November 2021

Efficiency of Teaching Sight Words in Similar vs Dissimilar Sets

Jensen Chotto

EFFICIENCY OF TEACHING SIGHT WORDS IN SIMILAR VS DISSIMILAR SETS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfilment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Jensen Chotto
B.S., Louisiana State University, 2019
December 2021

Table of Contents

Abstract

Early reading intervention can decrease the likelihood that children who struggle with reading develop long-term reading problems. Due to the prevalence of words that cannot be read phonetically in the English language, sight word instruction is required to supplement phonics instruction. In this study, we compared the effects of creating sets of sight words with the same starting letter (3 words per set, 3 total sets) versus distributing words with the same starting letter across sets when assessing acquisition of the combined set (9 words) in five 4-to-6-year-old children using a combined adapted alternating treatments design and pre-posttest design. All participants mastered the 3-word sets in both teaching conditions but did not master the control sets. With the exception of one 9-word set for 1 participant, all participants required teaching of the 9-word sets as a set (interspersed teaching). The total number of sessions to mastery of the 9-word sets varied across participants: two participants required substantially more sessions in the similar condition, two participants required approximately the same number of sessions in both conditions, and one participant required more sessions in the dissimilar condition. For the two participants who required substantially more teaching sessions in the similar conditions, they not only responded incorrectly more often during teaching but also made errors that suggest behavior was controlled by the first letter of the word rather the whole word. These findings are consistent with stimulus disparity research demonstrating that discrimination training is generally less efficient when comparison stimuli are similar.

Keywords: early intervention, literacy, reading, sight words, stimulus disparity

Efficiency of Teaching Sight Words in Similar vs Dissimilar Sets

Students who do not read proficiently are likely to have reading difficulties that persist and may lead to negative outcomes such as grade retention, school dropout (Darney et al., 2013), and emotional and behavioral problems (Daniel et al., 2006). Early reading intervention can decrease the likelihood that children who struggle with reading develop long-term reading problems (Velluntino et al., 2006). An important feature of effective reading instruction is phonics instruction, which teaches children the relationship between letters and sounds (phonemes). Phonemes are the smallest unit of words, which have no meaning (Bloomfield, 1926; Matos & Passos, 2006; e.g., the word *chat* includes the phonemes /ch/, /a/, and /t/). However, the English language contains many phonemic irregularities and words that cannot be read phonetically (e.g., *know, the*). Thus, reading these words must come under stimulus control of the entire word (Kirby et al., 1981; Kupzyk et al., 2011). Teaching reading in which the functional unit is the entire word is referred to as sight word instruction. Given the prevalence of irregularities in the English language, sight word instruction is required to supplement phonics instruction for early readers (McArthur et al., 2015).

A common method used to teach sight words is the use of flashcards (Kupzyk et al., (2011). Flashcards can be used to teach sight words by arranging a three-term contingency: the teacher presents the flashcard, the learner responds, and the teacher delivers a consequence. Common consequences include praise for a correct responses and error correction (e.g., modeling the correct response) for an incorrect response. Error correction procedures can impact the rate of acquisition for sight words (Kodak et al., 2016). Kodak et al. (2016) compared the efficiency of and preference for five error-correction procedures described in the research literature and used frequently in practice on sight word reading (4 participants) or tacting prepositions (1

participant) with five children with autism spectrum disorder (ASD). The error correction conditions included differential reinforcement (no error correction), a demonstration condition in which modeling of the correct response following incorrect responding was added to the differential reinforcement procedure, prompt delay in which reinforcers were added for responding to the model following incorrect responding, single response repetition in which the trial was re-presented following an error and the prompt delay procedures, and multiple response repetition in which the trial was re-presented three times following an error and the prompt delay procedures. Kodak et al. found that all error correction procedures led to mastery across participants, whereas the differential reinforcement condition did not produce mastery. The demonstration condition was generally efficient and preferred by some participants over the multiple response repetition and differential reinforcement conditions, but the prompt delay procedure was preferred for participants who were given the opportunity to select between the demonstration, prompt delay, and differential reinforcement procedures. An alternative error correction procedure assessed is fading in embedded pictures representing the word. Richardson et al. (2017) used a multielement design within a nonconcurrent multiple baseline design across word sets to compare the effects of using a text-only condition and two combined text-picture conditions on correct sight word reading with four children diagnosed with ASD. In one text-picture condition, pictures representing the word were embedded in the word during training sessions. The other text-picture condition involved fading in an embedded picture during error correction. Richardson et al. (2017) found that the embedded pictures during training impeded acquisition and when faded in as a prompt did not produce faster acquisition of the word sets, and the authors generally recommended against the use of embedded pictures in sight word reading instruction.

In addition to the type of error correction procedure chosen, the immediacy of when the correction is delivered also affects the rate of acquisition. Barbetta et al. (1994) compared using an immediate error correction versus a delayed error correction while teaching children sight words. The immediate error correction consisted of an experimenter repeating the correct word following an incorrect response. The delayed error correction consisted of the experimenter informing the participant that the response was incorrect and that the experimenter would return to that word later. At the end of three practice rounds, the experimenter displayed all the words to which the participants responded incorrectly. Immediate error correction was found to produce better performance than a delayed error correction. Barbetta et al. (1993) compared the use of whole-word error correction and phonetic prompting error correction in which individual phonemes were sequentially presented until the learner responded correctly on the number of words read correctly. They found that whole word error correction produced higher scores for learning sight words compared to phonetic prompting error correction.

In addition to flashcard instruction to teach sight words, some researchers have evaluated how the combination of flashcards with educational games can affect the rate of sight word acquisition. Kirby et al. (1981) evaluated the efficacy of word game bingo for the acquisition and maintenance of sight words with six third grade students using a modified multiple baseline design. Each participant was given a set of bingo cards. The experimenter presented a card and announced the word, and participants needed to find the word on their bingo cards and match the word with a token. The winners of the bingo game were given tokens that they could exchange for backup reinforcers on a weekly basis. Kirby et al. (1981) compared this intervention to a baseline condition that consisted of presenting sight word cards sequentially and asking participants to read each card. Results showed a significant increase in the percentage of words

read correctly compared to the participant's baseline performance (Kirby et al., 1981). More recently, Sperling et al. (2019) assessed the effectiveness of Reading Racetrack on sight word fluency (number of words read per minute), assessed via presentation of 30 sight word flash cards. A Reading Racetrack board was created for the study and consisted of 30 empty squares that were placed on a racecourse for either horses or cars depending on participants preference. Flashcards with sight words are set up face down on each space of the board and the child moves the game piece according to the number of the die. The experimenter turns over the card and asks the child to name the word; if the word is correct, the experimenter delivers praise to the child and the game continues. If the response is incorrect then the child gets corrected. Results of the Reading Racetrack intervention were compared to children's baseline reading fluency performance. During baseline, participants engaged in math activities for 15 minutes and then an experimenter assessed their reading fluency using 12 daily probes. Sperling et al. (2019) used a multiple baseline across subject design and found that all three of their participants increased the number of words they read per minute compared to baseline. Educational games to teach sight word reading are also available via mobile apps. Musti-Rao et al. (2015) evaluated the effectiveness of teacher-directed iPad® instruction using *The Sight Words: Kids Learn App* for iPad®, on first graders' sight word fluency using a multiple baseline design across participants. In baseline, students engaged in their typical reading routine, which consisted of independent reading time, partner reading, and small group reading directed by either the teacher, assistant teacher, student teacher, or assistant principal. During teacher-directed iPad® instruction, the teacher targeted five words for each session and prompted students to use the app to listen to the word, say the word, write the word, say the word, record the word, listen to the recorded word, repeat the word, and then move to the next word. This was repeated five times for each word,

and at the end of the instruction, students were allowed to play a game on the iPad®.

Experimenters administered weekly word reading probes composed of 30 probes each. During the weekly probes, experimenters pointed to a word and prompted the student to read the word; examiners then recorded responses as incorrect or correct. For all the participants, the number of correct responses per minute increased after intervention. Repeated reading practice of reading sight words with differential reinforcement and error correction procedures across various modalities (traditional flashcards, flashcards embedded into board games, flashcards presented via an app) has been shown to produce increases in sight word reading in children. However, the selection of the order in which specific words are taught as sight words and stimulus features of words taught as sight words has been given less consideration.

In early phonics reading instruction, words are often presented in "word families," which are words that only differ by the first letter (e.g., *bat*/*hat*; Saunders, 2011). However, sight words are typically presented based on their frequency in reading materials (e.g., Dolch© and Fry© sight word lists; Musti-Rao et al., 2015). Educators sometimes assume that discrimination between letters presented individually is transferred when these same letters are presented within whole words. However, responding effectively to letters presented in isolation does not necessitate responding effectively to letters presented within words. Stromer et al. (1993) conducted a delayed matching-to-sample task using letter-like nonsense stimuli with seven individuals diagnosed with intellectual disability. Stimuli were presented side by side on a touch-sensitive computer screen. After the stimuli disappeared, several stimuli were presented, including one of the two previous stimuli. Participants were instructed to choose the stimulus that was previously presented. Participants consistently responded correctly in the presence of one of the two original

stimuli and at chance levels in the presence of the other stimulus, demonstrating responding came under control of only part of the compound stimulus.

A couple of applied studies examining the discrimination between consonant-vowel-consonant (CVC) words that shared all but the first letter in common also suggest that responding is often controlled by part of a word rather than the whole word (Saunders et al., 2000; Yoo & Saunders, 2014). Both Saunders et al. (2000) and Yoo and Saunders (2014) evaluated young children's match-to-sample performance with CVC words that differed only by the first letter. Prior to the CVC word matching task, participants accurately responded to a match-to-sample task with letters presented individually. Both studies found that preschool children who identified letters with high accuracy when presented with letters in isolation were much less accurate on the word matching task. Further, following teaching procedures that produced correct responding to a single pair of words did not produce accurate responding to new pairs of words that differed only by the first letter. Yoo and Saunders demonstrated that teaching word matching by sequentially adding letters to the word following correct matching to the initial letter in isolation effectively taught responding to the whole word. Collectively, these findings are consistent with the phenomenon Ehri (1992) termed the partial-alphabetic phase. The partial-alphabetic phase is when a child identifies words by responding to parts in the printed word. Although these studies demonstrated stimulus control problems with different initial letters in words, the initial letter often controls responding in young children (Saunders, 2011). Words with the same starting letter are low in stimulus disparity, how different stimuli are from one another, which may hinder acquisition when taught together (Halbur, Caldwell, et al., 2021). However, if words with high stimulus disparity are taught together, perhaps features other than the whole word (e.g., word length, initial letter, ending letter) will control responding. In the present study, stimulus

disparity was manipulated between conditions by altering whether the first letter within a teaching set was the same or different. To our knowledge, this is the first study to systematically evaluate the effects of stimulus disparity in sight word instruction.

The purpose of this study was to determine the effects of creating sets of sight words with the same starting letter (3 words per set, 3 total sets) versus distributing words with the same starting letter across sets when assessing acquisition of the combined set (9 words). We measured the number of teaching sessions to mastery of each set of three words as well as the total number of teaching sets required for mastery of the combined set of 9 words. A secondary purpose was to assess the types of errors produced by each condition to potentially identify the source of stimulus control during teaching (Grow et al., 2011).

Method

Participants, Settings, and Materials

Eight 4- to 6-year-old children were recruited via word-of-mouth for parents who were seeking additional sight word practice for their children. Children were included in the study if they responded incorrectly to a set of 24, unique four-letter words during the preassessment (see screening procedures in below). Children also were required to correctly identify all 26 letters of the alphabet when presented, and emit the correct sound when asked, "What sound does this letter make?". Additionally, children were required to respond correctly when presented with CVC words and told, "Sound out this word". Five children were eligible for participation: Jill, Nico, Emmie, Jacob, and Cali. The participants' demographic information and DIBELS score is displayed in Table 1.

Table 1. Demographic Information

Participant	Gender	Age	Race/Ethnicity	DIBELS Score	DIBELS Range
Jill	F	5	White	411.53	Some Risk
Nico	M	6	Latino	371.9	Minimal Risk
Emmie	F	6	White	420.4	Negligible Risk
Jacob	M	6	White	386.4	Minimal Risk
Cali	F	4	White/ Latina	--	--

Note. Nico, Emmie, and Jacob's scores were compared to the benchmark goals for the middle of kindergarten. At the middle of year benchmark, scores of 371-392 are in the minimal risk range, and scores of 393 and higher are in the negligible risk range. Jill's score was compared to the benchmark goals for the end of kindergarten. At the end of year benchmark, scores of 406-419 are in some risk range.

Experimenters obtained parent informed consent for all participants prior to the start of the study and child assent before conducting research each day. We defined a research block as a 12-trial mastery probe followed by nine-trial teaching sessions for the dissimilar and similar sets. We conducted mastery probes and teaching sessions at an empty table in the participant's home. Cali's mother, a behavior analyst, conducted sessions with Cali and video recorded some

sessions for interobserver agreement. We conducted one probe and two to four teaching sessions per day, 2 to 4 days per week.

All words were printed on 12.7 cm X 6.35 cm white, laminated cards. Words were printed in Calibri size 28 font. Tokens were used as a reinforcer for correct responses across phases. The experimenter used a token board with nine hook-and-loop dots. Each token represented either 15 s of access to the tablet or a piece of candy or cereal. The tablet was preloaded with age-appropriate games. Cali, Nico, and Emmie chose between an edible and the tablet as their reinforcer at the beginning of each session. Jacob and Jill used the tablet as their reinforcer for the duration of the study.

Dependent Variables and Interobserver Agreement

Observers recorded all responses on a trial-by-trial basis using pen and paper data sheets. The primary dependent variable was the number of teaching sessions to mastery. This was calculated for each set of words within a condition and by adding all teaching and interspersed teaching sessions to calculate the total number of teaching sessions to mastery of the 9-word set in each condition. Observers also measured the percentage of correct responses during pre-and post-tests, mastery probes, and teaching sessions by dividing the number of correct responses per session by the total number of trials and multiplying by 100. Observers recorded a *correct response* when the participant said the word corresponding with the text on the stimulus card within 5 s of the stimulus card presentation. Additionally, observers calculated the percentage of different types of errors emitted across conditions by dividing the total number of each type of error made (incorrect, incorrect same set, incorrect different set, no response) in each condition by the total number of errors made and multiplying by 100 (Grow et al., 2011). Observers recorded an *incorrect same set response* when the participant said a word that was in the same

set and did not correspond to the presented stimulus. Observers recorded an *incorrect different set response* when the participant said a word that was in a different set. Observers recorded an *incorrect response* when the participant said anything other than the word presented on the card and if it did not correspond to a word in any set. Observers recorded *no response* when the participant did not say any word within 5 s of stimulus card presentation or indicated "I don't know."

We assessed interobserver agreement (IOA) by having a second observer independently collect data on sight word card presentation and the participant response. We defined an agreement as the experimenter and observer recording the same sight word card and participant response. We calculated IOA using the trial-by-trial exact agreement method. That is, we divided the number of agreements by the total number of trials per probe or session, multiplied that number by 100 to obtain a percentage and then averaged that number across probes and sessions for each condition. For the similar condition, we assessed IOA during 51%, 25%, 88%, 76%, and 52% of probes and sessions; mean IOA was 99%, 100%, 100%, 97%, and 100% for Jill, Nico, Emmie, Jacob, and Cali, respectively. For the dissimilar condition, we assessed IOA during 51%, 25%, 86%, 79%, and 42% of probes and sessions; mean IOA was 99%, 100%, 99%,99%, and 100% for Jill, Nico, Emmie, Jacob, and Cali, respectively. For the similar control condition, IOA was assessed during 48%, 19%, 92%, 77%, 50% of probes and sessions; mean IOA was 92%, 100%, 87%, 100%, and 100% for Jill, Nico, Emmie, Jacob, and Cali, respectively. For the dissimilar control condition, IOA was assessed during 59%, 24%, 92%, 81%, and 50% of probes and sessions; mean IOA was 90%, 100%, 100%, 97%, and 100% for Jill, Nico, Emmie, Jacob, and Cali, respectively.

Screening

DIBELS (8th Ed.)

We administered the DIBELS 8th Edition Letter Naming Fluency, Phonemic Segmentation Fluency, Nonsense Word Fluency, and Word Reading Fluency tests to four of the five participants to compare their performance to a normative sample of children in kindergarten. The other participant, Cali, was not yet in kindergarten and did not complete the DIBELS assessment. Three of the participants' scores were compared to the middle of the year benchmarks; Jill's score was compared to the end of the year benchmark due to the timing of when sessions were conducted within the school year. Based on their performance, students are categorized in one of four DIBELS benchmark sections: (1) students who just need core support with negligible risk of needing intervention, (2) students who need core support with minimal risk of needing intervention, (3) strategic support with some risk of needing intervention, and (4) intensive support students who are at risk and need intervention. Emmie scored within the core support (negligible risk) range, Jacob and Nico scored within the core support (minimal risk) range, and Jill scored within the strategic support range (some risk).

Preassessment and Word Set Creation

We conducted a preassessment screening to identify words to use in the teaching sessions. The preassessment included one, 24-trial session in which the experimenter presented one, four-letter sight word card at a time asked, "What does this say?" for 24 unique sight words. We collected data on correct and incorrect responses. Correct responses resulted in vocal praise (e.g., "Nice work!") and incorrect responses resulted in a neutral statement (e.g., "Okay"). In addition, the experimenter intermittently provided praise contingent on non-target behavior related to the session (e.g., "Good job paying attention;" "I like how you're sitting and looking at the card").

A pool of 61 sight words was created based on the Dolch Sight Word List©, the Fry Sight Word List©, and the DIBELS Word Reading Fluency section. Current sight word lists from the school were acquired from parents and those words were not included. Sight words were chosen for each participant based on words to which they responded incorrectly in the preassessment. If the child responded correctly to a word in the preassessment it was replaced with another word that began with the same letter. Twenty-four four-letter sight words were divided into similar, dissimilar, and control sets. Each similar and dissimilar set included three sets of three words. Word sets per participant are included in Table 2. In similar sets, each word began with the same starting letter within but not across sets. In dissimilar sets, each word began with a different starting letter within the set but had the same starting letter as one of the words in each of the other two sets. The similar control set contained three words with the same starting letter but a different letter than is used in any of the other sets. The dissimilar control set contained three words with different starting letters than any words from all other sets.

We used a combined adapted alternating treatments design (Sindelar et al., 1985) and pre-posttest design. During the multielement comparison of mastery probes, four conditions were alternated: similar, dissimilar, similar control, and dissimilar control. Additionally, the condition presented first during teaching sessions was counterbalanced across days, sessions were alternated between the two teaching conditions, and the experimenter conducted an equal number of teaching sessions for each condition per day.

General Procedure

Prior to all sessions, the experimenter stated the reinforcer options, allowed the participant to select the back-up reinforcer, and reminded the participant that they could earn tokens for correct responses and those tokens could be exchanged later for the back-up

reinforcer. Sessions were blocks of trials, following which the participants would have an opportunity to exchange tokens. A trial consisted of the presentation of a stimulus card, a participant response, and an experimenter-delivered consequence.

Table 2. Stimulus Sets

Participant	Similar Control	Dissimilar Control	Similar Sets	Dissimilar Sets
Jill	both, base, been	also, done, open	Set 1: same, side, such Set 2: mine, move, many Set 3: were, when, want	Set 1: high, life, cold Set 2: help, lake, cute Set 3: hard, left, cent
Nico	both, ball, been	your, done, take	Set 1: cold, cute, call Set 2: make, move, mine Set 3: hurt, hold, help	Set 1: same, last, with Set 2: stop, lake, were Set 3: such, left, want
Emmie	both, ball, been	your, done, take	Set 1: same, side, some Set 2: many, mine, move Set 3: were, went, want	Set 1: high, life, cold Set 2: head, lake, cute Set 3: hard, cent, lead
Jacob	both, base, been	also, done, open	Set 1: high, help, hard Set 2: life, lake, left Set 3: cold, cute, cent	Set 1: same, many, were Set 2: side, move, when Set 3: such, mine, want
Cali	both, ball, been	your, done, take	Set 1: same, soon, stop Set 2: cold, call, cute Set 3: lead, live, long	Set 1: mine, what, hold Set 2: move, were, hurt Set 3: many, want, help

Pre-test. Three pre-tests were conducted for each participant. Each pre-test included a similar test, in which the words in the similar word sets were shuffled and presented in a single nine-trial session, a dissimilar test, in which the words in the dissimilar word sets were shuffled and presented in a single nine-trial session, and a control test, in which the words in the control sets were shuffled and presented in a single six-trial session. Correct responses produced praise and a token exchangeable for one piece of Skittles® or 15 s access to tablet games at the end of the pre-

test. Incorrect and no responses produced the next trial. The experimenter delivered praise for on-task behavior (e.g., sitting still, looking at the card) after approximately every third trial.

Mastery probes. At the start of each session, we conducted a mastery probe session for stimulus sets currently being taught and the control sets. During mastery probes, each stimulus card was presented once in random order within each set. The order of sets presented during the mastery probe were randomized across days. The experimenter delivered praise and a token exchangeable for reinforcers at the end of the mastery probes contingent on each correct response. The experimenter moved to the next trial following incorrect responses. Stimuli assigned to control sets were only presented in the mastery probes.

A set was considered mastered when the participant responded correctly to all three stimuli in the set during a mastery probe across two days. Dissimilar and similar sets were taught sequentially. That is, after similar set 1 was mastered, similar set 2 entered teaching.

Teaching. Prior to the first teaching session of the day in each condition, the experimenter presented the card and modeled the correct response for all words to which the participant responded incorrectly during the mastery probe. If the participant did not imitate the model, the experimenter prompted the participant by asking, "What does it say?". If the participant provided a correct imitation, the experimenter delivered praise. If the participant responded incorrectly, the experimenter said, "Repeat after me" and repeated the model. An observing response procedure was implemented at session 32 for Jill due to off-task behavior such as climbing out of her seat and not looking at the card when responding. A sticky note was attached to the top of each sight word card, covering the word. The experimenter modeled how to lift the sticky note to see the word underneath and then respond. Jill then practiced using the

sight word cards in session that day and the observing response procedure was used for the remainder of the study.

Teaching sessions consisted of nine trials in which each word was presented three times in random order. The experimenter delivered praise and a token exchangeable for reinforcers at the end of each session contingent on each correct response. If the participant engaged in an incorrect or no response, the experimenter provided an error correction identical to the modeling procedure.

Post-test. Once the participant met the mastery criterion for all three sets in one condition, the experimenter conducted the post-test the following day. The similar and dissimilar post-tests were identical to the pre-test, in which all the words in the word sets were shuffled and presented in a single nine-trial session. The control post-test was conducted on the same day, after the second post-test. An interspersed set was considered mastered if the participant responded correctly to all stimuli in the post-test.

Interspersed teaching. If the participant responded incorrectly to any words during the post-test, the experimenter conducted interspersed teaching. The nine-word set was treated as a teaching set and the procedures for mastery probes, modeling, and teaching were implemented. During each interspersed teaching session, the experimenter presented each word once. Mastery criterion for the interspersed set was two consecutive mastery probes with 100% correct responding.

Treatment Integrity

We assessed treatment integrity by determining if the experimenter presented the correct sight word card and delivered the programmed consequence on each trial. We calculated treatment integrity by dividing the number of correct trials (i.e., correct sight word card

presentation and consequence) by the total number of trials per probe or session and multiplying that number by 100 to obtain a percentage. Treatment integrity was 100% across all conditions for all participants. The experimenter presented the correct sight word card and delivered the programmed consequence in all sessions for all conditions.

We assessed IOA for treatment integrity by having a secondary observer collect data on the sight word card presentation, participant response, and trial consequence. We defined an agreement as the experimenter and the observer recording the same sight word card, participant response, and consequence. We calculated IOA using the trial-by-trial exact agreement method. During the similar control condition, we assessed IOA for treatment integrity during 48%, 19%, 92%, 77% and 50% of probes and sessions for Jill, Nico, Emmie, Jacob, and Cali respectively. During the dissimilar control condition, we assessed IOA for treatment integrity during 59%, 24%, 92%, 81%, and 50% of probes and sessions for Jill, Nico, Emmie, Jacob, and Cali, respectively. During the dissimilar condition, we assessed IOA for treatment integrity during 51%, 25%, 86%, 79%, and 42% of probes and sessions for Jill, Nico, Emmie, Jacob, and Cali, respectively. During the similar condition, we assessed IOA for treatment integrity during 48%, 19%, 92%, 77%, and 50% of probes and sessions for Jill, Nico, Emmie, Jacob, and Cali, respectively.

Results

Table 3 depicts the number of sessions to mastery for each participant, including both the number of sessions to set mastery and number of sessions required for interspersed teaching, and Figure 1 depicts the number of correct responses emitted during across all mastery probes for each participant. Three different trends emerged from evaluating the number of sessions to mastery. Nico and Jill required substantially more sessions to set mastery for similar word set 1 compared to dissimilar word set 1. Emmie and Jacob mastered both the similar and dissimilar sets and interspersed sets in approximately the same number of sessions. Cali required more teaching sessions to set mastery for dissimilar word sets compared to similar word sets. All participants required interspersed teaching for both similar and dissimilar sets, except Jacob for the similar set. Nico and Cali required extensive interspersed teaching for the dissimilar condition. Jill and Emmie required a similar number of interspersed teaching sessions in both conditions.

Table 3. Number of Sessions to Mastery

Participant	Teaching Sessions to Mastery						Int. Teaching		Total	
	DS 1	S 1	DS 2	S 2	DS 3	S 3	DS	S	DS	S
Jill	7	20	10	9	9	5	7	6	33	40
Nico	4	13	10	11	4	3	11	6	29	33
Emmie	2	3	4	4	4	5	1	1	10	13
Jacob	5	4	4	7	5	3	2	0	16	14
Cali	6	4	9	6	3	2	11	1	29	13

Note. S = similar; DS = dissimilar

Figure 2 depicts the percentage of correct responses during the sequential set teaching and the interspersed probe teaching sessions. In general, participants' percentage of correct responses increased across teaching sessions. However, consistent with the mastery data, for

children who required more teaching sessions (Jill and Nico), word sets organized with the same first letter (similar condition) resulted in more errors during the teaching sessions.

Figures 3 and 4 display error analyses. Figure 3 displays the type of error made for each word for each participant across trials. Figure 4 shows a summary of the error types, depicting the percentage of each type of error emitted in the similar and dissimilar condition. The summary data reasonable describe the patterns observed across trials. That is, the types of errors within and across sets did not change other than to become less frequent after repeated teaching trials. Overall, there were two common errors: saying an incorrect word that is not in any set and saying an incorrect word in the same set. With the exception of Cali, all participants made considerably more errors by saying a different word in the same set in the similar condition than in the dissimilar condition. Participants rarely responded with words that were in different sets for both similar and dissimilar condition, but it was a slightly more common error made in the dissimilar condition. With the exception of Cali, participants rarely emitted no response in either condition.

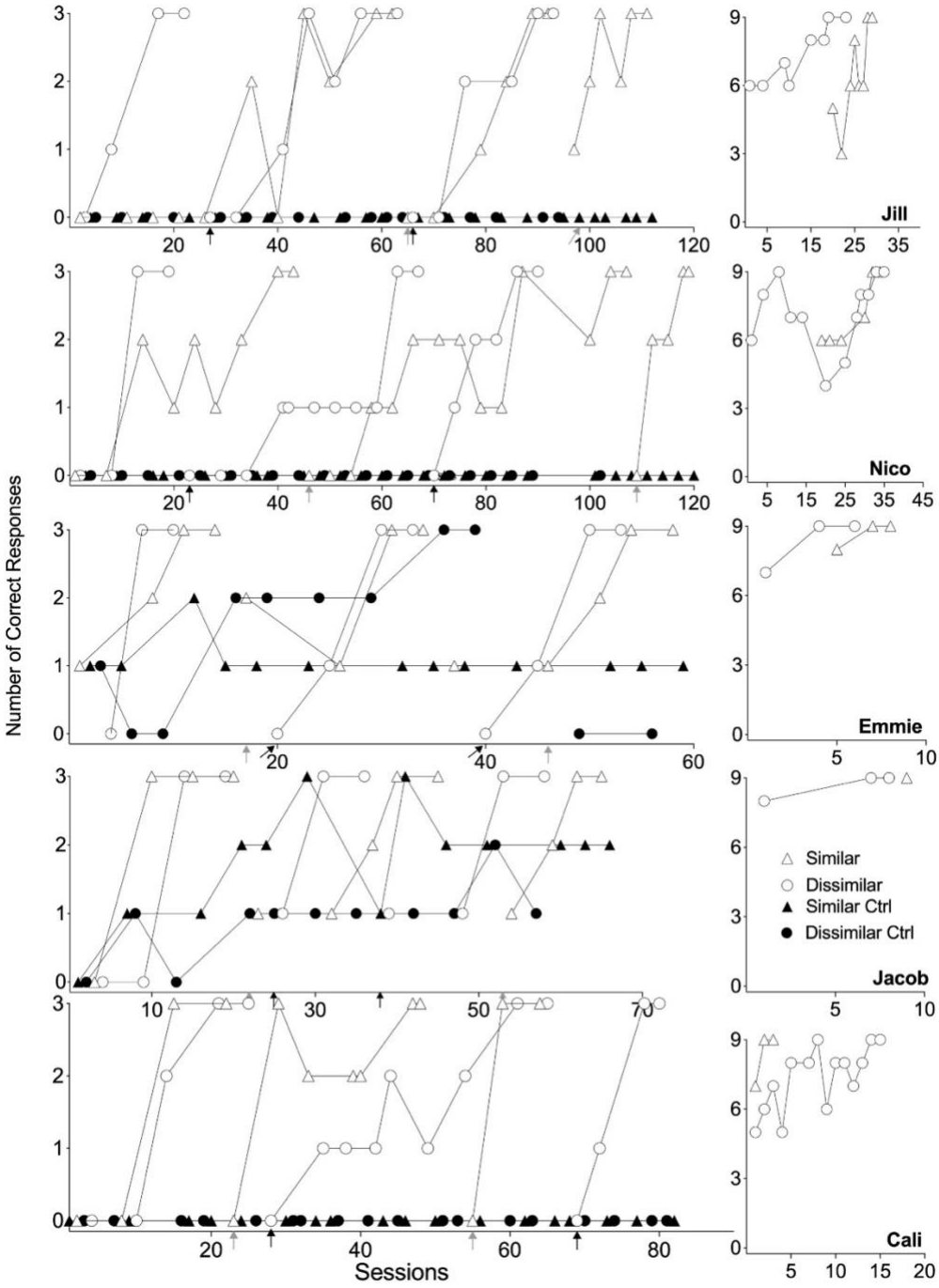

Figure 1. Sequential Set and Interspersed Mastery Probes

Note. Similar Ctrl = similar control; Dissimilar Ctrl = dissimilar control; grey arrows indicate a new similar set; black arrows indicate a new dissimilar set

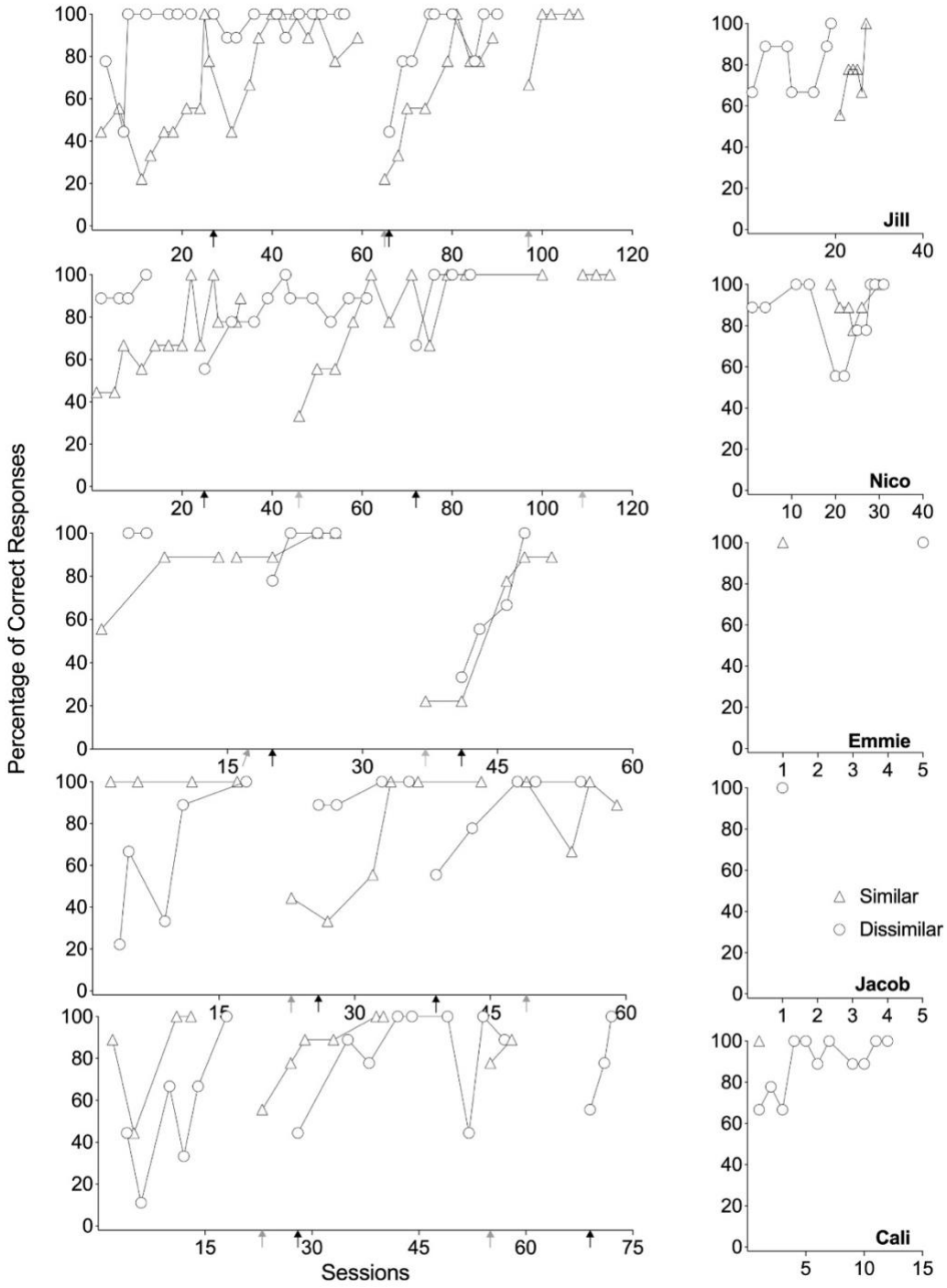

Figure 2. Sequential Set and Interspersed Teaching

Note. Grey arrows indicate a new similar set; black arrows indicate a new dissimilar set

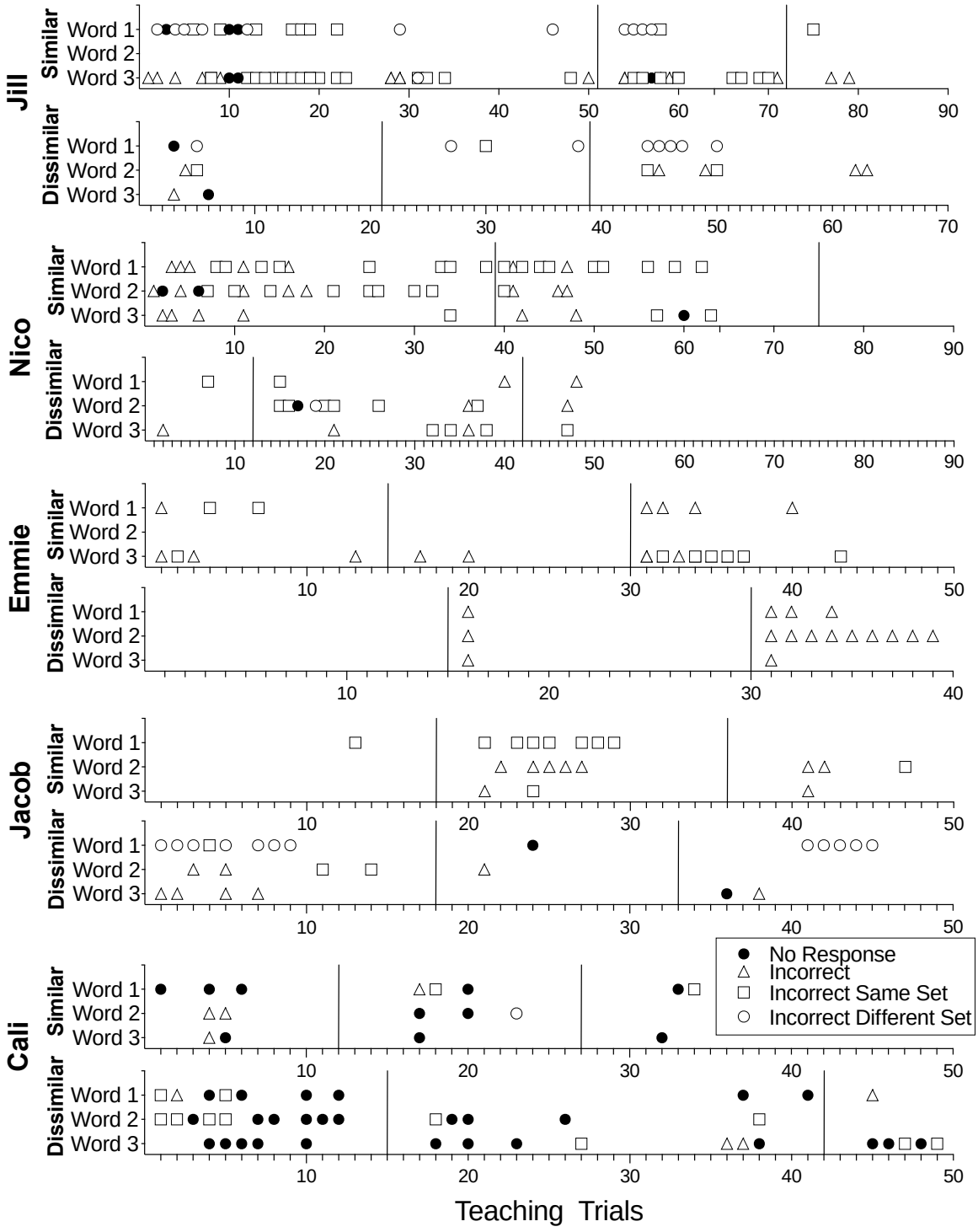

Figure 3. Types of Errors Made Across Trials for Each Participant

Note. Blank spaces depict correct responding. Vertical lines indicate a new set.

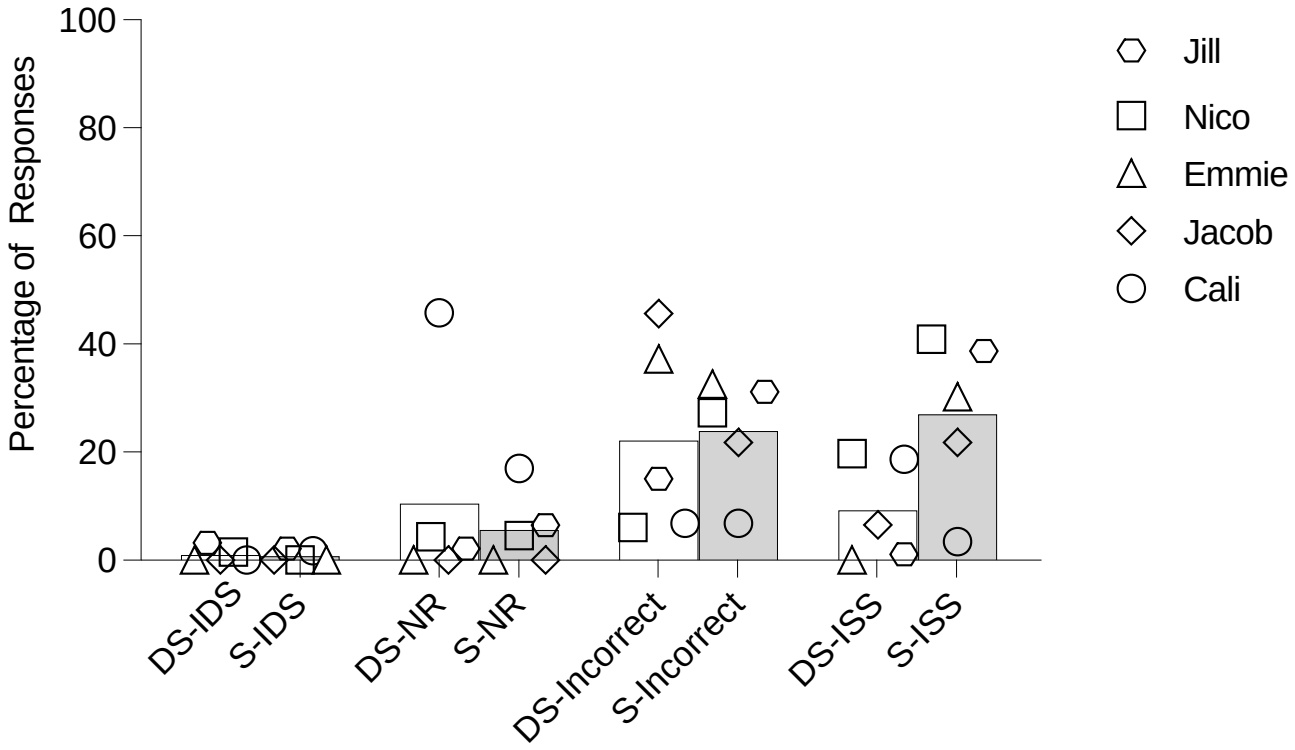

Figure 4. Types of Errors Emitted in Similar and Dissimilar Conditions

Note. DS = dissimilar; S = similar; IDS = in different set; NR = no response; ISS = in same set

Figure 5 depicts the number of correct responses during the pre-test and post-test for the control, similar, and dissimilar conditions. For all word sets, all the participants emitted zero correct responses to all stimuli on the pretest. On the dissimilar control post-test, 3 of the 5 participants emitted zero correct responses. Emmie emitted one correct response and Jacob emitted two correct responses. On the similar control post-test, 2 of the 5 participants emitted zero correct responses. One participant emitted two correct responses and two participants emitted one correct response. The mean number of correct responses on the post-tests were six for the dissimilar set and seven for the similar set. Emmie and Jacob both emitted a correct response for one more similar word than dissimilar word but received an equal amount of teaching sessions for both conditions. Cali also emitted a correct response for two more similar

words than dissimilar even though she received more dissimilar teaching sessions. Nico responded correctly to an equal number for both similar and dissimilar conditions despite receiving more similar teaching sessions. Jill emitted a correct response for one more dissimilar word than similar word but received more similar teaching sessions. Overall, 3 of the 5 participants emitted a correct response for at least one more similar word than dissimilar word on the post-test.

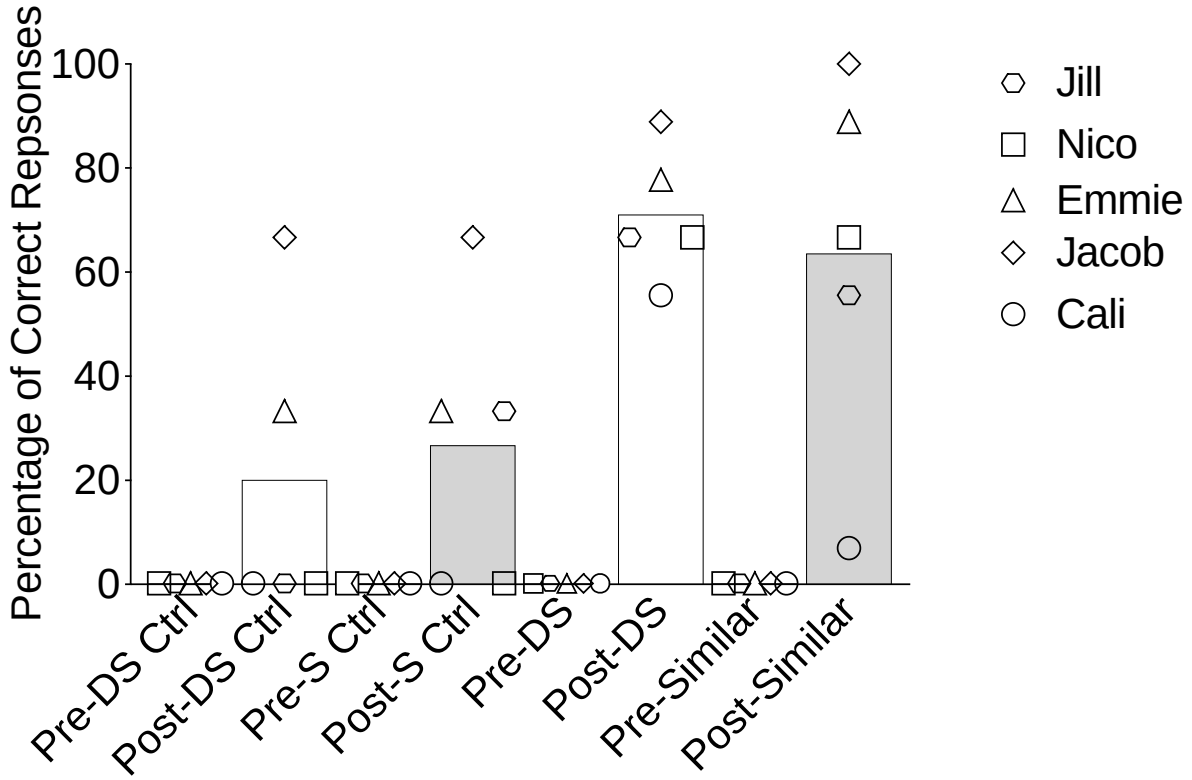

Figure 5. Number of Correct Responses in the Pre- and Post-Tests

Note. DS Ctrl = dissimilar control; S Ctrl = similar control; DS = dissimilar; S = similar

Discussion

We evaluated the effects of stimulus disparity in teaching sight words by arranging three-word sets with the same (similar condition) or different (dissimilar condition) starting letter on acquisition of a nine-word set with five young children. Additionally, we conducted an error analysis to evaluate the types of errors produced in each condition. Participants mastered 9-word sets of sight words in both intervention conditions and not in the control condition. Overall, the similar condition required approximately the same (two participants) or fewer (three participants) interspersed teaching sessions than the dissimilar condition. However, arranging sets with little stimulus disparity (i.e., the similar condition) produced slower acquisition and eventual mastery for the two participants most at risk for reading difficulties and who required the most sessions to mastery for both intervention sets. This finding is consistent with prior research on stimulus disparity demonstrating that discrimination training is generally less efficient when comparison stimuli are more similar (Halbur, Caldwell, et al., 2021; Halbur, Kodak, et al., 2021). Interestingly, the discrepancy primarily occurred in teaching the first 3-word set. Additionally, participants were more likely to make an error by saying a different word in the same set in the similar condition than in the dissimilar condition. Taken together, these findings suggest that some participants' responding was initially controlled by the first letter of the word and that repeated exposure to the teaching contingencies shifted control to the whole word.

Consistent with Saunders et al. (2000) and Yoo & Saunders (2014), low stimulus disparity produced errors in correct responding to sight words, further supporting the suggestion that teaching sight word reading requires arranging contingencies to ensure responding comes under the control of the whole word. The percentage of in same set errors made in the similar set suggest that responding was controlled by the first letter of each word. Anecdotally, even when

emitting incorrect responses, some participants were saying words that began with the same letter, but not words that were used in this study. For example, both James and Emmie emitted *said* for *side*. Both Nico and Jill required many sessions to master set 1 of the similar condition and fewer sessions in subsequent sets, perhaps suggesting stimulus control shifted from the first letter to more of the whole word. However, they both continued to make errors by saying a word from the same set in sets 2 and 3.

In general, more teaching sessions in one condition did not predict better responding on the post-test in that condition. One potential reason the similar set required fewer interspersed teaching sessions for Nico could be because he had more exposure to teaching sessions with those stimuli; however, Cali required more interspersed teaching sessions in the dissimilar condition but had less exposure to those stimuli. Alternatively, both Nico and Cali's responding may not have come under control of the whole word in the dissimilar set until the contingencies required it (i.e., during interspersed teaching). Both Cali and Nico emitted a higher percentage of in same set errors in the dissimilar condition compared to other participants.

Although clear differentiation between the teaching and control conditions was demonstrated for all participants, three of the participants responded correctly to some words in the control sets on the post-test, and Emmie mastered the initial dissimilar control set. These three participants were in school during the study and likely learned these words there, whereas the other two participants were not in school during the study.

Although we controlled starting letter, number of total letters, the absence of initial blends, and different second-letter vowels for words starting with the same letter, some words could have been more difficult to learn than others due to unprogrammed differences in word difficulty across sets. Three of the five participants had the word *cent* in one of their sets. Jacob

had the word *cent* in one of his similar sets, and he made three times the number of errors when presented with the word *cent* compared to the other words in that set (cold, cute), which could be because the *c* is a soft *c* which makes the /s/ sound, opposed to the hard *c* sound which makes a /k/ sound. However, Emmie had *cent* in one of her sets and made fewer errors when presented with that word. Future research examining stimulus disparity in sight word acquisition could examine the effects of auditory similarity within sets. Additionally, Wong et al. (2020) compared mastery at the level of the word versus the level of the set when teaching sight words and found that all the participants learned sight words faster in the word mastery condition. Future researchers may use a word level criterion as, this may be a more efficient way to teach words with little stimulus disparity.

Another potential limitation is that the adapted alternating treatments design may have obscured stimulus control problems that could have been present if each condition had been presented in isolation. Exposure to the similar condition, requiring control by more than the first letter, could have produced responding under the control of more than the first letter in the dissimilar condition as well. One way to potentially mitigate this issue in future research is to arrange for stimulus disparity in different components of the words across conditions (e.g., same initial letter in one condition, same ending letter in another condition). Also, all stimuli across all sets had little stimulus disparity in that they were all four-letter words. Future research that examines the effects of stimulus disparity on sight word acquisition may consider arranging for greater stimulus disparity between sets. Additionally, we did not include a category of errors for words with the same starting letter that were not in any sets. The extent to which errors in the general incorrect category began with the same letter as the presented word would have provided additional evidence to determine the source of stimulus control.

A third limitation is that arranging sets of three words may have been a less efficient teaching strategy than beginning with one larger set (i.e., the interspersed teaching condition; Kodak et al., 2020). However, the smaller set sizes were required to compare the effects of differing stimulus disparity on acquisition and the types of errors produced. Future researchers could include a comparison condition with a single large set combining stimuli of varying stimulus disparity. Additionally, future researchers may want to evaluate the extent to which stimulus disparity within larger sets of sight words affects stimulus control.

This study provides an initial step in evaluating how arranging sight words into different sets defined by stimulus features can affect acquisition and the types of errors that result from arranging word sets based on stimulus disparity. Given the necessity of sight word instruction for learning to read in English, research on effective and efficient sight word teaching methods is essential. In this study, we found that arranging sight words in sets with lower stimulus disparity resulted in more errors and therefore slower acquisition for some participants and provided no apparent benefit to the other participants. Popular approaches to teaching sight words do not arrange word sets in this way but rather by frequency in common books and other written materials students are likely to encounter. These findings do not suggest a change to that approach is necessary on the basis for concerns about the development of restricted stimulus control by the first letter. These findings do suggest, consistent with previous research, that restricted stimulus control by part of the word is likely for some individuals.

Appendix. IRB Approval

Print Close

Record Number	Record Owner	Record Title	Record Status
3905	Donaldson, Jeanne Marie	An Evaluation of Procedures to Increase Academic Fluency in Young Children	Approved

Detail Field Name	Detail Field Data
Record Title	An Evaluation of Procedures to Increase Academic Fluency in Young Children

References

Barbetta, P. M., Heward, W. L., & Bradley, D. M. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis*, *26*(1), 99-110. https://doi.org/10.1901/jaba.1993.26-99

Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A. D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis*, *27*(1), 177-178. https://doi.org/10.1901/jaba.1994.27-177

Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, *2*(3), 153-164. https://doi.org/10.2307/408741

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities*, *39*(6), 507-514. https://doi.org/10.1177%2F00222194060390060301

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology*, *51*(1), 117-128. https://doi.org/10.1016/j.jsp.2012.09.005

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri, & R. Treimam (Eds.), *Reading Acquisition* (pp. 107-143). Hillsdale, NJ: Erlbaum

Grow, L. L., Carr, J. E., Kodak, T. M., Jostad, C. M., & Kisamore, A. N. (2011). A comparison of methods for teaching receptive labeling to children with autism spectrum disorders. *Journal of Applied Behavior Analysis*, *44*(3), 475-498. https://doi.org/10.1901/jaba.2011.44-475

Halbur, M. E., Caldwell, R. K., & Kodak, T. (2021). Stimulus control research and practice: Considerations of stimulus disparity and salience for discrimination training. *Behavior Analysis in Practice*, *14*, 1-11. https://doi.org/10.1007/s40617-020-00509-9

Halbur, M., Kodak, T., Williams, X. A., Reidy, J., & Halbur, C. (2021). Comparison of sounds and words as sample stimuli for discrimination training. *Journal of Applied Behavior Analysis*, *54*(3), 1126-1138. https://doi.org/10.002/jaba.830

Kirby, K. C., Holborn, S. W., & Bushby, H. T. (1981). Word game bingo: A behavioral treatment package for improving textual responding to sight words. *Journal of Applied Behavior Analysis*, *14*(3), 317-326. https://doi.org/10.1901/jaba.1981.14-317

Kodak, T., Campbell, V., Bergmann, S., LeBlanc, B., Kurtz-Nelson, E., Cariveau, T., Haq, S., Zemantic, P., & Mahon, J. (2016). Examination of efficacious, efficient, and socially valid error-correction procedures to teach sight words and prepositions to children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, *49*(3), 532-547. https://doi.org/10.1002/jaba.310

Kodak, T., Halbur, M., Bergmann, S., Costello, D. R., Benitez, B., Olsen, M., Gorgan, E., & Cliett, T. (2020). A comparison of stimulus set size on tact training for children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, *53*(1), 265-283. https://doi.org/10.1002/jaba.553

Kupzyk, S., Daly III, E. J., & Andersen, M. N. (2011). A comparison of two flash-card methods for improving sight-word reading. *Journal of Applied Behavior Analysis*, *44*(4), 781-792. https://doi.org/10.1901/jaba.2011.44-781

Matos, M. A., & Maria de Lourdes, R. D. F. (2006). Linguistic sources of Skinner's Verbal behavior. *The Behavior Analyst*, *29*(1), 89-107. https://doi.org/10.1007/BF03392119

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2015). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities*, *48*(4), 391-407. https://doi.org/10.1177%2F0022219413504996

Musti-Rao, S., Lo, Y. Y., & Plati, E. (2015). Using an iPad® app to improve sight word reading fluency for at-risk first graders. *Remedial and Special Education*, *36*(3), 154-166. https://doi.org/10.11772F0741932514541485

Richardson, A. R., Lerman, D. C., Nissen, M. A., Luck, K. M., Neal, A. E., Bao, S., & Tsami, L. (2017). Can pictures promote the acquisition of sight-word reading? An evaluation of two potential instructional strategies. *Journal of Applied Behavior Analysis*, *50*(1), 67-86. https://doi.org.10,1002/jaba,354

Saunders, K. J., Johnston, M. D., & Brady, N. C. (2000). Identity matching of consonant-vowel-consonant words by prereaders. *Journal of Applied Behavior Analysis*, *33*(3), 309-312. https://doi.org/10.1901/jaba.2000.33-309

Saunders, K. J. (2011). Designing instructional programming for early reading skills. In W. Fisher, C. Piazza, & H. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 92-109) New York: Guillford.

Sindelar, P. T., Rosenberg, M. S., & Wilson, R. J. (1985). An adapted alternating treatments design for instructional research. *Education and Treatment of Children*, *8*(1), 67-76. https://www.jstor.org/stable/42898888

Sperling, M., Barwasser, A., & Grünke, M. (2019). The Effects of a Reading Racetrack Intervention on the Sight Word Fluency of Learning Disabled Elementary School Students with German as Second Language. *Insights into Learning Disabilities*, *16*(1), 79-90.

Stromer, R., Mcilvane, W. J., Dube, W. V., & Mackay, H. A. (1993). Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, *59*(1), 83-102. https://doi.org/10.1901/jeab.1993.59-83

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*, *39*(2), 157-169. https://doi.org/10.1177%2F00222194060390020401

Wong, K. K., Bajwa, T., & Fienup, D. M. (2021). The application of mastery criterion to individual operants and the effects on acquisition and maintenance of responses. *Journal of Behavioral Education*, *Published online ahead of print*. https://doi.org/10.1007/s10864-020-09420-3

Yoo, J. H., & Saunders, K. J. (2014). The discrimination of printed words by prereading children. *European Journal of Behavior Analysis*, *15*(2), 123-135. https://doi.org/10.1080/15021149.2014.11434509

Vita

Jensen Chotto, born in New Orleans, Louisiana, received her bachelor's degree from Louisiana State University in 2019. She plans to receiver her master's degree in December 2021. Her research interests include the assessment and reduction of problem behavior in young children, parent involvement in interventions, and increasing compliance on basic healthcare routines for children with intellectual and developmental disabilities. Upon completion of her master's degree, she will begin work on her doctorate.