

5-24-2022

The Effects of Paired Kinesthetic Movements and Embedded Pictures on Literacy Skills with Preschoolers

Erica Lozy

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Applied Behavior Analysis Commons](#), [Child Psychology Commons](#), and the [School Psychology Commons](#)

Recommended Citation

Lozy, Erica, "The Effects of Paired Kinesthetic Movements and Embedded Pictures on Literacy Skills with Preschoolers" (2022). *LSU Doctoral Dissertations*. 5878.

https://digitalcommons.lsu.edu/gradschool_dissertations/5878

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

THE EFFECTS OF PAIRED KINESTHETIC MOVEMENTS AND EMBEDDED PICTURES ON LITERACY SKILLS WITH PRESCHOOLERS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy

in

The Department of Psychology

by

Erica Lozy

MA, University of Maryland Baltimore County, 2017

BS, Western New England University, 2015

August 2022

Table of Contents

Illustrations	iii
Abstract	iv
Introduction.....	1
Visual Mnemonics.....	5
Kinesthetic Movements.....	14
Gaps in the Current Research.....	17
Purpose	18
General Method	19
Participants, Consent, Assent, and Setting.....	19
Materials.....	21
Dependent Variables	22
Interobserver Agreement.....	24
Treatment Integrity.....	25
Screening.....	26
Research Design.....	27
Pre-Post Tests (Treatment and Control Group).....	28
Letter Sets.....	31
Pre-Intervention Training.....	31
Procedure.....	33
Study 1: The Effects of Paired Kinesthetic Movements and Embedded Pictures on Letter-Sound Correspondence.....	37
Intervention Procedure	37
Results and Discussion.....	38
Study 2: Effects of Kinesthetic Movements versus Combined Kinesthetic Movements and Embedded Pictures on Letter-Sound Correspondence	45
Intervention Procedure	45
Results and Discussion.....	46
General Discussion	51
Appendix A. Study 1 Stimuli.....	64
Appendix B. Study 2 Stimuli	65
Appendix C. IRB Form.....	66
List of References	67
Vita.....	77

Illustrations

Tables

1. Participant Demographics	20
2. Number of Intervention Sessions per Condition and Preferred Condition	43
3. Number of Intervention Sessions per Condition.....	49

Figures

1. General Protocol	28
2. The Number of Correct Responses per Mastery and Follow-Up Probes.....	39
3. The Percentage of Training Sessions at 100% Accuracy	41
4. Pre- and Post-Test	42
5. Preference Assessment.....	44
6. The Number of Correct Responses per Mastery and Follow-Up Probes.....	46
7. The Percentage of Training at 100% Accuracy	47
8. Pre- and Post-Test Data	48
9. Preference Assessment.....	49

Abstract

Reading difficulties during childhood often continue during adulthood and result in adverse effects (e.g., unemployment, poverty). A common method to teach early literacy skills is via multisensory instructional programs, which use combinations of mnemonic devices, such as visual, auditory, tactile, and kinesthetic movements. The current literature on the effects of pairing visual mnemonics and kinesthetic movements with literacy skills, either in isolation or in combination, is sparse. The purpose of Study 1 and Study 2 was to compare the efficacy, generalization, and maintenance of and preference for letter sound interventions with and without mnemonic devices. Study 1 evaluated a traditional drill, a traditional drill paired with an embedded picture, and a traditional drill paired with kinesthetic movements flashcard methods. Study 2 evaluated a traditional drill paired with kinesthetic movements and a traditional drill paired with pictures and movements (combined mnemonics) flashcard methods. Both studies were conducted with preschool children using a single-subject design; Study 1 also used a pre- and post-test group design. In general, pairing letters with movements was the most effective method to teach letter sounds, but all intervention conditions were more effective than no intervention. Both studies demonstrated that all interventions were more effective than no intervention. The number of letters correct during the maintenance probes and generalization to more complex reading skills did not significantly differ between interventions conditions in either study. Nevertheless, there was a significant difference between intervention and control conditions on the generalization effects. Effectiveness of the intervention did not correlate with participant preference.

Introduction

Alphabet principle refers to the recognition and production of letter discrimination, letter name identification, and letter-sound correspondence. The acquisition of alphabetic principles are a targeted goal for preschool and kindergarten students because it is the best predictor of future success in reading, spelling, comprehension, and vocabulary skills (Cunningham & Stonovich, 1998; Hammill, 2004; National early Literacy Panel, 2008). Children who fail to acquire these skills are more likely to be diagnosed with a specific learning disability in reading (Gallagher et al., 2000) and are at a heightened risk for behavioral and emotional problems (Lonigan et al., 1999). Additionally, reading difficulties during childhood often continue into adulthood, increasing the probability for unemployment, poverty, crime, and early mortality (Morrisroe, 2014).

Over the past two decades, schools have received increased pressure to prepare young students for academic success (Greenberg et al., 2003). Two examples include the No Child Left Behind Act of 2002 and the Common Core State Standards (National Governors Association, 2010) outlining curriculum frameworks and literacy standards for preschool and kindergarten students. Despite these standards, approaches to teaching early literacy skills vary widely due to controversy over effective instruction (Piasta & Wagner, 2010). Forty percent of students enter kindergarten one or more years behind their peers in early literacy skills (U.S. Department of Education, 2016); thus, evaluating the effects of academic preschool programs is warranted.

Many academic programs are derived from the Orton-Gillingham approach developed in the early 20th century. Based on the belief that students with severe Dyslexia need to be taught through multiple channels (i.e., visual, auditory, and kinesthetic), the Orton-Gillingham approach uses a sequential, systematic, and multisensory approach to teach reading in a 1:1 or small group

format (Ritchey & Goeke, 2006). That is, reading skills are presented in a pre-determined, bottom-up format, and students are required to demonstrate fluency of each skill prior to receiving instruction on the next skill. In addition, instructions are presented via mnemonic devices to engage multiple senses. Mnemonic devices may include visual (seeing pictures), auditory (singing songs or saying alliterations), tactile (manipulating tangible objects), and kinesthetic movements (using gestures). Mnemonic devices are designed to increase the rate of acquisition, improve discrimination between similar items, and enhance short- and long-term retention and application of the information, especially when the information is unfamiliar, complex, abstract, and extensive (Levin, 1993).

Multisensory instructional programs use combinations of the aforementioned mnemonics and have been demonstrated as an effective method to increase literacy skills. For example, Dilorenzo et al. (2011) compared the current teaching instruction to Itchy's Alphabet[®] on letter identification, letter-sound correspondence, and phonological awareness with 61 kindergarteners. Current teaching instruction included the teacher (a) reading a book that had several words beginning with the target letter, (b) presenting picture cards beginning with the target letter, and (c) reading alliteration-style poems, and the students (a) brainstorming words beginning with the target letter (b) completing worksheets or art projects related to the target letter. Itchy's Alphabet[®] instruction included the teacher (a) reading a book that had several words beginning with the target letter and (b) presenting integrated visual letter cards, and the students (a) manipulating a toy related to the target letter, (b) listening to a song, and (c) playing interactive matching and memory games. Students in the class receiving instruction with Itchy's Alphabet[®] had significantly more correct responses than students in the comparison class on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS).

Additional research has demonstrated the efficacy of multisensory-based instruction (Labat et al., 2014) and programs (e.g., Jolly Phonics[®]; Callinan & Zee, 2010) to increase literacy skills for young children. Despite the efficacy of combining mnemonics in multisensory instruction, there is little empirical evidence for the use of each mnemonic in isolation or in varying combinations. Further, there is a lack of evidence supporting student preference for multisensory instruction. Classroom instruction relies heavily on teacher preference and school, district, or state requirements (Pianta et al., 2009). However, relying on the preference of others may not always be in the best interest of the student, and evidence should support the selection of instruction (Hanley et al., 1997; Hanley et al, 2005). Because Orton-Gillingham-based programs are commonly implemented in the United States (Ritchy & Goeke, 2006), it is important to identify the efficacy and preference of each mnemonic device at the individual level to inform classwide, small group, and individual instruction. Selection of the instruction should be balanced between efficacy and preference. First, classwide instruction can be selected based on the most effective and preferred instructional condition of the majority of the students. If two instructional conditions are equally preferred and do not hinder acquisition, instructional procedures can be alternated throughout the day. Second, the class can be divided into small groups based on either the most effective or preferred instructional condition, and instructional procedures can vary across groups. Third, the more effective and preferred instructional condition can inform individual interventions with students if warranted.

Previous research has evaluated multisensory instruction with large cohorts of students and has focused on pre- and post-test data to identify significant differences in literacy scores between groups (e.g., Dilorenzo et al., 2011; Callinan & Zee, 2010). Evaluating mnemonics with group designs are appropriate because they permit the use of inferential statistics to yield

independent effects sizes, demonstrating a larger effect of one literacy intervention on alphabetic principle compared to standard teaching conditions (Dilorenzo et al.) or of one intervention on phonics compared to another intervention (Callinan & Zee). In addition, the extent to which the results of a study can be generalized to other students and settings (external validity) increases when interventions result in significant increases across classrooms and schools. However, learning is an individual-level phenomenon. That is, letter-sound correspondence (and all reading skills) are not acquired at the group or class level but at the individual student level. Averaging data across participants prevents identification of learning processes and individual differences that may be particularly important for designing interventions that work for students most at risk for failing to acquire reading skills (Whitley & Kite, 2013). Evaluating mnemonics using single-subject designs has potential benefits. First, single-subject designs reduce threats to internal validity, such as history and group assignment, and external validity, such as interaction effects of selection biases and experimental treatment (Petursdottir & Carr, 2018). Second, they permit the identification of functional relationships between the independent and dependent variable, such as the effects of mnemonics on literacy skills. Third, they allow analyses within and across participants to identify the mechanisms influencing or hindering literacy acquisition (Kazdin, 2011).

A concurrent-chains procedure represents one single-subject research design appropriate to measure preference (Baum & Rachlin, 1969). During concurrent-chains procedures, two or more stimuli are simultaneously presented and are each correlated with a different consequence (initial link). Selection of one stimulus produces the respective consequence (terminal link), and more frequent selection of one stimulus allows preference among response options to be inferred. For example, Hanley et al. (2005) presented participants with three different colored switches

correlated with three different treatments effective at reducing problem behavior: the blue switch with functional communication training (FCT), the red switch with FCT plus punishment, and the white switch with punishment only. The experimenter implemented the treatment contingencies paired with the pressed switch for 2 min. Both participants allocated the majority of their responses to the red switch, suggesting a preference for FCT plus punishment. Concurrent-chains procedures have been demonstrated to reliably infer preference with preschool children (Fenerty & Tiger, 2010; Tiger et al., 2006a; Tiger et al., 2006b; Tiger et al., 2008) and individuals with limited vocal repertoires (Hanley et al., 1999). Because implementing instruction that is both effective and preferred for the learner may increase their engagement and long-term academic achievement (Harackiewicz et al., 2016), an evaluation of preference for mnemonic devices is warranted.

Visual Mnemonics

The method of pairing visual mnemonics with literacy intervention was the most common mnemonic technique implemented in practice and qualitatively researched in the 1970's (Higbee, 1979). Samuels (1967) conducted the first empirical investigation of visual mnemonics and evaluated the degree to which pictures, when presented with words, interfere with sight word acquisition. In the first experiment, Samuels assigned 30 kindergartners to one of three conditions: no-picture, black and white pictures, and colored pictures. In all groups, experimenters alternated training and test trials, and results demonstrated a significant difference between groups during test trials. Participants in the no-picture group responded with significantly more correct responses than participants in the picture groups. In the second experiment, Samuels evaluated whether pre-intervention reading ability affected sight word acquisition when training via paired mnemonics. Based on pre-assessment scores, he divided 26

first graders into an above- and below-average reading group. Participants in each group were assigned to either a no-picture or a black and white picture group, and procedures were identical to the first experiment. Participants in the below-average group read significantly more correct words during the no-picture condition compared to picture group; however, no significant differences between groups was found for participants in the above-average group. Samuels suggested that the degree to which pictures interfere with acquisition may depend on the abilities of the individual and that a further investigation into student motivation, attitude towards reading, and attention processes may further advance this research.

Samuels (1967) prompted researchers to investigate visual mnemonics on sight word acquisition in three domains: instructional procedures, type of visual mnemonic, and contributing mechanisms. First, researchers have demonstrated a null effect of instructional procedures on sight word acquisition for word-picture presentations (Didden et al., 2006; Didden et al., 2000; Harzam et al., 1976; Lang & Solman, 1979; Richardson et al., 2017). For example, presenting pictures after the sight word (Didden et al, 2000; Lang & Solman) or gradually fading the picture (Didden et al., 2006; Richardson et al.) does not affect the outcome: words presented alone result in more correct responses than words paired with pictures. Second, researchers have demonstrated variations that may reduce pictorial interference with sight word acquisition. These include increasing the size of the word (Lang & Solman; Singh & Solman, 1990) and increasing the saliency of the picture-word interaction (e.g., picture of a clock in the letter 'o'; Lippman & Shanahan, 1973). Third, Dittlinger and Lerman (2011) demonstrated that the detrimental effects of pictures on sight word acquisition may be due to overshadowing. That is, the correct response is controlled by the more salient stimulus (picture) and may account for the decrease of correct responses in word-only test trials compared to picture-word training trials. Although research has

demonstrated the interference of pictures on the acquisition of sight words, researchers have demonstrated a different effect of pictures on the acquisition of prerequisite reading skills.

Marsh et al. (1974) conducted the first study evaluating visual mnemonics on an essential prerequisite reading skill: letter-sound correspondence. Experimenters randomly assigned 40 preschool-aged children to a picture-picture, picture-letter, letter-picture, or letter-letter condition. In all conditions, the experimenter presented one stimulus (i.e., picture or letter card) and instructed the participant to respond with the matched stimulus (i.e., picture name or letter sound). For example, in the letter-picture condition, the experimenter showed the participant a picture and the correct response was the letter sound. Participants in the picture-picture group emitted the highest number of correct responses, followed by letter-picture, letter-letter, and picture-letter group. Thus, groups in which the letter sound was the correct response performed the worst. The authors attribute the difference in picture naming and letter-sound correspondence to a lack discrimination between responses rather than the presented stimuli. Further, the authors concluded that children have greater difficulty with responding with letter sounds than picture names. However, one reason for their finding may be due to the lack of correspondence between the pictures and the paired letter sounds (e.g., pairing the picture of a flower with 'f' despite beginning with a blended sound).

Marsh and Desberg (1978) extended the research by Marsh et al. (1974) by evaluating how the relationship between visual mnemonics and the response affect letter sound-correspondence. Across two experiments, experimenters randomly assigned 100 preschool- and kindergarten-aged children to one of the following conditions: letter-sound picture, letter-name picture, irrelevant picture, action picture, choice, and no-picture. In the picture conditions, the picture either began with the letter sound (letter-sound picture; e.g., pumpkin for /p/), started

with the letter (letter-name picture; e.g., glasses for /g/), did not begin with the letter name or sound (irrelevant picture), or depicted an action that produced the sound (action picture; e.g., a boy blowing out a candle, saying /p/). In the choice condition, the experimenters presented participants with two response options: letter-sound picture and action picture. Modeling and training were conducted with picture cards, and training sessions were immediately followed by test trials, conducted with letter-only cards to assess correct letter sounds in the absence of visual mnemonics. The authors did not report data on the allocation of responses for the choice group but reported consistent findings across experiments: all picture groups performed better than no-picture groups during training, but there was no difference between picture and no-picture groups during test trials. The authors attributed this finding to young learners not being cognitively mature enough to benefit from mnemonics. Alternatively, the mnemonics may have either been inadequate because they did not effectively link the visual stimulus to the response or were effective for some students but those effects were washed out in the group average.

Coleman and Morris (1978) evaluated this interpretation by developing integrated mnemonics that link the visual stimulus to the participant's response. Experimenters evaluated no-elaborator, stimulus elaborating, response elaborating, overlapping elaborating, and interactive elaborating mnemonics on letter-sound correspondence with 16 preschool-aged participants. Experimenters selected pictures based on the responses of 40 undergraduate students who were instructed to close their eyes and visualize images to printed letters (stimulus elaborating; e.g., the letter 'f' drawn like a candy cane) or to letter sounds (response elaborating; e.g., picture of a ticking clock for the letter 't'). In the overlapping and interactive elaborating conditions, the stimulus and response image were compounded (e.g., picture of a snake shaped as the letter 's') or intertwined (e.g., picture of a camel with humps shaped like an 'm' eating ice

cream saying “m-m-m”) into one picture. Similar to Marsh and Desberg (1978), modeling and training trials were conducted with picture cards and were immediately followed by test trials conducted with letter-only cards. The results differed from the previous experiments: participants responded with more correct responses when presented with letters trained via paired pictures compared to letters trained without picture during test trials. Specifically, the overlapping and interactive elaborator letters resulted in more correct responses than the no-elaborator, stimulus elaborator, and response elaborator letters during test trials. This study has two noteworthy implications. First, it is the first study to demonstrate the efficacy of mnemonics, compared to no mnemonics, for early literacy skills. Second, the results provide information on the type of mnemonic that may be the most effective to increase letter-sound correspondence. That is, mnemonics that use a picture beginning with the letter sound and is integrated with the letter may be more efficacious than presenting an unrelated picture next to the letter for letter sound acquisition.

Ehri et al., (1984) further expanded the literature on elaborated mnemonics. They hypothesized that the lack of efficacy for stimulus elaborator pictures, as demonstrated in Coleman and Morris (1978), was due to the lack of letter shape saliency in the picture. Thus, the experimenters evaluated the effects of embedded pictures, in which the letter was a salient visual feature in the (e.g., ‘w’ forming part of the wings of an insect). The experimenters randomly assigned 50 preschool-aged participants to an embedded picture group, a disassociated picture group in which the letter did not form part of the picture (e.g., picture of a plane with wings for ‘w’), and a no-picture control group (Experiment 2 only). Each group experienced a series of training sessions followed by test trials. Participants in the embedded group recalled significantly more letter sounds than participants in the disassociated and no-picture groups in both

experiments. In addition, participants in the no-picture group recalled more letters than participants in the disassociated group in Experiment 2. These results suggest that visual mnemonics may be effective when the picture is embedded within the letter shape rather than arbitrarily presented with the letter. Further, the detrimental effects of disassociated pictures on letter sound recall may have been due to overshadowing in which the picture was a more salient stimulus than the letter.

The procedures implemented by Ehri et al. (1984) included segmentation pre-training and phoneme segmentation training prior to the intervention. The objective of training was to teach participants how to segment the first sounds of picture names, and training consisted of frequent picture and letter drawings and lengthy explanations of picture-letter relations. Fulk et al. (1997) attempted to identify if phoneme segmentation pre-training was an essential component for the efficacy of embedded visual mnemonics on letter sound acquisition. In a multiple-baseline design across participants, experimenters compared the number of correct letter names and sounds during baseline (no-picture training) to treatment (embedded picture training). During both phases, test trials were conducted with the letter-only card. The experimenters demonstrated an increase in the number of correct responses for 3 of 3 participants during test trials compared to baseline. Although experimenters demonstrated an increase in acquisition without extensive segmentation training, they did not include a comparison group to determine if including components would have resulted in fewer training sessions. Nonetheless, Fulk et al. represents the first single-subject study of visual mnemonics. Argramonte and Belifore (2005) and Sener and Belifore (2005) replicated the procedures of Fulk et al. and obtained similar results: letter-name identification and letter-sound correspondence increased during visual mnemonic training compared to baseline.

Hetzroni and Shavit (2002) extended the research on visual mnemonics by comparing a dissociated picture and a no-picture group on Hebrew-letter identification. Their results differ from previous research in one major way: the dissociated picture group responded with more correct responses than the no-picture group. These findings differ from Ehri et al. (1984) and Marsh and Desberg (1978) who found that participants in the no-picture group responded with more correct sounds than those in the dissociated group. Shmidman and Ehri (2010) hypothesized that this inconsistency was due to the vast differences between Hebrew and English letters (i.e., Hebrew letters are generally square with few diagonals and differ only in curves or dots) and evaluated if dissociated pictures would also result in more correct responses than embedded pictures for Hebrew-letter identification. Results from Shmidman and Ehri were consistent with previous research: participants in the embedded group performed significantly better than participants in the dissociated group. However, the experimenters did not include a no-picture group. Based on variability in the findings of previous research (Ehri et al.; Coleman & Morris, 1978), the differential effects of no picture, dissociated picture, and embedded picture training on Hebrew-letter identification is unknown.

One important feature of Shmidman and Ehri (2010) is that participants had no previous exposure to Hebrew language. Their findings highlight the efficacy of using visual mnemonics to teach prerequisite reading skills for a secondary language. Manalo et al. (2013) further investigated how instructional procedures may affect visual mnemonics on English letter-sound correspondence for participants whose primary language was Japanese. The experimenters randomly assigned 140 sixth grade students to one of four conditions: embedded pictures with and without sound contrast instruction and no-pictures with and without sound contrast instruction. For conditions including sound contrast instruction, experimenters described how the

presented English letter sound differed from common sounds used in Japanese words with specific examples (e.g., /p/ is different from “pu” and “pa”). Participants in the embedded picture plus sound contrast instruction had the highest number of correct responses, suggesting that both variables are important when teaching a secondary language.

Research has also examined how instructional fading procedures for embedded pictures affect letter-sound correspondence. For example, Hoogeveen et al. (1989) and de Graaff et al. (2007) evaluated stimulus-fading procedures in which the picture gradually faded contingent on participants correct responding. Hoogeveen et al. evaluated a four-step fading procedure in a multiple baseline design across four participants. Results demonstrated an increase in correct letter sounds and generalization to reading novel words and sentences. In another study, de Graff et al. evaluated three conditions with 39 participants: embedded picture plus fading, embedded picture, and no-picture. Using a computer-assisted program, each condition was presented twice totaling six sessions per participant. Participants responded with the highest number of correct responses to the letters presented in the fading procedure followed by the embedded picture and no-picture condition. In addition, de Graff et al. examined the results across participants with high and low first-sound isolation scores, based on pre-assessment scores. No significant difference was found between groups in the fading condition; however, the high first-sound isolation group performed significantly better than the low first-sound isolation group in the embedded and no-picture conditions. These results suggest that fading procedures are effective for transferring the controlling response from the picture to the letter, especially for children with poor literacy skills.

Select researchers have extended their experiments to include retention sessions post-training. Argramonte and Belifore (2005) and Fulk et al. (1997) conducted two retention sessions

(1- and 3-weeks and 2- and 4-weeks, respectively), and Sener and Belifore (2005) and de Graff et al. (2007) conducted one retention session (1 week and 4 weeks, respectively) after training sessions ended. The number of correct responses remained stable 1, 2, and 3 weeks post training sessions (Argramonte & Belifore; Fulk et al.; Sener & Belifore) and began to decrease at 4 weeks (de Graff et al.; Fulk et al.). Results also demonstrated a higher number of correct responses for letters trained via faded embedded pictures compared to embedded and no pictures (de Graff et al.).

In addition to the acquisition and maintenance of letter-sound correspondence, the generalization of letter sounds to novel or more complex skills is essential for reading. Ehri et al. (1984) and Shmidman and Ehri (2010) evaluated how letter sound acquisition via disassociated and embedded pictures affects generalization with a series of pre- and post-tests. Tests included letter-sound recall, picture recall, and letter writing (Ehri et al.; Shmidman & Ehri), picture drawing and phonemic segmentation (Ehri et al.), and letter-sound identification and matching letter sounds to pictures (Shmidman & Ehri). Overall, the embedded picture group performed significantly better than the disassociated group on letter-sound recall, picture recall, and letter writing (Ehri et al.; Shmidman & Ehri), and letter-sound identification and matching letter sounds to pictures (Shmidman & Ehri). In addition, no significant differences were found between groups on picture drawing and phonemic segmentation (Ehri et al.). Shmidman and Ehri also conducted a series of transfer tasks to assess more complex literacy tasks. These included spelling and reading words, word acquisition, and matching written and spoken words. The embedded picture group performed significantly better on spelling and reading words and required significantly fewer trials for word acquisition compared to the disassociated picture group. Last, no significant differences were found between groups on matching written and

spoken words (Shmidman & Ehri). These results suggest that pairing embedded pictures with letters results in an immediate benefit (i.e., letter sound acquisition) and long-term advantages (i.e., generalization to reading skills).

Kinesthetic Movements

Kinesthetic movements in the classroom range from physical activity of moderate intensity, such as using an invisible jump rope during a math lesson (Take 10![®]; see Kibbe et al., 2011 for a review), to small motor movements, such as finger tapping during reading (Itchy's Alphabet[®]; Dilorenzo et al., 2011). Research has demonstrated a strong correlation between physical activity of moderate intensity and on-task behavior and academic achievement, supported by the Physical Activity Across the Curriculum project (see Donnelly & Lambourne, 2011 for a review). Further, the Center for Disease Control and Prevention funds state education agencies to implement the Whole School Whole Community Whole Child framework to address health in schools (Lewallen et al., 2015). One component of the model emphasizes a comprehensive physical activity curriculum to aid in students' academic achievement.

There is much less support for and research on the use of small motor kinesthetic movements in the classroom, despite being implemented in multisensory programs (e.g., Zoo-Phonics[®]; Jolly Phonics[®]). Thorpe and Borden (1985) conducted the first experiments comparing traditional teaching to paired kinesthetic movement teaching on on-task behavior and word reading accuracy. The experimenters compared both conditions with and without praise during 10-min sessions for groups of first grade students. During the traditional condition, participants repeated, segmented, and repeated the presented word. During the paired movement condition, participants repeated, traced with their finger, underlined, repeated, traced with their pencil, and underlined the presented word. On-task behavior was measured during sessions, and word

reading accuracy was measured via individual tests at the end of sessions. When the experimenters did not provide praise, the paired movement condition resulted in a higher percentage of on-task behavior and word reading accuracy than the traditional condition. However, when the experimenters provided praise, the traditional condition resulted in a higher percentage of word reading accuracy compared to the paired movement condition. The authors suggested that, in the absence of praise, paired kinesthetic movements should be implemented to increase literacy skills, but that praise is the important component across teaching conditions.

Lozy et al. (2020) found opposing results from Thorpe and Borden (1985): pairing kinesthetic movements with stimuli was more effective than the traditional approach for literacy skill acquisition. Lozy et al. compared a traditional drill flashcard condition, a paired kinesthetic movement flashcard condition, and a control condition on letter-sound correspondence with preschoolers in a multielement design. During the traditional drill condition, the experimenter only presented the letter card, and during the paired movement condition, the experimenter presented the letter card and emitted the corresponding Jolly Phonics® hand gesture. The experimenters defined mastery criterion as the participant emitting the correct response to all letters in the set in three probes across different days. Consistent findings emerged for the control condition: letter-sounds did not increase in 11 of 11 evaluations. However, results varied across intervention conditions: participants mastered the paired movement condition first in 6 evaluations, and participants mastered both intervention conditions with little differentiation between the number of training sessions in 5 evaluations. Experimenters conducted a post-hoc analysis and examined the correlation between engaging in the movement and the number of training sessions experienced and found a strong negative correlation between the percentage of engaging in the movement and the number of sessions to set mastery. These results suggest that

pairing kinesthetic movements with letter sounds may be more effective for increasing letter sound correspondence than a traditional drill format.

Lozy et al. (2020) also compared the retention of letter-sound correspondence for stimuli presented in the traditional, paired movement, and control conditions. Experimenters conducted follow-up tests at 1-, 3-, and 7-weeks post research sessions, and results demonstrated a higher number of correct responses for the paired kinesthetic movement condition across weeks. These results are consistent with Glenberg et al. (2004) who conducted reading retention tests 2 min after each session. Glenberg et al. compared retention across two groups; the paired movement group, in which participants were instructed to manipulate imaginary toys as they read the story, and the reread group, in which participants were instructed to read the story a second time, either silently or out loud. Participants in the paired kinesthetic movements group recalled a greater proportion of reading than participants in the control group.

Despite the concordance of results demonstrating the efficacy of kinesthetic movements on the acquisition and retention of literacy skills, there are several potential confounding variables that may have accounted for the findings. First, Thorpe and Borden (1985) equated the conditions on duration rather than word exposure. Because the experimenters immediately repeated each word during the kinesthetic movement condition, the number of presentations per word may have differed and accounted for the differentiation between conditions. Second, Glenberg et al. (2004) implemented a paired tactile mnemonics condition (i.e., manipulating toy objects while reading) prior to the paired kinesthetic movements condition (i.e., manipulating imaginary toy objects while reading). Thus, the history of pairing tactile mnemonics with story reading may have influenced that effect of kinesthetic movements on reading retention.

Research is also warranted on the generalization of and preference for pairing kinesthetic movements with literacy skills. In the only study evaluating the generalization of paired movements, Campbell et al. (2008) examined nonsense word reading in a multiple baseline design across participants. During baseline, experimenters implemented the participants' current reading program, and during treatment, experimenters paired kinesthetic movements with the current reading program (e.g., drawing the letter in the air, tapping out words). At the end of each session, participants read a list of nonsense words, and results demonstrated an increase in reading accuracy during treatment compared to baseline for 6 of 6 participants. Because baseline and treatment consisted of a combined letter sound, segmenting, word reading, and connected text instruction, it is unknown if pairing kinesthetic movements with one skill or the combination of skills accounted for the increase in nonsense word reading.

In the only study evaluating the preference for paired kinesthetic movements, Lozy et al. (2020) conducted a repeated concurrent-chains procedure with two initial link options: traditional drill and paired movements. One of 6 participants selected the traditional drill condition during 100% of trials, and 5 of 6 participants selected the paired movements condition during an average of 61% (range, 23% - 96%) of trials, respectively. Thus, the authors suggest that the paired movement may have been the more preferred condition.

Gaps in the Current Research

The current literature on visual mnemonics is more extensive than the research on kinesthetic movements; however, both areas warrant more investigation. First, there is a lack of research evaluating the use of embedded pictures and kinesthetic movements at the individual level. Given that acquisition of reading skills is an individual-level phenomenon, this is a particularly important area for additional research. Second, there is a lack of evidence supporting

learner preference for paired mnemonic instruction. Third, studies evaluating the maintenance of literacy skills taught via paired mnemonics are scarce, and those studies typically include few sessions with a short delay from training to maintenance. Fourth, more research is warranted on the generalization of literacy skills when using paired mnemonic instruction. Letter-sound correspondence is an essential pre-requisite reading skill and thus it is important to evaluate how letter sound instruction may affect more complex reading skills.

Purpose

The general purpose of these studies was to extend the literature on the multisensory teaching approach by evaluating the effects of single and combined mnemonics in a single-subject design on the acquisition of early literacy skills. Specifically, the first purpose was to evaluate the effects of pairing mnemonic devices with letters on letter-sound correspondence with preschool students. Study 1 compared an embedded picture traditional drill intervention, a kinesthetic movement traditional drill intervention, a traditional drill intervention (no mnemonic), and a control (no intervention) condition. Study 2 compared a single mnemonic intervention condition (kinesthetic movements, based on the results of Study 1), a combined mnemonic traditional drill intervention (embedded pictures plus kinesthetic movements) and a control condition. The second purpose was to evaluate the generalization of letter-sound correspondence, acquired via each intervention, on other pre-reading tasks using a pretest-posttest design. The third purpose was to evaluate participant preference for each intervention condition using a concurrent chains procedure (Hanley et al., 1997). The fourth purpose was to evaluate the maintenance of letter-sound correspondence post formal research training sessions.

General Method

Participants, Consent, Assent, and Setting

Experimenters recruited students from a public preschool and kindergarten center in southeast Louisiana and from word-of-mouth. Participants were selected if their primary teacher and caregiver (if recruited from the school) or if their caregiver (if recruited via word-of-mouth) provided consent for participation and the child assented to participate. The experimenter reviewed the study procedures with a group of four teachers selected by the principle at the preschool and kindergarten center. The teachers distributed consent forms to all students in their classroom, and parents had the opportunity to call the experimenters for any questions. The experimenter obtained verbal assent from the participant each day research sessions were conducted. The experimenter asked the participant if they wanted to work for games on the iPad[®]. If the participant refused, research sessions were not conducted that day.

Children who (a) responded with fewer than 12 correct responses during the pre-assessment and (b) responded with the correct imitation for at least 12 letters during the imitation assessment were included in this study. If recruited from the school, children who were regularly in class were selected to be in the treatment group.

Teachers and caregivers referred 24 children for participation in Study 1, and 14 children were eligible. Seven participants were assigned to the intervention group and seven participants were assigned to the control group. The experimenters intended for the control group to be a wait-list control group as opposed to a no-treatment control group, however, due to the COVID-19 pandemic, the school year ended prior to the scheduled date and research sessions were discontinued. Teachers and caregivers referred 3 children for participation in Study 2. All

children were eligible and were assigned to the intervention group. Participant demographics and group assignment are listed in Table 1.

Table 1. Participant Demographics

Participant	Study	Group	Age	Race/Ethnicity	Diagnoses	Services	Avg. Correct Responses: Pre-Assessment
Nico	1	Tx	5	Caucasian/Hispanic	-	Reading	0
Dale*	1	Tx	4	Caucasian/Hispanic	DD	Speech	0
Amara	1	Ctrl	4	Biracial	-	-	1
Sloan	1	Tx	4	African American	-	-	1
Riley	1	Ctrl	4	Biracial	-	Speech	2.5
Calla	1	Tx	4	Caucasian	ASD	OT	2
Nila	1	Ctrl	4	Caucasian	-	-	6.5
Milo	1	Tx	4	Caucasian/Hispanic	-	-	3.5
Kora	1	Ctrl	4	Caucasian	-	-	6.5
Cain*	1	Tx	4	Caucasian/Hispanic	-	-	5
Wyatt	1	Ctrl	5	African American	ADHD/ODD	-	8
Quade	1	Tx	4	African American	-	-	9.5
Alice	1	Ctrl	4	Caucasian	-	-	10
Cassie	1	Ctrl	4	Caucasian	-	-	12.5
Amara	2	Tx	4	Biracial	-	-	4
Ginny	2	Tx	3	Caucasian	-	-	0
Lyle	2	Tx	5	Caucasian	-	-	3

Note. Ctrl = Control; Tx = Treatment; ADHD = Attention/Deficit Hyperactivity Disorder; ASD = Autism Spectrum Disorder; DD = Developmental Delay; OT = Occupational Therapy. * = English Language Learner

We conducted sessions with Nico (Study 1), Ginny (Study 2), and Lyle (Study 2) at the dining room table in their homes. For all other participants, we conducted sessions at a table in an empty classroom or the cafeteria (outside of mealtimes) in the participants' preschool. We conducted two to five sessions per day, 3 to 4 days per week in a single session block with 2-min breaks between sessions.

Materials

Intervention session materials included stimulus cards, discriminative stimulus session cards, and reinforcers. Letter cards were printed in black ink with Arial font on 7.62 cm x 7.62 cm white laminated cards. Letters were printed on the front and back of the traditional drill cards. On the front of the embedded picture cards, the corresponding picture from the Zoo-Phonics[®] learning program was printed. On the back of the embedded picture and kinesthetic movement cards, the modeling and error correction procedure was printed. Additionally, on the back of the kinesthetic movement cards, a description of the physical movement was printed in Arial font, size 24, in black ink.

Five discriminative stimulus session cards, denoting each condition, were printed on 5.08 cm x 15.24 cm white laminated cards. Each intervention card displayed the letters a, b, c, and d, in black ink with Ariel font, and the control card was blank. The traditional drill session card displayed letters printed in size 140 font. The kinesthetic movement session cards displayed letters printed in size 100 font and Zoo-Phonics[®] animated figures engaging in each movement. The embedded picture session cards displayed the Zoo-Phonics[®] animal cards in size 140 font. The combined mnemonics session card displayed the Zoo-Phonics[®] animal cards in size 100 font and Zoo-Phonics[®] animated figures engaging in each movement.

Reinforcers included access to games on an iPad[®] for all participants and edibles for two participants in Study 1 (Dale and Milo, beginning at session 54 and 34, respectively) and for one participant in Study 2 (Ginny). When the iPad[®] was the selected reinforcer, the experimenter used a token board with eight Velcro[®] hook-and-loop dots to signal the duration of access to the iPad[®] at the end of session; each token represented 15 s of access. The iPad[®] had a wide variety of applications with children's games; participants could select any game or switch games

throughout their access period. When the participant selected edibles as the reinforcer, the experimenter presented a menu of three edibles options for the participant to select. During sessions, the experimenter implemented an accumulated reinforcer arrangement, in which they placed a dime-sized piece of food on a plate per each correct response and participants were given access to the plate at the end of session. An accumulated reinforcer arrangement was selected because it is just as effective as a distributed arrangement and often more preferred (DeLeon et al., 2014; Ward-Horner et al., 2017).

Pre- and posttests included test-specific materials. Stimulus cards for Test 1: Stimulus Generalization were letter cards printed in black ink with STCaiyun font on 7.62 cm x 7.62 cm white cards. Stimulus cards for Test 2: Letter Sound Identification and Test 5: Spelling Nonsense Words were identical to the intervention stimuli (i.e., letter cards printed in black ink). Letter cards were also used for Test 3: Match Pictures to Letters in addition to colored pictures printed on 7.62 cm x 7.62 cm cards. Stimuli for Test 4: Reading Nonsense Words included nonsense words printed in black ink in Ariel font, size 50, on 2.54 cm x 8.89 cm white cards.

Dependent Variables

Data collectors used paper and pencil data sheets to record the participants' responses during all phases of Study 1 and Study 2. During the pre-assessment, observers scored a correct response if the participant emitted the letter sound corresponding with the letter card within 5 s of the experimenter presenting the discriminative stimulus (stimulus card and instruction). For vowels, observers scored a correct response if the participant emitted the short vowel sound. Observers scored an incorrect response if the participant failed to respond or responded with anything other than the correct response. No response was recorded if the participant emitted the long vowel sound or the letter name, and the experimenter said, "That's one sound the letter

makes, but what's another sound," or "That's the name of the letter, but what sound does it make?", and scored the participant's subsequent response. During the imitation assessment, correct imitation was scored if the participant's response matched the sound emitted by the experimenter. Experimenters scored incorrect imitation if the participant did not respond or if the participant's response was anything other than the correct imitation.

During the pre-posttest phase, correct responses were specific to the test. For Test 1 and Test 4, a correct response was scored if the participant emitted the letter sound corresponding with the card within 5 s of the experimenter presenting the stimulus card and instruction. For Test 2, a correct response was scored if the participant selected the letter corresponding to the letter sound emitted by the experimenter. For Test 3, a correct response was scored if the participant selected the letter corresponding with the beginning letter sound of the picture card. For Test 5, a correct response was scored if the participant selected the letter corresponding with the beginning and ending letter sound of the nonsense word within 5 s of the experimenter emitting the nonsense word.

The primary dependent variable in the intervention comparison was the number and percentage of correct responses during mastery probes and training sessions, respectively. We calculated the percentage of correct responses by dividing the number of correct responses by the total number of trials and multiplying that number by 100. The secondary dependent variable was the percentage of correct movements emitted during training sessions for the kinesthetic movement conditions. Observers scored a correct movement if the participant emitted the physical movement that corresponded with the stimulus card within 5 s of the experimenter presenting the discriminative stimulus. An incorrect movement was scored if the participant responded with anything other than the correct movement, and no movement was scored if the

participant did not respond with any physical movement. The percentage of correct movements for each stimulus was calculated by dividing the number of correct responses by the total number of trials and multiplying that number by 100.

During the preference assessment, the main dependent variable was the selection response, defined as the participant picking up or pointing to the session card. The selection response was recorded as the condition selected.

Interobserver Agreement

A second observer independently collected in-vivo data on the card presentation and participant response and movement (when applicable) for all participants. An agreement was defined as the experimenter and observer recording the same card, participant response, and movement (when applicable). We calculated interobserver agreement (IOA) using the trial-by-trial exact agreement method. That is, we divided the number of agreements by the total number of trials per session, multiplied that number by 100 to obtain a percentage, and then averaged the percentages across sessions for each condition. During preference assessments, an agreement was defined as both observers recording the same selection response. IOA was calculated using the exact agreement method across sessions. That is, an agreement was coded as 1 and the total number of agreements was divided by the total number of selection responses and multiplied by 100 to obtain a percentage.

Study 1. During the pre- and post-tests, across all tests, IOA was assessed during an average of 35% of sessions and mean IOA was 100%. During the first evaluation, across all intervention conditions, IOA was assessed during 31% of sessions for Nico, 78% of sessions for Quade, 25% of sessions for Dale, 52% of sessions for Cain, 54% of sessions for Sloan, 48% of sessions for Calla, and 73% of sessions for Marvin. Mean IOA was 100% for Nico, Quade, Dale,

Cain, Sloan, and Milo; and 96% for Calla (range, 90 to 100%). During the second evaluation, across all conditions, IOA data were collected during 34% of conditions for Nico and was 100%. The experimenters assessed IOA during an average of 50% of preference assessment sessions, and IOA was 100%

Study 2. During the pre- and post-tests, across all tests, IOA was assessed during an average of 23% of sessions, and IOA was 100%. During the first evaluation, across all intervention conditions, IOA was assessed during 53% of sessions for Amara, 26% of sessions for Ginny, and 35% of sessions for Lyle. Mean IOA was 100% for all participants. During the second evaluation, across all conditions, IOA was assessed during 30% of conditions for Ginny and was 100%. The experimenters assessed IOA during an average of 40% of preference assessment sessions, and IOA was 100%.

Treatment Integrity

We assessed treatment integrity by determining if the experimenter delivered the programmed consequence on each trial. We calculated treatment integrity by dividing the number of correct trials (i.e., correct consequence) by the total number of trials per session and multiplying that number by 100 to obtain a percentage. A second observer collected data on trial consequences to assess IOA for treatment integrity. We defined an agreement as the experimenter and observer recording the same consequence. We calculated IOA using the trial-by-trial exact agreement method.

Study 1. During the first and second intervention and preference assessment, the experimenters assessed treatment integrity during 100% of EP, KM, TD, and probe only sessions. The experimenter delivered the correct consequence during 99% (range: 98% to 100%) of all sessions for all participants. During the first and second evaluation, the experimenters

assessed IOA for treatment integrity during 48% of EP sessions, 53% of KM sessions, 45% of TD sessions, and 41% of probe only sessions. Treatment integrity IOA was 100% across all sessions for all participants.

Study 2. During the first and second intervention and preference assessment, the experimenters assessed treatment integrity during 100% of EP, KM, TD, and probe only sessions. The experimenter delivered the correct consequence during 100% of sessions for all participants. During the first and second evaluations, the experimenters assessed IOA for treatment integrity during 34% of CM sessions, 35% of KM sessions, and 36% of probe only sessions. Treatment integrity IOA was 100% across all sessions for all participants.

Screening

Pre-assessment. An experimenter conducted a pre-assessment to identify letters to be included in the intervention comparison. The experimenter divided 26 letters into three sets of eight, nine, and nine letters and conducted two sessions for each set across different days. During sessions, the experimenter presented each card once and asked, “What sound does this letter make?” The experimenter delivered praise and a reinforcer (i.e., token for time on the iPad[®] or an edible) following correct responses and no consequence for incorrect responses. If the participant responded with the letter name or long vowel sound, the experimenter did not record the response and said, “That’s the name of the letter, but what sound does it make?” or “That’s one sound the letter makes, but what’s another sound?” The participant’s subsequent response was recorded and contingencies for emitting a correct and incorrect response was identical to the initial trial. Additionally, the experimenter made descriptive comments (e.g., “You’re working really hard!”) after approximately every third incorrect response.

In Study 1 participants were paired based on the number of correct responses and entered either the treatment or control group. No control group participant was paired with Nico, because research sessions were conducted during the summer prior to recruiting all other participants. Additionally, no treatment group participant was paired with Cassie due to an uneven number of treatment and control participants during the school year. In Study 2, all participants were assigned to a treatment group. The average number of correct responses across pre-assessments are listed in last column in Table 1.

Echoic assessment. An experimenter conducted an echoic assessment to determine if participants could correctly imitate the letter sound for letters to which they responded incorrectly during the pre-assessment. The experimenter equally divided the incorrect stimuli into three equal sets and conducted one session for each set. The number of stimuli per set depended on the number of incorrect responses but ranged from four to seven. During sessions, the experimenter did not show the card and told the participant to repeat after them (e.g., “Say /p/”). Correct imitation resulted in praise, and the experimenter provided no consequence for incorrect imitation.

Research Design

We used a single-subject research design in which each participant served as their own control to evaluate the effects of mnemonics on letter-sound correspondence. Specifically, we used a multielement design in which the intervention and control conditions were rapidly alternated in a 6:1 or 4:1 ratio, depending on the number of intervention conditions for each participant. This design allowed for a comparison of the intervention effects while controlling for the effects of repeated measurement (Sidman, 1960). We also used a pretest-posttest design to evaluate complex literacy skills that were not directly targeted during the intervention. This

design allowed us to evaluate the extent to which each condition resulted in response and stimulus generalization. In Study 1, we used a group design, in which we assigned participants to either an intervention or control group to evaluate the intervention effects on selected literacy skills assessed in the pre- and post-tests compared to no intervention. In addition, we used a concurrent-chains procedure to assess participant preference for intervention conditions. During the concurrent chains assessment, the participant selected the intervention condition implemented each day from among three or four intervention options and one control condition option. The intervention protocol for each participants is denoted in Figure 1.

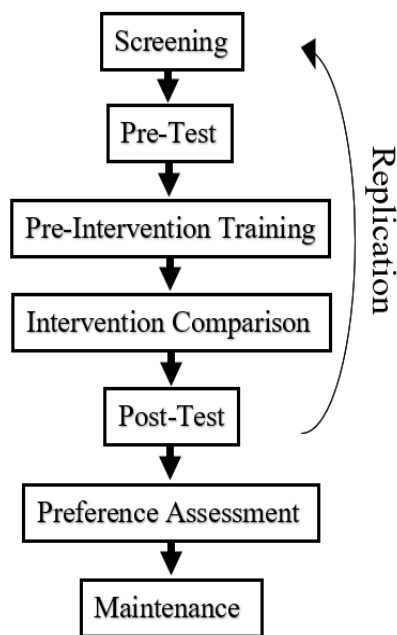


Figure 1. General Protocol

Pre-Post Tests (Treatment and Control Group)

The experimenter selected either 16 (4 vowels and 12 consonants) or 12 letters (3 vowels and 9 consonants) depending on the number of letters to which the participant responded incorrectly to during the pre-assessment and correctly imitated during the imitation assessment.

The experimenter divided the letters into three or four groups of four letters, each containing one vowel and three consonants. The experimenter conducted each test once for all letter sets, and tests were similar to those implemented by Shmidman and Ehri (2010). Tests included stimulus generalization (Test 1), letter sound recognition (Test 2), matching letter sounds to picture (Test 3), reading nonsense word (Test 4) and spelling nonsense word (Test 5). During each test, correct responses resulted in the delivery of a reinforcer, and no feedback was provided for incorrect responses. After Test 3, the experimenters attempted to match sets on letter difficulty; the experimenters re-assigned the letters to three or four four-letter sets based on the number of correct responses per stimulus during Tests 1, 2, and 3. That is, letters were split into sets that resulted in an approximately equal number of correct responses. In addition, letters were separated if they began with the participant's first or last name (Pence Turnbull et al. 2010) or shared similar features such as b and d (Treiman et al., 2006), and letters with challenging sounds such as v and j were separated (Justice et al., 2006). The experimenter conducted Test 4 and 5 with the new letter sets.

Test 1: Stimulus generalization. Experimenters assessed the extent to which participants could correctly emit the letter sound when letters were presented in a different font than the pre-assessment and intervention conditions to assess stimulus generalization (Wunderlich et al., 2014; Wunderlich & Vollmer, 2017). Prior to the first trial, the experimenter said, "I'm going to show you some letters, and I want to see if you know what sound each letter makes." During each trial, the experimenter presented one letter card and said, "What sound does this letter make?" Each session consisted of four trials.

Test 2: Recognize letters for sounds. Experimenters assessed the extent to which participants could identify the letter when given the letter sound, an essential skill for fluent

reading (NELP, 2008). Prior to the first trial, the experimenter said, “I’m going to say a letter sound, and I want you to pick which letter you think it is.” During each trial, the experimenter presented a field of four cards and said, “Pick up the letter that makes the [/g/] sound”. After each trial, the experimenter shuffled the card array, and each session consisted of four trials.

Test 3: Match letter sounds to pictures. Experimenters assessed participants’ ability to identify the correct beginning phoneme in words when presented with a visual cue. To conduct this test, one picture for each letter sound was created (e.g., if letters include u, f, m, and d, picture cards were umbrella, foot, mouth, and doctor). The experimenter selected pictures that would not be used in the embedded picture intervention condition. During each trial, the experimenter presented a field of four cards, held up the picture, said the picture name, told the participant to repeat the picture name, and said, “Pick up the letter that [picture] begins with.” After each trial, the experimenter shuffled the card array, and each session consisted of four trials.

Test 4: Stimulus generalization - reading nonsense words. Experimenters assessed the participants’ ability to recognize and identify the correct beginning phoneme in words with which they had no prior experience. The use of nonsense words increased the confidence of intervention effects because the chance of receiving outside instruction for the words was reduced (MacQuarrie et al., 2002). The experimenters created 9 or 12, three-letter consonant-vowel-consonant nonsense words. Prior to the first trial, the experimenter said, “I’m going to show you a word, and I want you to tell me the beginning sound.” Then, the experimenter presented a nonsense word (e.g., ‘cul’) and said, “What sound does this word start with?” Each session consisted of three trials.

Test 5: Spelling nonsense words. Experimenters assessed the participants' ability to recognize and identify the correct beginning and ending phonemes. The words were identical to the words presented in Test 4, and two trials were conducted for each word (i.e., six trials per session). Prior to each session, the experimenter said, "I want to see if you know the first and last sound of some words" and presented a field of four letter cards. Prior to the first trial, the experimenter said, "what letter does [cul] begin with?" and, prior to the second trial, said, "what letter does [cul] end with?". The experimenter did not shuffle the array between trials for each word but shuffled the array between words.

Letter Sets

The experimenter assigned letter sets to intervention and control conditions based on the number of correct responses during the pretests. The experimenter assigned the letter set with the highest number of correct responses to the control condition and randomly assigned the remaining sets to the intervention conditions.

In Study 1, the intervention conditions included a paired kinesthetic movement (KM) and a paired embedded picture (EP) condition. For two participants, a third traditional drill (TD) intervention condition was included. In Study 2, the intervention conditions included KM and a paired combined mnemonic (CM) condition.

Pre-Intervention Training

Pre-intervention training consisted of three phases: phoneme segmentation, picture, and kinesthetic movement training. During all phases, correct responses resulted in praise and the delivery of a reinforcer, and incorrect responses resulted in a phase-specific response. Each training phase ended after the participant independently emitted the correct response for each stimulus in the set.

Phoneme Segmentation Training. Phoneme segmentation was conducted to ensure participants could imitate and segment the first sound in words. Procedures were similar to those used by Ehri et al. (1984) and were conducted with words that began with the letters assigned to each condition. The experimenter selected words that were different from the picture and kinesthetic movement paired with the letter for the intervention comparison. Prior to training, the experimenter said, “I’m going to say a word and the beginning sound, and I want you to copy me.” During training, the experimenter said the word and the first letter sound. For example, the experimenter said, “/k/ kite /k/” and waited for the participant to imitate them. If the participant failed to respond after 5 s or responded with anything other than the correct response, the experimenter repeated the prompt.

Picture Training. Picture training was conducted to ensure participants could correctly identify the animal that would be presented in the EP condition. Prior to training, the experimenter said, “I want to see if you know these animals.” During training, the experimenter presented each animal that was going to be used in the EP condition and asked, “What animal is this?”. If the participant responded with the incorrect animal, the experimenter responded with the correct animal name and shuffled the card back into the pile until the participant emitted the correct response.

Kinesthetic Movement Training. Kinesthetic movement training was conducted to ensure participants could emit the physical movement that would be paired with the letter during KM and CM conditions. Prior to training, the experimenter said, “I’m going to do some movements and I want you to copy me.” During training, the experimenter emitted the movement and said, “Copy me.” An incorrect movement resulted in the experimenter physically guiding the participant to engage in the movement.

Procedure

Mastery Probes. Mastery probes were conducted with cards only containing the letter (i.e., no picture for the EP and CM letter set). Probes were conducted once at the beginning of each research block for all participants except Dale (Study 1). Due to inconsistent session assent, probes were conducted once a week for Dale, beginning at session 20, to minimize the aversive nature of probes when correct responses were unlikely (i.e., prior to responding correctly to any stimuli during training). Prior to each probe, the experimenter said, “I want to see if you know these letter sounds.” During the probe, the experimenter presented each letter once from each condition and said, “What sound does this letter make?”, for a total of 12 or 16 trials. Letters for each condition were grouped together within the probe, and the order of the conditions were randomized each day. Contingencies for emitting an incorrect response, the letter name, and the long vowel sound were identical to the pre-assessment screening. A correct response resulted in praise and the delivery of a reinforcer. If the participant emitted the correct response to all letters for one condition in the probe, the respective training session was not conducted. We defined the mastery criterion as the participant emitting the correct response to all condition stimuli during two consecutive mastery probes.

No Intervention (Control): Probe only. A set of letters was only presented in the mastery probe as a control to evaluate if correct responses increased due to outside instruction. That is, one set of letters never entered intervention.

Intervention. Each intervention condition was paired with a unique discriminative stimulus to increase the discrimination between conditions (see Appendix A and B). Prior to each session, the experimenter presented the discriminative stimulus, and stated a condition specifying statement (e.g., “We are going to learn letters when they have pictures in them”).

Additionally, the experimenter conducted modeling prior to every intervention session with letters that were incorrect in the mastery probe. Modeling procedures were dependent on the study conditions and are described in corresponding subsections of the Study 1 and Study 2 Method below.

All intervention sessions consisted of eight trials. At the start of session, the experimenter shuffled the training cards, presented each letter once, asked, “What sound does this letter make?”, shuffled the cards again, and then presented each card again. A correct response resulted in praise and the delivery of a reinforcer, and an incorrect response resulted in an error correction procedure identical to modeling for each condition. The error correction was repeated until the participant correctly imitated the experimenter. Contingencies for emitting the letter name and long vowel sound were identical to the pre-assessment screening.

If a participant did not meet the mastery criterion for one condition after 10 intervention sessions following mastery in another intervention condition, the experimenter presented the unmastered set according to the first mastered condition.

Post-Test. We evaluated to extent to which intervention training resulted in an increase in correct responses for related literacy skills. Five post-tests, identical to the pre-tests were conducted immediately after set mastery. In Study 1, when post-tests were conducted for a treatment group participant, post-tests were conducted for the matched-pair control group participant. Post-tests were not conducted with treatment group participants Dale, Milo (Study 1), and Amara (Study 2) and control group participants, Riley, Cassie, and Wyatt due to the COVID-19 pandemic school closures.

Replication. After the participant mastered each intervention set, we conducted a replication of the pre-assessment screening procedures with the probe only letters and letters

excluded from the initial evaluation. If the participant responded correctly to more than five letters during the replication pre-assessment, a replication intervention was not conducted. In Study 1, a replication was conducted with Nico, and in Study 2, a replication was conducted with Ginny.

Preference Assessment. We conducted a concurrent chains preference assessment to identify each participant's preference for each condition (Hanley et al., 1997). First, the experimenter conducted a mastery probe with the letter set previously assigned to the probe only condition. Second, the experimenter presented the initial link; they placed the discriminative stimuli in front of the participant, pointed to each stimulus while saying the condition specifying statement, and asked the participant about the conditions paired with each card. Third, the experimenter said, "I'm going to teach you the letters. Pick how you want to learn them" and then waited 5 s for the participant to make a selection. If the participant selected more than one card, the experimenter repeated the condition specifying statement and said, "Choose one." The terminal link consisted of the training procedure selected by the participant.

In Study 1, the experimenter presented four discriminative stimuli: KM, EP, TD, and No Intervention (excluding Nico). In Study 2, the experimenter presented five discriminative stimuli: KM, EP, CM, TD, and No Intervention. In addition, in Study 2, after three consecutive selections of the same stimulus, that stimulus was removed from the array to determine a preference hierarchy.

Maintenance. We evaluated the extent to which each condition resulted in retention of mastered letter sounds. Maintenance probes were identical to the mastery probe. To equate the delay from set mastery to maintenance probes between conditions, maintenance probes were

conducted based on the timing of mastery of each set rather than mastery of both letter sets.

Maintenance probes were conducted 1, 5, 7, and 9 weeks post set mastery.

Study 1: The Effects of Paired Kinesthetic Movements and Embedded Pictures on Letter-Sound Correspondence

The purpose of Study 1 was to compare paired kinesthetic movements (KM), embedded pictures (EP), and a traditional drill (TD) flashcard procedure on letter-sound correspondence. In addition, we compared the interventions to a control condition in which letters were only presented in the probe (probe only).

Intervention Procedure

For each condition, a correct response resulted in praise and the delivery of a reinforcer, and an incorrect response resulted in an error correction procedure identical to modeling for each condition.

Kinesthetic Movement. Prior to KM modeling, the experimenter said, “We are going to learn letters with movements” and presented the relevant contingency correlated stimulus. During modeling, the experimenter presented the card, said the letter sound, named the animal paired with the movement, and said the letter sound while engaging in the movement. Then, the experimenter told the participant to imitate them. For example, if the experimenter presented the letter card ‘a,’ they said “/a/, alligator, /a/” while extending and opening and closing their arms as if they were an alligator.

Embedded Picture. Prior to EP modeling, the experimenter said, “We are going to learn letters with pictures” and presented the relevant discriminative stimulus. During modeling, the experimenter presented the letter card, said the letter sound, named the picture, said the letter sound, and told the participant to imitate them. For example, if the experimenter presented the letter card ‘b,’ they said, “/b/, bear, /b/.”

One participant, Dale, consistently provided assent for the EP condition but not for the KM condition (i.e., asked to return to the classroom when the experimenter began a KM

session). The experimenters removed the KM condition and implemented an EP fading condition at session 60. Fading consisted of three steps: EP, EP presented in 100% saturation per word processor (pictures appeared faded), EP presented as TD (i.e., no picture). The experimenters defined the fading criteria as 100% correct responses during training across two days.

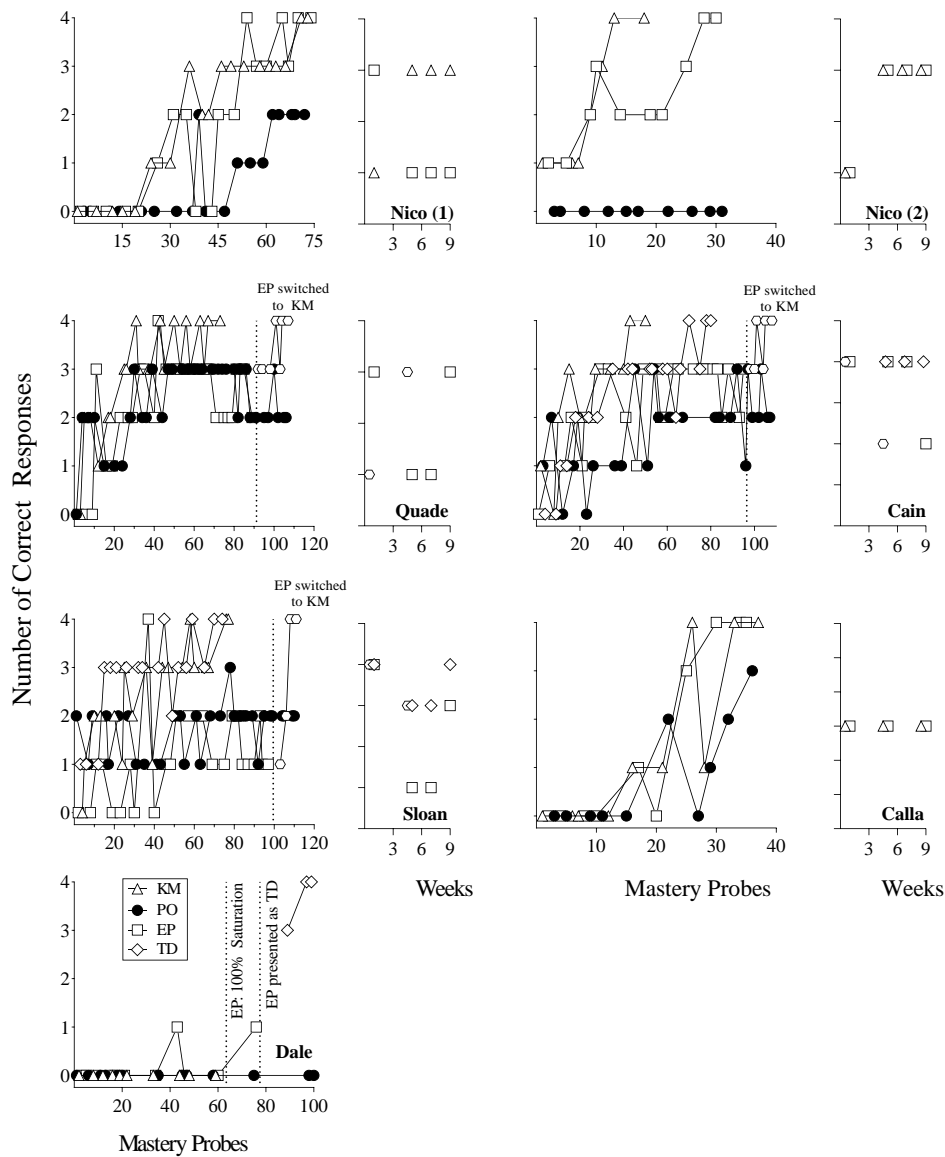
Traditional Drill. Prior to TD modeling the experimenter said, “We are going to learn letters.” During modeling, the experimenter presented the letter card, said the letter sound twice, and told the participant to imitate them. A TD intervention set was presented in Sloan and Cain’s intervention comparison.

Three participants, Quade, Sloan, and Cain did not meet the mastery criterion for the EP condition 10 intervention sessions following mastery in the KM condition. Thus, the EP stimuli were presented using the KM procedure.

Results and Discussion

Figure 2 depicts the number of correct responses during mastery and follow-up probes for the KM, EP, TD, EP presented in KM, and probe only conditions. Of the 6 completed evaluations, 4 participants mastered the stimulus set in the KM condition prior to the EP condition (see Quade, Cain, and Sloan’s evaluations and Nico’s second evaluation) and 2 participants mastered the EP and KM sets at approximately the same time (see Calla and Nico’s first evaluation). Of the 4 participants who met the mastery criterion in the KM condition first, 3 participants met criteria for the EP stimuli to be presented in the KM condition (see Quade, Cain, and Sloan’s evaluation). Two evaluations included a TD stimulus set: Sloan and Cain. Sloan mastered the TD and KM sets at approximately the same time, and both TD and KM were mastered prior to mastering letters originally assigned to the EP condition that were subsequently mastered in the KM condition. Cain mastered the stimulus sets in the following order: TD, KM,

and EP presented in KM. In addition, in 6 of 6 evaluations, participants did not meet the mastery criterion for the probe-only set.



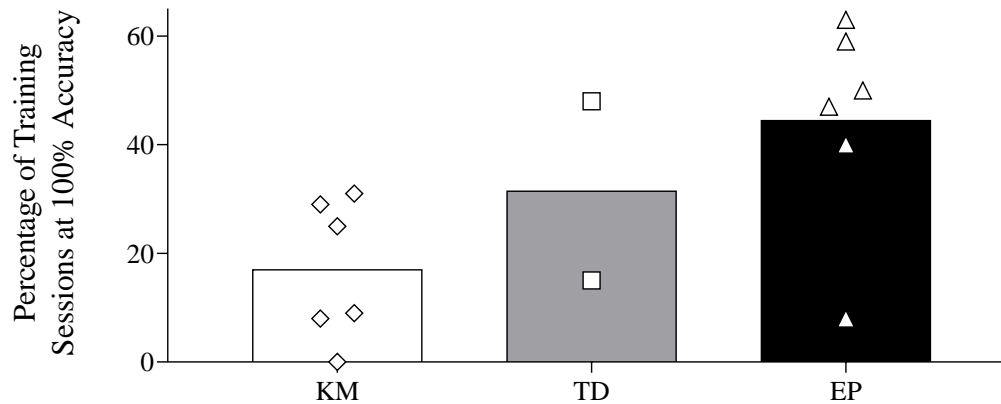
Note. EP = Embedded Picture; KM = Kinesthetic Movement; PO = Probe Only; TD = Traditional Drill

Figure 2. The Number of Correct Responses per Mastery and Follow-Up Probes

For KM training sessions, we examined the correlation between the percentage of correct movements and the number of stimulus trials to meet the mastery criterion. We calculated the percentage of correct movements for each evaluation by dividing the total number of trials in

which the participant emitted the correct movement by the total number of trials and multiplying that number by 100. Quade, Nico (evaluation 1), Calla, Cain, Nico (evaluation 2), and Sloan emitted the correct movement during an average of 1%, 17%, 35%, 41%, 44%, and 48% of trials across each evaluation, respectively. They mastered each set after 40, 50, 22, 20, 8, and 36 training trials per stimulus, respectively. The experimenters conducted a one-tailed Pearson correlation test, and results yielded a moderate negative correlation between the number of sessions required for participants to meet the mastery criterion and the percentage of engagement in the correct movements ($r = -.63$). As the percentage of correct movements emitted increased, the number of intervention sessions to mastery decreased. This finding is consistent with Lozy et al. (2020), which found a strong negative correlation between engaging in movements during training and the number of sessions until participants met the mastery criterion.

Figure 3 depicts the percentage of correct responses during training sessions for KM, EP, TD, and EP presented in KM conditions. Across all evaluations, participants engaged in training sessions at 100% accuracy without mastering the letter set during an average of 42% of EP sessions (range, 8% to 63%), 31% of TD sessions (range, 15% to 48%), and 14% of KM sessions (range, 0% to 31%). That is, during almost half of EP training sessions responding was at 100% accuracy despite not meeting the mastery criteria. Using a dependent samples t-test, we compared the percentage of training sessions in which responding was at 100% accuracy during EP ($M = 42$) to the percent of responding at 100% accuracy during KM ($M = 14$). Results yielded a significant difference between participants' responding at 100% accuracy during KM and EP, $t(7) = 7.33, p < .05$.



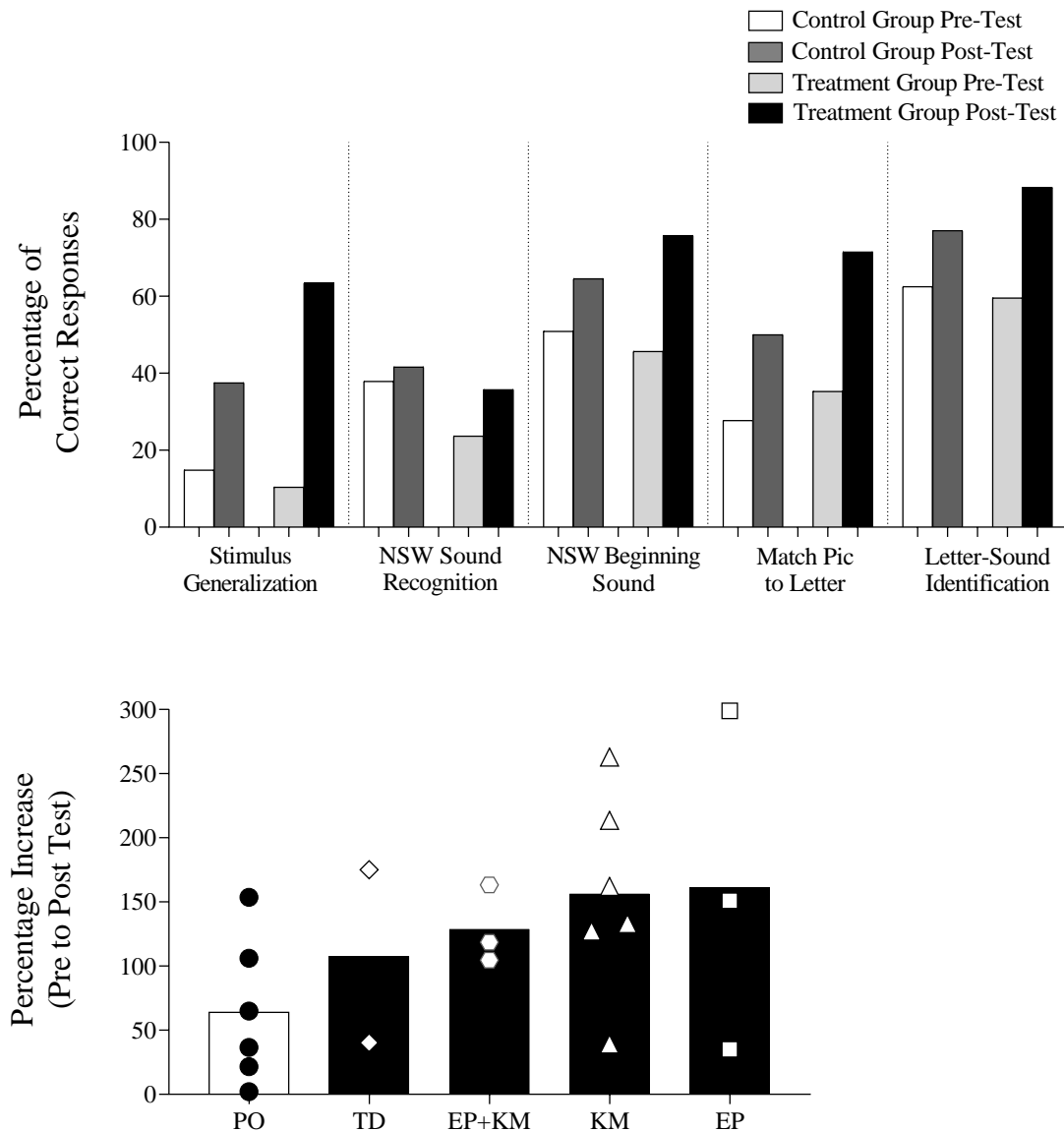
Note. EP = Embedded Picture; KM = Kinesthetic Movement; TD = Traditional Drill. Each data point represents each participant

Figure 3. The Percentage of Training Sessions at 100% Accuracy

The top panel of Figure 4 depicts the percentage of correct responses during pre- and post-tests for control and intervention groups per test. On average, the intervention group responded with fewer correct responses than the control group during the pre-tests and more correct responses than the control group during the post-tests. We conducted a two-way repeated measure analysis of variance (ANOVA) between groups (treatment and control) at the level of the test (pre and post). Results generated a significant interaction between groups at the pre-and-post-test, $F(1, 8) = 12.71, p > .05$. The main effect of test level (pre-and-post-test), $F(1, 5) = 20.65, p > .05$, and groups (treatment and control), was significant, $F(1, 8) = 42.44, p > .05$.

The bottom panel of Figure 4 depicts the percentage increase from pre- to post-tests for the probe only and intervention conditions for the treatment group participants. The average increase was 64% for the probe only condition (range: 2%-153%), 107% for the TD condition (range: 40%-175%), 129% for the EP with KM condition (range: 104%-118%), 129% for the KM condition (range: 39%-262%), and 156% for the EP conditions (range: 35%-299%). Using a repeated measures t-test, we compared the percentage of correct responses during all pre-tests (M

= 7.75) to the percentage of correct responses during all post-tests ($M = 60.02$). Participant responses significantly differed from pre- to post-tests, $t(14) = 6.40, p < .05$.



Note. NSW = Nonsense Words; EP = Embedded Picture; KM = Kinesthetic Movement; TD = Traditional Drill. Each data point represents one participant

Figure 4. Pre- and Post-Test

Table 2 indicates the number of intervention sessions for each condition during the choice evaluation. On average, participants experienced 14 KM training sessions (range: 4-10), 24 EP training sessions (range: 11-35), 5 EP presented with KM training sessions (range: 3-6),

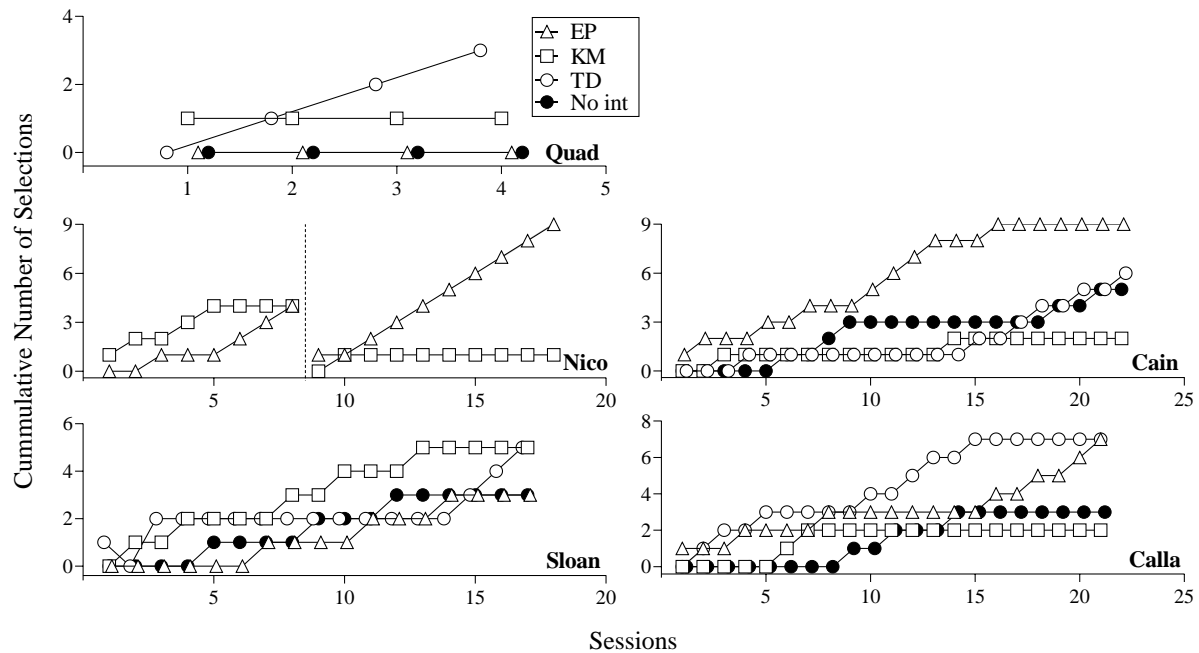
and 17 TD training sessions (13-21). A dependent samples t-test yielded a significant difference between the number of sessions participants experienced prior to mastering KM ($M = 5$) and the number of sessions participants experienced prior to mastering EP ($M = 24$), $t(7) = 2.65$, $p < .05$.

Table 2. Number of Intervention Sessions per Condition and Preferred Condition

Participant	Evaluation	Intervention Sessions per Condition				Preference
		KM	EP	TD	EP:KM	
Nico	1	25	24	-	-	EP
	2	4	13	-	-	EP
Quade	1	19	34	-	7	TD
Cain	1	10	35	21	3	TD & EP
Sloan	1	18	31	13	6	Unable to Assess
Calla	1	10	11	-	-	-
Dale	1	-	-	-	-	EP

Note. KM = Kinesthetic Movement; EP = Embedded Picture; TD = Traditional Drill

The last column in Table 2 denotes the preferred condition per participant, and Figure 5 indicates cumulative intervention selections per session for Nico, Quade, Sloan, and Calla. Of the four participants for whom KM was more effective than EP, preference was not identified for one participant (Calla), TD was preferred for one participant (Quade), EP and TD were preferred for one participant (Cain), and EP was preferred for one participant (Nico, second evaluation). Preference was likely not identified for Calla due to lack of discrimination between the conditions. Of the two participants in which there was little differentiation between KM and TD, EP was preferred for one participant (Nico, first evaluation) and we were unable to assess preference for one participant (Sloan). Anecdotally, Sloan would rotate between interventions each day, with the general statement, “I already did this one, so I’ll do this one next,” and no clear preference for any condition. Although Dale did not master the sets in either the EP or KM conditions, he demonstrated a clear preference for EP by only providing assent for the EP condition.



Note. KM = Kinesthetic Movement; EP = Embedded Picture; TD = Traditional Drill
Figure 5. Preference Assessment

The results of Study 1 demonstrate that pairing letters with movements was more effective than pairing letters with pictures, and that both interventions, as well as a traditional flashcard method, were more effective than no intervention. However, the number of letters correct during the maintenance probes and the generalization to more complex reading skills did not significantly differ between interventions conditions. Nevertheless, there was a significant difference between intervention and control conditions on the generalization effects and a significant difference between the treatment and control group on generalization effects. With respect to each intervention condition, participants had a higher percent increase for letters presented with movements and pictures compared to letters presented alone. Results also demonstrated that effectiveness of the intervention did not correlate with participant preference.

Study 2: Effects of Kinesthetic Movements versus Combined Kinesthetic Movements and Embedded Pictures on Letter-Sound Correspondence

Multisensory instructional programs often use the combination of mnemonics, such as presenting a letter with a picture and movement. Thus, the purpose of Study 2 was to compare the effects of combined mnemonics (KM and EP; CM) to a single mnemonic, kinesthetic movements (KM), on letter-sound correspondence. KM was selected as the single mnemonic condition, as opposed to EP, based on the results from Study 1. That is, in 4 of 6 evaluations, KM was mastered with fewer training sessions than EP, and in the remaining 2 of 6 evaluations, KM and EP were mastered at relatively the same time. Additionally, we compared both mnemonics conditions to a control condition in which letters were only presented during mastery probes (probe only).

Intervention Procedure

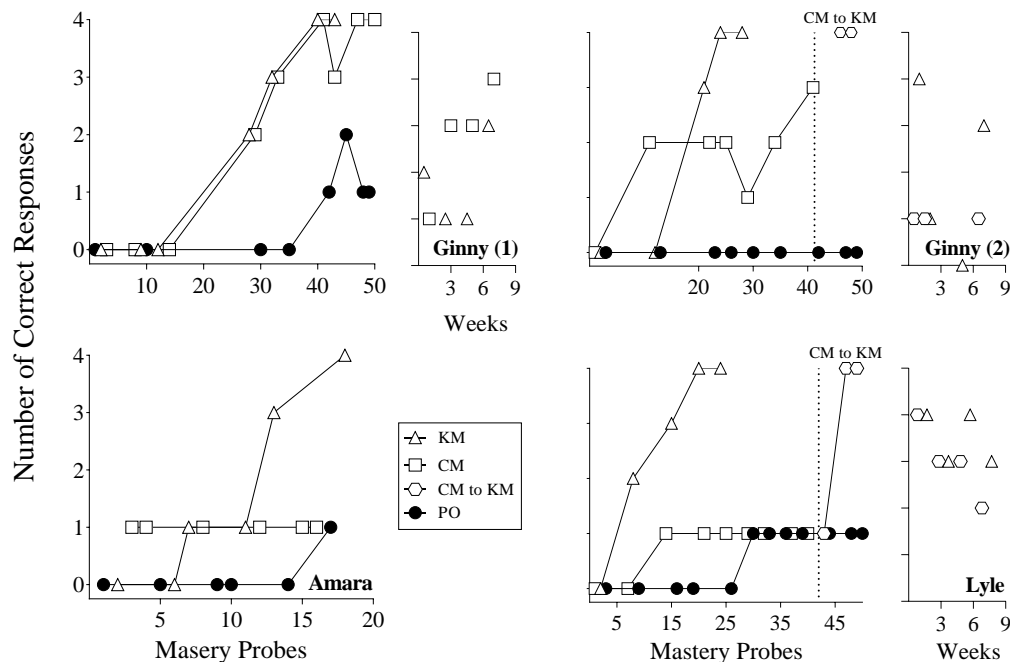
For each condition, a correct response resulted in praise and the delivery of a reinforcer, and an incorrect response resulted in an error correction procedure identical to modeling for each condition.

Kinesthetic Movement. The contingency specifying statement and modeling procedures for KM were identical to the procedures described in Study 1.

Combined Mnemonics. Prior to CM modeling, the experimenter said, “We are going to learn letters with pictures and movements” and presented the relevant discriminative stimulus. During modeling, the experimenter presented the letter card, said the letter sound, named the animal picture, and said the letter sound while engaging in the movement. Then, the experimenter told the participant to imitate them. For example, if the experimenter presented the letter card ‘a,’ they would say “/a/, alligator, /a/,” while extending their arms and opening and closing their arms as if they were an alligator.

Results and Discussion

Figure 6 depicts the number of correct responses during mastery and follow-up probes for the KM, CM, CM presented in KM, and probe only conditions. Of the 3 completed evaluations, 2 participants mastered the stimulus set in the KM condition prior to the EP condition (Lyle and Ginny's first evaluation) and 1 participant mastered the KM and CM sets at approximately the same time (Ginny's first evaluation). Despite the incomplete evaluation for Amara, when she mastered the KM set, there was no increase in the number of correct letter sounds demonstrated for the CM set. Of the 2 participants who met the mastery criterion in the KM condition first, both participants met criteria for the CM stimuli to be presented in the KM condition. In addition, in 4 of 4 evaluations, participants did not meet the mastery criterion for the probe-only set.

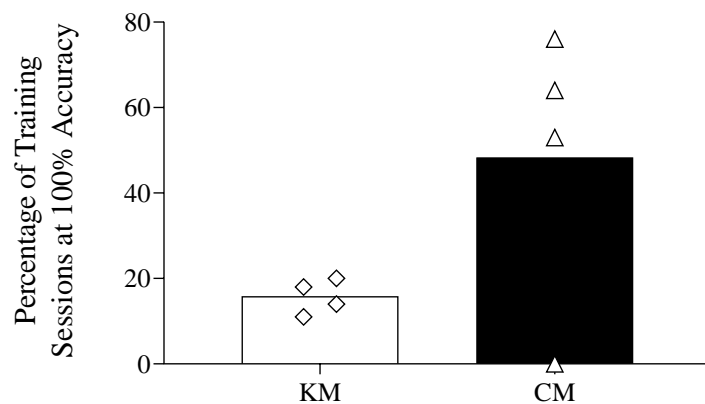


Note. KM = Kinesthetic Movement; CM = Combined Mnemonics; PO = Probe Only
Figure 6. The Number of Correct Responses per Mastery and Follow-Up Probes

Participants' number of correct responses during follow-up probes remained relatively consistent across time. That is, participants responded correctly to an average of 2.67 stimuli at

week 1 (range: 2-4), 2 stimuli at week 5 (range: 1-3), 2.2 stimuli at week 7 (range: 0-4), and 2.6 stimuli at week 9 (range: 1-4). In addition, participants responded with approximately equal correct responses for all stimulus sets during all weeks. At weeks 1, 5, 7, and 9, participants responded correctly to an average of 2.6 KM stimuli; 2.8 CM stimuli; and 2.2 CM:KM stimuli, respectively.

Figure 7 depicts the percentage of correct responses during training sessions for KM and CM conditions. Across all evaluations, participants engaged in training sessions at 100% accuracy without mastering the letter set during an average of 16% of KM sessions (range, 11% to 20%) and 48% of CM sessions (range, 0% to 76%). That is, almost half of CM training sessions were at 100% accuracy despite not meeting mastery criteria.

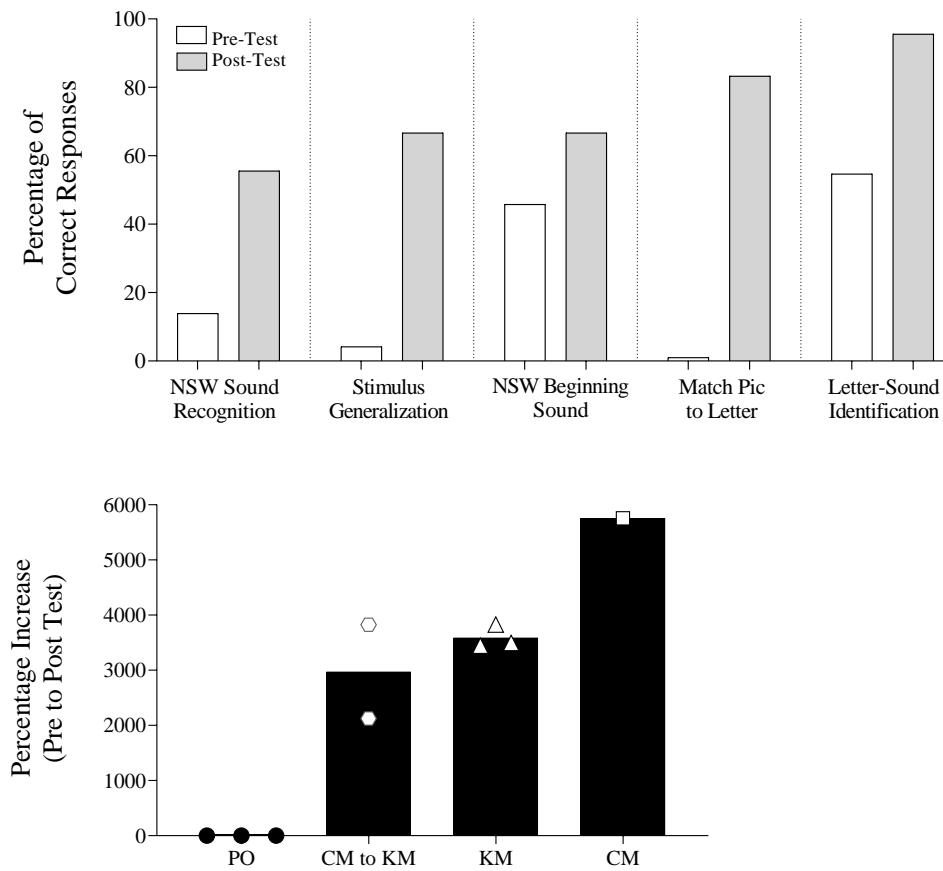


Note. KM = Kinesthetic Movement; CM = Combined Mnemonics. Each data point represents a participant

Figure 7. The Percentage of Training at 100% Accuracy

The top panel on Figure 8 depicts the percentage of correct responses during pre- and post-tests for the participants per test. On average, participants responded with more correct responses during the post-test compared to the pre-test for all tests, with the greatest percentage increase occurring in the match picture to letter test, and the smallest percentage increase occurring in identifying the nonsense word beginning sound. The bottom panel of Figure 8

depicts the percentage increase from pre- to post-tests for the probe only and intervention conditions. The average increase was 397% for the probe only condition (range: -18%-813%), 2975% for the CM in the KM condition (range: 0%-9900%), 3590% for the KM condition (range: 33%-9900%), and 5740% for the CM condition (range: 300%-9900%). Using a repeated measures t-test, we compared the percentage of correct responses during all pre-tests ($M = 7.75$) to the percentage of correct responses during all post-tests ($M = 60.02$). Participant responses significantly differed from pre- to post-tests, $t(14) = 6.40, p < .05$.



Note. KM = Kinesthetic Movement; CM = Combined Mnemonics; PO = Probe Only. Each data point represents a participant

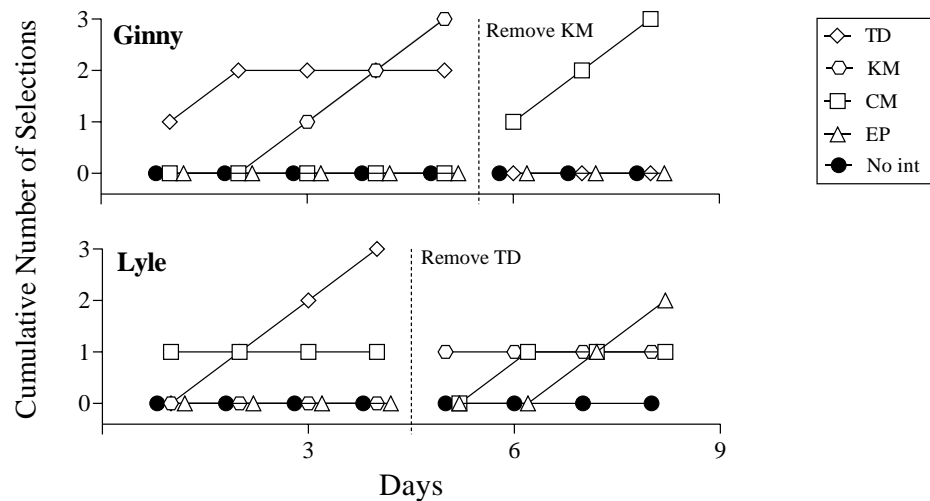
Figure 8. Pre- and Post-Test Data

Table 3 indicates the number of intervention sessions for each condition during the choice evaluation. On average, participants experienced 11 KM training sessions (range: 5-19), 20 CM training sessions (range: 17-22), and 5 CM presented in KM training sessions (range: 4-6). Of the 3 participants, a preference assessment was not conducted with one participant (Amara) and preference varied for the remaining two. Figure 9 indicates cumulative intervention selections per day for Ginny and Lyle, and the remove of a condition is denoted by a dashed line. Ginny initially demonstrated a preference for KM and, after KM was removed from the selection array, demonstrated a preference for CM. Lyle initially demonstrated a preference for TD and, after TD was removed from the selection array, began to demonstrate a preference for EP.

Table 3. Number of Intervention Sessions per Condition

Participant	Evaluation	KM	CM	CM:KM
Ginny	1	19	17	-
	2	11	21	4
Amara	1	5	-	-
Lyle	1	7	22	6

Note. KM = Kinesthetic Movement; CM = Combined Mnemonics



Note. CM = Combined Mnemonics; EP = Embedded Picture; KM = Kinesthetic Movement; TD = Traditional Drill

Figure 9. Preference Assessment

The results of Study 2 demonstrated that pairing letters with movements was more effective than pairing letters with both pictures and movements, and that both interventions were more effective than no intervention. Similar to Study 1, the number of letters correct during the maintenance probes and the generalization to more complex reading skills did not significantly differ between interventions conditions, but a significant difference was observed between intervention and control conditions.

General Discussion

We evaluated the efficacy, generalization, and maintenance of, and preference for, mnemonics paired with a traditional drill flashcard intervention on literacy skills with preschoolers. Across both studies, the results demonstrate that pairing kinesthetic movements with a traditional drill flashcard method for letter sound acquisition is more effective than presenting letters alone, presenting letters with embedded pictures, and presenting letters with movements and pictures. The results also demonstrate that any of the interventions evaluated were more effective than no intervention. Despite the efficacy of presenting letters with movements on letter sound correspondence, there was no significant difference between mnemonic interventions on generalization post-tests and maintenance data. However, there was a significant difference between post-tests for the control and intervention conditions for the treatment group, and a significant difference between post-tests for the treatment and control group. Further, results demonstrate that participants preferred intervention conditions that were presented with pictures (i.e., EP and CM). Thus, when selecting an intervention targeting literacy skill acquisition, clinicians should be attentive to the costs and benefits of each intervention.

Although participants required fewer training sessions to master letters presented with movements, participants demonstrated a preference for letters presented with pictures. This could be because participants experienced a denser schedule of reinforcement during EP and CM training sessions due to a higher percentage of correct responding during those training conditions compared to KM training sessions. One way to increase the density of the reinforcement schedule during training with kinesthetic movements is to intermittently present already acquired targets with unknown targets, as in incremental rehearsal (Tucker, 1988) and strategic incremental rehearsal (Kupzyk et al., 2011). In a direct comparison between a

traditional drill and strategic incremental rehearsal flashcard intervention for literacy skills, Lozy et al. (2019) demonstrated that strategic incremental rehearsal was either just as effective or more effective than traditional drill. Future research should examine the effects of use a strategic incremental rehearsal flashcard method, in which the target skill is paired with kinesthetic movements, on the preference for kinesthetic movements compared to embedded pictures. Although EP and CM resulted in a slightly higher percentage increase during the post-tests compared to KM, the difference is not significant and should not deter the use of pairing stimuli with movements.

These results are consistent with previous research, demonstrating that pairing kinesthetic movements with letters is effective at increasing literacy skill acquisition (Lozy et al., 2020) that result in some degree of maintenance (Argamonte & Belifore, 2005) and generalization (Campell et al., 2008). These results also extend research conducted on pairing mnemonics with academic skills by first evaluating the efficacy of embedded pictures and the combination of pictures and movements in a single-subject design, permitting an analysis at the individual level. Second, this study is the first to evaluate the maintenance of literacy skills with a long delay from training to follow-up sessions; previous research evaluated the maintenance of skills after a maximum of 4 weeks (Argamonte & Belifore). Third, only a few studies to date evaluated the emergence of untaught literacy skills post training sessions and the skills evaluated were limited (Campell et al.). This study demonstrates the emergence of a range of literacy skills after letter sound correspondence training.

Experimenters also implemented one possible fading procedure for embedded pictures with one participant, Dale. Previous research has demonstrated the efficacy of fading embedded pictures on letters using a four-step (Hoogeveen et al.1989) and six-step procedure (de Graff et

al., 2007) that moved from one step to the next based on an arbitrary number of sessions. In this study, we successfully faded out the embedded picture using a systematic three-step procedure, in which Dale moved to the next step after 100% accuracy in two consecutive training sessions. Additional research should determine fading procedures that are sensitive to when the transfer of stimulus control has occurred to improve efficiency in fading.

The efficacy of pairing letters with kinesthetic movements can be explained by many different theoretical approaches. First, embodied cognition describes a class of theories on the influence of sensorimotor experiences via bodily interactions with the environment during acquisition (Borghi & Cimatti, 2010). Theories of embodied cognition are on a spectrum ranging from disembodiment, which negates the effects of sensorimotor experiences on cognitive processing (Meteyard et al., 2012), to embodiment, which emphasizes the effects of sensorimotor experiences on cognitive processing. Embodiment states that higher order and offline cognitive processing (i.e., when removed from the learning environment) involve re-enactment of the bodily states from previous experiences (Foglia & Wilson, 2013). Embodied learning refers to the integration of the embodiment theory in educational settings to develop curricula predicated on presenting information via multiple sensory inputs to increase understanding and retention (e.g., pairing tactile mnemonics or kinesthetic movements with instruction; Nguyen & Larson, 2015). Kinesthetic learning is one component of embodied learning and refers to the learner's physical interaction with the environment (Kosmas & Zaphiris, 2018), and researchers have suggested that embodied and kinesthetic learning are the most effective for acquisition and retention when the sensorimotor task, such as body movements, is integrated with the learning material (Wilson & Golonka, 2013; Johnson-Glenberg et al., 2014). Because the embedded picture condition did not involve bodily movement, and the movement in the kinesthetic

movement condition was strongly related to the letter (e.g., an action related to an animal beginning with that letter sound), this may account for the efficacy of pairing movements with letters as opposed to pictures.

Second, presenting letters with pictures was less effective than presenting letters with movements for some participants, which may be related to the Gestalt law of continuity. The law of continuity states that humans continue to see shapes and forms past their actual stopping or breaking point (Moore & Fitz, 1993). In relation to letter identification, research has demonstrated correlations between letter font (easy- versus hard-to-read) and typographic variables (e.g., italicized, bolded, underlined) with speed of letter identification, memory of letters, and perceived effort for letter identification (Pelli et al., 2009; Diemand-Yauman et al., 2011, Keage, 2014), with harder-to-read, italicized fonts resulting in less speed, worse memory, and higher effort for participants. Research has also demonstrated that the degree of perturbation (i.e., disconnect) is inversely proportional to letter identification efficiency (Pelli et al.). Thus, it is likely that the degree of picture overlap and the number of breaks in the letter via the picture accounted for some participants' difficulty in mastering the embedded picture condition.

Although we were unable to determine if there is a relationship between letter mastery and the number of breaks per letter due to the limited number of participants and the variability of letters assigned to the EP condition across participants, future researchers should examine if the number of breaks per letter is related to mastery of the letter. Contingent on a significant correlation between breaks per letter and mastery, future researchers should also consider the number of breaks per letter when assigning letters to sets and attempting to equate set difficulty.

A third reason presenting letters with pictures was not effective for all participants may be due to the blocking and/or overshadowing effect (Johnson & Cumming, 1968; Singh &

Solman, 1990). Blocking and overshadowing refer to the stimulus (e.g., picture) controlling the learner's response in the presence of a compound stimulus (e.g., picture and letter), with prior learning history being a critical component of the blocking effect. For example, it is possible that, if the participant had an extensive learning history with the picture, this may have blocked the letter from gaining stimulus control over their response. Although we collected data on whether or not the participant responded correctly to the animal-only picture during the training phase, it is not enough information to determine whether participants had an extensive history with the animal. However, it is also possible that the picture was more salient stimulus than the letter and thus controlled participant responding (i.e., overshadowing). For example, in a picture-word stimulus, Dittlinger & Lerman (2011) demonstrated that the correct response was controlled by the more salient stimulus (picture) rather than the target stimulus (word). Specific to this study, overshadowing may explain why some participants responded at 100% accuracy for an average of 50% of training sessions but failed to respond correctly to the letter when presented alone in the mastery probe.

A fourth reason participants may have failed to master the embedded picture letter set despite responding with accuracy during training may be due to joint stimulus control (see Sidener, 2006 for a review). Joint stimulus control occurs when different stimuli (e.g., picture and letter) evoke the same response and arise at the same time (Lowenkron, 1998). For a stimulus to acquire joint stimulus control, a unique response and selection are paired, thus increasing discrimination among stimuli and the probability that the learner will select the correct response (Tu, 2006). The most common scenario of the phenomenon involves control of the selection response by the match in topography of two separate but interrelated verbal operants, typically an echoic and a tact. However, the phenomenon can be applied more

generally to any discriminated and self-repetitious responses under many experimental arrangements (Hutchison, 2013). Different forms of joint stimulus control may explain why embedded pictures hindered the acquisition of letter-sound correspondence for some participants and also explain why paired kinesthetic movements were more effective.

When a stimulus acquires joint control, the selection response is evoked by a single event: the combination of similar or dissimilar topographical stimuli (Lowenkron, 1998). Thus, participant responding during the embedded picture training condition may have been evoked by a single stimulus, the combination of the letter and picture, rather than the picture or the letter. Additionally, joint stimulus control is more likely to occur when a strong relation exists between the two stimuli (Lowenkron); the picture and letter were strongly related because the picture was embedded within the letter and the picture began with the letter sound. This adds further support for the notion that joint stimulus control acquired during training prevented participants from responding to the letter when presented alone.

Joint stimulus control may also involve a mediating response: a response that occurs between the presentation of the controlling discriminative stimulus and the response (Michael et al., 2011). A mediating response may act as the controlling variable for the response and can either be covert or overt. Such examples include key pressing (Torgrud & Holborn, 1989) and hand signs during match-to-sample tasks (Lowenkron, 1988). In the latter example, the experimenter's hand sign controls the topography of the participant's imitated hand sign, and the participant's hand sign is allowed only in the presence of the comparison stimuli (Lowenkron, 1998). Thus, the imitated topography specifies the correct response. Emitting a unique kinesthetic movement and vocal response in the presence of a verbal stimulus during discrete trial training has been conceptualized as a mediating behavior because it specifies the conditions

under which the response will be reinforced (Lozy et al., 2020). It is likely that the kinesthetic movement during training acted as a mediating response, increasing the probability that the participants would emit the correct response.

One explanation that may account for the efficacy of both paired kinesthetic movements and embedded pictures on letter-sound correspondence can be discussed via a differential observing response (DOR). A DOR is a procedural modification commonly used during conditional discrimination and match-to-sample tasks to increase correct responding. When a DOR is implemented, a unique response is paired with each stimulus to increase the saliency of critical stimulus features (Urcuioli & Callender, 1989). For example, Constatine and Sidman (1975) demonstrated an increase of correct responses on a match-to-sample task when they instructed participants to emit the stimulus name prior to making the response. DORs are commonly implemented with individuals with developmental disabilities to decrease restricted stimulus control (attending to isolated features of complex stimuli; Stromer & Dube, 1994), but may be inadvertently implemented when mnemonics are paired with stimuli because the experimenter typically emits the picture name followed by the target stimulus (e.g., letter sound or word) and prompts participants to imitate both responses (Dilorenzo et al., 2011).

For the few participants who mastered the embedded picture letter set, the picture and picture name may have provided a DOR and increased the saliency of the letter (Urcuioli & Callender, 1989). For example, during the modeling and error correction, the experimenter presented the embedded picture letter card and emitted the letter sound, animal name, and letter sound, and prompted the participant to repeat the sequence. Therefore, it's possible that the picture name may have acted as a DOR for some participants and increased correct responding when the picture was removed. For the participants who mastered the kinesthetic movement

letter set, the animal name and movement emitted during modeling and error corrections may have also acted as a vocal and non-vocal DOR (Lozy et al., 2020). Future research should directly compare whether emitting the animal name during modeling and error correction have an effect on the mastery of each condition.

Neuroimaging has been used to investigate the correlation between brain region activity, multisensory instruction, and the development of reading skills in children. Neuroimaging, such as functional magnetic resonance imaging and electroencephalography, have permitted researchers to identify brain regions related to specific reading skills (reading network) as well as the relation among regions (connectivity; see Cattinelli et al., 2013 and Martin et al., 2015 for a review). The degree of connectivity within the reading network has been shown to be correlated with reading skills (Horowitz-Kraus et al., 2016; Wang et al., 2013). For example, a high interaction between distal hierarchically segregated network clusters is associated with better performance on rhyming tasks (Wang et al.). Thus, reading is correlated with the interactions between brain regions, and researchers suggest that these interactions are dependent on the task or training modality (Smith et al., 2018). Multisensory instruction may increase literacy skills compared to traditional approaches because multiple brain regions are being simultaneously activated, which results in increased connectivity among the reading network (Byrge et al., 2014; Willis, 2008). In addition, multisensory instruction may increase the efficiency of retention because activation of one area of memory storage (e.g., visual cortex) activates another area such as the temporal lobe (due to auditory training), the parietal lobe (due to tactile training), or the cerebellum and basal ganglion (due to kinesthetic training; Rivera et al., 2005).

Recent neuroimaging studies have investigated brain regions involved in mnemonic processing during learning and recall tasks. It is well-known that the hippocampus is essential for

supporting memory (Scoville & Milner, 1957), but an extensive history with mnemonic training may support subcortical structures that result in advanced memory (Dresler et al., 2017; Guida et al., 2013). Brain activation for memory athletes and expert mnemonics, who have used mnemonics techniques for years, demonstrate less activation in the medial prefrontal cortex and right dorsolateral prefrontal cortex and more activation in structures involved in long-term memory compared to control participants. This research supports the hypothesis that an extensive history with learning via paired mnemonics results in reorganizing the brain's functional network organization to support superior memory (Dresler et al.) and support the use of presenting early literacy skills with mnemonics.

Despite the continuity of intervention results across Study 1 and Study 2, participants in Study 1 demonstrated a greater percentage increase from pre- to post-tests compared to participants in Study 2. This is likely because participants in Study 1 were receiving consistent classroom instruction on letter sounds, and Study 2 was conducted during the summer when participants were less likely to be receiving additional instruction. The differentiation in percentages may also be related to the small number of participants in Study 2. Study 1 had more participants than Study 2 and therefore more variability in pre- and post-test scores is reflected in the group average. Although Study 2 did not include a control group, the lack of classroom instruction for participants in Study 2 mitigated the need for a control group.

This study has several limitations that warrant discussion. First, not all participants in Study 1 experienced all the same intervention conditions (i.e., two of six participants experienced three conditions as opposed to two). It is possible that the addition of another intervention condition slowed acquisition for two reasons. One, research has demonstrated that a smaller number of sets with more stimuli per set is more effective than a larger number of sets with

fewer stimuli per set (Kodak et al., 2020). Two, increasing the number of letter sets decreased the degree of control of letter selection per set. The experimenters created letter sets by balancing letters across sets with respect to vowels and consonants, letters that shared similar features (e.g., b and d), were developmentally appropriate, and whether they began with the participants first or last name. Increasing the number of letters selected from the available letters based on the pre- and echoic assessments limited the degree of equity across sets. Thus, it is possible that the participants with more intervention conditions may have had more difficult letter sets with respect to the aforementioned variables (e.g., similar features).

A second limitation is missing data. Post-test and follow-up data were not collected for one intervention participant and three control participants in Study 1 and one treatment participant in Study 2. Therefore, the limited number of participants and lack of control group (Study 2), differentiation of intervention conditions experienced (Study 1), and missing data (Study 1 and Study 2) limited our ability to conduct more statistical analyses to determine if significant difference exists between intervention conditions, treatment and control groups, and maintenance data. For example, the number of participants and group assignment in Study 1 allowed us to conduct a two-way repeated measure ANOVA between groups (treatment and control) at the level of the test (pre and post), demonstrating a significant interaction between groups at each level. The limited sample size and lack of group assignment did not permit analyses to determine whether significant difference existed between intervention condition or pre- and post-test scores. Maintenance data could be examined using a multilevel model (MLM) to determine if the delay from mastery (i.e., weeks) predicted the number of correct responses, with participants as the random factor (Lozy et al., 2020). MLM is appropriate for single-subject designs because it can be applied with few parameters, measurements can range from 1 to

infinity, data are nested within participants, and it accounts for dependent error structures, heterogeneous variance, and moderating effects (Moeyaert et al., 2014). MLM would permit experimenters to determine if the number of correct responses significantly differs with respect to weeks and intervention condition. Future research should conduct a power analysis prior to recruiting participants to determine a necessary sample size per group assignment to detect a significant effect between conditions and groups.

Third, the letter sets differed from pre- to post-test for tests 1 through 3. That is, in order to equate the number of correct pre-test responses across letter sets used in the intervention, the experimenter re-organized the letters in each set after pre-test 1 through 3 and, after test 4 and 5, assigned the letter set with the highest number of correct responses to the control condition to increase the confidence in the post-test results between control and intervention conditions. However, equating letter sets based on pre-test responding means that the letter set presented in the pre-test differed from the letter set presented in the post-test for post-tests 1 through 3 and may have affected participant responding. A difference in responding is less likely for Test 1 (Stimulus Generalization) because letters were presented one at a time, but more likely for Test 2 (Recognize Letters for Sounds) and Test 3 (Match Letter Sounds to Pictures) because letters were presented in an array. Future researchers should equate letters based on pre-test responding but conduct post-tests with the same stimuli and determine intervention efficacy per letter as opposed to set.

A fourth limitation is that experimenters were only able to conduct within-participant replications for two of nine participants, and the replication yielded different results for both participants. Specifically, Nico and Ginny's first evaluation demonstrated both intervention conditions to be equally effective, and their second evaluation demonstrated paired kinesthetic

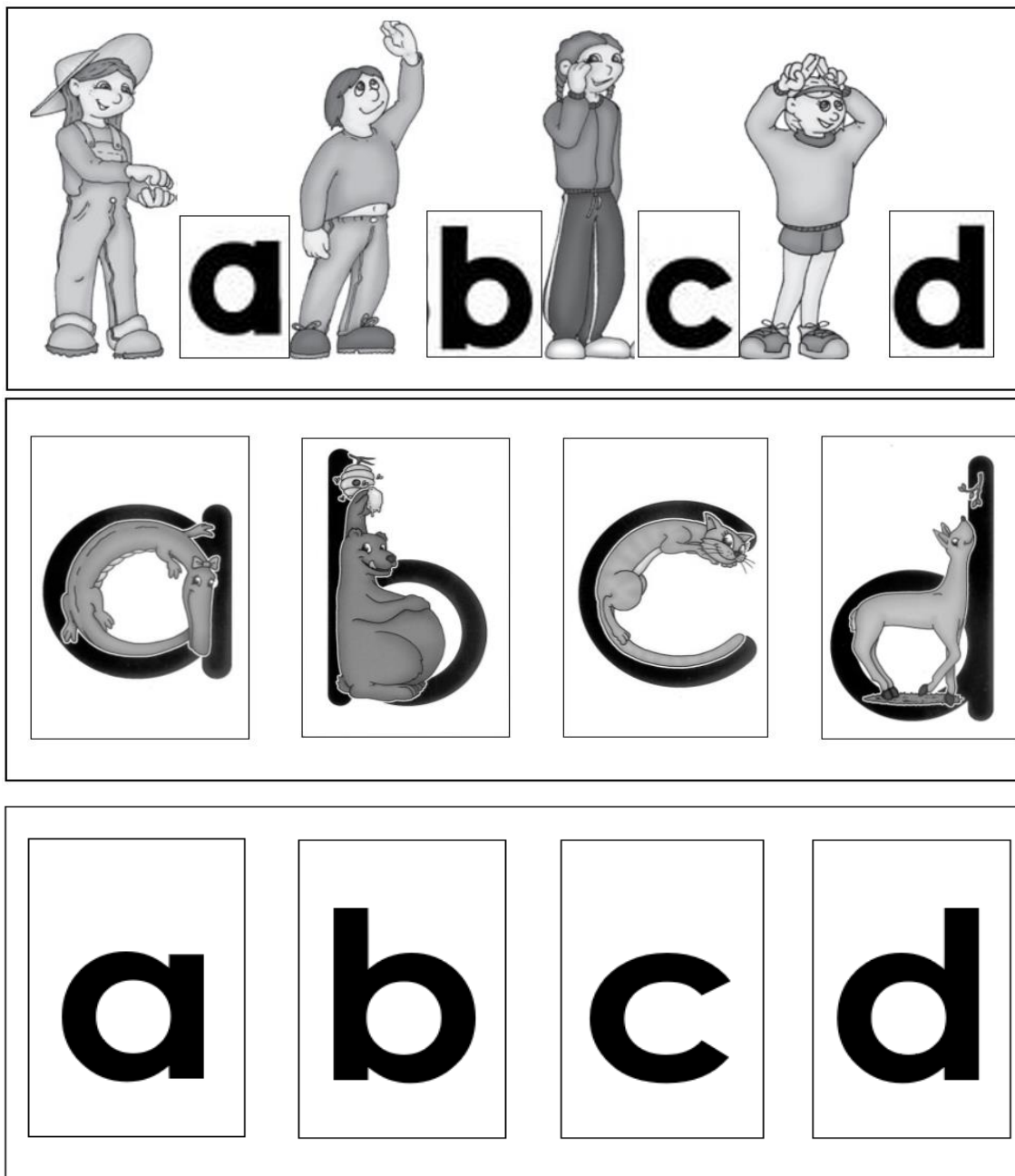
movement condition to be more effective. In addition, Nico and Ginny experienced fewer training sessions in the second evaluation compared to the first. One possible explanation for the differences in findings and number of training sessions may be due to the learning-to-learn phenomena (Green, 2001; Shepley & Grisham-Brown, 2019). Both evaluations were conducted during the summer in which participants experienced few structured periods of academic activity. Thus, it is possible that the first evaluation targeted stimulus control during structured periods of opportunities to respond to stimuli and contingencies for reinforcement. The learning-to-learn phenomena occurring during the first evaluation may have masked the differences between condition efficacies. Future researchers may also consider pairing mnemonics with arbitrary stimuli to evaluate the basic effect of mnemonic devices on acquisition. Although the use of arbitrary stimuli would decrease chance of outside instruction, it would also limit the applied value of teaching an important early literacy skill.

Despite the data demonstrating that presenting letters with movements was more effective than presenting letters alone, presenting letters with pictures, and presenting letters with pictures and movements, participants rarely selected this option during the preference assessment. In an extreme example of preference, one participant refused to engage in sessions that were not conducted in the embedded pictures condition. Preference assessment data and patterns of assent to participate in research highlight the importance of balancing efficacy and preference when selecting instructional methods. Although one condition may hinder acquisition (i.e., embedded pictures or combined mnemonics), it should still be considered when selecting intervention conditions because it may afford more opportunities to respond. Future research should evaluate the efficacy of alternating between preferred and effective paired mnemonic conditions on early literacy skills at both the individual and small-group level.

Although research has demonstrated the efficacy of combining multiple mnemonics for literacy skills (DiLorenzo et al., 2011), some mnemonics may hinder acquisition, such as embedded pictures, as demonstrated in these studies. Additionally, research in this area is primarily conducted in group designs and doesn't allow for analyses at the individual level or an evaluation of preference. Lozy et al. (2020) was the first study to compare a traditional drill and paired kinesthetic movement flashcard method in a single-subject design, and results demonstrated kinesthetic movements as being an effective procedure to increase letter-sound correspondence. The present studies attempted to further evaluate kinesthetic movements when compared to and paired with embedded pictures, and results demonstrated that pairing kinesthetic movements was generally more effective than both alternative methods as well as no intervention. The results also demonstrate the lack of continuity between the most effective and preferred condition, highlighting the importance that both variables should be considered when selecting instructional programs.

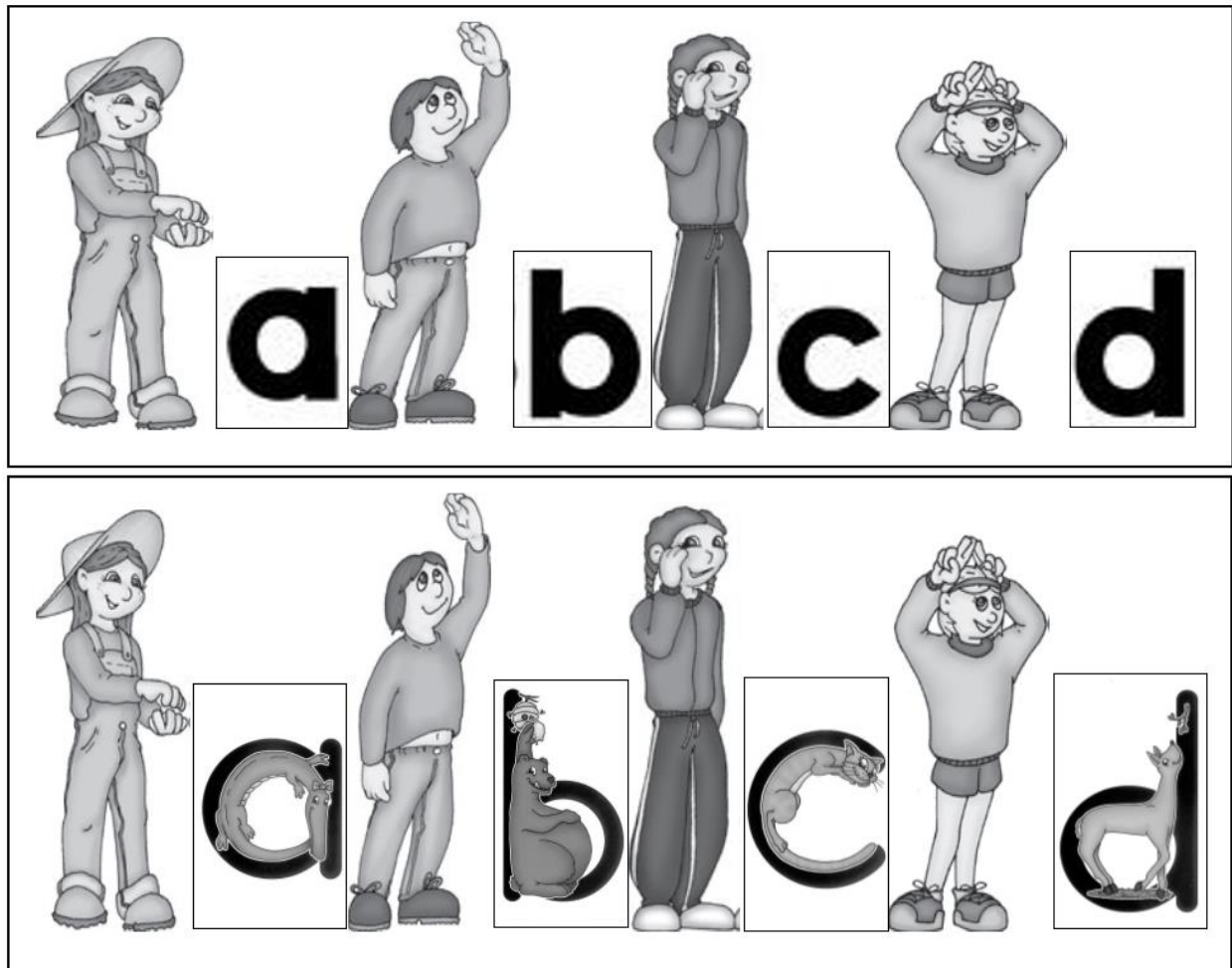
Appendix A. Study 1 Stimuli

The discriminative stimuli for the paired kinesthetic movement (KM), paired embedded pictures (EP), and traditional drill (TD) intervention conditions in Study 1.



Appendix B. Study 2 Stimuli

The discriminative stimuli for the kinesthetic movement (KM) and combined mnemonics (CM) intervention conditions in Study 2.



Appendix C. IRB Form

ACTION ON PROTOCOL APPROVAL REQUEST



Institutional Review Board
Dr. Dennis Landin, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/research

TO: Jeanne Donaldson
Psychology

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: August 31, 2017

RE: IRB# 3905

TITLE: An Evaluation of Procedures to Increase Academic Fluency in Young Children

New Protocol/Modification/Continuation: New Protocol

Review type: Full ☐ Expedited ☒ Review date: 8/22/2017

Risk Factor: Minimal ☒ Uncertain ☐ Greater Than Minimal ☐

Approved ☒ Disapproved ☐

Approval Date: 8/31/2017 Approval Expiration Date: 8/30/2018

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 30

LSU Proposal Number (if applicable):

Protocol Matches Scope of Work in Grant proposal: (if applicable)

By: Dennis Landin, Chairman 

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –

Continuing approval is **CONDITIONAL** on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc.**

**All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at <http://www.lsu.edu/irb>*

List of References

- Agramonte, V., & Belfiore, P. J. (2002). Using mnemonics to increase early literacy skills in urban kindergarten students. *Journal of Behavioral Education, 11*(3), 181-190. <http://dx.doi.org/1053-0819/02/0900-0181/0>
- Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior, 12*(6), 861-874. <http://dx.doi.org/10.1901/jeab.1969.12-861>
- Borghi, A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia, 48*(3), 763-773. <http://dx.doi.org/10.1016/j.neuropsychologia.2009.10.029>
- Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Sciences, 18*(8), 395-403. <http://dx.doi.org/10.1016/j.tics.2014.04.010>
- Callinan, C., & van der Zee, E. (2010). A comparative study of two methods of synthetic phonics instruction for learning how to read: Jolly Phonics and THRASS. *The Psychology of Education Review, 34*(1), 21-31.
- Campbell, M. L., Helf, S., & Cooke, N. L. (2008). Effects of adding multisensory components to a supplemental reading program on the decoding skills of treatment resisters. *Education and Treatment of Children, 31*(3), 267-295. <http://dx.doi.org/10.1353/etc0.0003>
- Cattinelli, I., Borghese, N. A., Gallucci, M., & Paulesu, E. (2013). Reading the reading brain: A new meta-analysis of functional imaging data on reading. *Journal of Neurolinguistics, 26*(1), 214-238. <http://dx.doi.org/10.1016/j.jneuroling.2012.08.001>
- Coleman, E. B., & Morris, G. (1978) Generalization tests: A terminology that focuses attention on fixed-effect restrictions. *Journal of Reading Behavior, 10*(4), 377-392. <http://dx.doi.org/10.1080/10862967809547291>
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934-945. <http://dx.doi.org/10.1037/0012-1649.33.6.934>
- de Graaff, S., Verhoeven, L., Bosman, A. M., & Hasselman, F. (2007). Integrated pictorial mnemonics and stimulus fading: Teaching kindergartners letter sounds. *British Journal of Educational Psychology, 77*(3), 519-539. <http://dx.doi.org/10.1348/00070906X160011>
- DeLeon, I. G., Chase, J. A., Frank - Crawford, M. A., Carreau - Webster, A. B., Triggs, M. M., Bullock, C. E., & Jennett, H. K. (2014). Distributed and accumulated reinforcement arrangements: Evaluations of efficacy and preference. *Journal of Applied Behavior Analysis, 47*(2), 293-313. <http://dx.doi.org/10.1002/jaba.116>

- Donnelly, J. E., & Lambourne, K. (2011). Classroom-based physical activity, cognition, and academic achievement. *Preventive Medicine*, 52(1), 36-42.
<http://dx.doi.org/10.1016/j.ypmed.2011.01.021>
- Didden, R., de Graaff, S., Nelemans, M., Vooren, M., & Lancioni, G. (2006). Teaching sight words to children with moderate to mild mental retardation: Comparison between instructional procedures. *American Journal on Mental Retardation*, 111(5), 357-365.
- Didden, R., Prinsen, H., & Sigafoos, J. (2000). The blocking effect of pictorial prompts on sight-word reading. *Journal of Applied Behavior Analysis*, 33(3), 317-320.
<http://dx.doi.org/10.1901/jaba.2000.33-317>
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, 118(1), 114-118. <http://dx.doi.org/10.1016/j.cognition.2010.09.012>.
- DiLorenzo, K. E., Rody, C. A., Bucholz, J. L., & Brady, M. P. (2011). Teaching letter-sound connections with picture mnemonics: Itchy's alphabet and early decoding. *Preventing School Failure: Alternative Education for Children and Youth*, 55(1), 28-34.
<http://dx.doi.org/10.1080/10459880903286763>
- Dittlinger, L. H., & Lerman, D. C. (2011). Further analysis of picture interference when teaching word recognition to children with autism. *Journal of Applied Behavior Analysis*, 44(2), 341-349. <http://dx.doi.org/10.1901/jaba.2011.44-341>
- Dresler, M., Shirer, W. R., Konrad, B. N., Müller, N. C., Wagner, I. C., Fernández, G., ... & Greicius, M. D. (2017). Mnemonic training reshapes brain networks to support superior memory. *Neuron*, 93(5), 1227-1235. <https://doi.org/10.1016/j.neuron.2017.02.003>
- Fenerty, K. A., & Tiger, J. H. (2010). Determining preschoolers' preferences for choice-making opportunities: Choice of task versus choice of consequence. *Journal of Applied Behavior Analysis*, 43(3), 503-507. <http://dx.doi.org/10.1901/jaba.2010.43-503>
- Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3), 319-325. <http://dx.doi.org/10.1002/wcs.1226>
- Friedel, J. E., DeHart, W. B., Foreman, A. M., & Andrew, M. E. (2019). A Monte Carlo method for comparing generalized estimating equations to conventional statistical techniques for discounting data. *Journal of the Experimental Analysis of Behavior*, 111(2), 207-224.
<http://dx.doi.org/10.1002/jeab.497>
- Fulk, B. M., Lohman, D., & Belfiore, P. J. (1997). Effects of integrated picture mnemonics on the letter recognition and letter-sound acquisition of transitional first-grade students with special needs. *Learning Disability Quarterly*, 20(1), 33-42.
<http://dx.doi.org/10.2307/1511091>

- Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(2), 203-213. <http://dx.doi.org/10.1017/S0021963099005284>
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, 96(3), 424-436. <http://dx.doi.org/10.1037/0022-0663.96.3.424>
- Green, G. (2001). Behavior analytic instruction for learners with autism: Advances in stimulus control technology. *Focus on Autism and Other Developmental Disabilities*, 16(2), 72-85. <http://dx.doi.org/10.1177/108835760101600203>
- Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, 58(6-7), 466-474. <http://dx.doi.org/10.1037/0003-066X.58.6-7.466>
- Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain and Cognition*, 79(3), 221-244. <https://doi.org/10.1016/j.bandc.2012.01.010>
- Hutchison, W. (1997). We also need complete behavioral models. *Journal of the Experimental Analysis of Behavior*, 67(2), 224-228. <http://dx.doi.org/10.1901/jeab.1997.67-224>
- Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children*, 70(4), 453-469. <http://dx.doi.org/10.1177/001440290407000405>
- Hanley, G. P., Iwata, B. A., & Lindberg, J. S. (1999). Analysis of activity preferences as a function of differential consequences. *Journal of Applied Behavior Analysis*, 32(4), 419-435. <http://dx.doi.org/10.1901/jaba.1999.32-419>
- Hanley, G. P., Piazza, C. C., Fisher, W. W., Contrucci, S. A., & Maglieri, K. A. (1997). Evaluation of client preference for function-based treatment packages. *Journal of Applied Behavior Analysis*, 30(3), 459-473. <http://dx.doi.org/10.1901/jaba.1997.30.459>
- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis*, 38(1), 51-65. <http://dx.doi.org/10.1901/jaba.2005.6-04>
- Hanley, J. A., Negassa, A., Edwardes, M. D. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American Journal of Epidemiology*, 157(4), 364-375. <http://dx.doi.org/10.1093/aje/kwf215>

- Harzem, P., Lee, I., & Miles, T. R. (1976). The effects of pictures on learning to read. *British Journal of Educational Psychology*, 46(3), 318-322.
<http://dx.doi.org/10.1111/j.2044-8279.1976.tb02328.x>
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 220–227. <http://dx.doi.org/10.1177/2372732216655542>
- Hetzroni, O. E., & Shavit, P. (2002). Comparison of two instructional strategies for acquiring form and sound of Hebrew letters by students with mild mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities*, 37(3), 273-282.
<http://dx.doi.org/jstor.stable/23880005>
- Higbee, K. L. (1979). Recent research on visual mnemonics: Historical roots and educational fruits. *Review of Educational Research*, 49(4), 611-629.
<http://dx.doi.org/10.3102/00346543049004611>
- Hoffmann, A. N., Samaha, A. L., Bloom, S. E., & Boyle, M. A. (2017). Preference and reinforcer efficacy of high - and low - tech items: A comparison of item type and duration of access. *Journal of Applied Behavior Analysis*, 50(2), 222-237.
<http://dx.doi.org/10.1002/jaba.350>
- Hoogeveen, F. R., Smeets, P. M., & Lancioni, G. E. (1989). Teaching moderately mentally retarded children basic reading skills. *Research in Developmental Disabilities*, 10(1), 1-18. [http://dx.doi.org/10.1016/0891-4222\(89\)90025-5](http://dx.doi.org/10.1016/0891-4222(89)90025-5)
- Horowitz-Kraus, T., Buck, C., and Dorrman, D. (2016). Altered neural circuits accompany lower performance during narrative comprehension in children with reading difficulties: An fMRI study. *Annals of Dyslexia*, 66(3), 301–318.
<http://dx.doi.org/10.1007/s11881-016-0124-4>
- Johnson, D. F., & Cumming, W. W. (1968). Some determiners of attention. *Journal of the Experimental Analysis of Behavior*, 11(2), 157–166.
<http://dx.doi.org/10.1901/jeab.1968.11-157>
- Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, 106(1), 86-104. <http://dx.doi.org/10.1037/a0034008>
- Keage, H. A., Coussens, S., Kohler, M., Thiessen, M., & Churches, O. F. (2014). Investigating letter recognition in the brain by varying typeface: An event-related potential study. *Brain and Cognition*, 88, 83-89. <http://dx.doi.org/10.1016/j.bandc.2014.05.001>
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

- Kibbe, D. L., Hackett, J., Hurley, M., McFarland, A., Schubert, K. G., Schultz, A., & Harris, S. (2011). Ten Years of TAKE 10!®: Integrating physical activity with academic concepts in elementary school classrooms. *Preventive Medicine*, 52(1), 43-50. <http://dx.doi.org/10.1016/j.ypmed.2011.01.025>
- Kodak, T., Halbur, M., Bergmann, S., Costello, D. R., Benitez, B., Olsen, M., et al. (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>
- Kosmas, P., Ioannou, A., & Zaphiris, P. (2019). Implementing embodied learning in the classroom: effects on children’s memory and language skills. *Educational Media International*, 56(1), 59-74. <https://doi.org/10.1080/09523987.2018.1547948>
- Kosmas, P. & Zaphiris P. (2018). Embodied cognition and its implications in education: An overview of recent literature. *International Journal of Educational and Pedagogical Sciences*, 12(7), 970-976. <http://dx.doi.org/10.1999/1307-6892/10009334>
- Kupzyk, S., Daly, E. J., III, & 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. <http://dx.doi.org/10.1901/jaba.2011.44-781>
- Labat, H., Ecalte, J., Baldy, R., & Magnan, A. (2014). How can low-skilled 5-year-old children benefit from multisensory training on the acquisition of the alphabetic principle? *Learning and Individual Differences*, 29(1), 106-113. <http://dx.doi.org/10.1016/j.lindif.2013.09.016>
- Lang, R. J., & Solman, R. T. (1979). Effect of pictures on learning to read common nouns. *British Journal of Educational Psychology*, 49(2), 138-149. <http://dx.doi.org/10.1111/j.2044-8279.1979.tb02408.x>
- Lewallen, T. C., Hunt, H., Potts-Datema, W., Zaza, S., & Giles, W. (2015). The whole school, whole community, whole child model: A new approach for improving educational attainment and healthy development for students. *Journal of School Health*, 85(11), 729-739. <http://dx.doi.org/10.1111/josh.12310>
- Levin, J. R. (1993). Mnemonic strategies and classroom learning: A twenty-year report card. *The Elementary School Journal*, 94(2), 235-244. <http://dx.doi.org/10.1086/461763>
- Lippman, M. Z., & Shanahan, M. W. (1973). Pictorial facilitation of paired-associate learning: Implications for vocabulary training. *Journal of Educational Psychology*, 64(2), 216-222. <http://dx.doi.org/10.1037/h0034602>
- Liu, X., Liu, C. H., & Li, Y. (2020). The effects of computer-assisted learning based on dual coding theory. *Symmetry*, 12(5), 700-713. <http://dx.doi.org/10.3390/sym12050701>

- Lonigan, C. J., Bloomfield, B. G., Anthony, J. L., Bacon, K. D., Phillips, B. M., & Samwel, C. S. (1999). Relations among emergent literacy skills, behavior problems, and social competence in preschool children from low-and middle-income backgrounds. *Topics in Early Childhood Special Education*, 19(1), 40-53. <http://dx.doi.org/10.1177/027112149901900104>
- Lowenkron, B. (1988). Generalization of delayed identity matching in retarded children. *Journal of the Experimental Analysis of Behavior*, 50(2), 163–172. <http://dx.doi.org/10.1901/jeab.1988.50-163>
- Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior*, 69(3), 327-354. <http://dx.doi.org/10.1901/jeab.1998.69-327>
- Lozy, E. D., Holmes, S. C., & Donaldson, J. M. (2020). The effects of paired kinesthetic movements on literacy skills acquisition with preschoolers. *Journal of Applied Behavior Analysis* <http://dx.doi.org/10.1002/jaba.677>
- Lozy, E. D., & Donaldson, J. M. (2019). A comparison of traditional drill and strategic incremental rehearsal flashcard methods to teach letter–sound correspondence. *Behavioral Development*, 24(2), 58--73. <https://doi.org/10.1037/bdb0000089>
- MacQuarrie, L. L., & Tucker, J. A. (2002). Comparison of retention rates using traditional, drill sandwich, and incremental rehearsal flash card methods. *School Psychology Review*, 31(4), 534-595.
- Mahdi, H. S., & Gubeily, M. A. I. (2018). The effect of using bizarre images as mnemonics to enhance vocabulary learning. *Journal of Social Studies*, 24(1), 113-135. <http://dx.doi.org/10.20428/JSS.24.1.5>
- Manalo, E., Uesaka, Y., & Sekitani, K. (2013). Using mnemonic images and explicit sound contrasting to help Japanese children learn English alphabet sounds. *Journal of Applied Research in Memory and Cognition*, 2(4), 216-221. <http://dx.doi.org/10.1016/j.jarmac.2013.09.003>
- Marsh, G., & Desberg, P. (1978). Mnemonics for phonics. *Contemporary Educational Psychology*, 3(1), 57-61. [http://dx.doi.org/10.1016/0361-476X\(78\)90009-7](http://dx.doi.org/10.1016/0361-476X(78)90009-7)
- Marsh, G., Desberg, P., & Farwell, L. K. (1974). Stimulus and response variables in children's learning of grapheme-phoneme correspondences. *Journal of Educational Psychology*, 66(1), 112-116. <http://dx.doi.org/10.1037/h0035810>
- Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping*, 36(5), 1963-1981. <http://dx.doi.org/10.1002/hbm.22749>

- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63(8), 760-769. <http://dx.doi.org/10.1037/0003-066X.63.8.760>
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8(4), 357-371. <http://dx.doi.org/1040-726X/96/1200-0357509.50>
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320. <http://dx.doi.org/10.1037/0022-0663.90.2.312>
- Meteyard, L., Cuadrado, S. R., Bahrami, B., and Viglicco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48(7), 788–804. <http://dx.doi.org/10.1016/j.cortex.2010.11.002>
- Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior*, 27(1), 3-22. <http://dx.doi.org/10.1007/BF03393089>
- Moore, P., & Fitz, C. (1993). Gestalt theory and instructional design. *Journal of Technical Writing and Communication*, 23(2), 137-157. <https://doi.org/10.2190/G748-BY68-L83T-X02J>
- Morrisroe, J. (2014). Literacy Changes Lives 2014: A New Perspective on Health, Employment and Crime. *National Literacy Trust*.
- National Early Literacy Panel. (2008). *Developing early literacy: Report of the national early literacy panel*. Washington, DC: National Center for Family Literacy
- National Governors Association. (2010). Common core state standards. *Washington, DC*.
- Nguyen, D., J. & Larson, J. B. (2015). Don't forget about the body: Exploring the curricular possibilities of embodied pedagogy. *Innovative Higher Education*, 40(4), 331–344. <http://dx.doi.org/10.1007/s10755-0159319-6>
- Pelli, D. G., Majaj, N. J., Raizman, N., Christian, C. J., Kim, E., & Palomares, M. C. (2009). Grouping in object recognition: The role of a Gestalt law in letter identification. *Cognitive Neuropsychology*, 26(1), 36-49. <https://doi.org/10.1080/13546800802550134>
- Petursdottir, A. I., & Carr, J. E. (2018). Applying the taxonomy of validity threats from mainstream research design to single-case experiments in applied behavior analysis. *Behavior Analysis in Practice*, 11(3), 228-240. <http://dx.doi.org/10.1007/s40617-018-00298-6>

- Phillips, B. M., Piasta, S. B., Anthony, J. L., Lonigan, C. J., & Francis, D. J. (2012). IRTs of the ABCs: Children's letter name acquisition. *Journal of School Psychology, 50*(4), 461-481. <http://dx.doi.org/10.1016/j.jsp.2012.05.002>
- Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological science in the public interest, 10*(2), 49-88. <http://dx.doi.org/10.1177/1529100610381908>
- Piasta, S. B., Phillips, B. M., Williams, J. M., Bowles, R. P., & Anthony, J. L. (2016). Measuring young children's alphabet knowledge: Development and validation of brief letter-sound knowledge assessments. *The Elementary School Journal, 116*(4), 523-548. <http://dx.doi.org/10.1086/686222>
- Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly, 45*(1), 8-38. <http://dx.doi.org/10.1598/RRQ.45.1>
- Raschke, D., Alper, S., & Eggers, E. (1999). Recalling alphabet letter names: A mnemonic system to facilitate learning. *Preventing School Failure: Alternative Education for Children and Youth, 43*(2), 80-83. <http://dx.doi.org/10.1080/10459889909603306>
- 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. <http://dx.doi.org/10.1002/jaba.354>
- Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham-based reading instruction: A review of the literature. *The Journal of Special Education, 40*(3), 171-183. <http://dx.doi.org/10.1177/00224669060400030501>
- Rivera, S., Reiss, A., Eckert, M. & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex, 15*(11), 1779-1790. <http://dx.doi.org/10.1093/cercor/bhi055>
- Samuels, S. J. (1967). Attentional process in reading: The effect of pictures on the acquisition of reading responses. *Journal of Educational Psychology, 58*(6p1), 337-342. <http://dx.doi.org/10.1037/h0020045>
- Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*(1), 55-63. <http://dx.doi.org/10.1044/jsdh.3701.55>
- Scoville, W. B., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. *The Journal of Neuropsychiatry and Clinical Neurosciences, 12*(1), 103-113. <https://doi.org/10.1176/jnp.12.1.103-a>

- Sener, U., & Belfiore, P. J. (2005). Mnemonics strategy development: Improving alphabetic understanding in Turkish students, at risk for failure in EFL settings. *Journal of Behavioral Education, 14*(2), 105-115. <http://dx.doi.org/10.1007/s10864-005-2705-x>
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology, 22*(2), 261-273. <http://dx.doi.org/10.1080/00335557043000203>
- Shepley, C., & Grisham-Brown, J. (2019). Applied behavior analysis in early childhood education: An overview of policies, research, blended practices, and the curriculum framework. *Behavior Analysis in Practice, 12*(1), 235-246. <http://dx.doi.org/10.1007/s40617-018-0236-x>
- Shmidman, A., & Ehri, L. (2010). Embedded picture mnemonics to learn letters. *Scientific Studies of Reading, 14*(2), 159-182. <http://dx.doi.org/10.1080/10888430903117492>
- Sidener D. W. (2006). Joint control for dummies*: An elaboration of Lowenkron's model of joint (stimulus) control. *The Analysis of Verbal Behavior, 22*(1), 119-122. <http://dx.doi.org/10.1007/bf03393033>
- Sidman, M. (1960). *Tactics of experimental research: Evaluating experimental data in psychology*. New York: Basic Books.
- Singh, N. N., & Solman, R. T. (1990). A stimulus control analysis of the picture-word problem in children who are mentally retarded: The blocking effect. *Journal of Applied Behavior Analysis, 23*(4), 525-532. <http://dx.doi.org/10.1901/jaba.1990.23-525>
- Smith, G. J., Booth, J. R., and McNorgan, C. (2018) Longitudinal task-related functional connectivity changes predict reading development. *Frontiers in Psychology, 60*(1), 1-13. <http://dx.doi.org/10.3389/fpsyg.2018.01754>
- Stromer, R., & Dube, W. V. (1994). Differential observing of complex sample stimuli and delayed matching performance: A brief report. *Experimental Analysis of Human Behavior Bulletin, 12*(1), 17-20.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. [http://dx.doi.org/10.1016/0364-0213\(88\)90023-7](http://dx.doi.org/10.1016/0364-0213(88)90023-7)
- Thorpe, H. W., & Borden, K. S. (1985). The effect of multisensory instruction upon the on-task behaviors and word reading accuracy of learning disabled children. *Journal of Learning Disabilities, 18*(5), 279-286. <http://dx.doi.org/10.1177/002221948501800507>
- Tiger, J. H., Hanley, G. P., & Heal, N. A. (2006)^a. The effectiveness of and preschoolers' preferences for variations of multiple-schedule arrangements. *Journal of Applied Behavior Analysis, 39*(4), 475-488. <http://dx.doi.org/10.1901/jaba.2006.48-06>

- Tiger, J. H., Hanley, G. P., & Hernandez, E. (2006)^b. An evaluation of the value of choice with preschool children. *Journal of Applied Behavior Analysis*, 39(1), 1-16. <http://dx.doi.org/10.1901/jaba.2006.158-04>
- Tiger, J. H., Hanley, G. P., & Larsen, K. M. (2008). A practical variation of a multiple-schedule procedure: Brief schedule-correlated stimuli. *Journal of Applied Behavior Analysis*, 41(1), 125-130. <http://dx.doi.org/10.1901/jaba.2008.41-125>
- Torgrud, L. J., & Holborn, S. W. (1989). Effectiveness and persistence of precurent mediating behavior in delayed matching to sample and oddity matching with children. *Journal of the Experimental Analysis of Behavior*, 52(2), 181-191. <http://dx.doi.org/10.1901/jeab.1989.52-181>
- Tu, J. C. (2006). The role of joint control in the manded selection responses of both vocal and non-vocal children with autism. *The Analysis of Verbal Behavior*, 22(1), 191-207. <http://dx.doi.org/10.1007/BF033930>
- Tucker, J. A. (1988). *Basic flashcard technique when vocabulary is the goal. Unpublished teacher material*. Chattanooga, TN: University of Tennessee at Chattanooga.
- U.S. Department of Education, National Center for Education Statistics. (2016). *The Condition of Education 2016*
- Urcuioli, P. J., & Callender, J. (1989). Attentional enhancement in matching-to-sample: Facilitation in matching acquisition by sample-discrimination training. *Animal Learning & Behavior*, 17(3), 361-367.
- Wang, J. X., Bartolotti, J., Amaral, L. A., and Booth, J. R. (2013). Changes in task-related functional connectivity across multiple spatial scales are related to reading performance. *PLoS One*, 8(3), 1-13. <http://dx.doi.org/10.1371/journal.pone.0059204>
- Whitley, B. E., & Kite, M. E. (2013). *Principles of research in behavioral science*. Routledge.
- Willis, J. (2008). *Teaching the brain to read: Strategies for improving fluency, vocabulary, and comprehension*. Association for Supervision and Curriculum Development.
- Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4(1), 58-65. <http://dx.doi.org/10.3389/fpsyg.2013.00058>
- Wunderlich, K. L., & Vollmer, T. R. (2017). Effects of serial and concurrent training on receptive identification tasks: A Systematic replication. *Journal of Applied Behavior Analysis*, 50(3), 641-652. <http://dx.doi.org/10.1002/jaba.4010>
- Wunderlich, K. L., Vollmer, T. R., Donaldson, J. M., & Phillips, C. L. (2014). Effects of serial and concurrent training on acquisition and generalization. *Journal of Applied Behavior Analysis*, 47(4), 723-737. <http://dx.doi.org/10.1002/jaba.152>

Vita

Erica Deborah Lozy was born in North Smithfield, Rhode Island. She received her bachelor's degree in Psychology from Western New England University, and subsequently her master's degree with a focus in Applied Behavior Analysis from University of Maryland Baltimore County. Erica is interest in conducting research and clinical work with young children who have difficulties with academic achievement and/or engage in externalizing problem behavior Upon completion of her doctorate from Louisiana State University, she will join a private practice in Tampa, Florida.