Task Interspersal and Performance of Matching Tasks by Preschoolers with Autism

by

CHRISTIAN A. BENAVIDES

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of

the requirements for the degree of Doctor of Philosophy,

The City University of New York

2007

UMI Number: 3278420

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Dr. Claire L. Poulson Chair of Examining Committee
Date	Dr. Joseph Glick Executive Officer

Dr. Nancy M. Hemmes

Dr. Bruce L. Brown

Dr. Marilyn K. Rousseau

Dr. Robert W. Allan

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

TASK INTERSPERSAL AND PERFORMANCE OF MATCHING TASKS BY PRESCHOOLERS WITH AUTISM

by

Christian A. Benavides

Adviser: Professor Claire L. Poulson

The current study examined the effects of task interspersal on the performance of matching-to-sample tasks by three children with autism. A pre-baseline assessed each child's mastery level of a large body of matching stimuli. These matching tasks included matching identical and non-identical animals, numbers, letters, and shapes. Through this assessment *mastered* and *non-mastered* matching-to-sample stimuli were determined empirically. Following a *baseline* condition that presented only non-mastered stimuli in succession, *treatment* was introduced in a multiple baseline design across children. During the treatment condition, trials with mastered stimuli were interspersed with trials with non-mastered stimuli. For all three children, the percentage of correct matching responses to the non-mastered stimuli increased systematically with the introduction of the interspersal procedure. Following treatment, a third condition was conducted that reduced the total number of reinforcers available per session to baseline levels. The data demonstrated that all three participants maintained treatment levels of correct responding during this third condition. Thus the increased reinforcement density during treatment was not needed for maintenance of correct responding. The discussion addresses additional control procedures that would be needed to evaluate the role of reinforcement density during treatment.

Acknowledgements

My sincerest and most heartfelt gratitude to Dr. Claire Poulson, who has been a well of strength and wisdom during the construction of this manuscript. This project would not have been possible without her guidance.

A warm and earnest thank you as well to Dr. Bruce Brown, Dr. Nancy Hemmes, Dr. Marilyn Rousseau, and Dr. Robert Allan who have all provided their considerable knowledge and expertise to the shaping of this project.

Thanks as well to the many teachers and staff members who allowed me to conduct this research. Special thanks to Dr. Carolyn Ryan, Rosanna Delladera, Adam Whiting, Marie Nicole Timal, and Marlen-Maria Constantinou for their time and effort. In addition, I thank the participants and their families for their time and their willingness to accommodate me.

Finally, thanks to my family and friends for their love, support, and understanding. In particular, thanks to my wife Lauren, for her patience and invaluable moral and logistic support.

Table of Contents

List of Tables

List of Figures

TASK INTERSPERSAL AND PERFORMANCE OF MATCHING TASKS IN PRESCHOOLERS WITH AUTISM

A large body of research has been devoted to the study of task interspersal as a variable in the presentation of training tasks (see Killu, 1999; Skinner, 2002; Belfore, Lee, Scheeler, & Klein, 2002; and Cates, 2005 for informative reviews). Such studies have been conducted with a variety of populations, including middle school students with mental retardation (Browder & Shear, 1996), young children with autism (Charlop, Kurtz, & Milstein, 1992; Dunlap & Koegel, 1980), adults with mental retardation (Neef, Iwata, & Page, 1980); and the victim of a severe stroke (Koegel & Koegel, 1986).

One subset of this research features interventions designed to improve responding on a variety of adaptive tasks. Some of the tasks targeted have included the identification of sight words (Browder & Shear, 1996; Knight, Ross, Taylor, & Ramasamy, 2003), spelling (Neef, Iwata, & Page, 1980; Cooke, Guzaukas, Pressley, & Kerr, 1993), multiplication & division (Cooke, & Reichard, 1996), and naming pictures (Rowan & Pear, 1985). In addition, latency to begin working on such tasks has been studied by Belfiore, Lee, Vargas, & Skinner (1997). This particular line of research is typified by early studies conducted by Dunlap & Koegel, (1980) and Neef, et al. (1980).

Neef et al. (1980) compared the effects of interspersed task presentation with a high-density reinforcement procedure on the acquisition of spelling words. Three young adults with mental retardation participated. During a pre-test, a large set of spelling words was presented twice to each participant. Words were categorized as learned (spelled correctly both times), unlearned (spelled incorrectly both times), or discarded (spelled correctly once).

The dependent variable of primary concern was the rate at which the unlearned spelling words were mastered. A particular word met mastery criterion when the word was written with correct spelling 5 consecutive times. Following a baseline condition in which 10 unlearned words were presented consecutively per session, two methods of task presentation were compared. During the *interspersal* condition, the presentation of learned words was interspersed with unlearned words on a 1:1 basis. During the *high-density reinforcement* condition, rate of reinforcement was increased non-contingently. Unlearned words were presented consecutively as they had been in baseline, however unlike baseline, ten presentations of non-contingent social reinforcement were included in each session. In all conditions, words that became mastered were replaced with new unlearned words after meeting mastery criterion.

Interspersal and high-density reinforcement were compared using a multi-element design. Results indicated that unlearned words met the mastery criterion faster during the interspersal condition when compared with both the baseline and the high-density reinforcement condition. In a post-experiment retention test, higher retention rates for words learned during interspersal conditions were found. An additional feature of the study was the presentation of four opportunities for the participants to choose their intervention (either interspersal or reinforcement) during randomly selected sessions. The results of the preference assessment showed interspersal training was chosen over high-density reinforcement during 11 of 12 choice opportunities across subjects.

Dunlap & Koegel (1980) similarly investigated the effects of interspersed task presentation on the responding of two children with autism (ages 7 & 5 years). Multiple target tasks were randomly selected from each participant's curriculum of objectives.

Examples of target tasks included identifying the "first" of a sequence of two actions and the use of possessive pronouns. Additional curriculum tasks were selected to be included during interspersal. The primary dependent variable was the percentage of all responses performed correctly. A constant task condition in which a target task was presented repeatedly in isolation during baseline was compared to an interspersal task condition during treatment using a multiple-baseline across tasks design.

The results generally demonstrated improved responding during interspersal; however, the conclusions that can be drawn from the data are limited by certain factors. For participant 1, the difference between performances during the two presentation methods was present for only 2 of the 3 target tasks (2 of the 3 legs of the multiple-baseline design). For participant 2, the increased correct responding present in the interspersal condition was limited to 1 data point for each of only two target tasks, thus, experimental control was weak. In addition, the session length was varied across sessions with no reported criteria for session termination. Thus, any changes in correct responding could be a function of session length rather than task interspersal. Also, no measures of teacher behavior were collected. Therefore, errors in treatment integrity might have been confounded with task interspersal. Finally, data on correct responding to non-target tasks were not separated from target task data. Consequently, one cannot evaluate the extent to which task interspersal might have influenced responding on target tasks alone.

Later studies added to the findings of Neef et al. (1980) and Dunlap & Koegel (1980). For example, Rowan & Pear (1985) examined the importance of using mastered tasks to intersperse as opposed to non-mastered tasks. Three children with mental retardation (ages 7-11) were taught picture naming during the study. Two conditions were compared using a reversal design: an *interspersal* condition in which mastered pictures were interspersed with target non-mastered pictures, and a concurrent condition that interspersed non-mastered pictures with the non-mastered target pictures. The results demonstrated improved rates of acquisition during interspersal over concurrent picture presentation.

Cooke & Reichard (1996) investigated mastered-to-non-mastered task ratios of interspersal in an effort to identify ratios that are likely to produce the greatest improvement in performance. Mastered-to-non-mastered ratios of 7:3, 1:1, and 3:7 were used with multiplication and division problems presented to five 5th grade students with learning disabilities and one 5th grader with behavioral-emotional handicaps. The results suggested that a mastered-to-non-mastered ratio of 3:7 was generally the most efficient of those tested.

A related study by McCurdy, Skinner, Grantham, Watson, & Hindman, (2001) measured the effects of task interspersal on the on-task behavior of a typically developing 9-year old girl. On-task behavior was measured while she completed two types of math assignments: a control assignment and an experimental assignment with easier problems interspersed with control problems. Both assignment types included problems from the current classroom curriculum. A multi-element design was used to assess the effects of the interspersal procedure present in the experimental assignment. Results indicated consistently more on-task behavior during interspersed assignments. A later study by Calderhead, Filter, & Albin (2006) yielded similar findings.

While the study of task interspersal has progressed to include various dependent measures, populations, and experimental parameters, its use with children with autism has received relatively little attention since Dunlap & Koegel (1980). The current study addresses some of the limitations of earlier experiments on interspersal with this population. Specifically, the present study established a clear delineation between data reported on non-mastered tasks and mastered tasks (used during interspersal). In addition, the session length was set to a definite number of response opportunities and the performance of instructors presenting the tasks was measured. Also, the difference between mastered and non-mastered tasks was established empirically in a pre-baseline assessment. Finally, the current study attempts to evaluate the effects of decreased reinforcement density during a maintenance procedure that followed the interspersal procedure. The interspersal procedure itself had been confounded with an increase in density of reinforcement over the baseline condition. Suggestions for further analysis of this confound are made in the discussion.

Method

<u>Subjects and Setting</u>

Three children with autism, *Jose*, *Zach*, and *Nick* participated in the current study. Jose and Zach were five while Nick was three years old. The children all attended a self-contained preschool for children with autism. The school features classrooms of 6 or 8 children taught by 3-4 staff members per room. Staff members typically use many ABA-based teaching procedures including discrete-trial and incidental teaching, shaping, and chaining.

The children were selected based on two criteria: consistent performance of sitting and following simple instructions and inconsistent performance on visual, two-dimensional matching tasks. Prospective participants were selected based on reports by their classroom teachers. Several of these prospects were evaluated in a pre-baseline assessment of matching skills. Those students who demonstrated deficits in the target skills then became participants.

At the onset of the study, sessions for all three students took place in the school lunch room during off-hours. For Jose and Zach, sessions eventually moved to their homes after they graduated from the aforementioned preschool. Sessions were conducted and observed by trained associates of the experimenter. For the majority of sessions one associate presented the trials and provided reinforcement while another observed and collected data. During several sessions a second observer was present to collect interobserver agreement data.

Parents of all participants signed a consent form approved by the City University of New York Internal Review Board.

Tasks and Materials

Four categories of visual stimuli to be described below were used in the matching trials: animals, letters, numbers, and shapes. In total, 837 different visual stimuli were available for use in the study. The breakdown of the stimuli presented in each category is presented in Table 1. The table further breaks down the categories into the sub-categories of *subsets* and *exemplars*. The definitions of subsets and exemplars are dependent on the category of stimuli, as follows.

For animals, *subsets* included the 15 different animals presented (e.g. cat, dog, duck) and *exemplars* refers to the 12 different pictures of each subset of animal. The pictures of animals all came from a commercially available software package (Clickart 400,000). For letters, the 26 subsets are the 26 letters of the alphabet, while the 12 exemplars refer to pre-determined font/color combinations. For numbers, 16 subsets represented the numbers 0-15, and the exemplars were font/color combinations similar to the letters. Finally, shape subsets were 17 different geometric shapes (e.g. circle, square, triangle, while the exemplars are the same shape in 9 different colors. Shape/color combinations were created using Microsoft Word 2003. A sampling of the stimuli used is available in Appendix A. These stimuli were made into picture cards approximately 2x2 inches in size.

The aforementioned pictures were used to create a large number of matching-to-sample trials. Trials were presented in an activity binder that featured several pages of 1 to 3 comparison pictures on the left side and 1 sample picture on the right. For example, if a red circle was the target stimulus on the right page, the three choice stimuli on the left page might be a blue circle, a yellow square and a black triangle. The pictures were attached to the various binder pages using Velcro. To determine which stimuli would be used in the experimental analysis, performance on these tasks was measured in the pre-baseline assessment.

In addition to the target stimuli, several larger pictures of preferred stimuli were also used. These stimuli were often favorite characters from books or cartoons identified by interviewing members of the children's teaching staff. Picture cards were made from these stimuli that were over twice the size of the target stimuli cards (the preferred stimuli cards were approximately 4 inches x 4 inches). These pictures were presented in the same matching format used for target stimuli except there was only one distracter on the right.

Table 1

Total Numbers of Visual Stimuli available by Categories, Subsets, and Exemplars

Category	Subsets	Exemplars per Subset	Total Stimuli
Animals	15	12	180
Letters	26	12	312
Numbers	16	12	192
Shapes	17	9	153
Total			837

Measurement and Reliability

The primary dependent variable was the percentage of correct responses to non-mastered matching-to-sample tasks. Matching stimuli were categorized as mastered or non-mastered using a pre-baseline assessment phase to be described below. Correct matching responses were defined as independently picking up the matching stimulus from the left-side page and attaching it to the right-side page directly under the target stimulus from the same subset (e.g. a yellow circle is selected and placed beneath a red circle). This response needed to occur within 10 sec of the presentation of the matching task (i.e. the page being turned to the current task) to be considered correct.

In addition, during the experimental phase, the duration to complete each session was measured. The start of each session was defined as opening the task book to the first item. Each session ended when the last trial was completed by the placement of the correct picture onto the right page either by a child's correct response or by the instructor's prompt. Duration was measured in minutes and seconds using an electronic timer.

Data were collected in vivo by a trained observer. A second observer collected interobserver agreement data on 26 – 32% of each child's sessions (10 of 38 sessions for Nick, and 12 of 38 for Zach & Jose). For correct responding to matching-to-sample items by the participants, interobserver agreement was taken by dividing all the agreements between the two observers in a given session by the total number of items in that session. For this measure, interobserver agreement yielded a range of 83-100% with an overall mean of 98% across all participants. Interobserver agreement for session duration was measured as the difference in time to complete a session recorded by the two observers. The differences ranged from 0 to 8 seconds with a mean of 1.85 seconds. The breakdown of interobserver agreement measures is presented in Table 2.

Finally, the correct application of consequences by the instructor was also measured for treatment integrity purposes. Each matching trial resulted in one of three consequences: a *prompt* if the student made no response after 10 sec, *praise* if the student made a correct match or *correction* of an incorrect match. A prompt consisted of physically guiding the student's hand to pick up the appropriate picture from the left page, placing it beneath the appropriate matching picture, and letting go. Praise was defined as an enthusiastic statement indicating that the correct response was made by the student. Correction consisted of removing the incorrectly placed picture back to its original position and then presenting a prompt as described above. Across all sessions, treatment integrity ranged from 67 to 100% with a mean of 98%. Treatment Integrity measures are presented in Table 3.

Table 2

Interobserver Agreement on Student Responses and Differences in Recorded Session Durations

	Percent of Responses Correct		Duration (sec)	
	Mean	Range	Mean	Range
Jose	98.96%	91.67-100%	1.56	0-7
Zach	98.61%	83.33-100%	1.67	0-8
Nick	95.83%	83.33-100%	1.60	0-8

Table 3

Percentage of Consequences Delivered Correctly by Instructors Working with Three Participants

	Mean	Range
Jose	99.06%	91.67-100%
Zach	96.39%	66.67-100%
Nick	97.72%	83.33-100%

Assessment

During the pre-baseline assessment, a random sample of each subset was assessed with each participant. Four exemplars of each subset were randomly selected to be target stimuli in the matching task described above. Matching trials were presented on consecutive pages in a letter-sized activity binder. During this initial phase of the assessment, the randomly selected exemplar was presented alone on the right page, while a non-identical exemplar of the same subset was presented on the left page. Along with this non-identical exemplar, 2 distracter stimuli of different subsets of the same category were present on the left page. For example, if a red circle had been randomly selected to appear on the right page, a circle of a different color would appear on the left page along with 2 other shapes of different colors

This phase of the assessment consisted of 296 trials presented over seven sessions. Each subset that was performed correctly on 3 or 4 of the 4 trials presented was considered mastered and not presented to that particular participant for the remainder of the study.

Following this initial phase, a second phase of the assessment was presented to identify potential mastered items for use during the interspersal phase of the experiment. Several familiar and preferred stimuli were selected for each child to serve this purpose. These stimuli consisted of cartoon and storybook characters recommended by the student's classroom teachers. Each stimulus selected this way was presented six times. Stimuli matched correctly on 5 or more presentations were considered mastered. Using this method, 24 mastered stimuli were identified for both Jose and Zach. This method did not reveal any mastered stimuli for Nick. Afterwards, this task was further simplified for Nick by removing the distracter and subsequently 24 mastered stimuli were identified for him as well. Following assessment, the experimental phase of the study began with baseline.

Baseline

Prior to the start of the baseline condition, eight sequences of 12 non-mastered stimuli were randomly selected for each student. These eight sequences would subsequently be presented to the students for the remainder of the study. One of each of these sequences was used as the target stimuli for each session. The order of sequence presentations was block randomized so that each sequence was presented exactly once before the next block of randomized sequences was presented.

During baseline, the student was seated in front of a table and given a task book from an instructor seated directly behind him. One or two observers were seated off to the side of the table, approximately four feet away. The task book included 12 matching-to-sample tasks, each one featuring one of the 12 non-mastered stimuli that had been pre-selected. Similarly to the previously described assessment procedure, 2 randomly selected distracter stimuli featuring different subsets of the same category filled out each trial. For Zach and Jose, the non-identical matching stimulus was presented along with two distracter stimuli. For Nick, the matching stimulus was presented with only one distracter. The target criterion response was to match a non-identical stimulus of the same subset to the target non-mastered stimulus.

Before the start of each session the instructor gave the participant a generalized instruction to begin such as “let’s do your activity book” or “time to start your book.” Each session began when the child opened the book to the first page and ended when the last trial was completed. Prior to the start of a session, each child had an opportunity to choose a play activity from a choice board that he would work towards during the session. Each child had to earn 12 tokens to gain access to his chosen activity. If fewer than 12 tokens were earned during the session, the child was allowed to earn additional tokens following the end of the session by correctly responding to simple 1-step instructions such as "point to your nose" or "point to your head." When baseline was completed, the interspersal condition began.

Interspersal

Procedures for the interspersal condition were similar for the most part to those described for baseline. The notable exception was the inclusion of 12 trials that featured mastered stimuli that had been previously identified during assessment. These 12 mastered trials were interspersed among the same 12 non-mastered trials that. For Zach and Jose, the target response during these trials was to match non-identical pictures of the same character from a field of two distracters. For Nick, the target response was identical matching of the same picture with only one distracter.

Each session now consisted of 24 trials, 12 non-mastered stimuli and 12 mastered stimuli. Each trial still held an opportunity for reinforcement. The same token reinforcement system from baseline was used for interspersal. If twelve tokens were earned prior to the end of the session, the session was stopped for 2-3 minutes so the child could engage in the activity he had chosen. The interspersal condition was followed by a third condition, described below, which attempted to control for the possible effects of additional reinforcement opportunities available during interspersal.

Interspersal/Low Density Reinforcement

The interspersal condition altered the heterogeneity of the difficulty of task items presented in baseline, but it also may have altered the density of reinforcement for correct responding. In order to control for the increase in reinforcement density available in the interspersal condition, an interspersal/low density condition (ILD) was run.

Procedures for the ILD condition were similar in all ways to the interspersal condition with one notable exception. During ILD only 12 of the 24 trials run per session held an opportunity for reinforcement. For every session, 6 of 12 non-mastered or target trials and 6 of 12 mastered trials were allotted for reinforcement availability. The child still needed to make the correct response to receive reinforcement during these pre-determined trials. The 12 trials specified for reinforcement availability were pre-determined with a random number generator.

Experimental Design

A multiple-baseline-across-subjects experimental design was used to compare the effects of interspersal with baseline. The design staggered the onset of the interspersal condition for first Jose, then Zach, then Nick. This design was followed by a third condition that investigated the effects of reinforcement density on percentage correct responding by recreating the reinforcement density established during baseline.

Results

Figure 1 presents the percentage of correct matching-to-sample responses to target (non-mastered) trials by the three participants across the three conditions. Overall, all three children demonstrated a systematic increase in percentage of correct responding with the introduction of the interspersal procedure in phase 2. This level of responding was maintained during the interspersal/ low-density reinforcement (ILD) phase.

Jose started baseline responding correctly to 33% of the trials presented. For the next 10 sessions of baseline, his correct responding ranged between 17% and 59% and maintained a level trend. When interspersal was introduced with session 11, Jose responded correctly to 42 % of matching-to-sample items. He increased to 92% in session 16 and remained fairly stable at that high level for the remainder of both the interspersal and ILD conditions.

Zach's percent correct ranged between 0% - 33% during 18 sessions of baseline. His data remained fairly stable between 33% - 42% over the next 4 sessions of interspersal. During the 5th session of interspersal, Zach's percent correct increased to 75% and then to 92% the session afterwards. From there, Zach's data remained fairly stable for the remainder of both the interspersal and ILD conditions, with a range of scores between 42% - 92% for the remainder of the study.

For Nick, baseline produced a level trend with a range of 0% - 42%. Correct responding then displayed an upward trend, from 42% to 83 % after the introduction of interspersal. As with Jose and Zach, data remained at that high level for the remainder of interspersal and ILD, with a range of scores between 58 % - 92%.

Figure 1. Percentage of matching-to-sample responses correct under Baseline (12 opportunities for reinforcement and correction procedure); Interspersal (Mastered Items Interspersed, 24 opportunities for reinforcement); Interspersal/ Low-Density Reinforcement (Interspersal condition with only 12 opportunities for reinforcement).

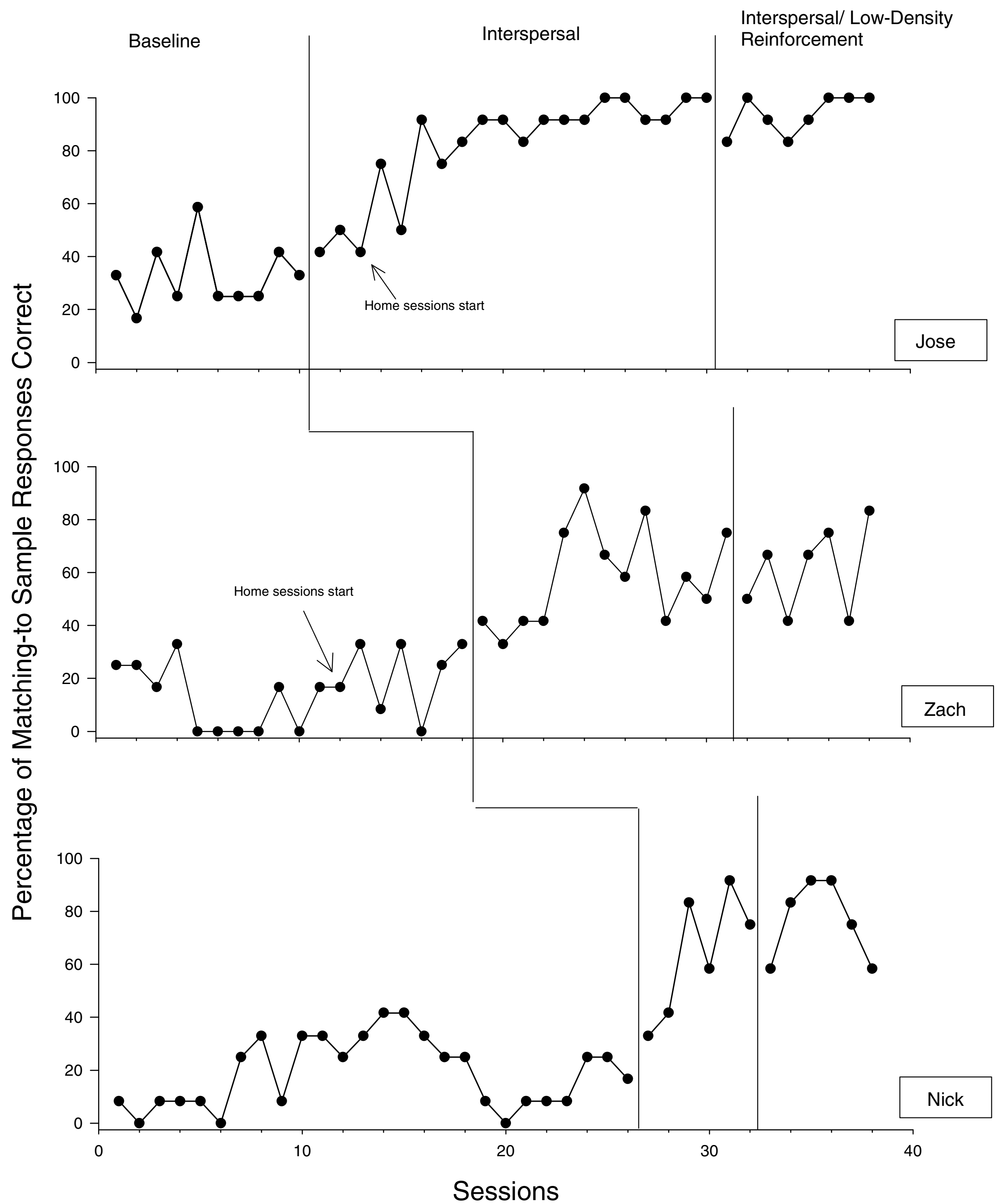

Baseline
Interspersal
Interspersal/ Low-Density Reinforcement
Percentage of Matching-to Sample Responses Correct
Home sessions start
Jose
Home sessions start
Zach
Nick
0
20
40
60
80
100
0
10
20
30
40
Sessions

Fig. 1

Session duration data are shown in Table 4. Overall, for all three children, session duration was shortest during baseline and increased when more trials were added during the interspersal and ILD conditions.

Figure 2 shows the rate of reinforcement for each experimental condition for each of the three children. This figure shows that the rate of reinforcement delivered was consistently lower for the baseline and the interspersal/ low-density reinforcement conditions than for the interspersal condition.

The rates of reinforcement remained stable at a low level for all three children during baseline. Jose earned the most reinforcement of the three, with a high rate of 3.07 reinforcers per minute during session 5 with a low of 0.85 during session 2. Zach and Nick both had sessions with zero reinforcers earned during baseline. Nick, however, had a slightly higher peak (1.72 rpm during session 11) than Zach (1.15 rpm during session 2).

With the introduction of interspersal, rate of reinforcement increased immediately for Jose, and followed an upward trend throughout the condition (peaking at 10.91 during session 25). Zach demonstrated a gradual upward trend throughout interspersal, starting at 1.46 rpm during session 19 and peaking at 3.99 during session 27. Nick's rate per minute of reinforcement increased quickly to 3.27 rpm shortly after the introduction of interspersal, and remained at that high level for the remainder of the condition.

The rate of reinforcement then decreased during the interspersal/ low density reinforcement condition for all three children. With reinforcers available at the same rate as baseline, Jose's actual rate per minute of reinforcement earned decreased quickly to 4.09 at the start of ILD. This level of reinforcers earned per minute maintained as a stable trend for the remainder of the ILD condition. Likewise, Zach's rate of reinforcement decreased to 1.49 at the start of ILD and remained at that level, showing very little variability (a low of 0.90 in session 37). Finally, Nick's rate of reinforcement also dropped `immediately to 0.84 rpm and remained stable for the remainder of ILD (range: 0.79-1.98).

Table 4

Mean Duration of Sessions

	Mean Duration (sec)		
	Baseline	Interspersal	ILD
Jose	152.78	241.95	210.50
Zach	243.88	548.17	444.43
Nick	246.39	507.20	499.00

Figure 2. Rate of reinforcement (reinforcers per minute) delivered during three experimental conditions.

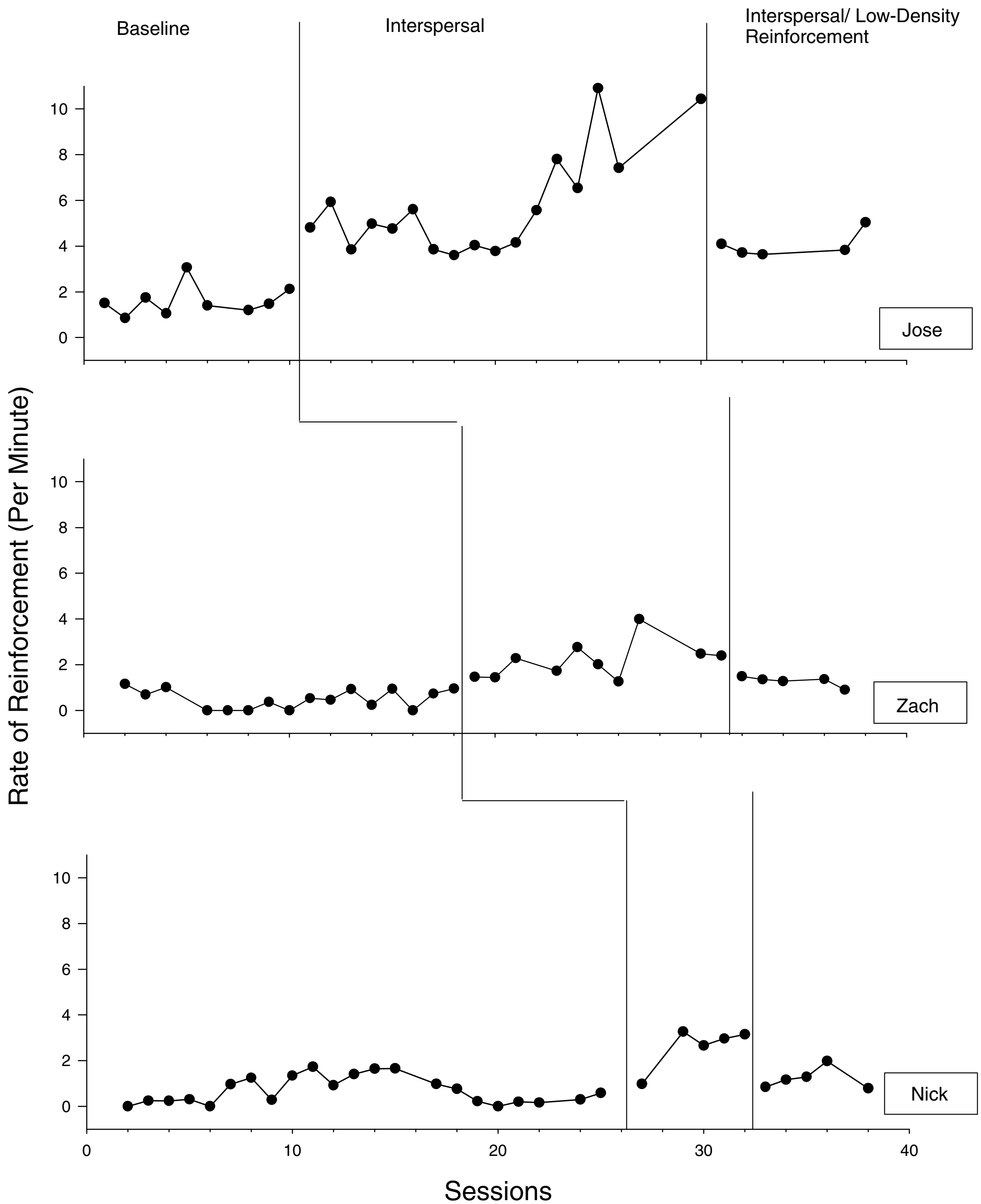

Baseline
Interspersal
Interspersal/ Low-Density Reinforcement
Rate of Reinforcement (Per Minute)
Jose
Zach
Nick
0
2
4
6
8
10
0
10
20
30
40
Sessions

Fig. 2

Discussion

Because there was a systematic increase in the percentage of correct responding to non-mastered matching-to-sample trials with the introduction of the interspersal procedure, one may conclude that the interspersal procedure caused the increase in the percentage of responses correct. As displayed in Figure 1, the percent correct responding to non-mastered matching-to-sample tasks increased systematically when interspersal was introduced to each of the three children. This result is a clear demonstration of the effectiveness of the interspersal procedure in teaching the matching-to-sample tasks to three participants. In addition, correct responding remained at the same high level when reinforcement was reversed to baseline levels. This demonstrates that decreased reinforcement rates during the third condition were not associated with decreased accuracy.

This finding fits into a large body of research demonstrating the effectiveness of task interspersal manipulations for facilitating a number of positive effects (Belfiore, et al. 2002). In particular, the current study builds on the work of Neef et al. (1980), Dunlap & Koegel (1980), and others who have interspersed tasks based on their previous mastery level. The current study extends the findings of Neef et al. (1980) by keeping a specific number of response opportunities consistent within experimental conditions, giving a clearer picture of the effectiveness of interspersal on acquisition over time. In addition, mastery level of the various stimulus subsets in the current study was established during an empirical assessment prior to the study.

The interspersal/ low density reinforcement condition served as a maintenance condition that kept the interspersed presentation used during treatment constant, while the rate of reinforcement was reduced to resemble baseline levels. As shown in Figure 2, the rate per minute of reinforcement reverted to near-baseline levels with the introduction of the ILD condition. As seen in Figure 1, this decrease in reinforcement density did not affect the level of correct responding. While these data demonstrate that decreased reinforcement rates did not affect accuracy during ILD, they do not address whether the increased reinforcement rate found during interspersal was necessary for the gains in accuracy made during that condition. For this particular confound to be addressed, reinforcement rate will have to be held constant in future studies by using a response-independent schedule of reinforcement that can be tightly controlled by the experimenters.

Another confound in this study is the temporal position of the earned reinforcer breaks. While the reinforcement contingency and exchange rate was held constant across all three conditions, the timing of the earned break varied across conditions. For all three children, these breaks were never earned till after the session was over during baseline. During interspersal, the additional reinforcement opportunities allowed these breaks to be earned in the middle of every session. During ILD, breaks were only earned at the end of every session once again for Zach and Nick; however, Jose earned mid-session reinforcer breaks during 4 of 8 ILD sessions. This confound could also be controlled in future studies by using a response-independent reinforcement schedule.

The extent to which an increase in reinforcement density is necessary for the acquisition of skills learned using task interspersal likely varies across participants and target responses. This is an interesting question for future research that is tied to our understanding of how interspersal procedures produce their effects.

Many explanations of such effects have been mentioned in the interspersal literature. Horner, Day, Sprague, O'Brien, & Heatherfield (1991), proposed that stimulus generalization could account for improvements in desirable behavior when high probability requests were interspersed with low probability requests.

Horner et al. conducted an interspersal study that featured three youths with a history of self-injury or aggression (ages between 12-14 years old). Two dependent measures were obtained for each participant: problem behavior (either self-injury or aggression), and attempts to complete a presented "hard" task. Hard tasks included putting on a shirt, sorting silverware, and following two-step instructions. "Easy" tasks were also identified for each participant for use in an interspersal intervention. Hard and easy tasks were empirically determined during a pre-baseline assessment. In addition to these tasks, "easy" tasks were also identified for each participant. Within a reversal design, separate experimental conditions presented hard tasks either alone or interspersed among easy tasks. Results indicated that during the interspersal condition, all three participants markedly decreased problem behavior and two of three participants increased attempts to perform hard tasks. The 3rd participant already had attempted to perform hard tasks during 100% of trials during the non-interspersed condition

Horner, et al. posited that interspersal was effective because compliance with interspersed requests was followed immediately by reinforcement and, therefore, compliance as a response class was strengthened. Thus, reinforcement for responding to varied instructions increased the likelihood of responding to the entire stimulus class of instructions

Although the target response class in the current study differs from the compliance responses measured by Horner et al. and many other researchers in the field of high-probability requests, it follows that stimulus generalization of matching-to-sample tasks may have been similarly strengthened in the interspersal condition. Reinforcement of the low difficulty matching-to-sample tasks may have increased attempts to perform the more difficult target tasks, resulting in the increased levels of correct responding.

A similar explanation for effects of task interspersal on preference of interspersed presentation known as the discrete task completion hypothesis is proposed by Skinner (2002). This hypothesis states that completion of a discrete task within a longer assignment functions as a conditioned reinforcer. Therefore the inclusion of additional, supposedly easier tasks for interspersal improves the overall rate of reinforcement (completion) within an extended task. While the results of the current study are consistent with the discrete-task-completion hypothesis, the presentation of extrinsic reinforcement for each correct response precludes the current findings from being considered as strong evidence in support of it.

Another limitation of the current study is the absence of generalization data on related tasks. To date, response generalization has received relatively little attention in interspersal research (Cates 2005). An investigation of interspersal procedures on generalization of a variety of tasks is a direction for future research.

Another direction for future research is an exploration of the fundamental theoretical underpinnings for the success of task interspersal interventions. The discrete-task-completion hypothesis is consistent with the stimulus control account for interspersal effects mentioned by Horner et al. A study that provides evidence for these accounts of task interspersal effects would be an important contribution. A study similar to the current experiment that limits extrinsic reinforcement to one reinforcer delivered at the end of an extended task would be more suited to testing the discrete task completion hypothesis.

In addition, further research should examine the effectiveness of interspersal procedures in teaching a larger variety of skills. Although many skill areas have been reported in the literature, areas of particular clinical importance such as the teaching of social and play skills to children with autism can be targeted.

The current study is further support for the use of interspersing low difficulty tasks to facilitate the learning of higher difficulty tasks for children with autism. The possibility that the effects observed during the interspersal condition were caused by the higher density of reinforcement associated with the interspersal condition in the present experiment could not be ruled out.

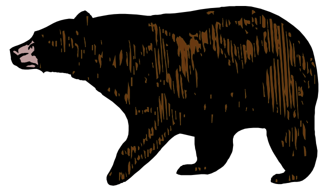
Appendix A

Sample Target Stimuli used for Matching-to-Sample Items

Bear exemplars

Bear 1

Bear 2



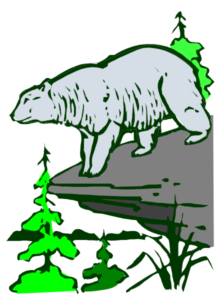
Bear 3

Bear 4

Bear 5

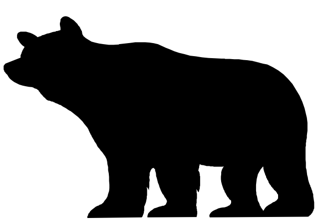
Bear 6

Bear 7

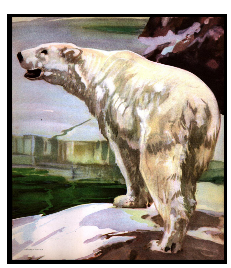


Bear 8

Bear 9



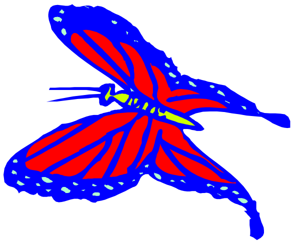
Bear 10



Bear 11

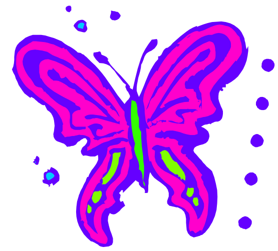
Bear 12

Butterfly exemplars

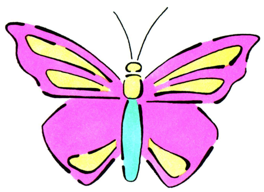


Butterfly 1

Butterfly 2

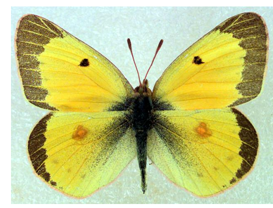


Butterfly 3



Butterfly 4

Butterfly 5



Butterfly 6

Butterfly 7

Butterfly 8

Butterfly 9

Butterfly 10

Butterfly 11

Butterfly 12

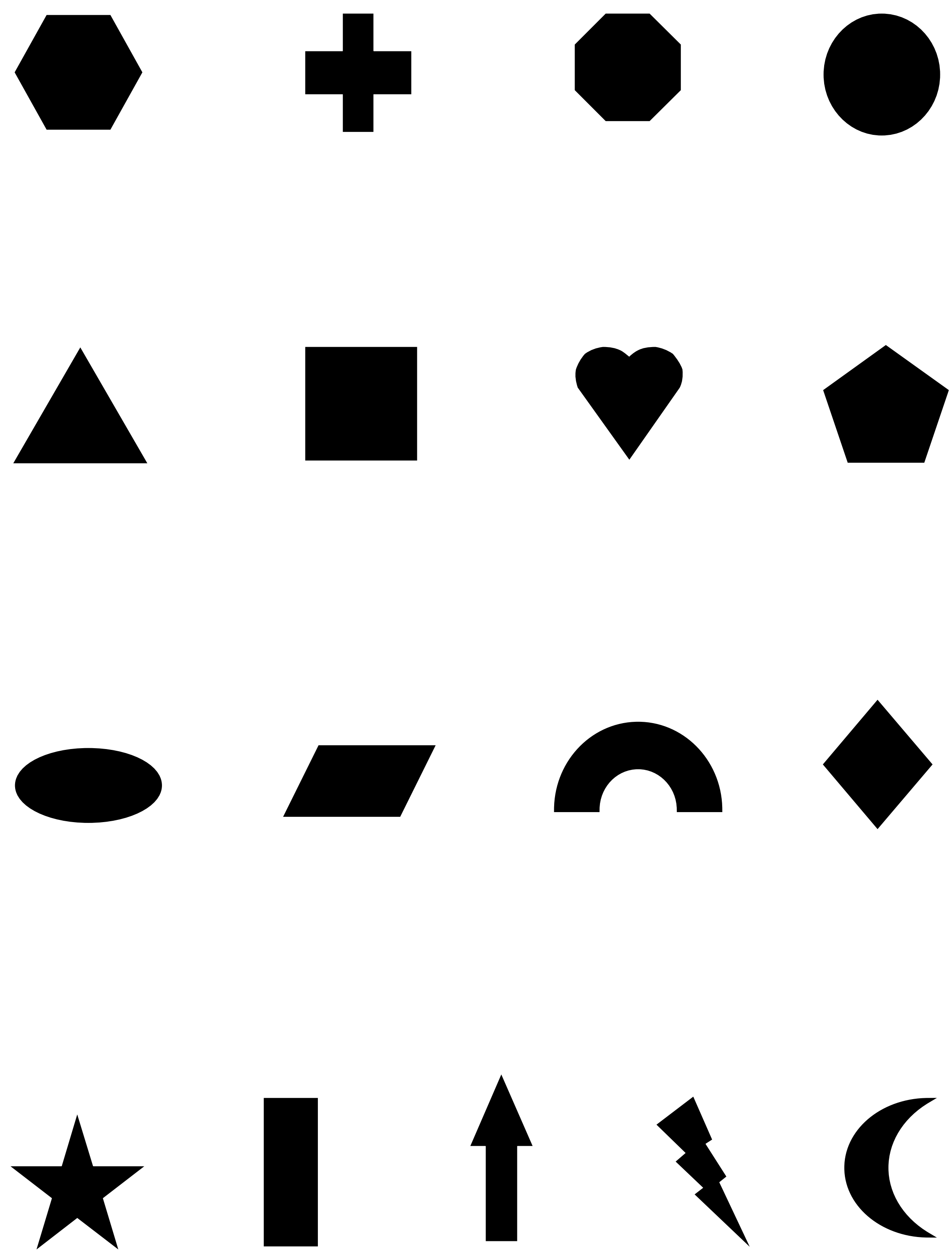
Letter Fonts

A B C D 1	A B C D 2
A B C D 3	A B C D 4
A B C D 5	A B C D 6
A B C D 7	A B C D 8
A B C D 9	A B C D 10
A B C D 11	A B C D 12

Number Fonts

2 3 4 5 A	2 3 4 5 G
3 8 4 0 B	7 3 4 5 H
4 3 9 5 C	1 3 4 5 I
5 3 4 9 D	5 3 4 0 J
6 3 4 5 E	6 3 4 5 K
8 3 4 5 F	7 3 4 5 L

Shapes



Appendix B
The Effects of Heterogeneity in Task Sequencing: A Review

The Effects of Heterogeneity in Task Sequencing: A Review

CHRISTIAN A. BENAVIDES

Queens College and the Graduate Center of the City University of New York

Abstract

A large body of applied research has been conducted on procedures that introduce heterogeneity to a sequence of tasks in order to promote desirable effects. Heterogeneity has been introduced to sequences of tasks by interspersing tasks that differ on some stimulus dimension. Research has been reported on interspersed tasks with varying probabilities of compliance, difficulty level, preference level, and mastery level. Effects of these interspersal procedures have included increases in compliance to requests, increased rates of acquisition of responses, altering preferences for tasks, and decreasing latency to respond to task presentation. This paper reviews the body of applied literature on heterogeneous task sequences, categorized into four areas representing dimensions of task heterogeneity. The major findings of each area are summarized. Some common interpretations of the findings including behavioral momentum, stimulus generalization and the discrete-task completion hypothesis are discussed.

THE EFFECTS OF HETEROGENEITY IN TASK SEQUENCING: A REVIEW

Procedures that present various tasks in an interspersed order have been reported to produce a variety of desirable effects according to a growing body of literature. This body of literature has described many distinct areas specific to particular behavioral goals and intervention procedures.

One area has focused on increasing compliance with tasks with which participants are typically unlikely to comply (low-probability requests). Increases in compliance with such requests have been obtained when they are immediately preceded by requests with which the same participants are likely to comply with (high-probability requests; Killu, 1999). Related research has demonstrated that interspersing tasks that have already been acquired (maintenance tasks) can lead to improved responding during acquisition of new tasks (Neef, Iwata, & Page, 1980; Browder & Shear, 1996; Koegel & Koegel 1986). A third area of research has shown that interspersal of less difficult tasks within an assignment consisting of more difficult tasks affects the acceptability of the assignment (as measured by questionnaires) when compared to assignments consisting of difficult tasks alone (Billington & Skinner, 2002; Wildmon, Skinner, & McDade, 1998).

What these and other areas of research have in common is the purposeful application of heterogeneity among the sequence of tasks that are presented. Each of the studies mentioned above feature an antecedent intervention procedure that presents tasks, requests, or instructions in a sequence that is mixed across some stimulus dimension. These dimensions have included probability of compliance (Ardoin, Martens, & Wolfe, 1999; Mace et al. 1988; McComas, Wacker, & Cooper, 1997), task difficulty (Cates et al., 1999; Wildmon et al., 1998), preferred or non-preferred activities (Belfiore et al., 1997), and acquisition or maintenance tasks (Charlop, Kurtz, & Milstein, 1992; Cooke & Reichard, 1996; Neef et al., 1980). In addition to the dependent measures mentioned in the previous paragraph, these studies have also targeted decreases in problem behavior (Horner, Day, Sprague, O'Brien, & Heatherfield, 1991), decreases in latency to respond (Wehby & Hollahan, 2000), altering preferences for tasks (Cates et al., 1999), and improving acquisition of new responses (Neef et al., 1980).

Despite the inherent similarities in each of these areas of study, the procedures mentioned above have been given a number of different labels by various researchers. Labels such as 'high-probability request sequence,' 'interspersal procedure,' 'errorless compliance training,' and others pertain to variations in target behavior and intervention procedures. Whereas such distinctions can be functional, these different interventions all produce their effects by introducing heterogeneity into the sets of tasks presented to the subjects.

What follows is a review of the procedures and results from this body of applied literature, and a discussion of the behavior analytic principles that may govern these effects. The present review of previous applied investigations is categorized by the stimulus dimension along which heterogeneity has been introduced. It should be noted that while this variable can serve as a valid basis for categorization, it has been inconsistently described in the literature. In some studies the only empirically established dimension on which demands are varied is topography (for example Dunlap & Koegel, 1980).

In addition to consideration of the stimulus dimension of heterogeneity, the review will also describe the populations studied, the target behavior, independent variables of interest, and results. Following the review, different interpretations of the effectiveness of interspersal procedures made by researchers will be described. These theories will be discussed in terms of how they help our understanding of the phenomena.

Review of the Applied Research

Probability of Compliance

Varying instructions associated with different probabilities of compliance is a commonly used interspersal procedure (Killu, 1999; Davis, 1993). Typically, instructions are varied between those with a high probability of compliance (high-p's) and a low probability (low-p's). This procedure is often referred to as the high-probability request sequence (Davis et al., 1992, Ducharme & Worling, 1994; Killu, Sainato, Davis, Davis, Ospelt, & Neely Paul, 1998) or the high-probability command sequence (Mace et al., 1988).

High- and low-p's for an individual participant are often established empirically by presenting him or her with a number of requests and categorizing those that meet or exceed a particular compliance percentage as high-p's and those that fail to meet a cutoff percentage as low-p's (Mace et al., 1988; Zarcone et al., 1994). In other studies, requests are categorized as low- or high-p's based on reports from family members or caregivers who know the participant well (Horner et al., 1991, Singer et al., 1987).

In many studies that feature the high-probability request sequence the dependent measure of interest is the percentage of low-p's with which the participant complies (Ardoin et al., 1999, Davis et al., 1992, Davis & Reichle, 1996; Houlihan, Jacobson, &

Brandon, 1994; Rortvedt & Milltenberger, 1994). The procedure in such studies is used as a treatment for noncompliance.

In one such study, Davis et al. (1992) used the high-p request sequence to increase compliance by two boys with behavior disorders and severe disabilities. Low- and high-p's were found by taking a group of tasks directly related to each boy's IEP (individual education plan) and presenting them each 10 times. Requests with which children failed to respond to 5 or more times were considered low-p. Requests with which the children responded during at least 80% of trials were considered high-p's. During baseline, low-p's were presented alone on a variable schedule averaging 1per min. During treatment, the low-p's were presented on the same schedule as in baseline but each presentation was preceded by 3 to 5 high-p requests. During both baseline and treatment the requests were delivered by a group of four trainers. A multiple baseline experimental design across trainers was used to evaluate the effects of the intervention procedure. The results indicated that the participants dramatically increased responding to low-p's after the first trainer began to use the high-p procedure. One participant began to generalize responding to low-p's with the fourth trainer in baseline after the third trainer entered intervention. The other participant generalized once the second trainer entered treatment.

The high-probability request sequence has also been adapted into a comprehensive procedure designed to systematically improve compliance in children who commonly demonstrate oppositional behavior. Known as errorless compliance training (Ducharme & Popynick, 1993), this approach uses a multi-phase intervention that begins with only high-p's and systematically fades in low-p's over four levels of treatment (Ducharme, Atkinson, & Poulton, 2000; Ducharme, Davidson, & Rushford, 2002; Ducharme, Lucas, & Pontes, 1994; Ducharme, Pontes, Guger, Crozier, Lucas, & Popynick, 1994; Ducharme, Popynck, Pontes, & Steele, 1996; Ducharme, Spencer, Davidson, & Rushford, 2002). Errorless compliance training starts with an assessment of a broad range of requests. This assessment is accomplished first through parent ratings questionnaires that consist of 120 requests. This parental rating is followed by an empirical assessment of selected requests. During the empirical assessment phase, participant's responses to each request are observed and measured. The requests are then arranged in a hierarchy according to their probability of compliance. Compliance levels are divided into 4 'levels' depending on their probability of compliance (Level 1: 76-100%, Level 2: Level 3 51-75%, 3: Level 4 26-50%, 4: 0-25%). The exact compliance percentages associated with each level may be skewed depending on the severity of the participant's compliance problems (Ducharme, 1996). At first only items with highly compliant requests are presented (Level 1 requests). Requests from levels 2, 3, and 4 are introduced in subsequent phases once a high compliance criterion has been met. Each phase associated with a particular level is preceded by a transition phase in which requests from the two adjoining levels are interspersed. Over 7 phases of treatment, requests associated with lower levels of compliance are introduced at a rate that maintains child compliance. Results have demonstrated that by the end of treatment, requests that yielded low levels of compliance prior to treatment have been faded in and are producing high levels of compliance (see Ducharme, 1996 for a review).

Although increasing compliance has been a common focus of high-probability request research, other dependent variables such as latency to respond (Ardoin et al., 1999), duration of task completion (Mace et al., 1988), occurrence of problem behavior

(Zarcone, Iwata, Mazaleski, & Smith, 1994), and occurrence of desired correlated behavior (Wehby & Hollahan, 2000) have been measured and discussed as well.

One of the seminal studies in the field of high-p request sequence research (Mace et al., 1988) is a prime example of this diversity. Mace and his colleagues reported on 5 studies that examined a number of parameters and measures of the high-p procedure. In all the studies, the subjects were adults with developmental disabilities. The first study demonstrated the effectiveness of a high-p procedure to increase compliance with requests to complete simple tasks (" . . . put your lunch box away"). Each task (low-p) was presented in both a "Do" and a "Don't" format in alternating phases of a reversal design. The results indicated that compliance to the low-p's increased dramatically when they were preceded by a high-p sequence. The second study examined the possibility that attention in the form of social statements ("Were going bowling this afternoon") delivered in a manner similar to the high-p sequence (3-4 statements delivered at short intervals prior to the low-p) may produce similar increases in compliance. The high-p sequence and the "attention control" condition were evaluated using a reversal design. Results showed that compliance to low-p's only increased when the high-p sequence was presented before the low-p. In the third study, the experimenters varied the interprompt time (IPT), the interval between the final high-p and delivery of the low-p request. The Results showed increases in compliance only when a short (5 sec) IPT was used; the long IPT interval (20 sec) was ineffective. In the fourth study a new dependent variable was measured: latency to begin a task following an instruction. The low-p's were requests to do household chores (emptying the trash, sweeping the bathroom floor). Using a multi-element design, the high-p request sequence effectively reduced the latency to begin tasks for both of the subjects involved. Experiment 5 measured another new variable: duration to complete a task. In this case the task was taking a shower. Showering was broken down to three segments including preparing for shower, showering, and getting dressed afterwards. The high-p request sequence was applied to all 3 segments and evaluated using a multi-element design. High-p requests included simple social responses such as "give me five," or "shake my hand." Results indicated a decrease in duration to complete all three segments when the high-p procedure was used.

In addition to increasing compliance, and decreasing latency to begin and duration to complete tasks, research on high-p procedures has also measured effects on correlative behavior both desirable (Davis et al., 1994; Sanchez-Fort, Brady, & Davis, 1995; Wehby & Hollahan, 2000) and undesirable (Zarcone et al., 1994, Mace & Belfiore, 1990).

For example, Wehby & Hollahan (2000) measured the effects of the high-p procedure on two variables: latency to begin math problems, and time spent engaged with task. The participant was a 13-yr old girl identified as learning disabled in math and written language. The experimenters presented a high probability request sequence prior to making requests to "begin independent seat work." In addition to measuring latency to begin task, the authors calculated the percentage of time each session that the subject was engaged with the task materials. A reversal design was used to compare 3 experimental procedures: low-p requests only (A), high-p sequence (B), and neutral social comments (C). The "neutral social comments" condition was conceptually similar to the "social statements" condition used in experiments 2 of Mace et al. (1988). Hollahan & Wehby's results showed a decrease in latency and an increase in percentage of time engaged during the high-p condition when compared with both the low-p alone and neutral comments conditions.

Zarcone et al. (1994) similarly measured the effects of the high-p procedure on undesirable behavior. Two adult men with developmental disabilities participated in this study. Both engaged in different forms of self-injurious behavior (SIB). A functional analysis was conducted on the SIB exhibited by both participants. According to the analysis, SIB was most likely maintained by escape from task demands for both men. Following the functional analysis, the authors measured 3 dependent variables in the next part of the experiment. The three variables were: percentage of trials (low-p presentations) with SIB, rate per minute of SIB, and percentage of compliance with low-p's. These variables were measured in 3 conditions alternated in a reversal design. The first condition, baseline, consisted of low-p's presented on a fixed-time schedule. The second, condition, *high-p sequence*, preceded each low-p presentation with a high-p sequence; instances of SIB in this condition led to the termination of the trial (removal of demands). The third condition, *high-P + extinction*, was similar to the 2nd condition except SIB no longer led to termination of the trial. The data demonstrated that the high-p procedure alone was ineffective in increasing compliance to low-p's or in lowering the occurrence of SIB; in some instances the procedure seemed counter-therapeutic. When extinction was added, SIB decreased dramatically for one of the two subjects (although both eventually reached near zero levels). Compliance to low-p's was also improved in one of the two subjects only when extinction was added to the treatment package. This study helped to point out some potential limitations of high-p based procedures, particularly when competing problem behavior is involved.

As mentioned in the descriptions above, high-probability request research has been conducted and demonstrated favorable effects with a wide variety of populations. These include typically developing children (Ardoin et al., 1999), teens using augmentative communication devices (Davis, Reichle, Southard, & Johnston, 1998), children with autism (Houlihan, Jacobson, & Brandon, 1994), teens with mental retardation (Mace, Mauro, Boyajian, & Eckert, 1997), and adults with developmental disabilities (Zarcone et al., 1994).

Task-Mastery Level

Task-mastery level is a second stimulus dimension along which heterogeneity has been introduced to positive effect. Studies in which this variable is manipulated intersperse mastered tasks under maintenance conditions with tasks under acquisition (Cooke, Guzaukas, Pressley, & Kerr, 1993; Neef et al., 1980). Typically, this interspersal method has been used to facilitate the acquisition of new responses (Charlop et al., 1992). Procedures that intersperse acquisition and maintenance tasks have been called 'task interspersal' (Charlop et al., 1992), 'interspersal training' (Cooke et al., 1993), or 'interspersal procedure' (Rowan & Pear, 1985).

These procedures have proven to be effective with a variety of populations and target responses. Populations have included young children with autism (Charlop et al., 1992), fifth graders with learning disabilities (Cooke & Reichard, 1996), stroke victims (Koegel & Koegel, 1986), and adults with mental retardation (Neef et al., 1980). Target responses have included spelling words (Neef et al., 1980), picture identification (Rowan & Pear, 1985), sight-word identification (Browder, & Shear, 1996), and multiplication problems (Cooke et al., 1993).

In an important early study, Neef et al. (1980) compared the effects of interspersed task presentation (interspersing maintenance with acquisition tasks) with a high-density reinforcement procedure on the acquisition of spelling words. Three young adults with mental retardation participated. The dependent variable was the cumulative number of spelling words mastered. Mastery criteria for a word occurred when the word was written with correct spelling 5 consecutive times. Prior to the experiment, words were pre-tested (presented twice each) to determine whether they were 'known' (correct written spelling twice), 'training items' (incorrect spelling twice) or to be discarded (correct once). The effects of three conditions on acquisition of spelling items were compared: baseline, high-density reinforcement, interspersal training, and retention testing. During baseline 10 words classified as training items were presented once each per session. In this and subsequent conditions, items were replaced with new training items after they had met mastery criteria. In the high-density reinforcement condition, social reinforcement was delivered once during each of ten trials for paying attention, writing neatly and "trying hard." During interspersal training, ten mastered items were alternated with training items. In the first phase of the experiment, only baseline sessions were conducted. The first phase was followed by a second phase in which high-density reinforcement and interspersal training sessions were alternated within a multi-element design. Results indicated that more responses were mastered during the interspersal condition when compared with baseline and high-density reinforcement. In a post experiment retention test, higher retention rates for words learned during interspersal conditions were found. In addition, each of the three participants was allowed to choose their intervention during 4 sessions (either high density-reinforcement or interspersal training). Interspersal training was chosen over high-density reinforcement during 11 of 12 choice opportunities.

In a later study, Charlop et al. (1992) compared 3 reinforcement contingencies used when interspersing acquisition with maintenance tasks. Five children with autism participated in the study. Target acquisition tasks were identified individually for each child. These included following simple directions ("raise arms"), and arranging items ("next to," "left/right," "first/last"). Maintenance tasks were identified as those that each child had performed at 80% or higher accuracy during the 3 months prior to the start of the study. In all conditions, a set of 27 trials (15 acquisition & 12 maintenance) was presented in a random order. The 3 conditions were: baseline, no-reinforcers, and praise only. In baseline, acquisition tasks were reinforced with praise and food both on a continuous reinforcement schedule (CRF), while maintenance tasks were reinforced with praise on a CRF and food on a VR3 schedule. In the no-reinforcers condition, acquisition tasks were on a CRF for praise and food, while maintenance tasks were not reinforced at all. Finally, during the praise only condition, acquisition tasks were on a CRF for praise and food, while maintenance tasks were reinforced with praise only on a CRF. For all participants, baseline conditions were applied first, consisting of 6-18 sessions for each participant. Following baseline, either the no-reinforcers or the praise-only condition was applied. Results indicated that all children failed to learn acquisition tasks during baseline. Meanwhile, both 'no-reinforcers' and 'praise only' conditions successfully produced acquisition (90% correct responding over 20 trials). The additional presumptive reinforcers delivered during baseline apparently interfered with acquisition.

Charlop et al. (1992) tested an important parameter of acquisition/ maintenance interspersal: the relative amount of reinforcement given for successful task completion. Other research has studied the optimal acquisition/ maintenance ratio (Cooke & Reichard, 1996). Cooke & Reichard (1996) tested maintenance-to-acquisition ratios of 7:3, 1:1, and 3:7. They found learning of multiplication and division problems by 5th grade students to be generally most efficient at a 3:7 acquisition-to-maintenance ratio. Still other research by Rowan & Pear (1985) explored the question of whether acquisition tasks interspersed with other acquisition tasks would have similar effects to acquisition/ maintenance interspersal. In this study, the authors found that acquisition/ maintenance interspersal was more effective in teaching picture naming to three children with mental retardation than was acquisition/ acquisition interspersal.

Task Complexity

A number of interventions have interspersed tasks that varied by their relative complexity (Billington & Skinner, 2002; Cates, Skinner, Watkins, Rhymer, McNeil, & McCurdy, 1999; McCurdy, Skinner, Grantham, Watson, & Hindman, 2001). In these cases the dimension of heterogeneity will be referred to as task complexity. Intervention procedures in these studies have commonly been called the ‘interspersal procedure’ (Cates et al., 1999; Robinson & Skinner, 2002) or ‘interspersal technique’ (Skinner et al., 1999; Wildmon et al., 1999). Tasks are commonly defined as high or low effort by features such as how many digits are involved in a mathematics problem (McCurdy, et al., 2001) or which mathematical operations are involved (such as presence vs. absence of multiplication in word problems, (Wildmon et al., 1999). The dependent variables of interest in these studies have included assignment choice, problem completion rates

(Billington & Skinner, 2002), percentage of problems completed correctly (Cates et al., 1999), on-task behavior (McCurdy, et al., 2001), and perceptions of tasks as measured by post-task questionnaires (Skinner et al., 1999; Wildmon & Skinner, 1998). Many studies feature several of these variables (Skinner et al., 1999; Wildmon & Skinner, 1998; Robinson & Skinner, 2002). It is also important to note that unlike other areas of interspersal research, many of these studies feature group designs.

This area of research is exemplified by Skinner, Robinson, Johns, Logan, & Belfiore (1996). In the first experiment of this study, 51 college students were presented with two sets of math problems. One set (the control assignment) contained sixteen 3 digit x 2 digit multiplication problems. The other set (experimental assignment) had 16 similar 3x2 problems interspersed with 6 1x1 problems. The task sequence in the experimental assignment was arranged so that a 1x1 problem was followed by three 3x2 problems. Each student was given 305 sec to finish the assignment. Using a within group design, the authors found several differences between the two assignments. Significantly more problems were completed in the experimental assignment (although there was no significant difference in the number of 3x2 problems completed). In addition, students' perception of the experimental assignment was favorable. Significantly more students reported the experimental assignment was less difficult, would require less time to complete, less effort to complete, and would choose the experimental over the control assignment if they had to do another one again. Experiment Two replicated these findings with 31 new students while controlling for possible effects of novelty of the brief problems by adding a third assignment. The third assignment was similar to the experimental assignment from the previous experiment except that the 1x1 multiplication problems were replaced by 3-digit/ 2-digit division problems. Experiment Three was an attempt to determine if the differences found between assignments in experiment 2 were caused by any qualities inherent to the specific 3x2 multiplication problems used. Ruling out such novelty effects was accomplished by presenting the three assignments without the interspersed 1x1 or 3/2 problems to a new group of 27 students. The three assignments were not significantly different in number of problems completed or in accuracy.

Other studies examined variations of the interspersal procedure including using math word problems (Wildmon et al., 1999), requiring assignments to be completed (Billington & Skinner, 2002), and tasks with escalating difficulty (Robinson & Skinner, 2002). In addition, Skinner et al. (1999) found that the relative problem-completion rate affected perception and choice.

The interspersal procedure was also adapted into an applied study by McCurdy, et al., (2001). This study is notable for its distinct dependent measure (on-task behavior) and its use of a single subject design. A 9-year old girl attending a regular education classroom participated. Her school psychologist reported the participant as frequently engaging in off-task behavior such as talking to peers, looking around the room and playing with pencils and paper. Her on-task behavior was measured during two types of math assignments: a control assignment and an experimental assignment with easier problems interspersed with control problems. Both assignment types included problems from the current classroom curriculum. A multi-element design was used to assess the effects of the interspersal procedure. Results indicated consistently more on-task behavior during interspersed assignments.

Task Preference

A final method for creating heterogeneity in a task sequence that has received relatively little attention is interspersing highly preferred tasks with target non-preferred tasks (Belfiore et al., 1997; Belfiore, Lee, Scheeler & Klein, 2002). Thus far the only target behavior of interest looked at in these studies has been latency to begin task (Belfiore et al., 1997; Belfiore et al., 2002).

In a study by Belfiore et al. (1997) two teenage students enrolled in a community-based alternative education school were measured on their latency to begin academic tasks. The tasks were 1x1 digit, or 3x3 digit multiplication problems. A forced choice preference assessment was conducted to identify the students' preference for either single-digit or multi-digit math worksheets. As expected, the assessment revealed 1x1 digit problems to be preferred. During baseline 5 consecutive non-preferred problem cards were presented. Intervention was similar to traditional high-p interventions, in that 3 preferred (1x1 problems) were inserted between each non-preferred card (3x3). An ABAB reversal design was used to assess the effects of intervention. The results demonstrated that the interspersal intervention successfully reduced the mean latency to begin each problem.

Discussion: How does Task Heterogeneity Improve Performance?

Whereas the review in the previous section was organized by the manner in which heterogeneity was introduced, it should be noted that the dimensions can, and probably do co-vary in some cases. A particular intervention designed to intersperse tasks using one of the dimensions mentioned may in fact be interspersing other dimensions as well. For example, maintenance tasks might also be highly preferred, whereas acquisition tasks might be non-preferred. Difficult tasks might also be associated with low compliance, whereas easy tasks might be associated with high compliance in some cases. Belfiore et al. (2002) used probability of compliance and preference for task interchangeably when describing their intervention. This raises the questions, how different are these areas of research? How are they related? While not offering a definitive answer, this discussion section will consider these questions in an effort to understand how interspersal interventions work. These questions will be addressed in the context of some commonly cited theories that explain the effects of task heterogeneity.

Behavioral Momentum

A commonly cited theoretical basis for many of the interventions reviewed earlier is behavioral momentum (Nevin, 1996; Nevin, Mandell, & Atak 1983; Nevin 2002). The term ‘Behavioral momentum’ refers to a metaphor that relates the momentum of an organism’s responding to the momentum of moving objects according to Newton’s second law of physical motion. Response rate and resistance to change of behavior are equated with velocity and mass of objects. Many of the intervention procedures from the preceding review are reported as having been derived from behavioral momentum or have been labeled as behavioral momentum procedures (Mace et al. 1988; Mace & Belfiore 1990; Belfiore et al. 2002; Romano & Roll 2000).

Behavioral momentum has been developed from a body of basic research on the persistence of responding following some disruption (see Nevin 1992; Nevin & Grace 2000 for a review). A basic paradigm for testing such resistance to change is described by Nevin (1996). Subjects learn to respond under a multiple schedule of reinforcement with different schedules of S^R associated with each of 2 S^Ds (‘components’). These components may be separated by a time-out period to minimize interaction. Data are taken on both components until a stable baseline is reached. Resistance to change can be evaluated by introducing a disrupter (such as prefeeding) during the timeout period. Responding after the disrupter is then compared to responding under the immediately preceding baseline session (separately for each stimulus). Comparisons are made within subjects and sessions. Smaller changes as compared with baseline are indicative of a greater resistance to change.

Nevin (1996) describes four conclusions from this body of work that comprise behavioral momentum theory. The first conclusion is that resistance to change of a discriminated operant depends on the rate of reinforcement obtained by the target response class. The second is that resistance to change maintained by a given reinforcement rate increases if additional reinforcers are allocated to an alternative concurrent response or are provided noncontingently. Third, resistance to change maintained by a given reinforcement rate is inversely related to the rate of reinforcement obtained by other, successive discriminated operants. Finally, the fourth conclusion is that resistance to change is independent of the steady-state baseline rate of the target response.

While acknowledging the relation between basic momentum research and high-p applied studies, Nevin (1996) goes on to caution applied researchers from making direct correlations between variables in their studies to those identified in the basic literature. Some key differences between the typical high-p intervention paradigm and the procedures used in basic resistance to change research are cited as reasons. The first difference is the lack of a well-defined alternated stimulus situation in high-p studies that corresponds to the alternated components of a multiple schedule found in basic studies. The second difference, echoed by Plaud & Gaither (1996), is that basic research establishes a steady state free operant responding rate against which to compare the effects of a disrupter. In high-p studies, the dependent variable of interest is not a free operant response and hence does not yield a steady-state baseline for comparison with the effect of a disrupter. Finally, the nature of the disrupter in high-p research is different from the pre-feeding or extinction disrupters typically used in basic studies. In high-p studies the disrupter is a more demanding request (low-p). This disrupter may not be independent from the contingencies that maintain responding as another (e.g. pre-feeding) would be. Nevin goes on to address each of these concerns and argues that these differences do not necessarily preclude a behavioral momentum interpretation of high-p research. Such interpretations, he adds, do not come without a fair amount of speculation.

In a behavioral momentum account of high-p research, compliance to requests must be treated as a discriminated operant, with a change in demand level being treated as a disrupter. The momentum metaphor does not fit other forms of task interspersal as neatly. Interventions that introduce heterogeneity through acquisition/ maintenance tasks, or task complexity may be altering the discriminated operant when tasks are interspersed between maintenance and acquisition or low and high complexity tasks. While the change between types of task sequences in these interventions may coincide at times with changes between low and high-probability tasks (for example, a high-difficulty task may have a relatively low probability of compliance) the responses of interest are not a percentage of compliance in all cases. For example, in the case of task heterogeneity studies featuring interspersed task-mastery level, the dependent variable of interest is typically correct responding to novel tasks, while the interspersed tasks have been previously mastered (maintenance tasks). The discriminated operant in this case, may not fall under the common response class of "compliance to requests" for both acquisition and maintenance tasks. Therefore, the momentum account, while very successful in stimulating applied research, for our purposes may be limited to interspersal procedures featuring probability of compliance as the heterogeneous dimension.

Discrete-Task Completion Hypothesis

As described by Skinner (2002), the "discrete-task completion hypothesis" proposes that completion of a discrete task within a longer assignment functions as a conditioned reinforcer. This hypothesis, to be described below, has been used to explain the effects of interventions based on interspersing brief, easy tasks with longer, more difficult tasks (Skinner et al. 1996; Skinner et al. 1999). As mentioned earlier, this area of research has commonly focused on the effects of such interspersal procedures on the preferences of participants for particular sets of assignments. A regular finding of this research is that preferences for assignments are strengthened by using the interspersal procedure (Cates et al. 1999; Skinner et al. 1999). The discrete-task completion hypothesis accounts for this effect by claiming that the increased number of tasks completed during the interspersal conditions also means that more reinforcement is available during interspersal conditions. Thus, according to the matching law (Herrnstein 1961; Herrnstein 1970), participants would be more likely to choose the assignment associated with relatively more reinforcement.

A review of eight studies that featured the interspersal procedure was conducted by Skinner (2002). The analysis revealed a linear relation between relative problem completion rates and choice of assignment. This finding supports the discrete-task completion hypothesis. Skinner goes on to describe three assumptions that the hypothesis makes. The first is that people generally have a learning history of reinforcement for assignment completion. The second is that completed assignments as well as completed tasks within those assignments constitute stimulus events. Finally, the third assumption is that discrete task completion becomes a reinforcing event through association with reinforcement for assignment completion.

The discrete-task completion hypothesis is specific to explaining response allocation in interspersal procedure studies. It cannot account for results from other areas of task heterogeneity research in which choice making is not the dependent variable of interest. Nevertheless, the same principles of reinforcement associated with task completion and classical conditioning of discrete task completion are relevant when considering a third hypothetical account for the effectiveness of task heterogeneity-based procedures, stimulus generalization.

Generalization of Stimulus Control

Horner et al. (1991) posits a hypothesis for the effectiveness of an interspersed-requests procedure based on stimulus generalization of a response class that consists of "instruction following." Before considering this hypothesis, we'll review the experiment that generated it.

In the first of two experiments reported by Horner et al., three youths (ages between 12-14 years old) with mental retardation participated. Each participant had a history of self-injury or aggression. Two dependent measures were obtained for each participant: problem behavior (either self-injury or aggression), and attempts to complete a presented “hard” task. Attempts were defined as initiating the first response in the requested response chain. Hard tasks included putting on a shirt, sorting silverware, and following two-step instructions. In addition to these tasks, “easy” tasks were also identified for each participant. Hard and easy tasks were empirically assessed using the following criteria: hard tasks were completed correctly on fewer than 33% of trials while easy tasks were completed correctly on 70% or more of trials. In separate experimental conditions hard tasks were presented either alone or interspersed among easy tasks. These conditions were presented using a reversal design. In all conditions, the instructor physically interrupted occurrence of self-injury or aggression and the participant was subsequently re-directed to the task. Results indicated that during the interspersal condition, all three participants markedly decreased problem behavior and two of three participants increased attempts to perform hard tasks. The 3rd participant already had attempted to perform tasks during 100% of trials during the non-interspersed condition. In the second experiment, similar experimental conditions were presented, but data were collected over a more protracted period of time. Results were similar but less dramatic than in the first experiment.

In the discussion of these experiments, Horner et al. hypothesize that the interpersal procedures were effective because compliance with the interspersed requests was followed immediately by reinforcement and, therefore, compliance as a response class was strengthened. Thus, reinforcement for responding to varied instructions increased the likelihood of responding to the entire stimulus class of instructions. This spread of effects theoretically occurred according to the principle of stimulus generalization (Guttman & Kalish 1956). Horner et al. asserted that stimulus generalization predicts that reinforcement for responding to less-difficult interspersed requests will increase attempts to perform the more difficult tasks.

Other researchers do not refer explicitly to stimulus generalization as the underlying principle explaining their research outcomes, but the explanation comes readily to mind when reviewing their studies. One such study was reported by Davis & Reichle (1996). In this study, the dependent variable of interest was the percentage of correct responses made to requests. The composition of the high-p sequence was manipulated so that it consisted of invariant high-p requests in one condition and variant requests in another. In the invariant condition, the same high-p would be repeated three times before each low-p, whereas in the variant condition, three different high-p's would be presented. These two conditions were compared in a reversal design, against a baseline in which low-p requests were presented alone using a reversal design. Results demonstrated that both the invariant and variant high-p interventions improved responding to low-p's, but the improvement was greater and more persistent in the variant high-p condition. It should also be noted that responding to high-p's decreased in the invariant condition. These results may be predicted by the generalization of stimulus control account of task interspersal if we consider the superiority of training multiple exemplars in producing generalization effects (Stokes & Baer 1977). The invariant-requests condition may not have presented sufficient exemplars of instructions when compared to the variant condition. As a result, the response of following instructions failed to generalize as well in the invariant condition. Of course, behavioral momentum can account for the same result in a different way. When the responding to the high-p requests decreased in the invariant condition, reinforcement earned for following instructions declined; this in turn weakened the momentum of responding.

The stimulus-generalization account is attractive for its inclusiveness and parsimony. This hypothesis would explain interspersal effects regardless of the stimulus dimension along which heterogeneity was established, because all that is necessary for generalization to occur is the reinforcement of responding in a similar stimulus situation. The stimulus-generalization hypothesis is also consistent with the matching law, and thus with the discrete-task completion hypothesis. The discrete-task completion hypothesis is an assertion that reinforcement is often inherent for completion of discrete tasks within a larger assignment. The effects of such reinforcement in the presence of one stimulus (low-complexity or high-probability task) might spread its effects to another, similar stimulus (high-complexity or low-probability tasks).

Although the discrete-task completion hypothesis originates from a particular area of interspersal research that emphasizes task complexity as the dimension of heterogeneity (Skinner 2002), stimulus generalization may be a reason problem-completion rates increased when low-complexity tasks were interspersed in those studies. The increase in on-task behavior found in the 'task complexity' study by McCurdy et al. (2001) reviewed earlier may reflect the responding (on-task) to low complexity tasks generalizing to more difficult tasks. Nevertheless, this hypothesis has received relatively little attention in task-heterogeneity literature.

Summary & Conclusions

Heterogeneous task sequencing has become an established feature of applied interventions. Its prevalence is well supported by numerous studies (see reviews by Davis & Brady 1993, Killu 1999, and Skinner 2002). Several areas of research have emerged that measure heterogeneity in different ways, including manipulations of compliance probability, task-mastery level, task complexity, and task preference. It is argued that the common variable in these procedures is task heterogeneity, and that therefore, they can all be subject to the same theoretical analysis and scrutiny. This review has argued for the desirability of the term ‘task heterogeneity’ as a general descriptor for this area of research. Whereas the broad and reliable effects of these procedures have been amply reported, the underlying learning processes that govern these effects have not been clearly established. Among the questions to be answered are to what extent is rate of reinforcement confounded with task interpersal, to what extent does the physical similarity between interpersal and target tasks affect responding, and to what extent is task completion a reinforcing event. We have reviewed some theoretical accounts of the effectiveness of task heterogeneity procedures, including behavioral momentum, the discrete-task completion hypothesis, and stimulus generalization as explanations.

References

Ardoin, S. P., Martens, B. K., & Wolfe, L. A. (1999). Using high-probability instruction sequences with fading to increase student compliance during transitions. *Journal of Applied Behavior Analysis,* 32, 339-351.

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.

Belfiore, P. J., Lee, D. L., Scheeler, M. C., & Klein, D. (2002). Implications of behavioral momentum and academic achievement for students with behavior disorders: Theory, application, and practice. *Psychology in the Schools,* 39, 171-179.

Billington, E. J., & Skinner, C. H. (2002). Getting students to choose to do more work: Evidence of the effectiveness of the interspersal procedure. *Journal of Behavioral Education,* 11, 105-116.

Browder, D. M., & Shear, S. M. (1996). Interspersal of known items in a treatment package to teach sight words to students with behavior disorders. Journal of Special Education, 29, 400-413.

Cates, G. L., Skinner, C. H., Watkins, C. E., Rhymer, K. N., McNeil, S. L., & McCurdy, M. (1999). Effects of interspersing additional brief math problems on student performance and perception of math assignments: Getting students to prefer to do more work. *Journal of Behavioral Education,* 9, 177-192.

Charlop, M. H., Kurtz, P. H., & 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, 795-808.

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 & Treatment of Children,* 16, 213-234.

Cooke, N. L., & Reichard, S. M. (1996). The effects of different interspersal drill ratios on acquisition and generalization of multiplication and division facts. *Education & Treatment of Children,* 16, 124-142.

Davis, C. A., & Brady, M. P. (1993). Expanding the utility of behavioral momentum with young children: Where we've been. Where we need to go. *Journal of Early Intervention,* 17, 211-223.

Davis, C. A., Brady, M. P., Williams, R. E., & Hamilton, R. (1992). Effects of high-probability requests in young children with behavior disorders. *Journal of Applied Behavior Analysis,* 25, 905-916.

Davis, C. A., Brady, M. P., Hamilton, R., McEvoy, M.A., & Williams, R. E. (1994). Effects of high-probability requests on the social interactions of young children with autism. *Journal of Applied Behavior Analysis*, 27, 619-637.

Davis, C. A., & Reichle, J. (1996). Variant and invariant high-probability requests: Increasing appropriate behaviors in children with emotional-behavioral disorders. *Journal of Applied Behavior Analysis,* 29, 471-482.

Davis, C. A., Reichle, J., Southard, K., & Johnston, S. (1998). Teaching children with severe disabilities to utilize nonobligatory conversational opportunities: An application of high-probability requests. *Journal of the Association for Persons with Severe Handicaps,* 23, 57-68.

Ducharme, J. M. (1996). Errorless compliance training: Optimizing clinical efficiency. *Behavior Modification,* 20, 259-280.

Ducharme, J. M., Atkinson, L., & Poulton, L. (2000). Success-based, noncoercive treatment of oppositional behavior in children from violent homes. *Journal of the American Academy of Child and Adolescent Psychiatry,* 39, 995-1004.

Ducharme, J. M., Davidson, A., & Rushford, N. (2002). Treatment of oppositional behavior in children of parents with brain injury and chronic pain. *Journal of Emotional and Behavioral Disorders,* 10, 241-248.

Ducharme, J. M., Lucas, H., & Pontes, M. (1994). Errorless embedding in the reduction of severe maladaptive behavior during interactive and learning tasks. *Behavior Therapy,* 24, 489-501.

Ducharme, J. M., Pontes, M., Guger, S., Crozier, K., Lucas, H., & Popynick, M. (1994). Errorless compliance to parental requests II: Increasing clinical practicality through abbreviation of treatment parameters. *Behavior Therapy,* 25, 469-487.

Ducharme, J. M., & Popynick, M. (1993). Errorless compliance to parental requests: Treatment effects and generalization. *Behavior Therapy,* 24, 209-226.

Ducharme, J. M., Popynick, M., Pontes, M., & Steele, S. (1996). Errorless compliance to parental requests III: Group parent training with parent observational data and long-term follow-up. *Behavior Therapy,* 27, 353-372.

Ducharme, J. M., Spencer, T., Davidson, A., & Rushford, N. (2002). Errorless compliance training: Building a cooperative relationship between parents with brain injury and their oppositional children. *American Journal of Orthopsychiatry,* 72, 585-595.

Ducharme, J. M., & Worling, D. E. (1994). Behavioral momentum and stimulus fading in the acquisition and maintenance of child compliance in the home. *Journal of Applied Behavior Analysis,* 27, 639-647.

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

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology,* 51, 79-88.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4, 267-272.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, 7, 179-182.

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

Houlihan, D., Jacobson, L., & Brandon, P. K (1994). Replication of a high-probability request sequence with varied interprompt times in a preschool setting. *Journal of Applied Behavior Analysis,* 27, 737-738.

Kennedy, C. H., Itkonen, T., & Lindquist, K. (1995). Comparing interspersed requests and social comments as antecedents for increasing student compliance. *Journal of Applied Behavior Analysis,* 28, 97-98.

Killu, K. (1999). High-probability request research: Moving beyond compliance. *Education and Treatment of Children,* 22, 470-494.

Killu, K., Sainato, D. M., Davis, C. A., Ospelt, H., & Neely Paul, J. (1998). Effects of High-Probability Request Sequences on Preschoolers' Compliance and Disruptive Behavior. *Journal of Behavioral Education,* 8, 347-368.

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

Mace, F. C., & Belfiore, P. (1990). Behavioral Momentum in the treatment of escape-motivated stereotypy. *Journal of Applied Behavior Analysis,* 23, 507-514. GET

Mace, F. C., Hock, M. L., Lalli, J. S., West, B. J., Belfiore, P., Pinter, E., & Brown, D. K. (1988). Behavioral momentum in the treatment of noncompliance. *Journal of Applied Behavior Analysis,* 21, 123-124.

Mace, F. C., Mauro, B. C., Boyajian, A. E., & Eckert, T. L. (1997). Effects of reinforcer quality on behavioral momentum: Coordinated applied and basic research. *Journal of Applied Behavior Analysis,* 30, 1-20.

McComas, J. J., Wacker, D. P., & Cooper, L. J. (1997). Increasing Compliance with medical procedures: Application of the high-probability request procedure to a toddler. *Journal of Applied Behavior Analysis,* 31, 287-290.

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, 23-32.

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

Nevin, J. A. (1992). An integrative model for the study of behavioral momentum. *Journal of the Experimental Analysis of Behavior,* 57, 301-316.

Nevin, J. A. (1996). The momentum of compliance. *Journal of Applied Behavior Analysis,* 29, 535-547.

Nevin, J. A. (2002). Measuring behavioral momentum. *Behavioural Processes,* 57, 187-198.

Nevin, J. A., & Grace, R. C. (2000). Behavioral momentum and the law of effect. *Behavioral and Brain Sciences,* 23, 73-130.

Nevin, J. A., Mandell, C., & Atak, J. R. (1983). The analysis of behavior momentum. *Journal of the Experimental Analysis of Behavior,* 39, 49-59.

Plaud, J. J., & Gaither, G. A. (1996). Behavioral momentum: Implications and development from reinforcement theories. *Behavior Modification,* 20, 183-201.

Robinson, S. L., & Skinner, C. H. (2002). Interspersing additional easier items to enhance mathematics performance on subtests requiring different task demands. *School Psychology Quarterly,* 17, 191-205.

Romano, J. P., & Roll, D. (2000). Expanding the utility of behavioral momentum for youth with developmental disabilities. *Behavioral Interventions,* 12, 99-111.

Rortvedt, A. K., & Miltenberger, R. G. (1994). Analysis of a high-probability instructional sequence and time-out in the treatment of child noncompliance. *Journal of Applied Behavior Analysis,* 27, 327-330.

Rowan, V. C., Pear, J. J. (1985). A comparison of the effects of interspersal and concurrent training sequences on acquisition, retention, and generalization of picture names. *Applied Research in Mental Retardation,* 6, 127-145.

Sanchez-Fort, M. R., Brady, M. P., & Davis, C. A. (1995). Using high-probability requests to increase low-probability communication behavior in young children with severe disabilities. *Education and Training in Mental Retardation and Developmental Disabilities,* 30, 151-165.

Singer, G. H., Singer, J., & Horner, R. (1987). Using pre-task requests to increase the probability of compliance for students with severe disabilities. *The Association for Persons with Severe Handicaps,* 12, 287-291.

Skinner, C. H. (2002). An empirical analysis of interspersal research evidence, implications, and applications of the discrete task completion hypothesis. *Journal of School Psychology,* 40, 347-368.

Skinner, C. H., Hall-Johnson, K., Skinner, A. L., Cates, G., Weber, J., & Johns, G. A. (1999). Enhancing perceptions of mathematics assignments by increasing relative problem completion rates through the interspersal technique. *The Journal of Experimental Education,* 68, 43-59.

Skinner, C. H., Robinson, S. L., Johns, G. A., Logan, P. & Belfiore, P. J. (1996). Applying Herrnstein's matching law to influence students' choice to complete difficult academic tasks. *Journal of Experimental Education,* 65, 5-17.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis,* 10, 349-367.

Wehby, J. H., & Hollahan, M. S. (2000). Effects of high-probability requests on the latency to initiate academic tasks. *Journal of Applied Behavior Analysis,* 33, 259-262.

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, 319-325.

Wildmon, M. E., Skinner, C. H., & McDade, A. (1998). Interspersing Additional brief, easy problems to increase assignment preference on mathematics reading problems. *Journal of Behavioral Education,* 8, 337-346.

Zarcone, J. R., Iwata, B. A., Mazaleski, J. L., Smith, R. G. (1994). Momentum and extinction effects on self-injurious escape behavior and noncompliance. *Journal of Applied Behavior Analysis,* 27, 649-658.

Bibliography

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.

Belfiore, P. J., Lee, D. L., Scheeler, M. C., & Klein, D. (2002). Implications of behavioral momentum and academic achievement for students with behavior disorders: Theory, application, and practice. *Psychology in the Schools,* 39, 171-179.

Browder, D. M., & Shear, S. M. (1996). Interspersal of known items in a treatment package to teach sight words to students with behavior disorders. *Journal of Special Education,* 29, 400-413.

Cates, G. L. (2005). A review of the effects of interspersing procedures on the stages of academic skill development. *Journal of Behavioral Education,* 14, 305-325.

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, 53-67.

Charlop, M. H., Kurtz, P. H., & 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, 795-808.

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 & Treatment of Children,* 16, 213-234.

Cooke, N. L., & Reichard, S. M. (1996). The effects of different interspersal drill ratios on acquisition and generalization of multiplication and division facts. *Education & Treatment of Children,* 16, 124-142.

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

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

Killu, K. (1999). High-probability request research: Moving beyond compliance. *Education and Treatment of Children,* 22, 470-494.

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

McCurdy, M., Skinner, C. H. (2001). Increasing on-task behavior in an elementary student during mathematics seatwork by interspersing additional brief problems. *School Psychology Review,* 30, 23-32.

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

Rowan, V. C., Pear, J. J. (1985). A comparison of the effects of interspersal and concurrent training sequences on acquisition, retention, and generalization of picture names. *Applied Research in Mental Retardation,* 6, 127-145.

Skinner, C. H. (2002). An empirical analysis of interspersal research evidence, implications, and applications of the discrete task completion hypothesis. *Journal of School Psychology,* 40, 347-368.

Skinner, C. H., Hall-Johnson, K., Skinner, A. L., Cates, G., Weber, J., & Johns, G. A. (1999). Enhancing perceptions of mathematics assignments by increasing relative problem completion rates through the interspersal technique. *The Journal of Experimental Education,* 68, 43-59.

Skinner, C. H., Robinson, S. L., Johns, G. A., Logan, P. & Belfiore, P. J. (1996). Applying Herrnstein's matching law to influence students' choice to complete difficult academic tasks. *Journal of Experimental Education,* 65, 5-17.

Teeple, D. F., & Skinner, C. H. (2004). Enhancing grammar assignment perception by increasing task demands: Extending additive interspersal research to students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders,* 12, 120-127.