed it with its corresponding card. The placement and order of the

flashcards were randomized between each trial. Five trials were conducted. All four participants responded correctly to each of the five trials and were therefore included in the study.

The second pretest was aimed at identifying both known and unknown words to be included in the study. Words from the Dolch Sight Word reading list (1936), the Fry instant word list (1980), and sight words or vocabulary from the child's curriculum (both at and below the child's current age or grade level) were tested for knowledge. Words below the child's current grade level were included because most participants were delayed in their academic progress. At the pretest start, the researcher provided a generalized instruction for the child to read, such as, "let's read these words." The researcher then presented each word individually, allowing five seconds for the child response before presenting the next word. Twenty words were presented during each session with a break provided between sessions. Noncontingent social reinforcement in the form of praise, and /or brief physical interaction was provided after approximately every 2 flashcards to maintain child participation in the task. After each session, a preferred tangible reinforcer in the form of an edible or toy, or access to a preferred activity (e.g., 5-10 min movie clip or preferred game) was provided. No feedback was provided for correct responses, nor was an error correction procedure implemented for errors or non-responses. The correct and incorrect response criteria during the pretest were consistent with the correct and incorrect response criteria during the study. Unknown words were those to which the child responded incorrectly on 3/3 or 4/5 opportunities during a pretest probe. Known words were those to which the child responded correctly on 3/3 or 4/5 opportunities. The testing continued until a minimum of thirty known and thirty unknown words were identified (see Appendix E for a list of the mastered words used for each participant as the interspersed words during the Interspersal Similar condition). Words that did not meet either criterion were not included in the study. These

procedures were implemented periodically during the study when new sets of unknown words were needed.

The third pretest was aimed at identifying gross and fine motor movements that the child was capable of consistently imitating. These were then used as the interspersed skills during the Interspersal Dissimilar condition. During this pretest the researcher would give a generalized instruction "Do this," or "You try this," or "can you do this?' while modeling the movement. The same schedule of reinforcement, procedures and criteria for known and unknown responses was used as during the second pretest (see Appendix F for motor imitation movements that were used for each participant).

Dependent Variables

Dependent variables for this study included the cumulative number of sight words mastered, the rate of challenging behaviors, the percentage of correct responses to the mastered interspersed tasks, generalization to an alternate reading activity, and maintenance of mastered words. The primary dependent variable was the cumulative number of new sight words mastered per condition. The criterion for a word to be mastered was correct responding to that word in three consecutive sessions, or in four out of five consecutive sessions. For the sight words (both target and mastered) a response was scored as correct if the child: a) accurately and independently pronounced the words to the best of his or her ability (i.e., any consistent articulation errors for that child (e.g., f for th) were not scored as an error), and b) responded within 5s of the presentation of the word card. A response was scored as incorrect if the child: a) did not respond within 5s, said "I don't know," or otherwise asked for help, b) the response was not accurate (including saying the wrong word, mispronouncing the word or omitting part of the word, such as the final letter sound), or c) if other vocal behavior preceded the response (the

child sounded the word out before reading it) (See Appendix G for a list of words mastered by each participant).

The second dependent variable was the rate of challenging behavior. The duration of each session was also recorded using a timer. To determine the rate of challenging behavior the frequency of behavior was divided by the duration of each session (frequency / duration). For the no interspersal condition, only the time spent during the sight word teaching was included in this calculation, not the time spent engaged in the independent activity.

Data were also collected on an auxiliary dependent variable, the percentage of correct responses to the mastered, interspersed tasks (either words and motor imitation, depending on the condition). Generalization to novel books was also measured and probe data were collected. The same criteria for correct and incorrect responding was used as above, except only the first trial of each word was recorded.

Interobserver Agreement and Treatment Fidelity

Interobserver agreement (IOA) was assessed by having a second observer independently score a videotape of the session. Data were recorded on the correct and incorrect response to the sight word and occurrence or non-occurrence of challenging behaviors. The observer held an undergraduate degree in ABA and was familiar with the IEIBT service provider but did not work directly with any of the children who were participating in the study, nor was she involved in the administration of the intervention procedures. Agreement was assessed for 11% of all sessions, distributed across all participants (range of 6-15% of the sessions for any given participant), and distributed across all conditions (range of 10-12% for any given condition) (see Appendix H for a breakdown of which sessions were scored for reliability). IOA for the scoring of the child's response to the sight word was calculated by totaling the number of agreements divided by the number of agreements plus disagreements and converting this ratio into a percentage. IOA for

the scoring of challenging behavior, which was generally a low frequency behavior, was calculated by dividing the smaller number of behaviors by the larger number of behaviors and multiplying by 100%. Agreement was also scored for the total duration of each session. The duration was scored as agreement if the time frames recorded were the same plus or minus 15 seconds. IOA was then calculated by dividing the total number of agreements plus disagreements by the total number of agreements and converting this ratio into a percentage. To control for observer drift, videotapes of the sessions were randomized for chronology and for participants before being scored for reliability.

To assess treatment fidelity, a second observer measured the integrity of implementation of the independent variables via videotapes of sessions, using a checklist and specifically scoring: a) correct presentation of the instruction (flashcard) using the correct materials, b) correct response consequence for the target words, (i.e., delivery of the identified reinforcer contingent upon correct response to the target sight word), c) incorrect response consequence for the target words, (i.e., contingent upon error to the target sight word during treatment conditions, researcher provided a model prompt, and prompted until the child responded correctly), d) correct response consequence for the mastered instructions, (social or brief physical reinforcement contingent upon correct responses to mastered interspersed trials on a VR 2 schedule), and e) frequency of social or physical reinforcement during mastered interspersed trials and during the independent task following the no interspersal condition. Treatment fidelity was assessed on 57 sessions, distributed across all participants and conditions.

Experimental Design

An alternating treatment design including baseline conditions was implemented across participants. After baseline measures were collected, two teaching conditions were alternated. To control for multiple treatment interference different words were taught in each condition, and the

materials used in each condition was assigned a distinct color code. Additionally, when greater sufficiency in experimental control was required, a reversal design was embedded within the alternating treatment design. The number of reversals and the duration of conditions were determined on an individual basis and through ongoing visual inspection of the data.

Independent Variables

Baseline.

During the baseline condition, flashcards with black words printed on a white background were used. In a session, each word in a 10 word set was randomized and presented once for a total of 10 trials. Similar to the pretest condition, no reinforcement or feedback was provided contingent upon correct responses, nor was an error correction procedure implemented for errors or non-responses. If the child initiated a response within 5s of the presentation of the flashcard, following the response and regardless of if it was correct or incorrect, the researcher presented the next flashcard. If the child did not respond, the researcher continued to present the card for five seconds and proceeded to the next card.

Noncontingent social reinforcement in the form of praise and or brief physical interaction was provided on average after every 2 flashcards to maintain child participation in the task, and a preferred tangible reinforcer in the form of an edible or access to a preferred activity, (e.g., a brief movie clip or toy) was provided at the conclusion of each sitting. Any words that were mastered during baseline, were removed from the list and replaced with different, unknown words.

Teaching conditions.

Across all teaching conditions, (Interspersal Similar, Interspersal Dissimilar, and No Interspersal), the total number of target words and the contingencies of reinforcement and error correction procedure for the target words were consistent. The 10 target (unknown) words

assigned to the condition were randomized and each was presented once per session. For each correct response to a target word, a reinforcer identified during the preference assessment was provided. For each incorrect response to a target word an instructive error correction procedure was implemented. The researcher prompted a correct response by stating the word and the child was prompted to orally respond (e.g., "This word is _____. You say _____."). If he or she was not already doing so, the researcher also prompted the participant to visually reference the word. A minimum response for this criterion was fleeting eye contact with the word card. If the child did not respond with the correct vocal response, or did not visually reference the word, the error correction steps were repeated until both responses were emitted. On most occasions, the child responded correctly to the first trial of the error correction procedure.

Interspersal Similar (IS).

All flashcards (both target and mastered words) in this condition were presented on a background of *blue* cardstock. During this condition, the interspersed skills were mastered sight words identified as known during the pretest. The same thirty mastered words were used for the duration of that child's participation, but the words were randomized each session. The target words were also presented in a randomized order, although the sequence of trials was always three mastered to each one target. For example, (MMM(T1)MMM(T2)MMM(T3)MMM(T4)MMM(T5)MMM(T6)MMM(T7)MMM(T8)MMM(T9)MMM(T10)) where M is a mastered word and T is a target word. Reinforcement in the form of brief praise and or physical interaction was provided on a VR 2 schedule for correct responses to the mastered sight words. For incorrect responses to mastered sight words, no correction was provided and the next word was presented.

Interspersal Dissimilar (ID).

All flashcards were presented on a background of *yellow* cardstock. The flashcards included 10 target sight words and 30 cards that were blank on one side (child's view) and had a mastered motor imitation response (researcher's view). The general procedures for all teaching conditions described above were implemented. In this condition, the interspersed skills were gross motor or fine motor imitation. The ratio of mastered skills to target words and the sequence of trials was the same as the Interspersal Similar condition, except that in the above example M is a mastered motor imitation and T is a target word.

No Interspersal.

All flashcards in this condition were presented on a background of *green* cardstock. The flashcards included 10 target sight words and 30 blank cards. The blank flashcards were included to ensure each condition was associated with an equal stack of cards, and they were used as a cue to the researcher for the schedule of reinforcement for the mastered task. The general procedures for all teaching conditions described above were implemented. In this condition, the 10 target words were presented once, followed by predetermined duration of a mastered task. The order of the target words was randomized across sessions. The sequence of trials was (T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, followed by a mastered task such as drawing imitation, math, or patterns). The duration of the mastered task was variable and dependent upon the duration of the previous interspersal session. In the no interspersal sessions the mastered task was used to ensure that session duration and frequency of social reinforcement was equal across conditions. While the participant was engaged in the mastered task, the researcher continued to flip through the flashcards at a pace similar to the interspersal condition and provided brief social or physical reinforcement approximately every 2-3 cards. In this way, sessions were yoked, such that the duration and frequency of social reinforcement of the last interspersal session indicated the

duration and frequency of reinforcement for the independent activity during the no interspersal session.

Generalization and maintenance measures.

At the conclusion of the study, but before maintenance probes were conducted, the researcher probed generalization of the words mastered on flashcards during the study to an easy reader or individualized book. The child's response (correct or incorrect) was scored for each word learned during the study. During the generalization probe the researcher assisted the child with other unknown words from the children's easy reader, and as necessary would point to the word that the child should read. For correct responses to the target words, social praise was provided continuously and tangible reinforcement was provided on a VR 2 schedule. For incorrect responses, the error correction procedure used during the study was implemented. For those words that were present multiple times in the easy reader, data were collected and reported for the first opportunity to read a targeted word, although the researcher provided reinforcement for subsequent correct responses and prompted any subsequent incorrect responses.

Maintenance of the words was probed approximately two weeks and four weeks after the conclusion of the study. Each word was tested 3-5 times, using the same criteria for identifying known and unknown words during the pretest. During the maintenance probes, words from all conditions were randomized and 20 words were presented in a session. Ten-minute breaks were provided between sessions. For correct responses to the target words, a continuous schedule of social praise or brief physical reinforcement was provided. Edibles were provided on a VR 2 schedule for correct responses. In order to ensure an accurate measure of maintenance, and because each word was tested between 3-5 times, no feedback or reinforcement was provided for errors until the conclusion of the maintenance probe.

Immediately following each maintenance probe, the words that were not maintained were reviewed and the error correction procedure was implemented until the child was able to respond correctly and independently upon presentation of the words. Once he or she demonstrated one correct independent response for each word, the session ended.

Concurrent chains preference assessment: Treatment condition.

Each day, after exposure to both of the current teaching conditions, the child was provided with the opportunity to choose which teaching condition would be conducted (e.g., the child could choose the third and the fourth sessions of that day, or the third and the fifth sessions that day). A concurrent chains arrangement was used in which the initial link was the selection of the color-coded stack of cards, and the terminal link was the implementation of the condition paired with the selected color (see Hanley, Piazza, Fisher, Contrucci, & Maglierei (1997) and Leaf, Sheldon, & Sherman (2010) as examples of preference assessment of instructional procedures). The different colored stacks of cards were presented face down to ensure that the child was not choosing based on the presented word, and the researcher presented a generalized instruction for the child to choose. For example “Which stack should we do?” or “You pick”. Once he or she selected the condition, that condition was implemented using the procedures described above.

Social Validity

In addition to the child assent procedures, and the measures of child preference for treatment conditions, the social validity of the overall study from the parents was also collected. Following the completion of the study a parent satisfaction survey was sent to the parents of each of the four participants via U.S. postal mail. A stamped return-addressed envelope was also sent. No personally identifiable information was requested, so the parents could respond anonymously if they chose. The survey solicited feedback regarding satisfaction with the goals and procedures,

the outcome, the research staff, and overall experience. A five point Likert-type scale was used with one representing the least satisfaction and five representing the most satisfaction for the qualitative feature (satisfied, likely, helpful). Three open ended questions were also included to assess what the parents liked most or least about the study, and a general question about changes in their child's ability to read as a result of the study (see Appendix I).

Results

The current study collected data on five dependent variables across two phases for each of four participants. Results will be presented to allow for both within and across participant comparisons. First, individual data will be provided on each dependent variable and for both phases of the study. Then those results will be summarized across participants. The child preference data are reported for each participant, but the graphs are summarized across participants and are presented only in the summary data section.

Participant 1

Number of words mastered.

Figure 1 displays the cumulative number of words mastered across baseline, interspersal similar (IS) and no interspersal conditions. The sessions are represented on the abscissa and the cumulative number of total words mastered is represented on the ordinate. Participant 1 met the mastery criteria for one word in each group during the first baseline (sessions 1-16). During the first comparison (s: 17-42) he mastered 15 words in the IS condition and 14 in the no interspersal condition. A reversal was conducted to determine if similar patterns of responding would occur with additional word sets. During the second baseline (s: 43-56), no words were mastered in either condition. During the second comparison (s: 57-88) participant 1 mastered twice as many words in the IS condition than the no interspersal condition, with 18 and 9 words respectively. Another reversal was conducted; again no words were mastered in either group during Baseline 3

(s: 89-106) and again, during comparison 3 (s: 107-132) participant 1 mastered more words under the IS condition than the no interspersal condition with 12 and 9 words respectively.

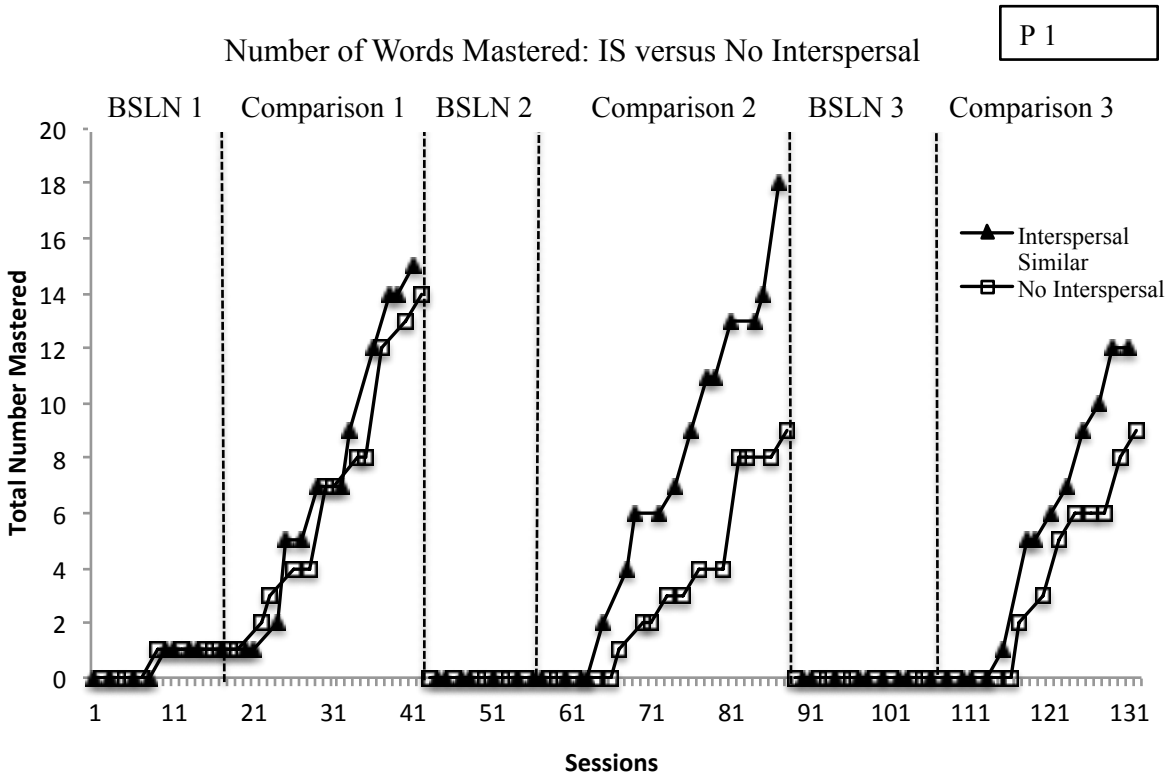

Figure 1. Cumulative number of words mastered during baseline and a comparison of two alternating teaching conditions, interspersal with similar skills (other mastered sight words) and no interspersal of mastered skills.

Figure 2 depicts the cumulative number of words across baseline, IS, and ID conditions. During baseline (s: 1-14), zero words were mastered in either group. During the comparison (s: 15-28) participant 1 mastered significantly more words when similar skills were interspersed (IS) than when dissimilar skills were interspersed (ID), 7 words compared to 2 words.

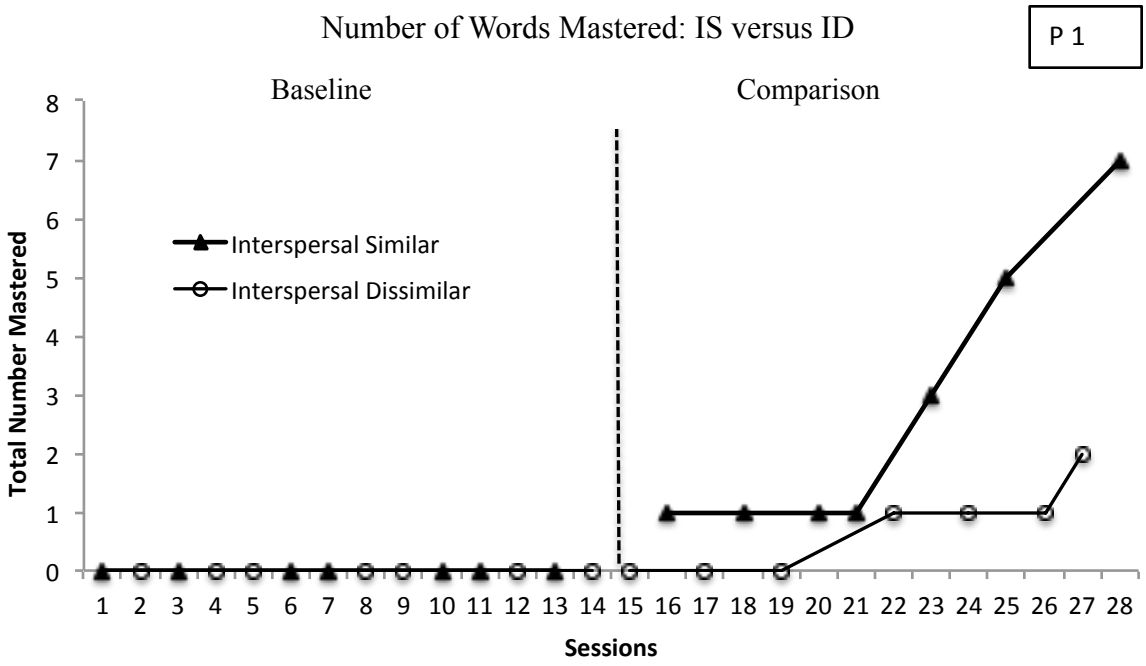

Figure 2. Cumulative number of words mastered by participant 1 during baseline and a comparison of two alternating teaching conditions, Interspersal Similar (mastered sight words) and Interspersal Dissimilar (mastered fine motor imitation).

Rate of challenging behavior.

Figure 3 depicts the rate of challenging behavior across baseline, IS, and no interspersal conditions. The mean number of behaviors per minute per condition is noted below the sessions. Overall, the rate of challenging behavior was variable, but fairly equally distributed across all conditions. During baseline 1 (s: 1-16) challenging behavior occurred at an average rate of 0.88 behaviors per minute (range 0-2.4 behaviors per minute). In comparison 1 (s: 17-42) the average Rate was 0.32 (range 0-1.07) in the IS condition and 0.59 (range 0-1.53 behaviors per minute) in the No Interspersal condition. During baseline 2 (s: 43-56) the average Rate was 0.83 (range 0-1.67). In comparison 2 (s: 57-88) the average Rate was 0.69 (range 0-1.80) in the IS condition and 0.86 (range 0-2.42) in the no interspersal condition. In baseline 3 (s: 89-106) the average rate was 0.74 (range 0-3.08), and finally during comparison 3 (s: 107-132) the average rates were

0.41 behaviors per minute (range 0-1.02) and 0.76 behaviors per minute (range 0.31-1.44) in the IS and no interspersal conditions respectively.

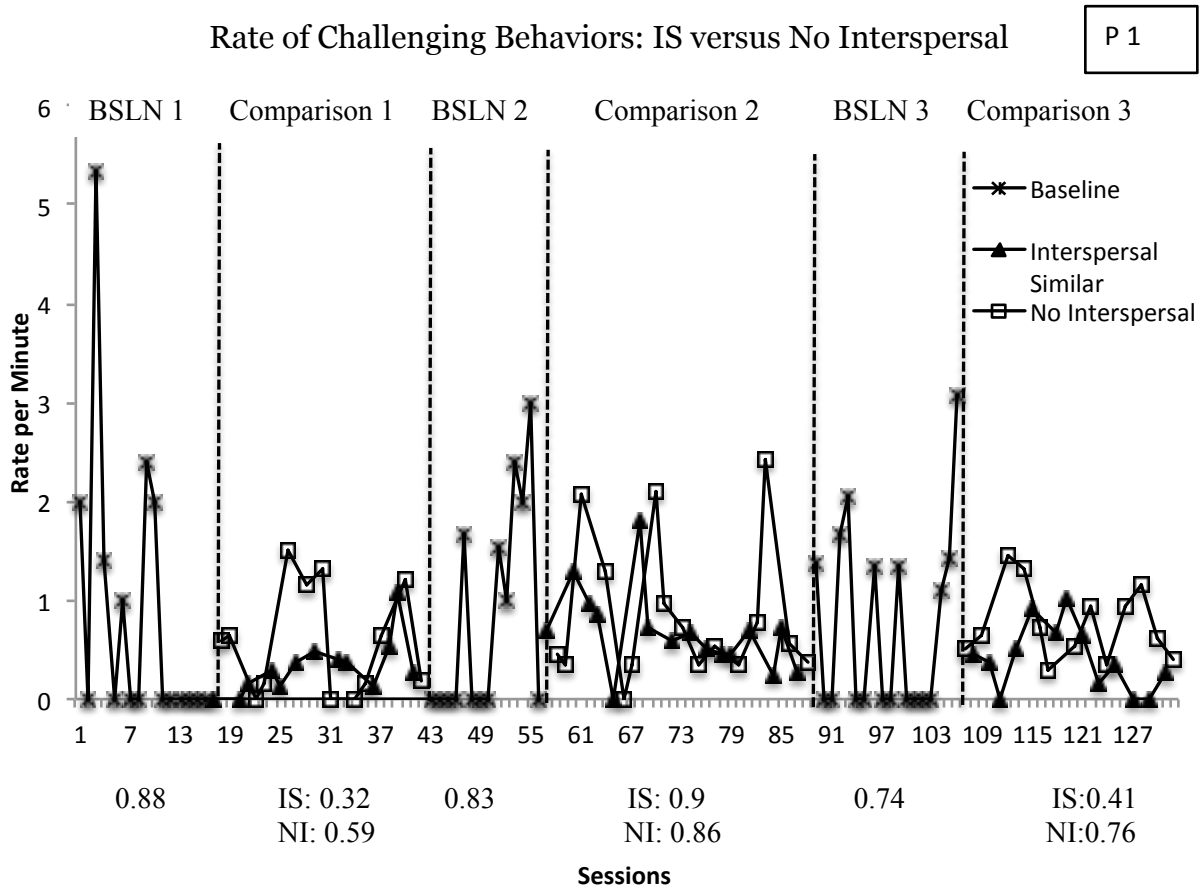

Figure 3. Rate per minute of challenging behavior demonstrated by participant 1 across baseline and a comparison of two alternating teaching conditions, IS and No Interspersal. The average rate per minute is indicated beneath each condition.

Figure 4 shows the rate of challenging behavior across baseline, IS, and ID conditions. Rates of challenging behavior during this phase of the study were lower for participant 1 than during the initial phase of the study. The scale of the ordinate has been adjusted to better display the data included of this phase. During baseline, (s: 1-14) the average rate of challenging behavior per minute was 0.43 (range 0-1.67). In the comparison (s: 15-28) the rates were again variable and fairly equally distributed initially. However, starting at session 24 an increasing trend can be seen in the ID condition, and a decreasing trend can be seen in the IS condition beginning with session 23. The end points are 1.17 and 0.23 behaviors per minute, respectively.

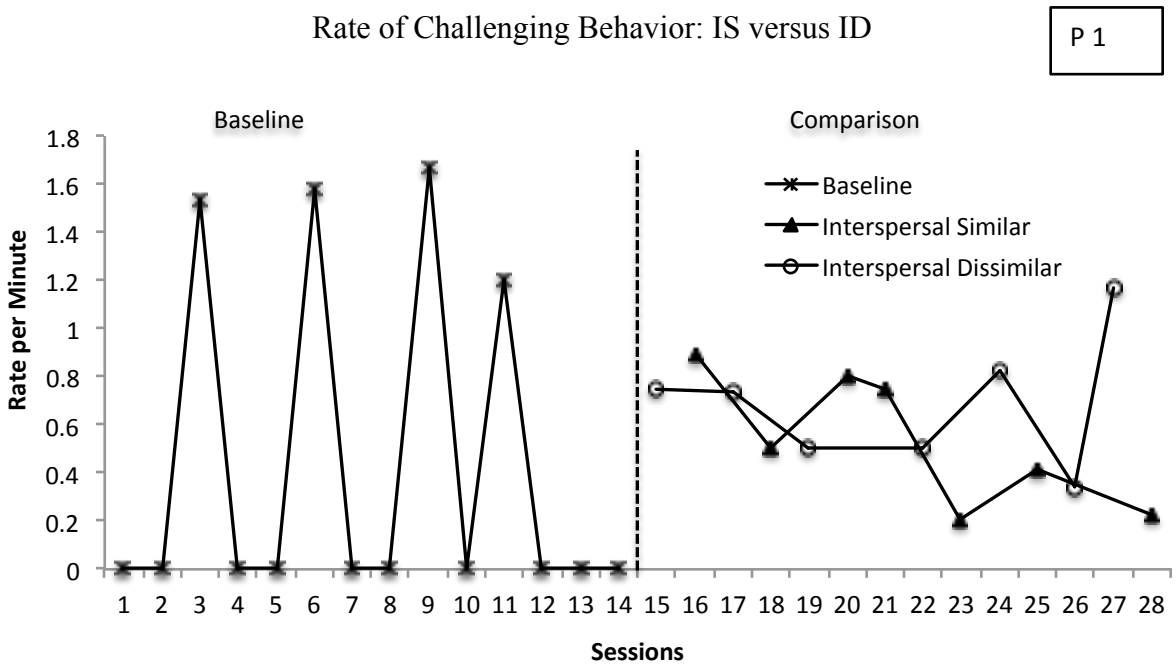

Figure 4. Rate of challenging behaivor per minute demonstrated by participant 1 across baseline and a comparison of two alternating teaching conditions, IS and ID.

Child preference.

When provided a choice between the no interspersal and the IS conditions, participant 1 demonstrated a preference for the interspersal condition (IS). When provided a choice between the IS and ID conditions, he demonstrated a preference for interspersal of similar skills (IS) (see Figures 21 and 22 for more detail).

Generalization and maintenance.

Figure 5 depicts the total number of words mastered in each condition, the total number of words that met the criteria for generalization to novel books, and the total number of words that met the criteria for maintenance during two maintenance probes. The first maintenance probe was conducted one week after the completion of the study with the second probe conducted three weeks after study completion. Generalization was highest for the words taught in the IS condition and lowest for the words taught in the ID condition.

Of the 51 words that were mastered in the IS condition, participant 1 demonstrated generalization for 31 words (61% of the total mastered), maintained 24 words (47%) during the first probe, and 25 words (49%) during the second probe. Of the two words that were mastered in the ID condition, he demonstrated generalization for 1 (50%), maintained 2 (100%) during the first probe, and one (50%) during the second probe. Of the 31 words that were mastered in the no interspersal condition, he demonstrated generalization for 18 (58%), maintained 14 (45%) during the first probe, and 22 (71%) during the second probe.

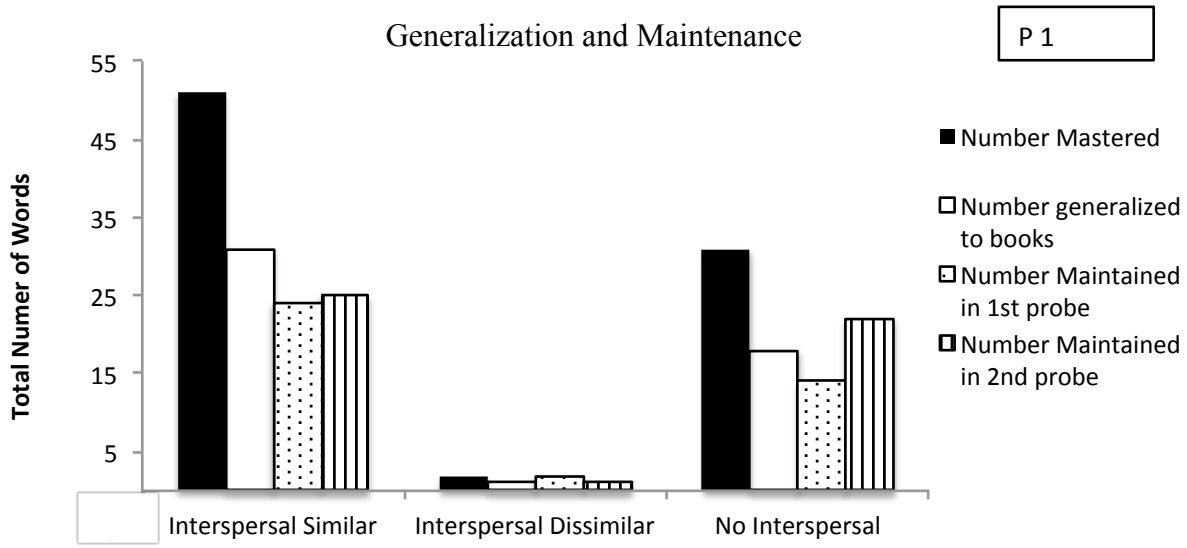

Figure 5. Total number of words in each condition that were mastered, generalized to novel books, maintained in the first maintenance probe, and those that were maintained in the second maintenance probe.

Participant 2

Number of words mastered.

Figure 6 illustrates the cumulative number of words mastered by participant 2 during baseline, IS, and the no interspersal conditions. No words were mastered during baseline (s: 1-12). During the comparison (s: 13-40), participant 2 demonstrated mastery of words in both conditions, but the data paths were consistently separated indicating a more rapid mastery of

words that were taught within the IS condition. She mastered a total of nine words in the IS condition and five in the no interspersal condition.

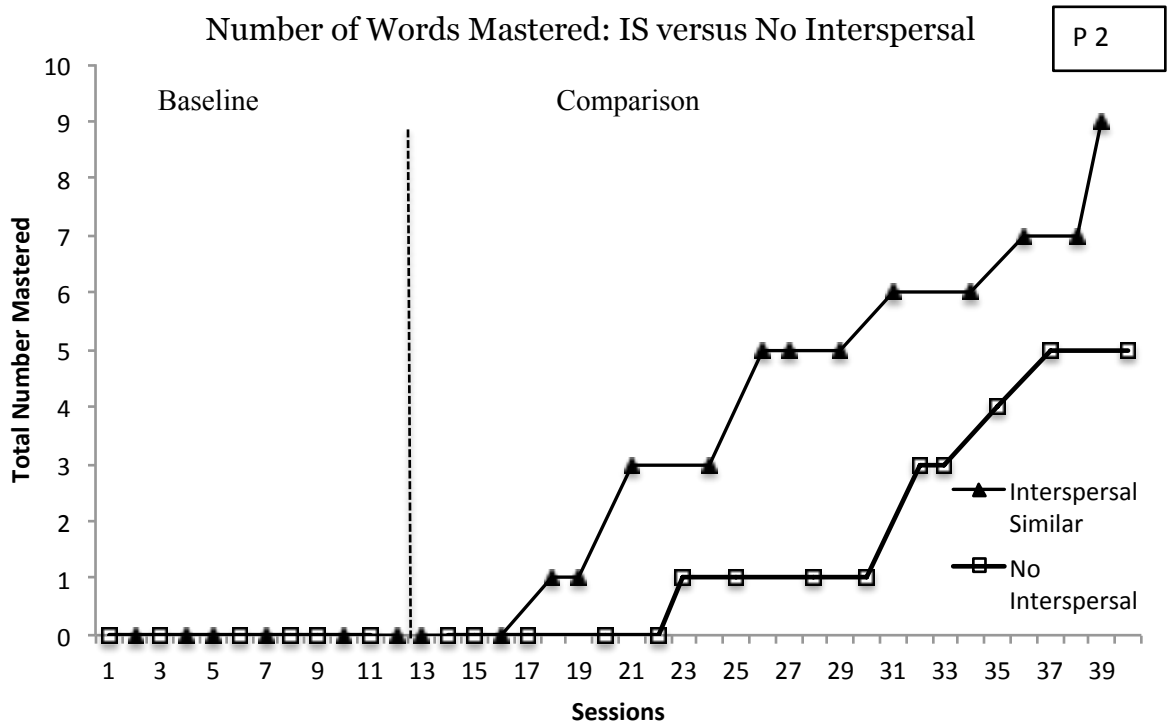

Figure 6. Cumulative number of words mastered by participant 2 during baseline and a comparison of two teaching conditions: IS and No Interspersal.

Figure 7 shows the cumulative number of words mastered by participant 2 during baseline, IS and ID conditions. No words were mastered during baseline 1 (s: 1-12). Rate of mastery was similar for both conditions during comparison 1 (s: 13-44) with 11 mastered in IS and 9 in ID. The data paths crossed at session 19 and remained within 2 points of each other throughout the comparison. Because there was little differentiation between the two conditions a reversal was conducted to determine if a similar pattern of responding would occur with new word sets. Again, no words were mastered during baseline 2 (s: 45-56) and responding was closely matched across both conditions during comparison 2 (s: 57-84). However in the second comparison, 12 words were mastered in the ID condition and 9 in the IS condition.

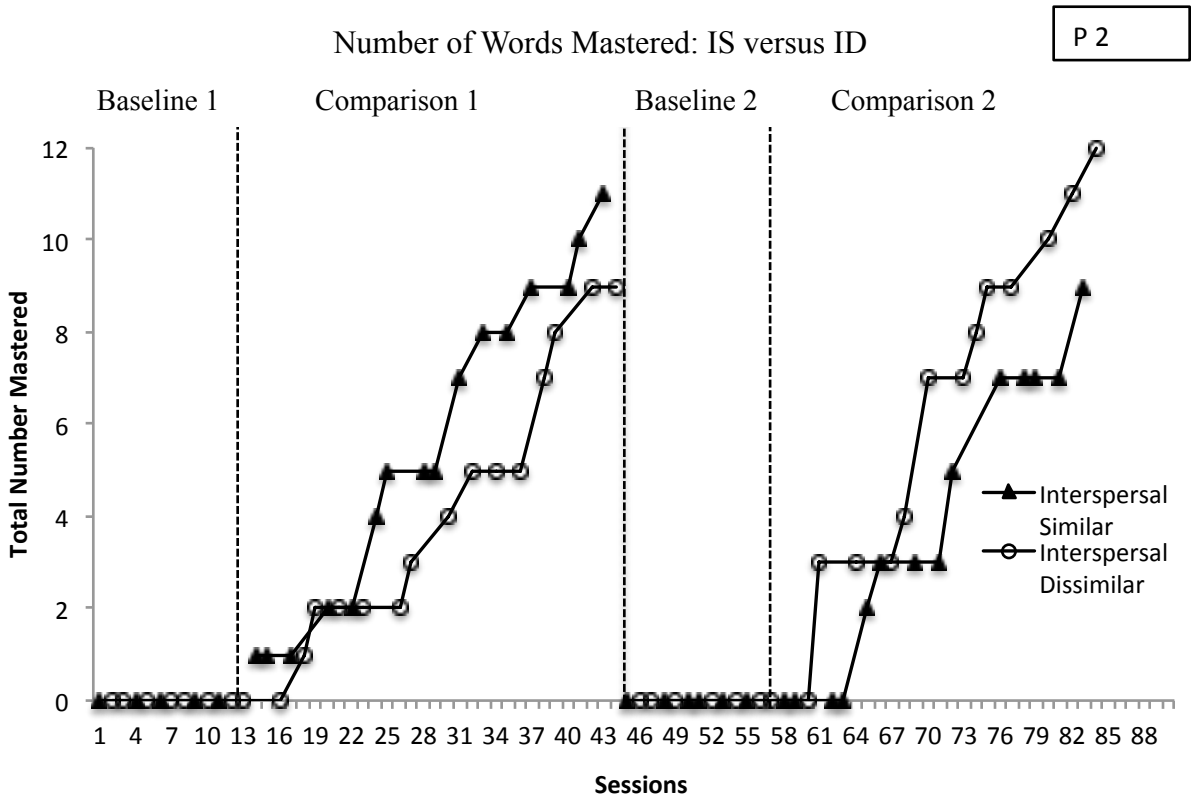

Figure 7. Cumulative number of words mastered by participant 2 during baseline and a comparison of two teaching conditions: IS and ID.

Rate of challenging behavior.

Figure 8 illustrates the rate of challenging behavior across baseline, the IS and the no interspersal conditions. Zero challenging behaviors occurred during baseline (s: 1-12). During the comparison (s: 13-40) challenging behavior was infrequent, only occurring in six of the 28 sessions, and these were fairly equally distributed across both conditions. The most common behavior was no response within 5 seconds, either due to inattention or slow responding. The average rate of challenging behavior was 0.09 across both of the teaching conditions, with a range of 0 to 0.34 behaviors per minute for the IS and a range of 0 to 0.63 for no interspersal.

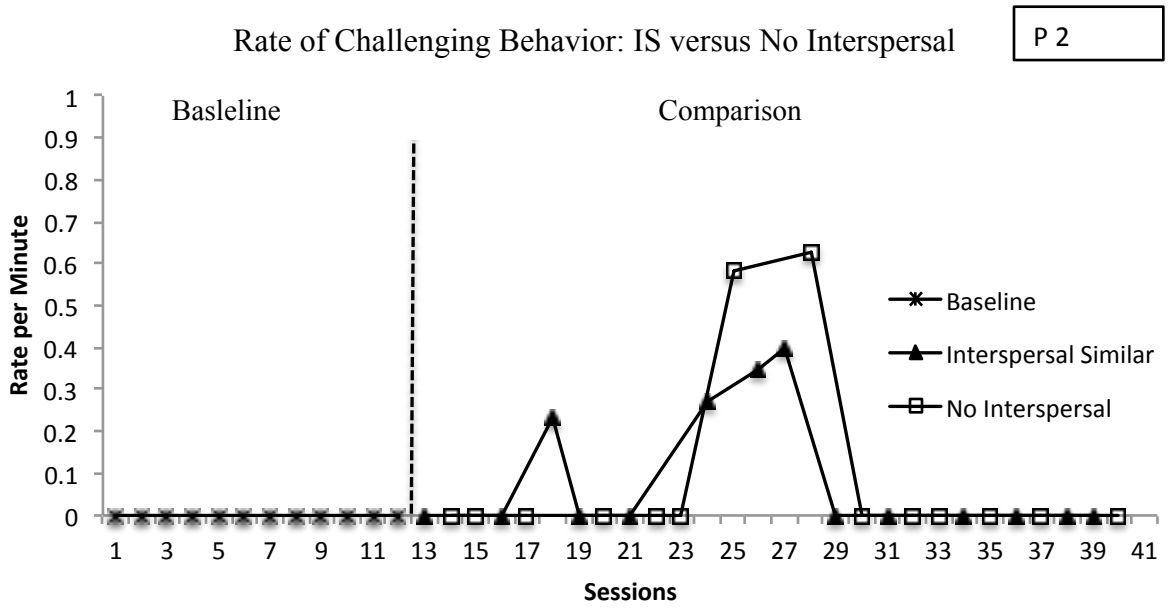

Figure 8. Rate of challenging behavior per minute demonstrate by participant 2 across baseline and a comparison of two alternating teaching conditions, IS and No Interspersal.

Figure 9 displays the rate of challenging behavior across baseline, and the IS and the ID conditions. Zero challenging behaviors were exhibited during baseline 1 (s: 1-12). During comparison 1 (s: 13-44), most challenging behaviors occurred during the ID sessions, but this was limited to four sessions. There was one session during which challenging behavior occurred during the IS condition. The average rate of challenging behaviors per minute was 0.02 (range 0-0.38) during the IS condition, and was 0.07 (range 0-0.33) for the ID. Again, no challenging behavior occurred during baseline 2 (s: 45-56). The challenging behavior that occurred during comparison 2 (s: 57-84) was again fairly equally distributed across both interspersal conditions but was more frequent than in the first comparison. During the IS condition, challenging behavior occurred at an average rate of 0.17 (range 0-0.76) and in the ID condition 0.09 (range 0-0.62).

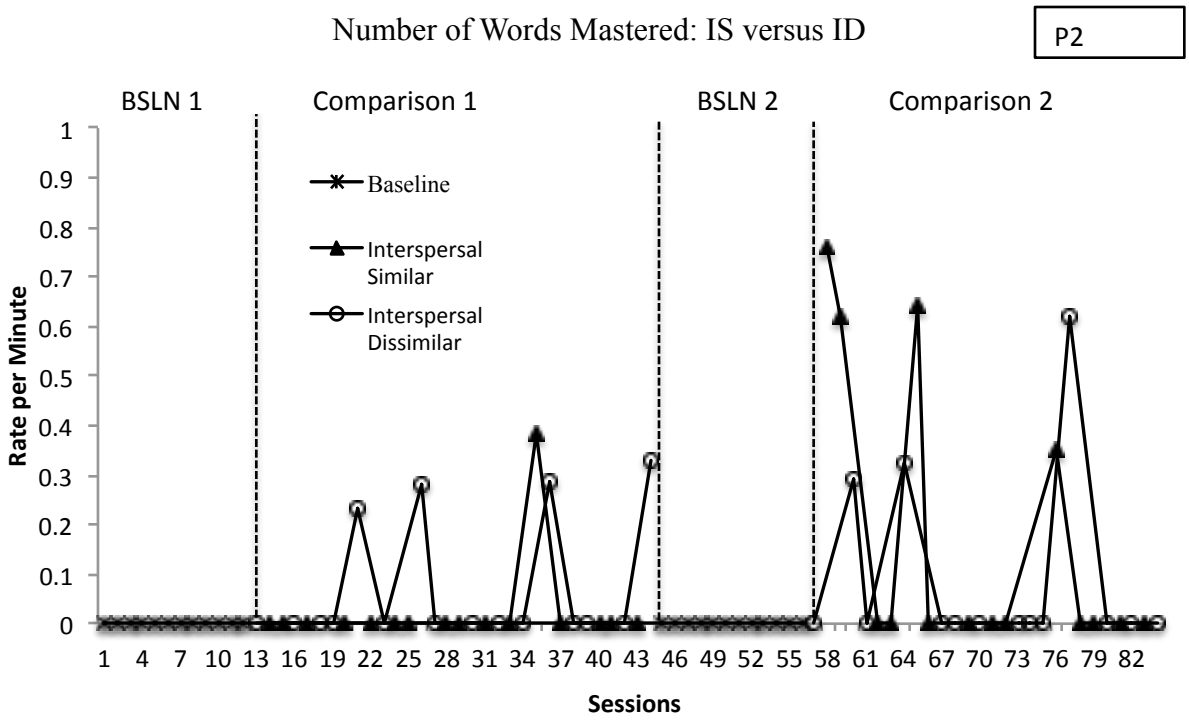

Figure 9. Rate of challenging behavior per minute demonstrated by participant 2 across baseline and a comparison of two alternating teaching conditions, IS and ID.

Child preference.

When given the opportunity to choose between the IS and no interspersal conditions, Participant 2 demonstrated no preference between the two. When she was given the choice of the IS or the ID conditions. She demonstrated a preference for the IS condition (see Figures 21 and 22 for more detail).

Generalization and maintenance.

Figure 10 illustrates the total number of words mastered per condition, the total number of words that met generalization and maintenance criteria by participant 2 during a generalization probe to novel books and during 2 maintenance probes. The first maintenance probe was conducted three weeks after the completion of the study with the second probe conducted 4.5 weeks after study completion. Participant 2 demonstrated generalization at a comparable level across the two interspersal conditions (69 and 67% of the total number words

mastered per condition). Specifically, she met the criteria for generalization for 20 of the 29 words mastered in the IS condition and for 14 of the 21 words mastered in the ID condition. Generalization was significantly lower for the words mastered in the no interspersal condition (20%). A similar pattern was also demonstrated during the first maintenance probe, when 62% and 57% of the words mastered in the interspersal conditions were maintained versus only 20% from the no interspersal condition.

Of the 29 words that were mastered in the IS condition, participant 2 demonstrated generalization for 20 of them (69%), maintained 18 (62%) during the first probe, and 24 (83%) during the second probe. Of the 21 words that were mastered in the ID condition, she demonstrated generalization for 14 (67%), maintained 12 (57%) during the first probe, and 21 (100%) during the second probe. Of the five words that were mastered in the no interspersal condition, she demonstrated generalization for one (20%), maintained one (20%) during the first probe, and five (100%) during the second probe.

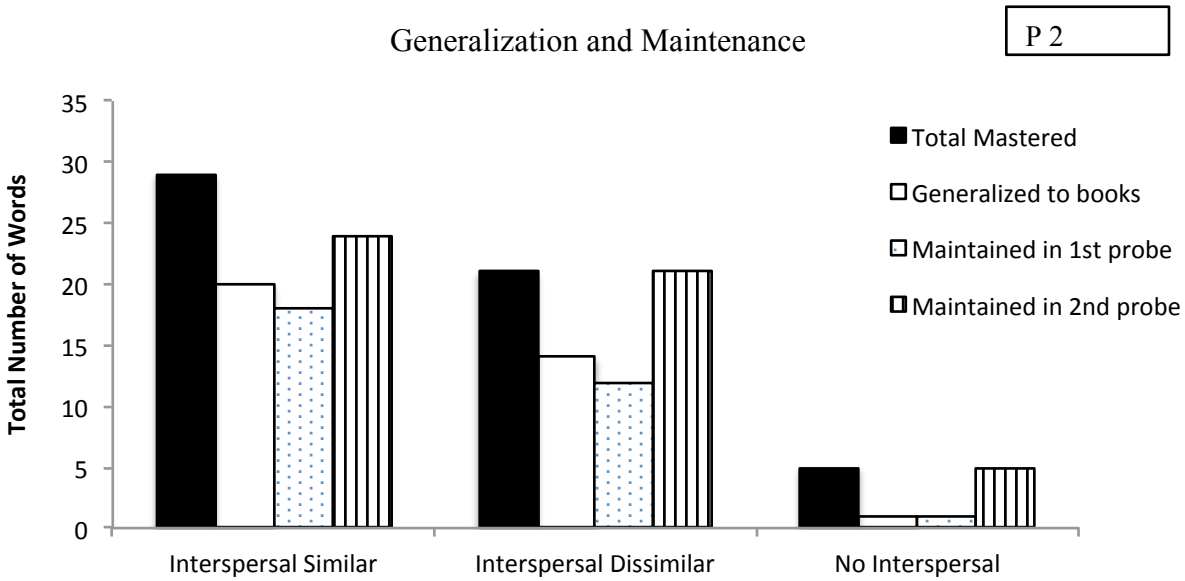

Figure 10. Total number of words that were mastered during the study in each condition, total number that were generalized to novel books, maintained in the first and second maintenance probes.

Participant 3

Number of words mastered.

Figure 11 displays the cumulative number of words mastered by participant 3 during baseline, IS and No Interspersal conditions. No words were mastered during baseline 1 (s: 1-12). During comparison 1 (s: 13-60) 8 words were mastered in the IS condition and 4 in the no interspersal. Because the results were inconsistent during the comparison, as indicated by the initial faster rate of mastery in the no interspersal condition and the crossing of the data paths at session 41, a reversal to baseline conditions was conducted. Again, no words were mastered during baseline 2 (s: 61-82). Finally, during comparison 2 (s: 83-126) the pattern of responding was similar to comparison 1. There was little differentiation between the data paths of the two conditions until session 112. Beginning with session 112 and for the rest of the comparison, Participant 3 again demonstrated faster mastery of novel sight words in the IS condition. He mastered a total of 9 words in the IS and 6 in the no interspersal condition.

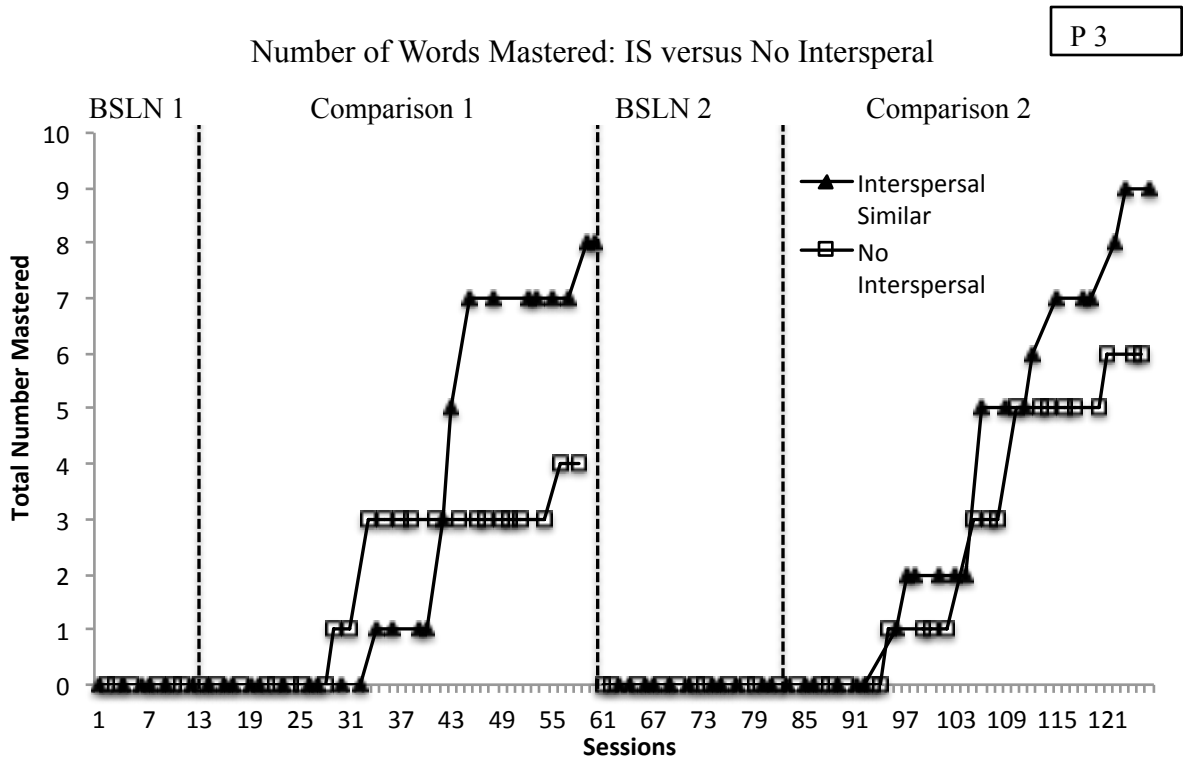

Figure 11. Cumulative number of words mastered by participant 3 during baseline and a comparison of two alternating teaching conditions: IS and No Interpsersal.

Figure 12 depicts the cumulative number of words mastered by participant 3 during baseline, IS, and ID conditions. During baseline (s: 1-20), one word from the IS word set was mastered. During the comparison, (s: 21-54) 6 words were mastered in the IS condition and 4 in the ID.

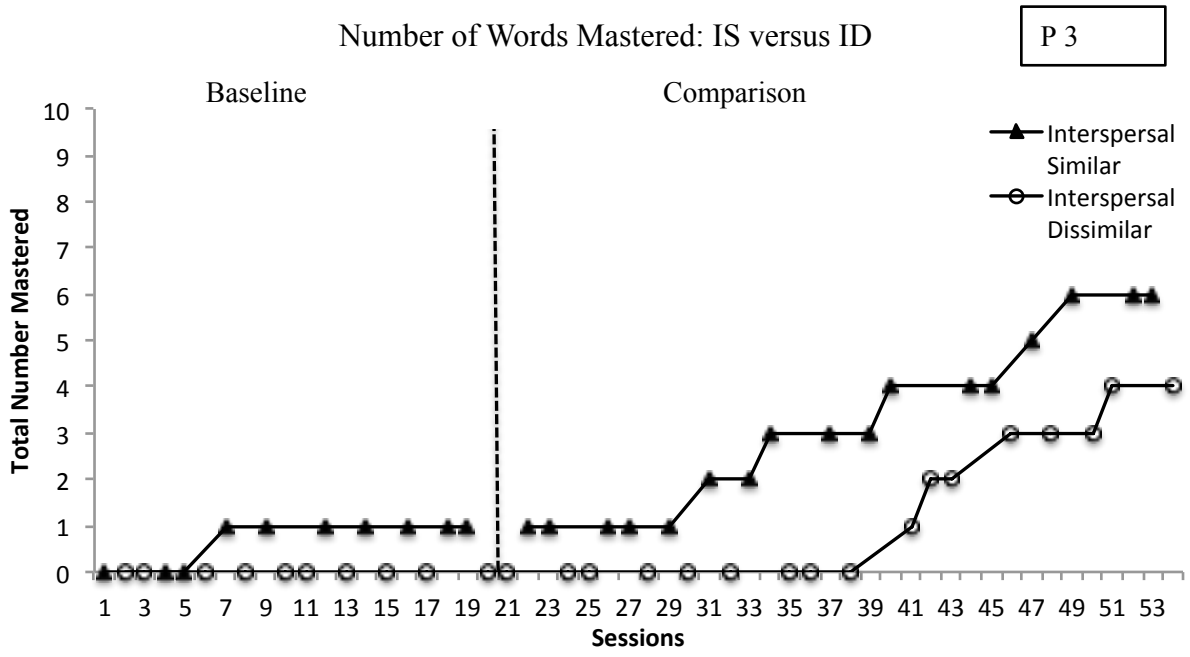

Figure 12. Cumulative number of words mastered by participant 3 during baseline and a comparison of two alternating teaching conditions: IS and ID.

Rate of challenging behavior.

Figure 13 illustrates the rate of challenging behavior during baseline, IS, and no interspersal conditions. Participant 3 engaged in the highest rates of challenging behavior during the baseline conditions. During baseline 1 (s: 1-12) the average number of challenging behavior per minute was 0.29 (range 0-1.53 behaviors per minute). During comparison 1 (s: 13-60), the average rates for IS and no interspersal were 0.08 (range 0-0.73) and 0.31 (range 0-3.08), respectively. During baseline 2 (s: 61-82), the average rate was 0.41 behaviors per minute (range 0-2.26). The rates during comparison 2 (s: 83-126) decreased but remained variable across sessions and both conditions. The average rate for the IS condition was 0.18 (range 0-0.84) and the average rate for no interspersal was 0.20 (range 0-0.99).

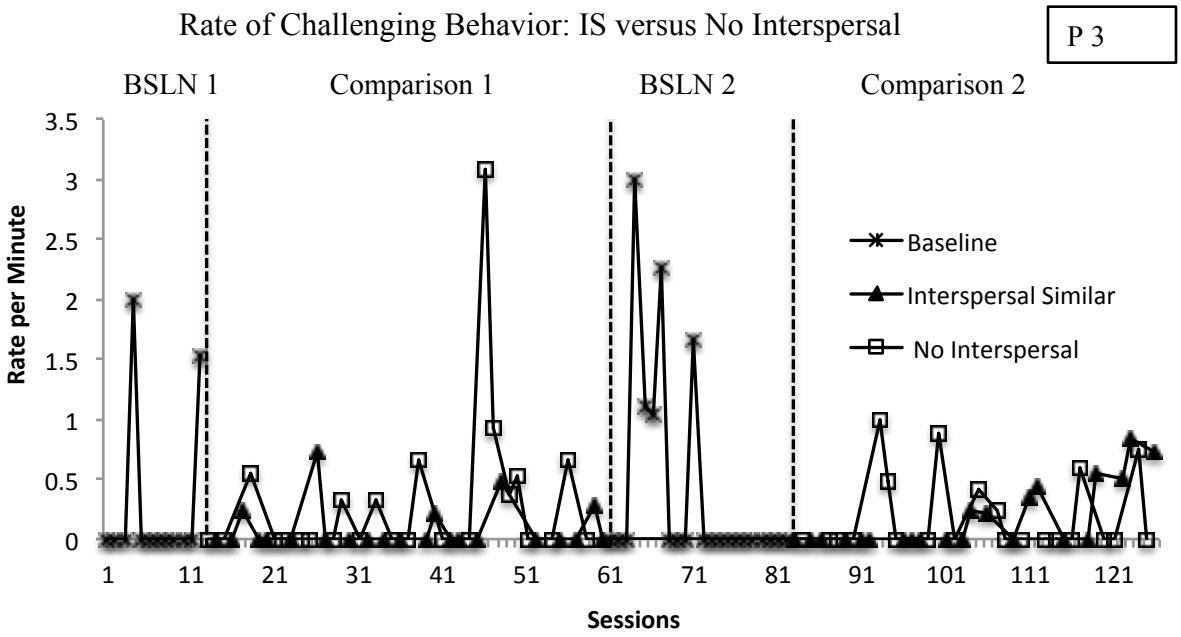

Figure 13. Rate of challenging behavior per minute of participant 3 across baseline and a comparison of two alternating teaching conditions: IS and no interspersal.

Figure 14 displays the rate of challenging behavior per minute demonstrated by participant 3 across baseline and 2 alternating treatment conditions: IS and ID. Similar to the previous figure, the sessions with the highest rates of challenging behavior occurred during the baseline condition. During baseline (s: 1-20) the average rate of challenging behavior per minute was 0.26 (range 0-2.14). Challenging behavior reduced to 0 at session 8 and remained at zero for the duration of baseline measures. During the comparison (s: 21-54) challenging behavior was variable at a low rate across both conditions. The average rate during IS was 0.20 (range 0-0.57); and the average rate during ID was 0.13 (range 0-0.71).

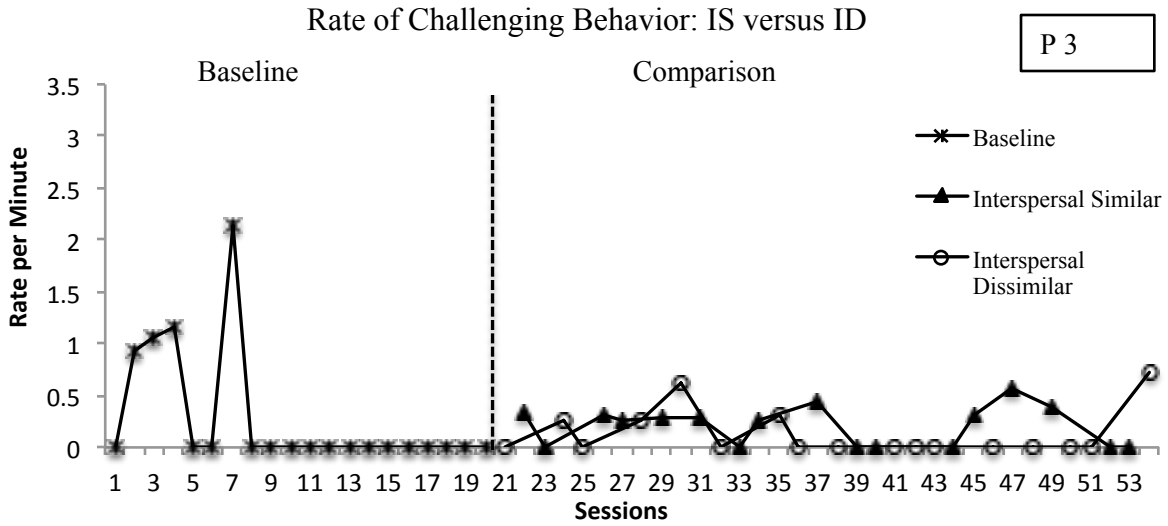

Figure 14. Rate of challenging behavior per minute of participant 3 across baseline and a comparison of two alternating teaching conditions: IS and ID.

Child preference.

When presented with the opportunity to choose between either the IS or no interspersal conditions, and when presented with the opportunity to choose between the IS or ID conditions, participant 3 demonstrated a stronger preference for the IS condition (see Figures 21 and 22 for more detail).

Generalization and maintenance.

Figure 15 depicts the total number of words that were mastered in each condition and the total number of these words that then met the generalization and maintenance criteria during a generalization probe to novel books and during 2 maintenance probes. The first maintenance probe was conducted two weeks after the completion of the study with the second probe conducted four weeks after study completion. The percentage of words mastered by participant 3 that were generalized to books was higher for the words taught in the interspersal conditions than for the words taught without interspersal (78% and 75% as compared to 40%). The percentage of words maintained at two and four weeks was highest for the words taught with ID (100%), although the total number of words mastered in that condition was less than the other conditions.

Of the 23 words that were mastered in the IS condition, participant 3 demonstrated generalization for 18 (78%), maintained 11 (48%) during the first probe, and 14 (61%) during the second probe. Of the four words that were mastered in the ID condition, he demonstrated generalization for three (75%), maintained four (100%) during the first probe, and four (100%) during the second probe. Of the 10 words that were mastered in the no interspersal condition, he demonstrated generalization for four (40%), maintained five (50%) during the first probe, and 5 (50%) during the second probe.

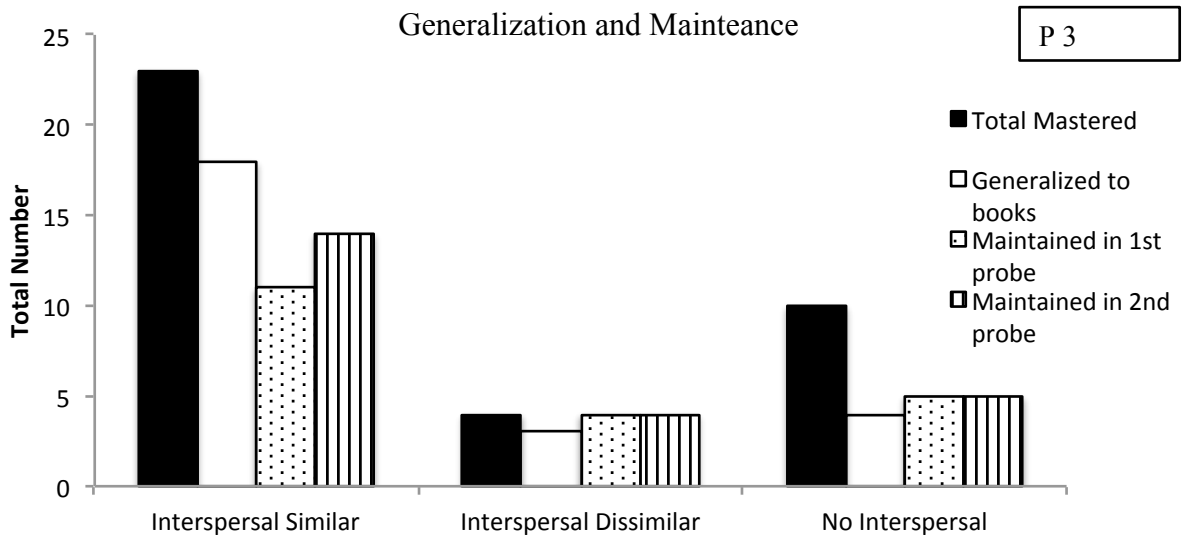

Figure 15. Total number of words that were mastered during the study in each condition, total number that were generalized to novel books, and maintained in the first and second maintenance probes.

Participant 4

Number of words mastered.

Figure 16 illustrates the cumulative number of words mastered by participant 4 in baseline, and the IS and no interspersal conditions. No words were mastered during baseline 1 (s: 1-14). During comparison 1 (s: 15-38) mastery rates were similar and undifferentiated across both conditions. He mastered nine words in the IS condition and eight in the no interspersal condition. Similar patterns of responding were demonstrated during the reversal. Again, no

words were mastered in baseline 2 (s: 39-50) and during comparison 2 (s: 51-76), seven words were mastered in the IS, and six in the no interspersal condition.

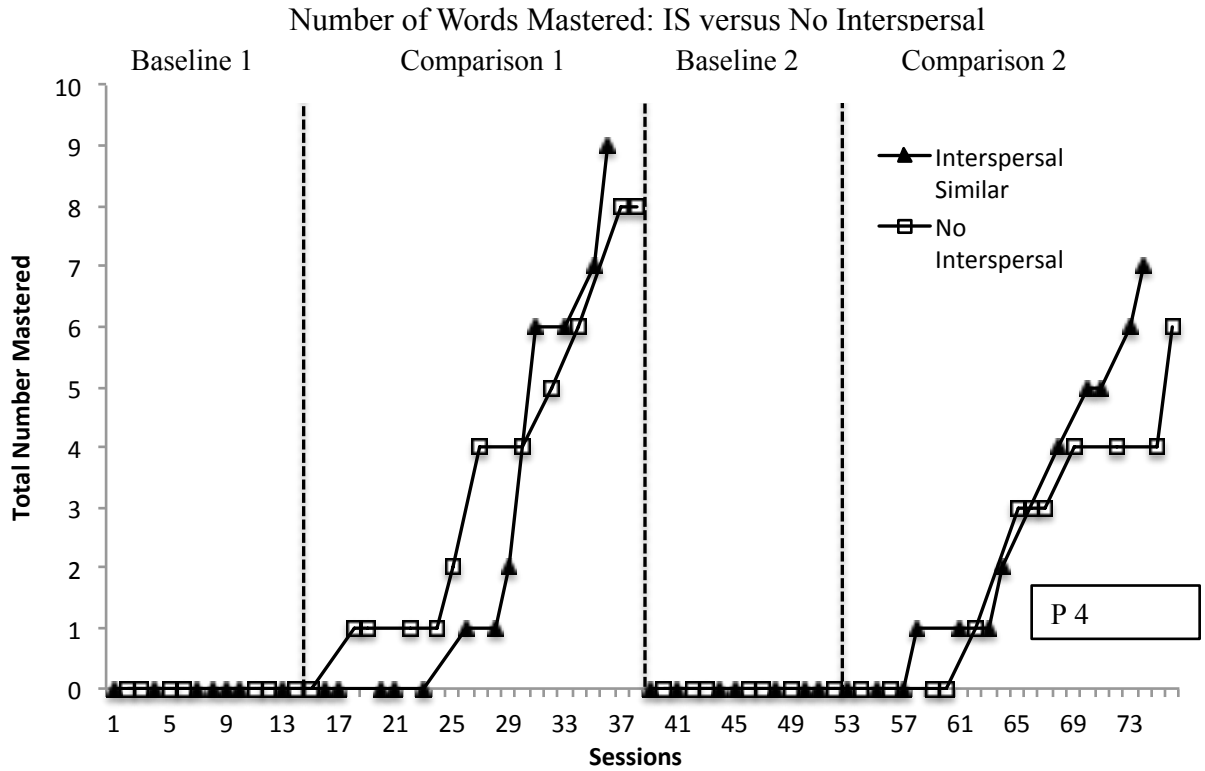

Figure 16. Cumulative number of words mastered by participant 4 during baseline and a comparison of two alternating teaching conditions: IS and No Interspersal.

Figure 17 portrays the cumulative number of words mastered by participant 4 during baseline, and the IS and ID conditions. One word was mastered during baseline 1 (s: 1-20). Responding during comparison 1 (s: 21-38) was similar across both conditions, after a delay, he mastered six words in the IS condition and seven in the ID. No words were mastered during baseline 2 (s: 39-52). Responding during comparison 2 (s: 53-84) was similar to that of comparison 1, mastery rates were undifferentiated across both conditions. He mastered 10 and 12 words in the IS and ID conditions respectively.

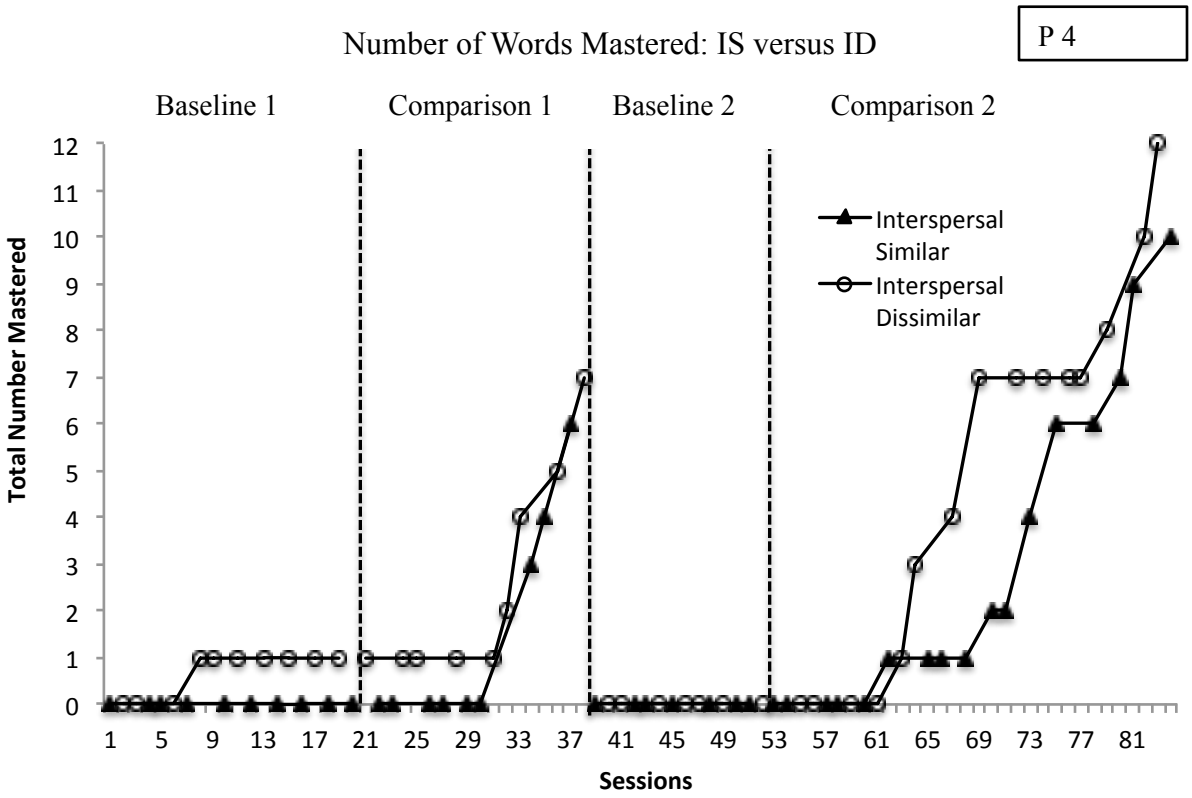

Figure 17. Cumulative number of words mastered by participant 4 during baseline and a comparison of two alternating teaching conditions: IS and ID.

Rate of challenging behavior.

Figure 18 depicts the rate of challenging behavior for participant 4 across baseline and the IS versus the no interspersal teaching conditions. No challenging behavior occurred during baseline 1 (s: 1-14). During comparison 1 (s: 15-38), the rate of challenging behavior during the IS condition was 0.10 (range 0-0.50), but these were confined to four sessions. No challenging behaviors occurred during the no interspersal condition. During baseline 2 there were two sessions during which challenging behavior occurred. The average rate for the condition was 0.25 (range 0-1.82). The frequency of challenging behaviors increased during comparison 2 (s: 51-76), occurring in three sessions for each condition. The rate of challenging behavior was equal across both conditions at 0.19 behaviors per minute the ranges were 0 to 0.89 for the IS condition 0 to 0.82 during the no interspersal condition.

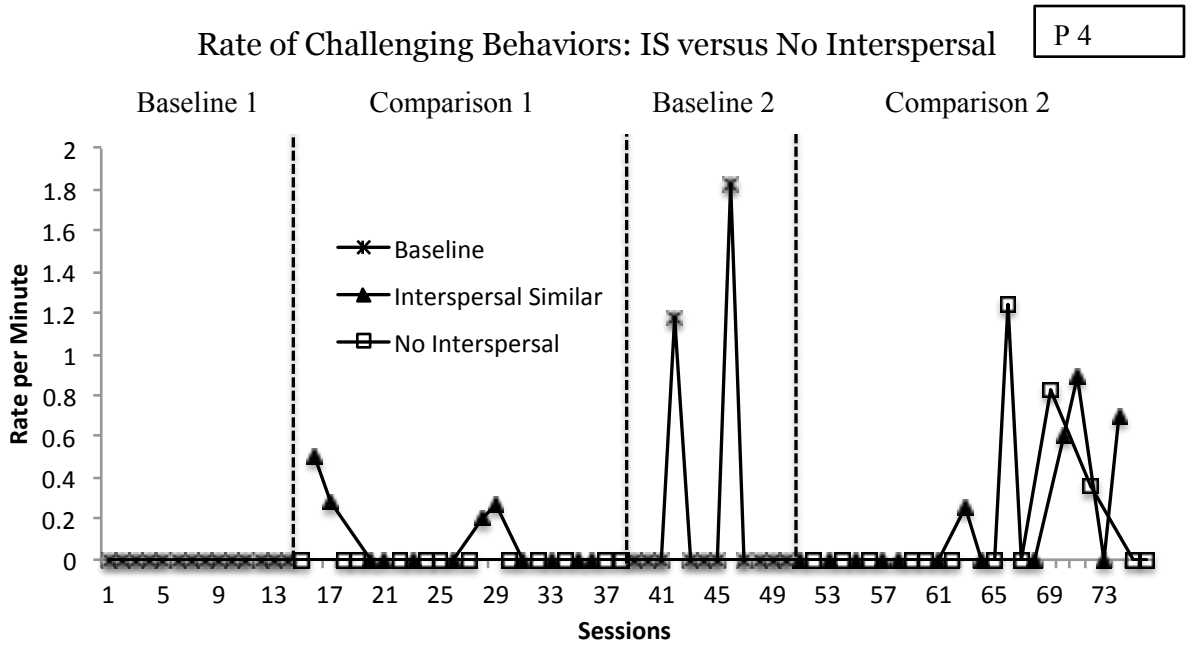

Figure 18. Rate of challenging behavior per minute of participant 4 across baseline and two alternating teaching conditions: IS and No Interspersal.

Figure 19 illustrates the rate of challenging behavior for participant 4 during baseline, IS and ID conditions. During baseline 1 (s: 1-20) there were three sessions in which challenging behavior was exhibited. The average rate of challenging behavior per minute was 0.19 (range 0-1.46). During comparison 1 (s: 21-38) there was one session during the IS condition when challenging behavior occurred. The average rate was 0.03 (0-0.26). There were two sessions during the ID condition during which challenging behavior occurred. The average rate was 0.05 (range 0-0.27). During baseline 2 (s: 39-52) there was one session that included challenging behavior. The average rate was 0.13 (range 0-1.82). During comparison 2 (s: 53-84), challenging behavior occurred more frequently in both conditions, during four sessions in the Interspersal Similar condition and five in the Interspersal Dissimilar condition. The average rate was 0.09 (range 0-0.57) and 0.14 (range 0-0.76) during IS and ID, respectively.

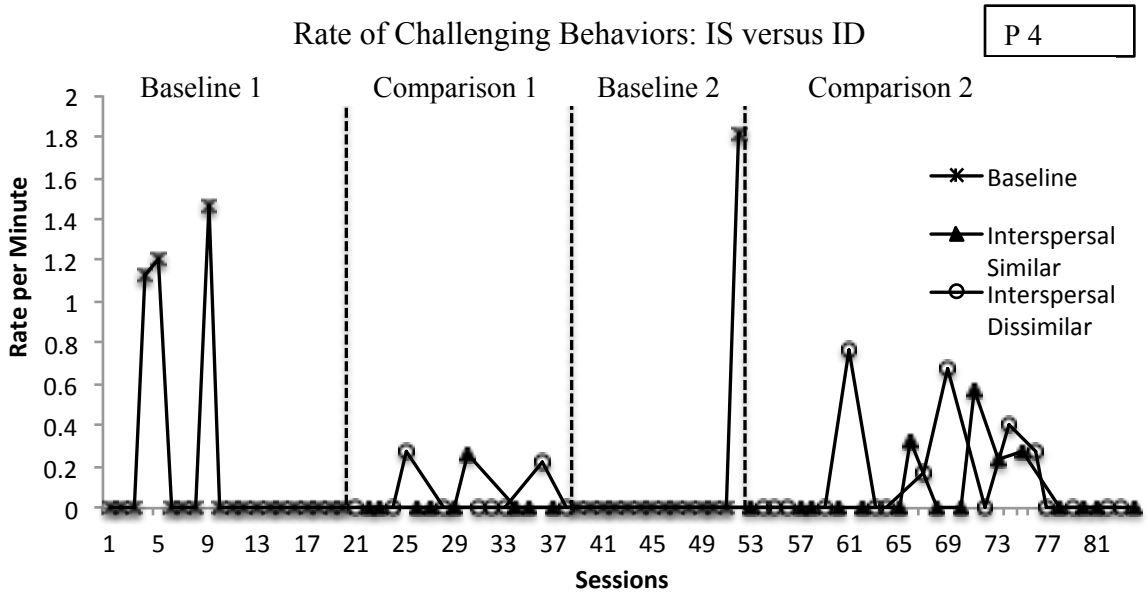

Figure 19. Rate per minute of challenging behavior of participant 4 across baseline and two alternating teaching conditions: IS and ID.

Child preference.

During both concurrent chains preference assessments, when participant 4 could choose either the IS or no interspersal condition, and when he could choose either IS or ID, he demonstrated a preference for the IS condition (see Figures 21 and 22 for more detail).

Generalization and maintenance.

Figure 20 displays the total number of words mastered per condition, and the total number of these words that met the generalization or maintenance criteria during a generalization probe with novel books and during two maintenance probes. The first maintenance probe was conducted two weeks after the completion of the study with the second probe conducted three weeks after study completion. The percentage of words that met the generalization and maintenance criteria was comparable across all conditions.

Of the 32 words that were mastered in the IS condition, participant 4 demonstrated generalization for 24 (75% of the words mastered in that condition), maintained 23 (72%) during the first probe, and 26 (81%) during the second probe. Of the 18 words that were mastered in the

ID condition, he demonstrated generalization for 13 (75%), maintained 14 (78%) during the first probe, and 18 (100%) during the second probe. Of the 14 words that were mastered in the no interspersal condition, he demonstrated generalization for 10 (71%), maintained 8 (57%) during the first probe, and 10 (71%) during the second probe.

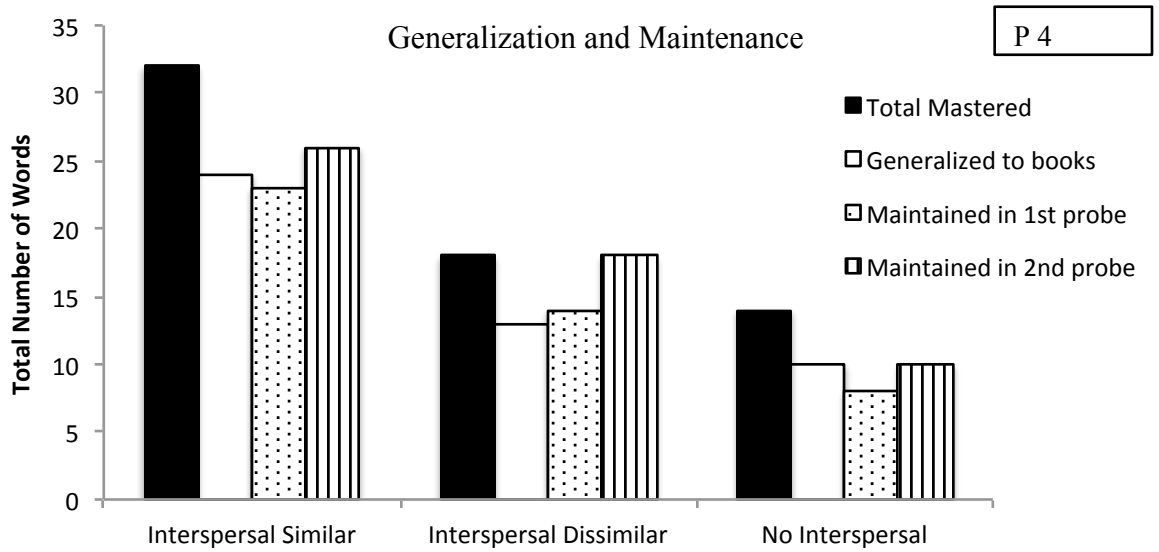

Figure 20. Total number of words that were mastered during the study in each condition, the total number that were generalized to novel books, and those that were maintained at the first and second maintenance probes.

Summary of Results Across All Participants

Table 3 summarizes the total number of words mastered within each condition across all participants. Across all participants during the first phase of the study, which compared the IS and the no interspersal conditions, the IS resulted in more words mastered. The magnitude of the difference between the two conditions ranged from four to nine words. During the second phase of the study, which compared the cumulative number of words mastered in the IS and ID conditions the results were equally distributed between the two conditions. Two of the participants (participant 1 and 3) mastered more words within the IS condition, and one participant (Participant 2) mastered more words in the IS condition during the first comparison

and in the ID condition during the second comparison. Participant 4 mastered more words in the ID condition in both comparisons.

Table 3

Total Number of Sight Words Mastered in Each Condition

Condition	Participant 1	Participant 2	Participant 3	Participant 4
Baseline	2	0	0	0
IS	15	9	8	9
No Interspersal	14	5	4	8
Reversal: Baseline	0	--	0	0
IS	18	--	9	7
No Interspersal	9		6	6
Reversal: Baseline	0	--	--	--
IS	12	--	--	--
No Interspersal	9			
Baseline	0	0	1	1
IS	7	11	6	6
ID	2	9	4	7
Reversal: Baseline	--	0	--	0
IS		9		10
ID	--	12	--	12

*Highlighted numbers indicate the conditions within that comparison with more words mastered.

Table 4 presents the duration (in weeks) of the participants' participation in the study from the first baseline to the final treatment condition, but excluding the generalization and maintenance probes.

Table 4
Study Duration Across Participants

	Duration of Participation (weeks)
P1	8
P2	9
P3	6
P4	6

Because the number of days per week that the participant was available for the study varied across participants, the total number of sessions conducted per condition and the total number of words mastered per condition is also presented in Table 5. The mean indicated is the average number of words mastered per session in each condition.

Table 5

Number of Sessions and Number of Words Mastered per Condition, Mean Number of Words Mastered per Session (total words mastered / total number of sessions)

	No I			IS			DS		
	Sessions	Words Mastered	Mean	Sessions	Words Mastered	Mean	Sessions	Words Mastered	Mean
P1	42	32	0.76	49	52	1.06	7	2	0.28
P2	14	5	0.36	44	29	0.65	30	21	0.7
P3	46	10	0.22	63	23	0.37	17	4	0.24
P4	25	14	0.56	51	32	0.63	24	19	0.79

Table 6 summarizes the percentage of correct responses for the mastered interspersed trials for each participant. The percent of correct responses would affect the rates of reinforcement within that session because social reinforcement was provided on a variable ratio schedule of 2 for correct responses to mastered tasks only. No data is provided for the no interspersal condition because no mastered skills were interspersed, and the rate of reinforcement for these sessions was yoked to the IS session preceding it. The percentage of correct responses

was comparable both within and across all participants. The outlier is the percentage of correct responding by participant 1 to mastered motor imitation skills.

Table 6

Mean Percentage of Correct Responses for Interspersed Trials Across Participants

	Interspersal Similar	Interspersal Similar vs. Interspersal Dissimilar	
	Mastered Sight Words	Mastered Sight Words	Mastered Motor Imitation
P 1	89% (range 73-100%)	87% (range 77-93%)	78% (range 67-93%)
P 2	88% (range 67-100%)	94% (range 83-100%)	96% (range 90-100%)
P 3	90% (range 77-100%)	93% (range 83-100%)	98% (range 93-100%)
P 4	87% (range 63-97%)	92% (range 83-97%)	93% (range 80-100%)

In order to draw conclusions about the participant's learning rate across conditions the duration of the instructional time was recorded. The mean duration of sessions across conditions is summarized in Table 7. The mean duration of an IS session was approximately 1.5 times that of a no interspersal session, which typically was equal to an additional 1 to 3 minutes.

Table 7
Mean Duration of Sessions (min) for Each Condition, including time required to complete the independent or mastered tasks that was yoked to shorter sessions

Condition	Participant 1	Participant 2	Participant 3	Participant 4
Baseline	1.31	0.79	0.55	0.61
IS	6.52	3.39	3.84	4.18
No Interspersal	4.27	1.74	2.48	2.63
Reversal: Baseline	0.94	--	0.73	0.58
IS	5.33	--	3.45	3.79
No Interspersal	3.26		1.81	2.38
Reversal: Baseline	0.73	--	--	--
IS	4.79	--	--	--
No Interspersal	2.76			
Baseline	0.71	0.77	0.72	0.61
IS	4.68	2.58	3.11	2.93
ID	6.02	3.88	3.33	3.93
Reversal: Baseline	--	0.80	--	0.60
IS		2.78	--	3.04
ID	--	3.25		3.80

Table 8 summarizes the rates of challenging behavior across all conditions for each participant. Overall, the rates of challenging behavior were low across all participants, averaging less than 1 behavior per minute. Rates were typically variable, but comparable across all conditions. The highest rate of challenging behavior within phase 1 and phase 2 of the study are highlighted for each participant below. During phase 1 (the comparison of IS to no interspersal) three of the four participants exhibited their highest average rates of challenging behavior during baseline conditions. The remaining participant had negligible rates of challenging behavior

during any of the conditions. During the second phase of the study (the comparison of IS to ID), two of the four participants again exhibited their highest average rates of challenging behavior during baseline conditions, one participant was during IS, and one participant was during ID.

Table 8

Rate of Challenging Behaviors (including the range) in Each Condition

Condition	Participant 1	Participant 2	Participant 3	Participant 4
Baseline	0.88 (0-2.4)	0	0.29 (0-1.53)	0
IS	0.32 (0-1.07)	0.09 (0-0.34)	0.08 (0-0.73)	0.10 (0-0.50)
No Interspersal	0.59 (0-1.53)	0.09 (0-0.63)	0.31 (0-3.08)	0
Reversal: Baseline	0.83 (0-1.67)	--	0.41 (0-2.26)	0.25 (0-1.82)
IS	0.69 (0-1.80)	--	0.18 (0-0.84)	0.19 (0-0.89)
No Interspersal	0.86 (0-2.42)		0.20 (0-0.99)	0.19 (0-0.82)
Reversal: Baseline	0.74 (0-3.08)	--	--	--
IS	0.41 (0-1.02)	--	--	--
No Interspersal	0.76 (0.31-1.44)			
Baseline	0.43 (0-1.67)	0	0.26 (0-2.14)	0.19 (0-1.46)
IS	0.54 (0.21-0.89)	0.02 (0-0.38)	0.20 (0-0.57)	0.03 (0-0.26)
ID	0.69 (0.34-1.67)	0.07 (0-0.33)	0.13 (0-0.71)	0.05 (0-0.27)
Reversal: Baseline	--	0	--	0.13 (0-1.82)
IS		0.17 (0-0.76)	--	0.09 (0-0.57)
ID	--	0.09 (0-0.62)		0.14 (0-0.76)

*Highlighted numbers indicate the conditions that had the highest rates of challenging behavior.

Figure 21 depicts the results of the concurrent chains preference assessment across all participants, where the initial link was the selection of the color coded stack of cards, and the terminal link was the implementation of the condition paired with the selected color (either the

IS or the no interspersal condition). In this figure, the sessions when the child was presented with the opportunity to choose is represented on the abscissa. This does not include baseline sessions or those sessions during which the researcher implemented the teaching sessions without offering the child a choice of conditions. The total number of opportunities to choose the treatment condition varied across participants. Participants 1 through 4 are on the ordinate. As seen, three of the four participants demonstrated a preference for the IS condition. One of the four participants chose the interspersal condition as often as the no interspersal condition. No participants demonstrated a stronger preference for the no interspersal condition. Individual results are as follows. Participant 1 demonstrated a strong preference for the IS condition, which he chose 16 of the 24 opportunities, over the no interspersal, which he chose on 8 of the 24 opportunities. Therefore his ratio of IS to no interspersal is 16:8. Participant 2 demonstrated no preference between the two. She chose each four times. Participant 3 demonstrated a slightly stronger preference for the IS condition, which he selected on 17 of the 28 opportunities. He chose the no interspersal condition on 11 of the 28 opportunities. Therefore the ratio of IS to no interspersal is 17:11. Participant 4 demonstrated a stronger preference for the IS condition, which he selected on 13 of the 17 opportunities. He chose the no interspersal condition on four of the 17 opportunities. Therefore the ratio of IS to no interspersal is 13:4.

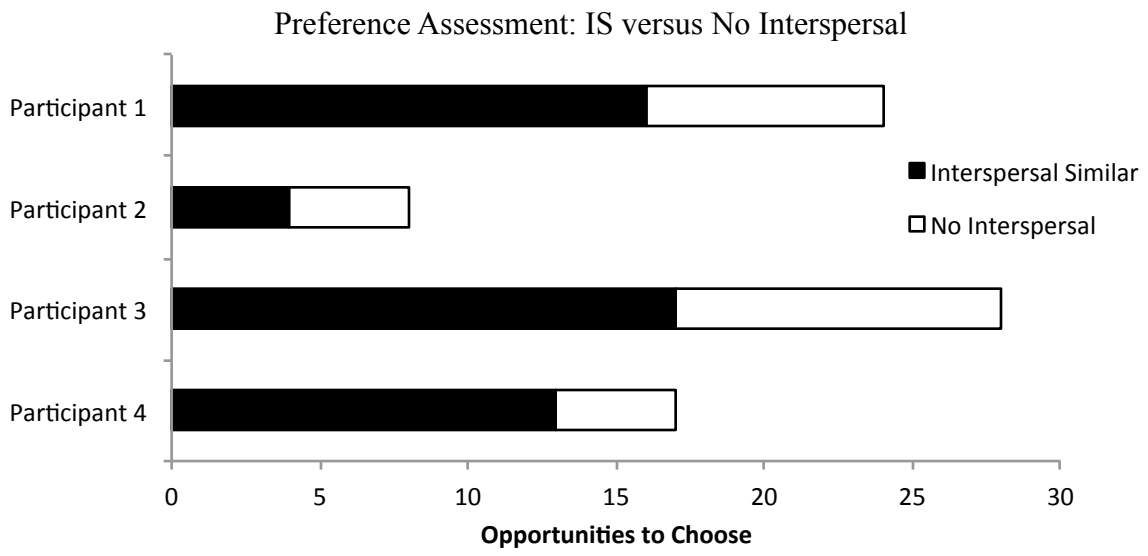

Figure 21. Results of the concurrent chains preference assessment for the Interspersal Similar and No Interspersal conditions.

Figure 22 portrays the results of the concurrent chains preference assessment when the terminal link was either the IS or the ID condition. All participants demonstrated a preference for the IS condition. Participants 1 and 3 chose the IS condition almost exclusively. Individual data is as follows. Participant 1 selected the IS condition 3 times and ID once. Participant 2 demonstrated a preference for the IS condition, which she chose on 14 of the 20 opportunities. Conversely, she chose the ID on only 6 of the 20 opportunities. Participant 3 selected the IS condition on 10 of the 11 opportunities. He chose the ID condition on 1 of the 11 opportunities. Therefore the ratio of IS to ID is 10:1. Participant 4 chose the IS condition on 11 of the 16 opportunities. He chose the ID condition on five of the 16 opportunities. Therefore his ratio of IS to ID is 11:5.

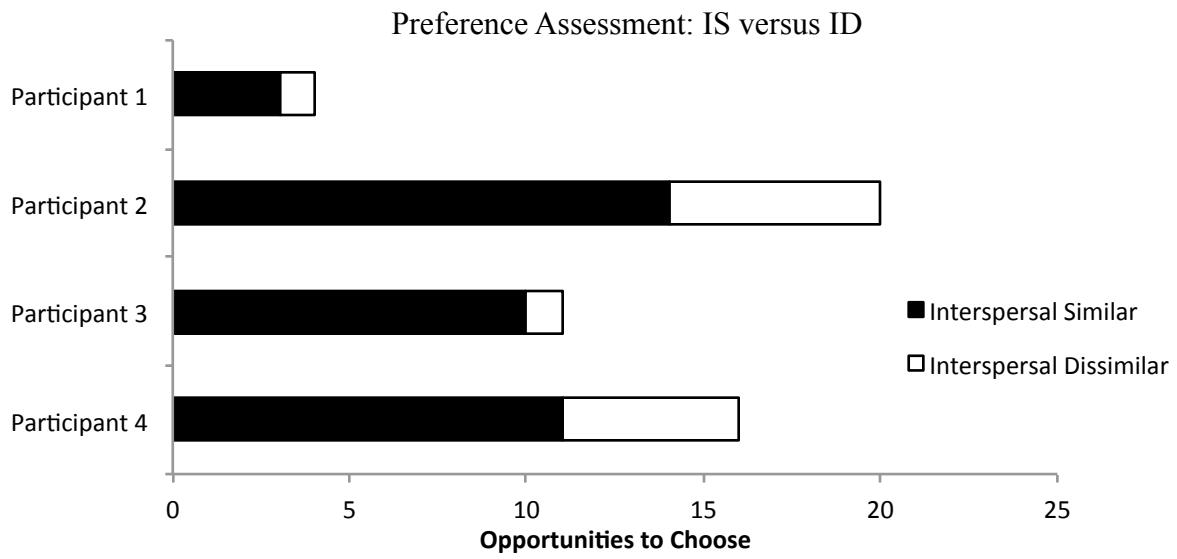

Figure 22. Results of the concurrent chains preference assessment for the Interspersal Similar and Interspersal Dissimilar conditions.

Table 9 summarizes the generalization and maintenance data across all participants. Data are presented by the total number of words mastered in each condition, then of that total, the number that were generalized to novel books, and the number that were maintained in each the first and second maintenance probes. For all participants and across all measures, the IS condition had the highest number of words mastered, generalized, and maintained. However, there was also more teaching trials conducted in that condition than the other two conditions. Because the total number of words mastered varied, and the total teaching trials varied across each condition and participant, the data are also presented as the percentage of the total mastered in each condition that were generalized to novel books, and maintained during first and second maintenance probes. All participants generalized the highest percentage of words that were taught in the IS condition. During the first maintenance probe, three of the four participants responded correctly to the highest percentage of words from the ID Condition. During the second maintenance probe, two of the four participants responded correctly to the highest percentage of words from the ID condition, one participant had equal responding across ID and no interspersal, and one participant maintained the highest percentage of words in the no interspersal condition.

Table 9

Generalization and Maintenance Data for All Participants

		Total Number Mastered	Number Generalized to books	Number Maintained in 1st probe	Number Maintained in 2nd probe	Percentage Generalized to books	Percentage Maintained in 1st probe	Percentage Maintained in 2nd probe
P 1	IS	51	31	24	25	61%	47%	49%
	ID	2	1	2	1	50%	100%	50%
	No Interspersal	31	18	14	22	58%	45%	71%
P 2	IS	29	20	18	24	69%	62%	83%
	ID	21	14	12	21	67%	57%	100%
	No Interspersal	5	1	1	5	20%	20%	100%
P 3	IS	23	18	11	14	78%	48%	61%
	ID	4	3	4	4	75%	100%	100%
	No Interspersal	10	4	5	5	40%	50%	50%
P 4	IS	32	24	23	26	75%	72%	81%
	ID	18	13	14	18	72%	78%	100%
	No Interspersal	14	10	8	10	71%	57%	71%

Interobserver Reliability

Interobserver agreement was scored for correct and incorrect responding to the target sight words, the frequency of challenging behavior, and the duration of the session. Mean agreement for participant responding to the target sight words was 99% (range 80-100%). Mean agreement for the frequency of challenging behavior was 99% (range 50-100%). Mean agreement for the duration of the session was 98% (see Appendix H for a breakdown of reliability scores across each participant and condition).

Treatment Fidelity

Treatment fidelity was scored for the presentation of the instruction, implementation of the correct response consequence for the target sight word, implementation of incorrect response consequence (error correction procedure) for the target sight word, correct and response

consequence for the interspersed trials, the frequency of reinforcement during the independent task. Of the sessions scored for treatment fidelity, the instruction was presented correctly, using the correct materials on 100% of the sessions. The correct response consequence was implemented on 100% of the opportunities. The error correction procedure was implemented correctly on 99% of the opportunities (range 86-100%). The implementation of the correct and the incorrect response consequence for the interspersed skills was scored for the interspersal similar condition only, as the video did not adequately allow for the independent observer to see both the model of the motor movement and the child's response to allow for accurate scoring for the ID condition. The correct response consequence for correct, mastered words was physical or social reinforcement on a VR 2 schedule, or for 50% of the correct responses. Each session was scored as meeting this criteria if the percent of correct responses that were followed by reinforcement was 40-60%. The correct response consequence for the mastered words was implemented correctly on 70% of the sessions. The 30% of the sessions that did not meet this criteria either had too rich or too lean a schedule of reinforcement. However, the average rate of correct trials that were followed by reinforcement was 47% (range 27-75%). The incorrect response consequence for the mastered words was 100%. The mean frequency of reinforcement during the independent task in the no interspersal condition was seven. The mean frequency of reinforcement provided for interspersed mastered skills during IS was 13, and for ID was 12.

Social Validity

Three of the four parent satisfaction surveys were returned. All responses for both families were neutral or favorable, as indicated with all scores ranging between 3 and 5, where 1 represented least and 5 represented most satisfied. The mean satisfaction score for each respondent was 3.5, 4.6, and 4.6. See Table 10 for specific responses. To the open ended questions about what they liked most, one parent responded that participation in the study did not

interfere with his treatment, and that she liked the consistency between the research sessions and treatment on the behavioral procedures. Another family indicated that participation in the study gave their child an opportunity to learn and improve his vocabulary, and that his pronunciation improved. The third family indicated that it helped them identify their daughter's challenges with sight words To the open-ended question about what they liked least, two parents indicated nothing, and the other parents indicated that more structure would allow more children and families to benefit. When asked if and how their child's ability to read had changed as a result of the study, one parent indicated not much, another family reported definitely, and noted improvements in his ability to break down and try to read new words, as well as improvements in his comprehension. The third family reported yes that the introduction of new words and the addition hours of reading practice through the study positively affected their daughter's ability to read.

Table 10

Social Validity Scores for Three Respondents

Item	Respondent 1	Respondent 2	Respondent 3
Child's progress	3	4	4
Helpfulness of information learned about child's learning style	3	4	5
Word selection	4	5	4
Teaching procedures	3	5	5
Schedule and location of sessions	4	5	5
Experience with researcher	4	5	5
Researcher competency	3	5	5
Researcher's understating of child performance	4	5	4
Overall experience	3	4	5
Recommend to another	4	4	4

Discussion

The study purpose was to evaluate the differential effects of various interspersal conditions on the learning of four children with autism. First, the researcher sought to evaluate the difference in the rates of sight word acquisition and challenging behavior when interspersal was included in the teaching versus when it was not. Second, the researcher sought to evaluate acquisition and challenging behavior when the interspersed skills were similar versus dissimilar to the target skill. Finally, the researcher sought to measure participant preference regarding the use of interspersal, and the type of skills interspersed. This section will first provide a general overview of the findings and how they contribute to the scientific knowledge of interspersal research. It will also include a further explanation of the data presented in the results, review study limitations, and provide conclusions and recommendation for future research.

All participants demonstrated faster mastery rates of sight words in the teaching conditions compared to baseline. This was replicated both within participants through the use of reversals, and across participants. This indicates that for these participants, repeated exposure and access to noncontingent reinforcement was not sufficient for learning. The data suggest that there was a delay in the mastery of words after the condition change from baseline to the teaching condition (interspersal similar (IS), interspersal dissimilar (ID), or no interspersal). This delay may be partially attributed to an artifact of the measuring system (number of words mastered). The mastery criteria required correct responding to the word across three consecutive sessions. Therefore, unless there was correct responding during baseline conditions the earliest that a word could be mastered after the participant was exposed to a teaching condition was three sessions after the condition change. The delay was longer if the child required one or more opportunities to have the word correctly modeled for him or her as a part of the instructive error correction. When correct responses occurred to a target word during baseline, it was extended to

determine if repeated exposure would lead to mastery. Once responding was determined to be either sufficiently variable or if the rate of correct responding returned to zero, a condition change would occur. In order to demonstrate experimental control despite the likely delay in mastery of new words concurrent with the condition change, this study maintained baseline conditions for longer durations. The data from this empirical study suggest that the treatment package, which included the use of contingent reinforcement, an instructive error correction, and different iterations of interspersal procedures were responsible for the participants' acquisition of novel sight words.

The efficacy of the treatment package can be measured through comparison of baseline to treatment, and the differential effects of the interspersal procedures can be measured through a comparison of the two alternating teaching conditions. The majority of participants in most comparisons learned more words in the IS condition than the no interspersal or ID conditions. The exception to this was participant 4, who learned approximately the same number of words in both conditions. Notably, his reading abilities at the onset of the study were stronger than those of the other participants. Although data were not recorded regarding the percentage of sight words read correctly of the total number of sight words presented during the pretest, participant 4 responded correctly to a majority of sight words for each of the first through fourth grade sight word lists. Therefore, the words identified for him for the purposes of the study were vocabulary words obtained from a generalized third grade curriculum, and these words were often more complex (more letters and syllables per word) than those of the other participants. These results may indicate that the effects of interspersal on sight word acquisition is greater when a child is beginning to build his or her sight word awareness, and lessens as he or she learns. Future research should investigate the possible relations between initial reading repertoires and the effects of interspersal research.

In the current study all participants demonstrated a preference for the IS compared to either the no interspersal or ID conditions, even when their acquisition rate was equal or slightly higher in the other condition (either no interspersal or ID). Reinforcement theory offers an explanation for this preference; the interspersal and no interspersal condition differed in both the schedule of reinforcement for correct responses and for incorrect responses, presuming that the avoidance of the error correction procedure functioned as reinforcement. During the no interspersal condition, it was possible that the child would experience multiple consecutive errors, would be exposed to the error correction procedure on consecutive trials, and that reinforcement of any kind would therefore be withheld for longer durations than in the interspersal conditions (even if it would be equal by the end of the session). In contrast, the interspersal of mastered responses could function as negative reinforcement, through avoidance of the increase in task demand associated with the implementation of the error correction procedure across consecutive words. In other words, because during the interspersal conditions three mastered instructions were delivered before each target word, there were few instances when multiple errors occurred consecutively, and no instances when the error correction procedure was implemented consecutively. These differences may explain why the participants would chose an interspersal condition over a no interspersal condition, but they do not explain why these four participants demonstrated a stronger preference for the interspersal of similar tasks versus the interspersal of dissimilar tasks.

Additionally, most participants had slightly higher percentages of correct responses to the mastered motor imitation than the mastered sight words. Despite this, all participants still demonstrated a clear preference for the IS condition. Perhaps the shifting attention that was needed to complete the motor imitation, then reading the sight word, then motor imitation, throughout the session represented a significant increase in the task demands for these

participants. Future studies evaluating the type of skill interspersed, the response effort required by the sequence of trials, and participant preference are needed to determine if similar patterns of responding would result for other participants. Such studies could also clarify the controlling variables of interspersal procedures.

The data on the patterns of challenging behavior yielded several points worth discussion. First, slightly higher rates of challenging behavior occurred during baseline conditions for three participants during phase 1, and for two participants during phase 2 (see Table 5), despite the fact that baseline conditions had the lowest number of demands (10 responses vs. 40 during teaching conditions), and the child was still able to access both social/physical reinforcement and preferred edibles or activities. However, reinforcement provided during baseline was not contingent on correct responses to the sight words, as it was during the teaching conditions. Social reinforcement was provided on average after every two flashcards to maintain the child's interest in the task and compliance with staying at the table during the session. Access to a preferred edible or activity was provided at the end of the session during baseline conditions contingent upon the child not leaving the work area. Therefore, the type of available reinforcement was comparable across baseline and intervention conditions, but the contingencies were different. Another possible contributor to the higher rates of challenging behavior during baseline was the possible aversive situation when the child was presented with successive unknown words. Indeed, participant 1 would often whine or throw the cards during baseline, or would respond, "I don't know" concomitant with challenging behavior. Participant 2 who had more age appropriate language and communication skills than the other participants, did not often engage in challenging behavior but stated after a baseline session "she felt like she didn't know these words and wasn't doing very well with them."

A second point related to the pattern of challenging behavior is that participants 2, and 4 demonstrated higher rates of challenging behavior toward the end of the study (see figure 9 for participant 2 and Figures 18-19 for participant 4). This may have been due to the extended duration of the study and the lack of task variation. On at least one occasion, each of these participants did ask, "Why do we only practice flashcards together?" possibly indicating their interest in an alternate teaching task (e.g., math or spelling), or their interest in an alternate reading activity (e.g., reading books or worksheets). However, for participant 2, there is at least one other explanation for the increase in recorded challenging behavior during the last comparison of the study. A significant portion of her challenging behaviors throughout the study were non-responses, or more specifically, she did not begin her response within five seconds of the presentation of the flashcard. As she mastered more words she was more likely to have longer response latency during the initial presentations of a new unknown word. However, this was also paired with observable behavior that had been associated with covert 'thinking' behavior for her in the past (facial expressions and pressing her lips together), which was less frequent at the beginning of the study. Therefore the increase in recorded challenging behavior, especially during the last comparison may have been more a reflection of her making more of an effort to think of the word (try to sound the word out in her head) before responding. Overall, however the rates of challenging behavior were variable across all conditions and participants, limiting the conclusions that can be made regarding the control of the treatment package on these behaviors.

This study provides a further analysis on the role of interspersal procedures on acquisition of new skills (previous findings have been mixed). Additionally, this study conducted a systematic manipulation of the interspersal procedures while controlling for other variables that often co-vary with the use of interspersal, such as the schedule of reinforcement and the duration

of the sittings. Controls were implemented for the complexity of the words and the matched groups were based on the number of letters and grade level of the word. Pretests were conducted to confirm that all responding to either the sight words and interspersed motor skills in the study were demonstrated to be known or unknown. The matching pretest also confirmed that each participant already had in his or her repertoire the visual discrimination skills that may facilitate reading. Previous studies have held constant either the opportunities to respond or the duration of the instructional sessions (Volpe, Mulé, Briesch, Joseph, Burns, 2011). In contrast this study's aim was to hold both variables constant by yoking the sessions across the two compared conditions. Additionally, the sessions were yoked to control for the rate of reinforcement across conditions. Controlling these variables allows for a clearer analysis of the impact that interspersal procedures have on skill acquisition, rates of challenging behavior, and child preference.

This study also sought to evaluate how the type of task which was interspersed with the target task affected acquisition. Of specific interest was evaluating the interspersal of similar versus the interspersal of dissimilar tasks. Motor imitation was selected as the dissimilar task because it was a non-language based instruction, did not require a flashcard presentation as the instruction, and did not involve a vocal response from the child. It was also selected because the duration necessary to respond was similar to the reading, and the duration of the sessions was thought to be an important variable to control. The motor tasks that were selected were responses that the child could complete while sitting in front of the experimenter. During implementation, a notable difference in the participants' responding was related to the necessary shifting of attention between the two types of tasks. The rapid presentation of the flashcards during the IS condition was often correlated with the child maintaining visual focus on the flashcards and responding quickly to the discriminative stimulus. This was also true for the consecutive trials of motor imitation. After the first trial of motor imitation, the child was visually focused on the

experimenter and seemed to be anticipating that the subsequent trials would also be motor imitation. Responding was generally more immediate under both of the above conditions. However, responding to the sight word after responding to the motor imitation task seemed more cumbersome. Responding was generally slower, even if it was within the five-second response criteria. Additionally, the participants were more likely to be distracted when the flashcard was presented in the ID condition than the IS. For example, the participants may have continued to look at their hands, continued to manipulate their fingers, pick at their fingers, or touch, rub or scratch their faces, arms, head. Although none of these behaviors would have been disruptive in a classroom setting, they did impede the participant from reorienting his or her attention to the sight word.

A combination of experimental designs was used to demonstrate experimental control. For those comparisons when the differential effect of the alternating treatments was not readily evident, the use of reversals was implemented to provide empirical evidence to further identify different patterns of responding across conditions. The reversals also allowed for a within-participant evaluation of the consistency of or replication of his or her responses. In the current study, without the use of reversals, conclusions cannot be made regarding the impact of the interspersal procedures for participant 4. However, because his responding was similar across multiple reversals, and distinct for baseline versus treatment, one can more confidently assert that for this participant the use interspersal procedures did not have a strong effect on his learning. Similarly, through the use of the combined experimental design, the efficacy of interspersal when compared with no interspersal for participant 3 becomes clearer.

Cates (2005) recommended that future research investigate the impact of interspersal procedures on generalization. “Because research under the interspersing model has not required students to perform the complex operations outside of the targeted academic task (e.g., reading

passage or mathematics worksheet), it is unclear as to what may occur if students are asked to demonstrate such skills in other situations or assignments (e.g., doing story problems after learning the operation on a worksheet)" (p. 324). The current study offers a model for measuring generalization of sight words learned on flashcards to the more complex task of reading those words in published age appropriate children's books. Including such a measure of generalization in this and future studies allows readers to draw conclusions about the generality of the behavior change, and may also increase the social validity of the studies.

Limitations

Although this study contributes to the literature, there are several limitations in the areas of internal validity, generality of findings, and experimental design that need to be addressed. This study included Dolch and Fry high-frequency sight words. It is possible that participants may have had exposure to these words outside of the controlled study session, which may be a threat to internal validity. Participants 1 and 2 were receiving class wide reading instruction in their classrooms and IEBT programs for participants 1, 2, and 3 included instruction on sight word reading and or comprehension of the written word. Efforts were made to ensure that the words targeted during the study were distinct from the home and school targets, however, it is possible that any or all other forms of instruction affected responding in this study. However, the previous, typical exposure of these words to each child had not resulted in acquisition of these words and there was a relatively short time frame between introduction of a new word and mastery of it. This suggests that the mastery of new words within the study was not a direct result of uncontrolled external variables.

Because this study evaluated the effects of interspersal procedures with one target skill area, sight words, it is difficult to predict if the same patterns of responding would emerge for these participants in other areas of their academic learning. Future research should investigate if

there are similar effects with the same participants across multiple target behaviors. This investigation did not teach or measure comprehension of the words included (either for target words or mastered words) so conclusions cannot be made regarding how the current treatment package impacted comprehension.

This treatment package included interspersal procedures, the use of contingent reinforcement, and an instructive error correction procedure, additionally teaching occurred across multiple words simultaneously. The results suggest that the treatment package had an impact on the acquisition of novel sight words for all of the participants. The total treatment package impacted child performance, however, the relative contributions of each component cannot be determined without a further component analysis.

Child performance on generalization and maintenance probes was moderate to strong (ranging from 45-100% with the exception of participant 2 in the no interspersal condition). In retrospect, another alternative method for measuring maintenance and generalization might have been to implement systematic generalization and maintenance probes after a specified time following the mastery of the word instead of, or in addition to, probes occurring after a designated time delay following the completion of the study. In some cases, these maintenance probes occurred nine weeks after the mastery of the first words, during which time that word would not have been systematically practiced. The intent of the current study was to evaluate the effect of teaching conditions on acquisition, maintenance and generalization. To maintain treatment integrity, parents and other caregivers were asked specifically not to target the words taught in the study until after the completion of the study. However, this may have impeded retention and generalization. An example of a procedure to facilitate maintenance is provided by Neef, Iwata, and Page (1980). As the spelling words were mastered they were then included as the interspersed words when teaching other novel words. This resulted in overall higher rates of

maintenance (63-96% of words learned) than was found in this study. Future studies might also evaluate the use of programmed generalization to natural activities with people relevant to the child's daily life (e.g., homework, story time with books targeting newly mastered words with parents and or other teachers).

Although this investigation did include a measure of generalization to everyday reading activities relevant to the child's life that many previous studies did not include, conclusions regarding the intervention's effect on generalization are limited by the use of a post-intervention probe only. A more precise measure of generalization might have included a pre-intervention assessment of the participants' ability to read these words in books.

The participants of this study had a high percentage of parents whose spoke English as a second language. Of the four participants, only one set of parents spoke English as their primary language, and the primary language of the other parents was Spanish, Urdu, and Mandarin. Three of the four children communicated solely in English, one of the four was bilingual in Urdu and English. It is possible that this cultural context may have had some impact on the acquisition of sight words by these children, either outside of or during the current study through limited exposure to these words (spoken or written) in their home environments. It is also possible that this cultural context may impact the generalization of these findings to other participants whose parents' primary language is commensurate with the language of the words targeted.

The study's experimental design demonstrated experimental control over the cumulative number of words mastered. However, on the measure of challenging behavior, there was little differential effect demonstrated between the teaching conditions. One possibility may have been carry over effects between the rapidly alternating conditions. Although there was a 10-minute break between each session, this may have been insufficient to reduce the carry over of generalized compliance or noncompliance from the previous session. A reversal design may be

another alternate design to evaluate the potential differential effects of interspersal procedures on challenging behaviors, especially when those behaviors occur at a low rate. The low rate of challenging behavior in the current study may suggest why less robust effects that were found on the measures for challenging behavior. In fact, other studies using the alternating treatments design have demonstrated reduced rates of challenging behavior using treatment packages that include interspersal when the challenging behavior occurred a high rate (Reed, Luiselli, Morizio, and Child, 2010). Finally, challenging behavior included both overt problem behavior (escape, aggression, property destruction), and non-responding. It is possible that the grouping of all of these behaviors together confounded the results. Future research should collect data on both overt problem behavior and non-responding but present these data separately to determine the effect of interspersal procedures on each response class.

Finally, while most of the concurrent chains preference assessments resulted in differential responding, these results would be more compelling if a control choice was also included. Future studies should include a control condition, for example one in which no demands and no reinforcement is provided, to enhance experimental control within the preference assessment.

Conclusions and Recommendations

This study offers a proven intervention using interspersal of similar tasks in teaching sight words to children with autism. However, efficacy is only one important variable that contributes to moving from application of scientific discovery to practical use. How efficiently a treatment creates behavior change is another key variable. Regarding the efficiency of teaching conditions in the current study, all participants learned more words in the IS condition than the no interspersal condition. However, in general, a session for the IS condition was longer in duration than that of the no interspersal condition. The discrepancy in session duration was

smaller when comparing IS to ID, but the average ID session was slightly longer in duration. This could be a practical limitation in the application of an interspersal procedure if instructional efficiency for the teaching of new skills is the priority. However, the interspersal of mastered tasks might have a positive effect on the maintenance of those skills.

There are several possible explanations for the behavioral process at work during the implementation of interspersal procedures. One hypothesized controlling variable is the higher rates of reinforcement relative to other conditions in some studies. However, the current study controlled the rates of reinforcement across conditions and the number of words mastered continued to be higher in the interspersal conditions than the no interspersal. Another possible behavioral mechanism for these results may be explained by behavioral momentum. In a study conducted by Mace et al. (1988), a high probability request sequence was used to affect the response class of compliance with requests. Several instructions with a high probability of compliant responding were presented immediately before an instruction that had a low probability of a compliant response. Use of this high probability request sequence resulted in increased compliance, decreased response latency, and decreased duration of the response for the targeted, low probability request. In this study, the response class of interest may be conceptualized as fluent reading. The implication of behavioral momentum seems especially evident for participant 2, who had a high number of errors due to self-corrected responses or sounding out the words versus responding fluidly within the five seconds. The frequency of this type of error was much less prevalent during the interspersal condition, compared to the no interspersal condition. Perhaps the practice of reading three mastered words fluidly, created the behavioral momentum to read the target word fluidly. Then this correct response would allow the participant to contact reinforcement, thus increasing the likelihood that she would continue to respond correctly to the target words. Put another way, the antecedent intervention of the

interspersal and the consequent intervention of the reinforcement both likely contributed to the learning that occurred in the current study.

Choice, as a motivational variable, has become increasingly prevalent as a component of effective intervention for children with autism. Given the language and communication challenges that many children with autism face, it continues to be important that we as clinicians and researchers strive to identify methods to identify and incorporate child choice as it is possible and appropriate. This study used preference assessments to identify potential reinforcers, and extends previous research on assessing preference as related to the instructional format (Hanley, Piazza, Fisher, Contrucci, & Maglieri, 1997; Leaf, Sheldon, & Sherman, 2010).

There are several areas of interspersal that warrant additional research. These include possible interactions between the procedure and the strength of the response class, the type of skill, the complexity of the skill, and the population. To determine if there are different effects of interspersal depending on the strength of the response class or the individual's existing repertoire, studies could compare the effects across a group of students that are in the early stages of sight word acquisition (limited repertoire) to other students who are in the later stages of sight word acquisition (larger repertoire). To determine if there are different effects depending on the type of skill, additional studies could use a multiple baseline design across skill areas but within participants (expressive labeling, receptive labeling, reading, math). This type of design would clarify if the impact of interspersal procedures is consistent across different skill areas for a given individual. Additional research that compares the effects of teaching with and without interspersal procedures on generalization, specifically including measures of generalization to more naturalized (less contrived) activities would enhance the social validity of the results.

In summary, the results of the current study contribute to interspersal literature with an experimental comparison of teaching that does and does not include interspersal and a

comparison of the type of task interspersed. This study also offer some strategies for isolating an antecedent intervention or contextual variable such as interspersal while controlling other important variables such as session duration, reinforcer density, and number of opportunities to respond. The results indicate a preference for teaching procedures that include interspersal over those that do not, and the interspersal of tasks that are similar to the target task over those that are different. Horner et al (1991) noted, "Interspersed requests may be a procedure for increasing the general likelihood that a student will attempt to follow instructions. Only when attempts occur do the traditional instructional procedures become operational" (p. 276). On the matter of preference, instructional efficacy and instruction efficiency; while research related to educational strategies is important for the continued improvement of education for all children, it is may be especially important for those who are experiencing academic difficulties. Those children who have a history of difficulty learning are at risk of developing a pattern of avoiding instructional situations, which could lead to further delays and continued difficulties learning. For this group of students specifically, treatment decisions based collectively on effectiveness (the likelihood that the intervention will lead to skill acquisition), efficiency (the rate of learning) and student preference (likely the less aversive instructional format) may be especially important.

References

Adcock, J., & Cuvo, A. J. (2009). Enhancing learning for children with autism spectrum disorders in regular education by instructional modifications. *Research in Autism Spectrum Disorders, 3,* 319-328. doi:10.1016/j.rasd.2008.07.004

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th. Edition. Washington, D.C.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of Applied Behavior Analysis, 10,* 405-416. doi:10.1901/jeab.1967.10-405

Baer, D. M., & Sherman, J. A. (1964). Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology*, *1*, 37-49. doi:10.1016/0022-0965(64)90005-0

Belfiore, P. J., Lee, D. L, Vargas, A. U. & Skinner, C. H. (1997). Effects of high-preference single-digit mathematics problem completion on multiple-digit mathematics problem performance. *Journal of Applied Behavior Analysis, 30,* 327-330. doi:10.1901/jaba.1997.30-327

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

Billington, E. J., Skinner, C. H., & Cruchon, N. M. (2004). Improving sixth-grade students perceptions of high-effort assignments by assigning more work: Interaction of additive interspersal and assignment effort on assignment choice. *Journal of School Psychology, 42*(6), 477-490. doi:10.1016/j.jsp.2004.08.003

Browder, D. M. & Shear, S. M. (1996). Interspersal of known items in a treatment package to

teach sight words to students with behavior disorders. *The Journal of Special Education, 29,* 400-413. doi:10.1177/002246699602900403

Burns, M. K. (2007). Comparison of opportunities to respond within a drill model when rehearsing sight words with a child with mental retardation. *School Psychology Quarterly, 22*(2), 250-263. doi:10.1037/1045-3830.22.2.250

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education & Treatment of Children, 28*(3), 237-249. http://www.educationandtreatmentofchildren.net/

Burns, M. K., Ardoin, S. P., Parker, D. C., Hodgson, J., Klingbeil, D. A., & Scholin, S. E. (2009). Interpersonal technique and behavioral momentum for reading word lists. *School Psychology Review, 38*(3), 428-434. http://www.nasponline.org/publications/spr/index.aspx?vol=40&issue=4

Burns, M. K., Boice, C. H. (2009). Comparison of the relationship between words retained and intelligence for three instructional strategies among students with below-average IQ. *School Psychology Review, 38,* 284-292.

Burns, M. K., & Dean, V. J. (2005). Effect of drill ratios on recall and on-task behavior for children with learning and attention difficulties. *Journal of Instructional Psychology, 32*(2), 118-126. http://www.projectinnovation.biz/jip_2006.html

Calderhead, W. J., Filter, K. J., & Albin, R. W. (2006). An investigation of incremental effects of interspersing math items on task-related behavior. *Journal of Behavioral Education, 15*(1), 53-67. doi:10.1007/s10864-005-9000-8

Cates, G. L., Skinner, C. H., Watson, T. S., Meadows, T. J., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based

decision making: An evaluation of interspersing procedures. *School Psychology Review, 32*(4), 601-616. http://www.nasponline.org/publications/spr/index.aspx?vol=40&issue=4

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. doi:10.1901/jaba.1992.25-795

Centers for Disease Control and Prevention (2012). *Autism Spectrum Disorders.* Retrieved February 23, 2012, from http://www.cdc.gov/ncbddd/autism/index.html

Chong, I. M., & Carr, J. E. (2005). An investigation of the potentially adverse effects of task interspersal. *Behavioral Interventions, 20*(4), 285-300. doi:10.1002/bin.202

Committee on Children With Disabilities (2001). Technical Report: The Pediatrician's Role in the Diagnosis and Management of Autistic Spectrum Disorder in Children. *Pediatrics,* 107, e85. http://www.pediatrics.org/cgi/content/full/107/5/e85

Cooke, N. L., Guzaukas, R., Pressley, J. S., & Kerr, K. (1993). Effects of using a ratio of new items to review items during drill and practice: three experiments. *Education and treatment of Children, 16,* 213-235. http://www.educationandtreatmentofchildren.net/

Cuvo, A. J., Davis, P. K., & Gluck (1991). Cumulative and interspersal task sequencing in self-paced training for persons with mild handicaps, *Mental Retardation, 29,* 335-342.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29,* 519-533. doi:10.1901/jaba.1996.29-519

Dolch, E. W. (1936). A basic sight vocabulary. *The Elementary School Journal, 36,* 456-460. http://www.jstor.org/stable/995914

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. http://www.sciencedirect.com/science/journal/00220965

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis, 13*(4), 619-627. doi:10.1901/jaba.1980.13-619

Durand, V. M., and Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. *Journal of Applied Behavior Analysis, 24,* 251-264. doi:10.1901/jaba.1991.24-251

Esperanto. (n.d.). In *Mirriam-Webster's online dictionary*. Retrieved from http://www.merriam-webster.com/dictionary/esperanto

Fry, E. (1980). The new instant word list. *The Reading Teacher, 34,* 284-289. http://www.jstor.org/stable/20195230

Glover, A. C., Roane, H. S., Kadey, H. J., and Grow, L. L. (2008). Preference for reinforcers under progressive and fixed-ratio schedules: A comparison of single and concurrent arrangements. *Journal of Applied Behavior Analysis, 41,* 163-176. doi:10.1901/jaba.2008.41-163

Hanley, G. P., Piazza, C. C., Fisher, W. W., Contrucci, S. A., & Maglierei, K. A. (1997). Evaluation of client preference for function-based treatment packages. *Journal of Applied Behavior Analysis, 30,* 459-.473. doi: 10.1901/jaba

Hawkins, J., Skinner, C. H., & Oliver, R. (2005). The effects of task demands and additive interspersal ratios on fifth-grade students' mathematics accuracy. *School Psychology Review, 34*(4), 543-555. Retrieved from www.csa.com

Horner, R. H., Day, H. M., Sprague, J. R., & O'Brien, M. (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

Ingvarsson, E. T., Kahng, S. & Hausman, N. L. (2008). Some effects of noncontingent positive reinforcement on multiply controlled problem behavior and compliance in a demand context. *Journal of Applied Behavior Analysis, 41,* 435-440. doi:10.1901/jaba.2008.41-435

Joseph, L. M., & Nist, L. M. (2006). Comparing the effects of unknown-known ratios on word reading learning versus learning rates. *Journal of Behavioral Education, 15*(2), 69-79. doi:10.1007/s10864-006-9008-8

Knight, M. G., Ross, D. E., Taylor, R. L., & Ramasamy, R. (2003). Constant time delay and interspersal of known items to teach sight words to students with mental retardation and learning disabilities. *Education and Training in Developmental Disabilities, 38*(2), 179-191. http://daddcec.org/Publications/ETADDJournal/tabid/70/IssueID/14/Default.aspx

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. doi:10.1901/jaba.1986.19-425

Koegel, L. K., Singh, A. K., & Koegel, R. L. (2010). Improving motivation for academics in children with autism. *Journal of Autism and Developmental Disorders, 40*(9), 1057-1066. doi:10.1007/s10803-010-0962-6

Leaf, J. B., Sheldon, J. B., & Sherman, J. (2010). Comparison of simultaneous prompting and no-no prompting in two-choice discrimination learning with children with autism. *Journal of Applied Behavior Analysis, 43,* 215-228. doi: 10.1901/jaba.2010.43-215

Logan, P., & Skinner, C. H. (1998). Improving students' perceptions of a mathematics assignment by increasing problem completion rates: Is problem completion a reinforcing event? *School Psychology Quarterly, 13*(4), 322-331. doi:10.1037/h0088988

MacQuarrie, L. L., Tucker, J. A., Burns, M. K., & Hartman, B. (2002). Comparison of retention rates using traditional, drill sandwich, and incremental rehearsal flash card methods. *School Psychology Review, 31*(4), 584-595. http://www.nasponline.org/publications/spr/index.aspx?vol=40&issue=4

Martin, J. J., Skinner, C. H., & Neddenriep, C. E. (2001). Extending research on the interspersal procedure to perceptions of continuous reading assignments: Applied and theoretical implications of a failure to replicate. *Psychology in the Schools, 38*(4), 391-400. doi:10.1002/pits.1027

Matson, J. L., Benavidez, D. A., Compton, L. S., Paclawskyj, T., & Baglio, C. (1996). Behavioral treatment of autistic persons: A review of research from 1980 to the present. *Research in Developmental Disabilities, 17,* 433-465. doi:10.1016/S0891-4222(96)00030-3

McCurdy, M., Skinner, C. H., Grantham, K., Watson, T. S., & Hindman, P. M. (2001). Increasing on-task behavior in an elementary student during mathematics seatwork by interspersing additional brief problems. *School Psychology Review, 30*(1), 23-32. http://www.nasponline.org/publications/spr/index.aspx?vol=40&issue=4

Montarello, S., & Martens, B. K. (2005). Effects of interspersed brief problems on students' endurance at completing math work. *Journal of Behavioral Education, 14*(4), 249-266. doi:10.1007/s10864-005-8649-3

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. doi:10.1901/jaba.1980.13-153

Neef, N. A., Iwata, B. A., & Page, T. J. (1977). The effects of known-item interspersal on acquisition and retention of spelling and sightreading words. *Journal of Applied Behavior*

Analysis, 10(4), 738-738. doi:10.1901/jaba.1977.10-738

Nist, L., & Joseph, L. M. (2008). Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders' word recognition, acquisition, maintenance, and generalization. *School Psychology Review, 37*(3), 294-308. http://www.nasponline.org/publications/spr/index.aspx?vol=40&issue=4

Reed, D. D., Luiselli, J. K., Morizio, L. C., & Child, S. N. (2010). Sequential modification and the identification of instructional components occasioning self-injurious behavior. *Child & Family Behavior Therapy, 32*(1), 1-16. doi:10.1080/07317100903539642

Rhymer, K. N., & Cates, G. L. (2006). Student performance on and preferences for mathematics word problems: An investigation of the effects of explicit timing and interspersing procedures. *School Psychology Quarterly, 21*(1), 34-45. doi:10.1521/scpq.2006.21.1.34

Rhymer, K. N., & Morgan, S. K. (2005). Comparison of the explicit timing and interspersal interventions: Analysis of problem completion rates, student preference, and teacher acceptability. *Journal of Behavioral Education, 14*(4), 283-303. doi:10.1007/s10864-005-8651-9

Richman, D. M., Wacker, D. P., Asmus, J. M., & Casey, S. D. (1998). Functional analysis and extinction of different behavior problems exhibited by the same individual. *Journal of Applied Behavior Analysis, 31,* 475-478. doi:10.1901/jaba.1998.31-475

Roberts, M. L. & Shapiro, E. S. (1996). Effects of instructional ratios on students' reading performance in a regular education program. *Journal of School Psychology, 34,* 73-91. doi: 10.1016/0022-4405(95)00026-7

Roberts, M. L., Turco, T. L., & Shapiro, E. S. (1991). Differential effects of fixed instructional ratios on students' progress in reading. *Journal of Psychological Assessment, 9,* 308-318.Robinson, S. L., & Skinner, C. H. (2002). *School Psychology Quarterly, 17*(2), 191-

205. doi:10.1521/scpq.17.2.191.20858

Rowan, V. C. & Pear, J. J. (1985). A comparison of the effects of interspersal and concurrent training sequences on acquisition, retention, and generalization of pictures names. *Applied Research in Mental Retardation, 6,* 127-145. doi: 10.1016/S0270-3092(85)80066-7

Schmidgall, M., & Joseph, L. M. (2007). Comparison of phonic analysis and whole word-reading on first graders' cumulative words read and cumulative reading rate: An extension in examining instructional effectiveness and efficiency. *Psychology in the Schools, 44*(4), 319-332. doi:10.1002/pits.20227

Skinner, C. H., Hurst, K. L., Teeple, D. F., & Meadows, S. O. (2002). Increasing on-task behavior during mathematics independent seat-work in students with emotional disturbance by interspersing additional brief problems. *Psychology in the Schools, 39*(6), 647-659. doi:10.1002/pits.10058

Szadokierski, I., & Burns, M. K. (2008). Analogue evaluation of the effects of opportunities to respond and ratios of known items within drill rehearsal of esperanto words. *Journal of School Psychology, 46*(5), 593-609. doi:10.1016/j.jsp.2008.06.004

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

Volpe, R. J., Mulé, C. M., Briesch, A. M., Joseph, L. M., & Burns, M. K. (2011). A comparison of two flashcard drill methods targeting word recognition. *Journal of Behavioral Education, 20*(2), 117-137. doi:10.1007/s10864-011-9124-y

Weber, R., & Thorpe, J. (1989). Comparison of task variation and constant-task methods for severely disabled in physical education. *Adapted Physical Activity Quarterly, 6,* 338-353.

Wildmon, M. E., Skinner, C. H., McCurdy, M., & Sims, S. (1999). Improving secondary students' perception of the "dreaded mathematics word problem assignment" by giving them more word problems. *Psychology in the Schools, 36*(4), 319-325. doi:10.1002/(SICI)1520-6807(199907)36:4<319::AID-PITS5>3.0.CO;2-T

Wildmon, M. E., Skinner, C. H., Watson, T. S., & Garrett, L. S. (2004). Enhancing assignment perceptions in students with mathematics learning disabilities by including more work: An extension of interspersal research. *School Psychology Quarterly, 19*(2), 106-120. doi:10.1521/scpq.19.2.106.33310

Winterling, V., Dunlap, G., & O'Neil, R. E. (1987). The influence of task variation on the aberrant behavior of autistic students. *Education and Treatment of Children, 10,* 105-119. http://www.educationandtreatmentofchildren.net/

Appendix A
Summary Table of Early Interspersal Literature

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Dunlap (1984)	5 children with autism in clinic	Simultaneous treatment	Varied tasks - Trials to criterion - Child affect	Constant-task NI I	Interspersal led to faster acquisition and more favorable ratings of child affect	n/a
Koegel & Koegel (1986)	boy with intellectual disabilities 1:1 in clinic	multiple baseline across skill areas	Varied-tasks - percentage of correct responses - child affect	NI I	Interspersal led to higher percentage of correct responses and more favorable ratings of child affect	n/a
Neef, Iwata, & Page (1977)	3 men with DD in classroom	multielement	Spelling and sight words - acquisition rate - maintenance	NI NI + high density sr+ IS	Interspersal led to better acquisition and maintenance for spelling and sight words	n/a
Neef, Iwata, & Page (1980)	3 men with DD in classroom	multielement	Spelling words - acquisition - maintenance	NI NI+high density sr+ IS + folding in learned words	Interspersal led to better acquisition and maintenance for spelling and sight words	overall preference for interspersal
Rowan & Pear (1985)	3 children with intellectual disabilities in university clinic	ABA within subjects, with counterbalancing	Expressive labels -cumulative number learned -trials to criterion -maintenance -generalization	NI NI+high density sr+ IS + folding in learned words	Interspersal led to more words learned with fewer trials to criterion Little difference in maintenance or generalization	n/a
Winterling, Dunlap, & O'Neill (1987) study 1	2 children with autism in clinic	repeated reversal	Various skills -Percentage of aberrant behavior	NI NI+high density sr+ IS + folding in learned words	Interspersal led to lower rates of aberrant behavior	n/a
Winterling, Dunlap, & O'Neill (1987) study 2	woman with autism in clinic	alternating treatment	Various skills -Percentage of aberrant behavior - trials to criterion	NI NI+high density sr+ IS + folding in learned words	Interspersal led to lower rates of aberrant behavior and faster acquisition of skills	n/a

Note. Abbreviations are used for no interspersal (NI), interspersal (I), interspersal similar (IS)

Appendix B
Summary Table of Interspersal Literature with the General Education Population

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Belfiore, Lee, Vargas, & Skinner (1997)	2 girls age 14 and 15 -classroom		Math -Response latency	NI IS	Interspersal led to reduced response latency.	n/a
Billington, Skinner, & Cruchon (2004)	44 6th grade students -classroom	within-groups design	Math -Student perception -Number of problems completed -Accuracy	NI IS	Interspersal led to more total problems completed was more favorably perceived. No significant difference in number of target problems completed or accuracy.	One group preferred interspersal. No difference in other group.
Burns, Ardoin, Parker, Hodgson, Klingbeil, &Scholin (2009)	46 4th grade students -1:1 in school	between groups design	Sight words -Number of word read correctly per minute	NI IS	Interspersal did not result in higher rates of words read correctly per minute.	n/a
Cates, Jackson, Meadow, Skinner, Watson, & Weaver (2003)	5 2nd grade students -classroom	alternating treatment design	Sight words -Number of words mastered -Learning rate	NI IS with unknown:known ratio 1:1 IS with 1:3 ratio	No interspersal led to the highest learning rate. No difference on the number of words mastered or maintained.	n/a
Hawkins, Skinner, and Oliver (2005)	52 5th grade students -classroom	within-groups design	Math -Accuracy on written and spoken target problems	NI IS with unknown:known ratio 1:1 IS with 1:3 ratio	Interspersal with 1:1 ratio led to increased accuracy on written math assignment. Interspersal with 1:3 ratio led to increased accuracy for spoken math problems.	n/a
Joseph & Nist (2006)	2 5th and 1 6th grade students -school psychologist's office	alternating treatment design	Sight words -Total number of words mastered and maintained -Cumulative learning rate	NI IS with unknown:known ratio 1:3 IS with 3:3 ratio	Similar number of words mastered across conditions, slightly higher in IS with 3:3 ratio. No interspersal resulted in the highest cumulative learning rate and slightly more words maintained.	n/a
MacQuarrie, Tucker, Burns, & Hartman (2002	25 3rd grade students; 26 7th grade students -classroom	within-groups design	Esperanto words -Number of words maintained	NI IS: Incremental rehearsal IS: drill sandwich	Incremental rehearsal led to highest number of words maintained.	n/a

Summary Table of Interspersal Literature with the General Education Population (continued)

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Martin, Skinner, & Neddenriep (2001)	48 7th grade students -1:1 classroom	within-groups design	Reading -Number of words read correctly per minute -Student perception	NI IS	No difference between conditions on number of words read correctly per minute, or on student perception of effort. Students perceived no interspersal as less time consuming.	No difference between conditions.
McCurdy, Skinner, Grantham, Watson, & Hindman (2001)	1 4th grade girl -classroom	alternating treatment design	Math -Percentage of on-task behavior	NI IS	Interspersal led to higher percentages of on-task behavior.	n/a
Montarello & Martens (2005)	4 5th grade students -school library	alternating treatment design	Math -Total correct -Correct per minute	NI IS IS + token sr+	Responding was undifferentiated across conditions.	Preference for interspersal.
Nist & Joseph ((2008)	5 1st grade students -classroom	multielement design without baseline	Sight words -Number of words mastered and maintained	NI IS IS: Incremental rehearsal	Incremental rehearsal led to the high number of words mastered and maintained.	All participants preferred no interspersal.
Rhymer & Cates (2006)	187 2nd grade students -classroom	within-groups design	Math -Accuracy -Task duration -Problem completion rate -Student perceptions	Explicit timing IS	Interspersal led to higher number of problems completed correctly and had more favorable student perceptions. No difference on number of target problems completed correctly or in problem completion rates. Explicit timing required less time to complete assignment.	No significant difference.
Rhymer & Morgan (2005)	45 3rd grade students -classroom	within-groups design	Math -number of problems completed -accuracy	Explicit timing IS	Interspersal lead to higher number of problems completed and more favorable student perceptions. Explicit timing led to a higher number of target problems completed. No difference in accuracy.	76% of students preferred interspersal.

Summary Table of Interspersal Literature with the General Education Population (continued)

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Roberts & Shapiro (1996)	42 2nd grade students -classroom	between groups design	Reading -Rate of words read correctly /incorrectly -Number of words learned	IS with differing drill ratios assessment only (control) group	Intervention did not lead to any significant difference in gains compared to the control.	n/a
Roberts, Turco, & Shapiro (1991)	42 2nd-5th grade students -school	between groups design	Reading -Rate of words read correctly /incorrectly -Number of words learned	IS with differing drill ratios	Interspersal ratios with higher percentages of unknown words resulted in greater number of words learned.	n/a
Robinson & Skinner (2002)	30 7th grade students -school	within-groups design	Math -Scaled scores -Percentage of correct responses	NI IS	No significant differences on multiplication subtest. Interspersal resulted in higher scaled scores in mental computation.	Preference for interspersal (approaching significance).
Schmidgall & Joseph (2007)	6 1st grade students -classroom	alternating treatment design	Reading -Accuracy -learning rate	NI IS other	Other (phonics) led to more words learned. NI led to highest cumulative learning rates.	More students preferred other (phonics).
Szadokierski & Burns (2008)	27 4th grade students -1:1 classroom	within-groups design	Esperanto words -Number of words maintained	All IS:Inremental rehearsal with differing drill ratios	No significant differences found based on drill ratio. Higher opportunities to respond resulted in increased rates of maintenance.	n/a
Volpe, Mule, Briesch, Joseph, Burns (2011)	4 1st grade students -1:1 classroom	multielement designs	Sight words - words mastered, maintained, generalized -Learning rate	NI IS: Incremental rehearsal	No interspersal led to the fastest learning rate. No differences on other measures.	No difference between conditions.
Wildmon, Skinner, McCurdy, Sims (1999)	76 high school students -school	within-groups design	Math -Total and target problems completed -Student perception	NI IS	Interspersal led to more problems completed favorable perceptions. No interspersal led to more target problems completed. No difference in accuracy.	Preference for interspersal.

Note. Abbreviations are used for no interspersal (NI), interspersal (I), interspersal similar (IS)

Appendix C

Summary Table of Interspersal Literature with Atypical Populations

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Adcock & Cuvo (2009)	3 children with autism -Therapy room	Multiple baseline across participants	Varied-tasks -Percentage of correct responses	NI treatment package with interspersal	Treatment package with interspersal led to increase in percentage of correct responses.	n/a
Benavides & Poulson (2009)	3 children with autism - School/home	Multiple baseline across participants	Matching -Percentage of correct responses	NI IS IS +low density sr+	Interspersal led to increase in percentage of correct responses	n/a
Browder & Shear (1996)	3 students with disabilities -special education classroom	Multiple probe	Sight words -Number of words correct -Percentage of correct/incorrect responses	IS: Incremental rehearsal	Teaching package with interspersal led to increases on all measures.	n/a
Burns (2007)	9 year old boy with intellectual disabilities -special education classroom	alternating treatment design	Sight words -Number of words mastered	IS: different drill ratios/number of opportunities to respond	The condition with the higher number of opportunities to respond resulted in higher number of words mastered	n/a
Burns & Dean (2005)	5 4th grade students -1:1	Within subjects group design	Esperanto words -Number of words maintained -Percentage of on-task intervals	NI IS: Incremental rehearsal with various drill ratios	Interspersal with higher percentage of known words had highest number of words maintained and percentage of intervals with on-tsk behavior.	n/a
Burns & Boice (2009)	20 4th grade students in special education 1:1 classroom	Within subjects group design	Esperanto words -Number of words maintained	NI IS IS: Incremental rehearsal	Incremental rehearsal led to most words maintained. No difference between interspersal and no interspersal.	n/a

Summary Table of Interspersal Literature with Atypical Populations (continued)

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Calderhead, Filter, Albin (2006	Middle school girl in special education -classroom	alternating treatment design	Math -Rate of on-task behavior -Percent of correct responses	NI IS with different drill ratios	Interspersal led to higher mean rates of on-task behavior. No difference in accuracy between the conditions.	n/a
Charlop, Kurtz, & Milstein (1992)	5 children with autism -Tutoring room	Multiple baseline across participants	Varied-tasks -Percent of correct responses to acquisition tasks	All Interspersal conditions with differing schedules of reinforcement	Interspersal with a richer schedule/quality of reinforcement led to higher percentages of correct response.	n/a
Cooke, Guzaukas, Pressley, & Kerr (1993) Study 1	4 adolescents with disabilities -Classroom	alternating treatment design	Spelling -Percent of correct responses -Number of words learned per minute	NI IS: Incremental rehearsal	Similar percentages of correct responses across both conditions. No interspersal led to higher learning rates.	Preference for interspersal
Cooke, Guzaukas, Pressley, & Kerr (1993) Study 2	3 students with learning disabilities 1:1 in class	alternating treatment design	Math -Accuracy rate for written and oral tests	NI IS: Incremental rehearsal	Interspersal led to higher accuracy rates	2 of 3 preferred interspersal
Cooke, Guzaukas, Pressley, & Kerr (1993) Study 3	3 boys with learning disabilities -in class with peer tutors	alternating treatment design	Reading -Number of words read correctly per minute	NI IS: Incremental rehearsal	No interspersal led to higher numbers of words mastered.	Preference for interspersal
Horner, Day, Sprague, O'Brien, & Heathfield (1991)	3 students with intellectual disabilities -group home	Reversal design	Varied-tasks Aggression and self injury -Percent of trials with attempt to respond	NI: easy NI: hard I with hard tasks	Interspersal led to lower rates of aggression than all hard tasks and higher percentage of trials with attempts to respond.	n/a
Reed, Luiselli, Moizio, & Child (2010)	9 year old student with autism -in classroom	Sequential modification with alternating treatment	Matching -Rate of self-injurious behavior	Difficulty and rate Interspersal Interspersal + task novelty	Interspersal conditions led to reduced rates of SIB	n/a

Summary Table of Interspersal Literature with Atypical Populations (continued)

Reference	Participants and setting	Design	Dependent Variables	Independent Variables	Outcomes	Preference Assessment
Skinner, Hurst, Teeple, & Meadows (2002)	4 students with emotional and behavioral disorders -classroom	Alternating treatment design	Math -Percentage of intervals of on-task behavior	NI IS	Data were variable with little differentiation in data paths. Overall percentage of intervals of on-task behavior was higher in Interspersal.	n/a
Knight, Ross, Taylor, & Ramasamy (2003)	4 students with disabilities -school	modified parallel treatment design	Sight words -Percentage of words read correctly	Constant time delay IS: Incremental rehearsal	Constant time delay resulted in higher percentage of words read correctly for 2 participants, little differentiation for 2.	n/a
Koegel, Singh, & Koegel (2010)	4 children with autism -home or after school program	nonconcurrent multiple baseline	Math/Writing -Response latency -Rate of assignment completion -disruptive behavior	treatment package that included interspersal	Implementation of treatment package resulted in improvements on all measures.	n/a
Volkert, Lerman, Trosclair, Addison, & Kodak (2008) Study 1	5 children with autism -school	multielement and nonconcurrent multiple baseline design	Expressive labeling -Percentage of correct responses	NI ID IS	Little differential responding across conditions.	n/a
Volkert, Lerman, Trosclair, Addison, & Kodak (2008)	3 children with autism -school	mulitelement and nonconcurrent multiple baseline design	Expressive labeling -Percentage of correct responses	NI I both compared with : low quality sr+ high quality sr+	High quality sr+: Interspersal led to faster acquisition for 1 participant. Undifferentiated for others. Low quality sr+:Inconsistent results, 2 participants had faster acquisition with interspersal.	n/a
Wildmon, Skinner, Watson & Garrett (2004)	56 7th and 8th grade students with math learning disability	within groups design	Math -Number of problems completed -Number of target problems completed -Accuracy of target problems	NI IS	Interspersal led to higher number of problems completed. No difference on number of target problems completed or target problem accuracy.	87% of students preferred interspersal

Note. Abbreviations are used for no interspersal (NI), interspersal (I), interspersal similar (IS)

Appendix D

List of Books Used in Generalization Probes

Feiffer, J. (1999). *Bark, George.* New York, NY: HarpersCollins Publishers.

Hall, K. (2004). *My new school.* New York, NY: Scholastic Library Publishing.

Hirsch, Jr., E.D. (Ed.). (2002). *What your third grader needs to know, Revised edition*. New York, NY: Random House Publishing Group.

Mayer, M. (2010). *Going to the firehouse.* New York, NY: HarpersCollins Publishers.

Mayer, M. (2010). *Going to the sea park.* New York, NY: HarpersCollins Publishers.

Mayer, M. (2010). *Just a little sick.* New York, NY: HarpersCollins Publishers.

Mayer, M. (2010). *Snowball soup.* New York, NY: HarpersCollins Publishers.

Mayer, M. (2010). *This is my town.* New York, NY: HarpersCollins Publishers.

Mayer, M. (2010). *To the rescue!* New York, NY: HarpersCollins Publishers.

Rosetta Stone (1975). *Because a little bug went ka-choo!* New York, NY: Random House, Inc.

Scarry, R. (2011). *I know my opposites!* Nashville, TN: The Clever Factory, Inc.

Seuss, Dr. (1960). *Green eggs and ham.* New York, NY: Beginner Books; Random House, Inc.

Seuss, Dr. (1960/1988). *One fish two fish red fish blue fish.* New York, NY: Beginner Books; Random House, Inc.

Simmons, J. (1997). *Come along, Daisy!* Great Britain: Orchard Books.

Ziefert, H. (2000). *Little red riding hood.* New York, NY: Viking and Puffin books.

Appendix E

Known words for each participant identified during pretests and listed in alphabetical order

Participant 1	Participant 2	Participant 3	Participant 4
A	A	Airplane	Bottom
Am	Anna	Apple	Breakfast
At	Ask	Ball	Brought
Bag	Are	Big	Busy
Ball	(child's last name)	Blowing	Desk
Block	Can	Blue	Different
Brown	Can	Book	Drop
Do	Cat	Chicken	Dream
Four	Come	Computer	Dust
Help	Dad	Cookie	Early
Here	Dog	Coughing	Herself
I	Hat	Duck	Himself
Is	I	Egg	Quiet
It	In	Hairbrush	Radio
Jump	Is	Heather	Ready
Like	It	Hunter	Remember
Look	Mom	Key	Rich
Me	Mop	Kissing	River
New	My	Laughing	Roof
No	Now	My	Sand
On	Pig	Orange	Save
Out	Pup	Pencil	Space
Penny	Red	Purple	Stairs
Play	Sarah	Scissors	Stand
Run	Stop	Want	Station
Small	The	Waving	Stay
The	To	Writing	Still
Three	Two	Yellow	Story
Tree	Us	Yes	Straight
Two	We	zoo	Wheel

Appendix F

List of known gross or fine motor imitation responses identified during pretesting

Participants 1-3	**Participant 4**
Sign language: k	Prayer hands (palms together)
Sign language: l	Touch left elbow
Sign language: d	Touch table with pointer
Sign language: t	Open and close both hands
Sign language: i	Knock on table
Prayer hands (palms together)	Touch chin
Fist on palm (rock)	Touch both eyes
Palm on palm (paper)	Both hands on head
Index and middle finger on palm (scissors)	Rub palms together
Rub palms together	Touch right elbow
Both thumbs up	Cross arms in front of chest
Fingers interlaced, both index fingers up	Rub right arm
Thumb to little finger	Arms out to side (airplane arms)
Open and close both hands	Raise right arm
Pinching with index and thumb	Pull ears
Peace sign (index and middle finger in a V)	Clap
Drum fingers on table	Both hands on stomach
Hook index fingers	Both hands on cheeks
Thumb to middle finger	Peace sign (index and middle finger in a V)
Index finger to thumb, both hands, hook together	Stick out tongue
Interlace fingers, stretch arms out	Raise both arms
Diamond with index fingers and thumbs	Point with both index fingers
Four fingers up	Blow
Both hands, fingertips together	Touch both shoulders
Touch palm	Raise left arm
Interlace fingers	Open and close mouth (chomp)
‘ok’ sign	Touch nose
Clasp hands together	Cover mouth
Clap	Hand over hand flat on table
Thumb to ring finger	Cover eyes

Appendix G

Words Mastered by Each Participant, listed chronologically (2 pgs)

Participant 1	Participant 2	Participant 3	Participant 4
After	Goes	First	Site
Soon	Does	Down	Energy
Ate	On	Pretty	Paragraph
White	From	By	Length
Will	She	After	Perhaps
Our	Said	Every	Held
Went	Ride	Would	Cells
Must	Use	Letters	Syllables
Old	Who	An	Poetry
Have	No	Here	Apache
Down	Funny	Did	Aqueducts
Fly	Know	May	Hemisphere
Over	What	Before	Geography
Cold	Our	Fly	Central
Again	Find	That	Isthmus
Away	Too	Live	Prosperity
Put	Going	Dishes	Record
Every	Could	Gave	Center
Ran	Before	Five	Distance
Just	Myself	Ate	General
Let	First	Always	Europe
Once	Together	Think	Region
May	Seven	Once	Paint
Make	Today	Were	Biography
Into	Start	Shapes	Iroquois
Thank	Throw	Vehicles	Platypus
Ask	Silver	Colors	Anasazi
There	Food	Heavy	Amphitheater
But	Smile	Rooms	Navajo
Please	Leave	Wash	Plateau
Good	Cloud	Circle	Congruent
Funny	Sand	Oval	Product
Want	World	About	Multiplication
How	Born	Characters	Tibia
Never	There	Good	Fibula
Write	Wash	Star	Vertebrate
Warm	Again	Bring	Regrouping
Keep	Sleep		Vertex
Drink	Better		Geometry
Call	Found		Invertebrate
Made	Light		Polygons
Soft	Eight		Amphibians
Very	Airplane		Aluminum

Ant	Fresh	Glaciers
Dry	Plant	Habitat
Could	Gold	Phases
Are	Left	Constellation
Pretty	Gift	Nutrients
Take	Slow	Chlorine
Ride	Fur	Yosemite
Find	People	Telescopes
She	Yard	Astronomers
Right	Banana	Absorbed
Why	Country	Conservation
Their	Began	Asteroid
Sit		Pollutes
Upon		Galileo
Does		Gravitational
Both		Recycling
Draw		Meteorite
Clean		Satellite
Better		Ecology
Hurt		Emissions
Clock		Attraction
Thin		
Show		
Done		
Only		
Grow		
About		
Fall		
Pick		
Fur		
Bug		
Kick		
Comb		
Food		
Broom		
Horse		
Rex		
Crayon		
Letters		
Dark		
Woody		

Appendix H

Percentage of Sessions Scored for IOA, and the IOA Scores Across Participants and Conditions

Participant 1

	Total number of sessions conducted	Total number of sessions IOR scored	Percentage of total sessions IOR scored	IOR of sessions scored
Baseline	62	9	14.52%	98% (80-100%)
No Interspersal	42	5	11.9%	100%
I- Similar	49	6	12.24%	100%
I- Dissimilar	7	1	14.29%	100%

Participant 2

	Total number of sessions conducted	Total number of sessions IOR scored	Percentage of total sessions IOR	IOR of sessions scored
Baseline	36	5	13.89%	100%
No Interspersal	14	4	28.57%	99% (90-100%)
I- Similar	44	7	15.91%	100%
I-Dissimilar	30	3	10.00%	100%

Participant 3

	Total number of sessions conducted	Total number of sessions IOR scored	Percentage of total sessions IOR scored	IOR of sessions scored
Baseline	54	4	7.41%	100%
No Interspersal	46	1	2.17%	100%
I-Similar	62	4	6.45%	100%
I- Dissimilar	17	2	11.76%	100%

Participant 4

	Total number of sessions conducted	Total number of sessions IOR scored	Percentage of total sessions IOR scored	IOR of sessions scored
Baseline	60	4	6.67%	98% (90-100%)
No Interspersal	25	3	12%	100%
I- Similar	50	7	14%	100%
I-Dissimilar	25	3	12%	100%

Appendix I

Social validity survey sent to the parents of participants after the conclusion of the study (2 pgs.)

The Role of Interspersal Procedures in Sight-Word Acquisition
Parent Satisfaction Survey

Your child recently participated in a study on the role of interspersal procedures in sight-word acquisition. Your feedback about the study goals, procedures, and outcomes is important. Please complete the survey and return to me with the pre-addressed, stamped envelope. Thank you!

Outcomes

How would you rate your satisfaction with the progress that your child made during his /her participation in the study?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How would you rate the helpfulness of the information learned through this study regarding your child's learning style?

1	2	3	4	5
extremely unhelpful	unhelpful		helpful	extremely helpful

Goals and Procedures

How satisfied were you with the words that were selected for the study? Words were pulled from the Dolch sight word list, a broad based curriculum, and/or by parent/therapy team requests.

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How satisfied were you with teaching procedures used during the sessions?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How satisfied were you with the schedule and location of the sessions?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

Research Staff

How would you rate the overall experience of working with the researcher?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How would you rate the researcher's competency?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How would you rate the researcher's ability to answer questions about your child's performance during sessions and in the study overall?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

Overall Experience

How would you rate the overall experience of participating in this study?

1	2	3	4	5
extremely dissatisfied	dissatisfied		satisfied	extremely satisfied

How likely would you be to recommend this study to another family?

1	2	3	4	5
extremely unlikely	unlikely		likely	extremely likely

What did you like most about the study?

What did you like least about the study?

Do you think your child's ability to read has changed as a result of this study?