



Efficacy of Picture-Text Compound Prompts During Sight-Word Instruction

Taylor K. Lewis¹ · Tom Cariveau¹  · Alexandria Brown¹ · Paige Ellington¹

Accepted: 30 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Researchers have continually sought to refine interventions targeting reading-related skills in an effort to improve their efficacy, efficiency, or social validity. Despite their prominence in early reading materials, pictures are often excluded from reading intervention procedures, likely because pictures have been shown to impede stimulus control by the textual stimuli when pictures and text are presented simultaneously. The current study describes the use of a novel prompt type, picture-text compound prompts, embedded in a strategic incremental rehearsal procedure to teach sight words to four children exhibiting reading challenges. Prompted opportunities resulted in the presentation of four picture-text compounds (e.g., the word *dog* appeared below a picture of a dog). To correctly respond, the participant was required to match the identical target textual stimulus to the element in the corresponding compound. Doing so then allowed the picture element of the compound stimulus to function as a tact prompt. Picture-text compound prompts produced mastery levels of responding in three or fewer instructional sessions and participants' responses during generalization and maintenance probes were generally high. The current findings suggest that picture-text compounds prompts are efficacious, although additional research is needed to evaluate the relative efficiency and social validity of similar arrangements.

Keywords Literacy · Picture-text problem · Reading · Response prompting

✉ Tom Cariveau
cariveaut@uncw.edu

¹ Department of Psychology, University of North Carolina Wilmington, 601 S. College Rd, Wilmington, NC 28403, USA

Introduction

Preventing or remediating early reading difficulties represents a fundamental concern in education (NICHHD, 2000; Snow et al., 1998), yet many children are still unable to read at a basic level (NAEP, 2022). Recommended reading interventions commonly emphasize explicit instruction to teach discrete skills and repeated opportunities for the learner to practice the skill (Carnine et al., 2017). Although the general curriculum may provide some of these opportunities, students might also require supplemental instruction to enhance reading outcomes (O'Connor et al., 2005; Vaughn et al., 2009). Consistent with tiered systems of support (see Fuchs & Vaughn, 2012; Vaughn & Fuchs, 2003), these supplemental interventions can be individualized to a particular learner's needs to better support their reading development and ensure that they can access the general curriculum. Individualizing instruction represents a significant challenge for instructors as they must (a) determine the specific needs of the student, (b) prepare or identify instructional materials to appropriately target the student's needs, and (c) deliver the supplemental instruction. Therefore, methods that reduce demands on instructors at any of these steps or can teach the requisite skills rapidly are invaluable in educational settings.

Attempts to deliver individualized instruction may contact significant barriers in educational settings and students may not receive the intensity of instruction that is necessary to produce sufficient gains in reading (McIntosh et al., 1993; McKenna et al., 2015). Given constraints on educators' time and the diverse needs of all students, it may be critical that instructional methods aimed to remediate reading deficits can be delivered without direct support from an instructor. Some forms of instruction seemingly necessitate the presence of an instructor to prompt and reinforce correct responding. Specifically, because reading requires vocal (or subvocal) responding to textual stimuli, instructors may commonly utilize echoic prompts to evoke the target vocal response (e.g., Kupzyk et al., 2011; Lozy & Donaldson, 2019; Rahn et al., 2015). Echoic prompts may be common in reading instruction simply due to their availability: an instructor can immediately prompt any vocal response. Nevertheless, other prompting methods might also be effective and confer other advantages (e.g., require less oversight by instructors).

A prompt is defined as any stimulus that currently evokes the target response and can be used to transfer control to the target conditions (Deitz & Malone, 1985). Although echoic prompts predominate in early reading instruction, other prompt types serve as viable alternatives. Tact prompts, which include the presentation of a nonverbal discriminative stimulus (e.g., picture) to evoke the target response (Feng et al., 2015; Vedora & Conant, 2015), may be used in early reading as a child is likely able to name considerably more stimuli than they can read. Tact prompts might confer additional benefits, such as being socially valid. Indeed, classrooms and early reading materials are replete with pictures (Hodkinson, 2017). Nevertheless, using pictures to transfer control to textual stimuli remains a source of controversy, particularly as pictures have been shown to impede control by the textual stimulus for some learners (Richardson et al., 2017; Samuels, 1967; Singh & Solman, 1990; see review by Kennedy & Cariveau, 2023).

The detrimental effect of pictures on word reading has been referred to as the picture-text problem (Kennedy & Cariveau, 2023). Research on the picture-text problem has typically followed a two-step design. First, instructional trials are presented, which include the simultaneous presentation of a single unknown textual stimulus (e.g., written word *cat*) and a representative picture (e.g., a picture of a cat). Following instruction, the textual stimuli are presented alone during test trials. Several authors have found that accurate responding is observed during training; however, performance during text-only probes remains at pre-training levels (Samuels, 1967) or is significantly affected compared to instruction with the text alone (Richardson et al., 2017; Singh & Solman, 1990). This finding suggests that the learner's performance was controlled exclusively by the picture in the compound. The finding that pictures hinder word-reading performances may discourage educators from using tact prompts; however, alternative stimulus arrangements might also prevent the picture-text problem.

Singh and Solman (1990) were the first to conceptualize the picture-text problem as overselective stimulus control by the elements of a compound stimulus. These authors suggested that overselective control resulted from the simultaneous presentation of picture and text, which allowed for the textual stimulus to be treated as redundant. When this is the case, the textual element of the compound stimulus would be described as the *undersampled* element (terms used by Broomfield et al., 2008, 2010). Several studies have described successful methods to produce greater control by the undersampled elements (e.g., Broomfield et al., 2008, 2010; Gomes-Ng et al., 2023; Reed et al., 2012). As one example, Dube and McIlvane (1999) showed greater control by the undersampled element when learners were first required to match the identical compound stimuli on every trial. Alternatively, arrangements that require only differential responding to the undersampled element (i.e., a differential observing response) might also be used to remediate previously overselective performances (e.g., Walpole et al., 2007). Such a preparation may be readily arranged in early reading materials and also allow for the use of tact prompts during instruction.

Strategic Incremental Rehearsal

Strategic incremental rehearsal is a trial-based instructional procedure characterized by an incrementing set size (Kupzyk et al., 2011). Specifically, during each instructional session, a subset of unknown targets (e.g., two) are initially presented and additional targets are added to the instructional session once the participant exhibits accurate responding to each target on consecutive presentations. Thus, the number of targets presented during instruction increases over the course of a session based on the learner's performance. Instruction continues until either (a) correct responding is observed to all targets in a set or (b) some time-based termination criterion is met (e.g., 8 min; Kupzyk et al., 2011). Several studies have found SIR to be efficacious in promoting reading-related performances (e.g., sight-word reading, letter-sound correspondence; Hathaway et al., 2021; January et al., 2017; Kupzyk et al., 2011; Phipps et al., 2022), two of which found SIR to be more efficient than

similar trial-based procedures (January et al., 2017; Kupzyk et al., 2011). Notably, descriptions of SIR typically include some reference to flashcards, although Lewis et al. (2024) suggested that the efficacy of SIR is likely not due to the instructional modality (i.e., flashcards) but instead to the systematic introduction of targets into an instructional set. In their work, Lewis et al. arranged SIR using a printed word list (SIR–WL; see example in Fig. 1) to teach novel sight words to four elementary-aged participants exhibiting reading challenges. During instruction, the experimenter used a blank sheet of paper to cover the word list so that only the target row of words was presented. Once the participant responded accurately to all the targets in the row, the experimenter moved the paper to reveal another row of words which included the previously presented target words and an additional novel target. Instruction continued until all targets were introduced and the participant correctly responded to the entire set or 3 min of instruction elapsed. The authors found that SIR presented in a word list was consistently effective and resulted in mastery and response maintenance in as few as two instructional sessions.

To date, every study on SIR has used echoic prompts during instruction. Therefore, the current study extended the work of Lewis et al. (2024) by embedding picture-text compound prompts during SIR–WL for elementary-age children exhibiting reading deficits. Like Lewis et al., we included measures of generalization and maintenance as additional metrics of instructional efficacy and efficiency.

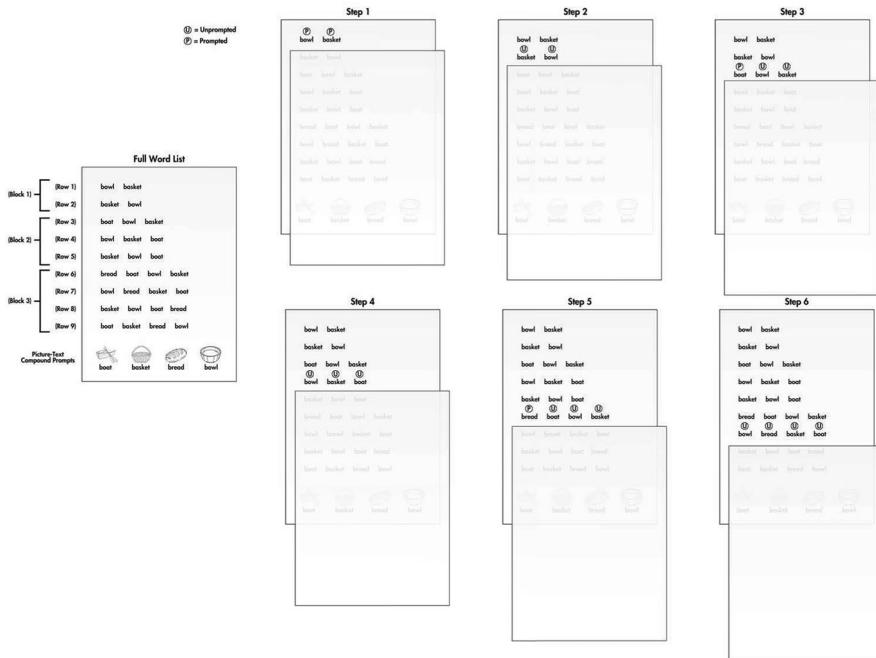


Fig. 1 Diagram of SIR–WL with picture-text compound prompts. SIR–WL= strategic incremental rehearsal–word list

Method

Participants and Setting

Four children exhibiting reading deficits participated. All participants attended a high-poverty (McFarland et al., 2019) elementary school in the southeastern US. Participants were referred by their teachers as being among the lowest performing students in their class. Participants' legal guardian provided permission to participate in the study and participants provided daily assent. If a participant did not provide assent for more than two consecutive days, they were no longer included in the study. No participant met this criterion during the study. All participants' word-reading performance was assessed using the easyCBM Word Reading Fluency (WRF) benchmark assessment (Alonzo & Tindal, 2010). Piper was a Black female enrolled in kindergarten. She performed at the 27th percentile on the WRF benchmark. Silas, Juliette, and Marlie were all enrolled in 1st grade. Silas was a Black male. He performed at the 20th percentile on the WRF measure. Juliette and Marlie were Black females and performed at the 18th and 4th percentiles on the WRF benchmark, respectively.

All sessions took place in a classroom with individual workspaces for each experimenter–participant dyad. The participants' desks were separated by at least 2 m and faced away from any other student in the room. The experimenter sat next to the participant.

Materials

The experimenter maintained data collection materials (e.g., pens, timers, clipboards) and tangible items. The participant received access to the tangible items during 2-min breaks following each instructional session. Participants selected tangible items from an array of toy bins at the beginning of each day. A participant-specific word list included four target words printed across nine rows on a 21.6 cm by 27.9 cm sheet of white paper (see Fig. 1). All words appeared in size-14 Futura font. This font was selected as it is a sans serif font and includes a single-story *a* (rather than double-story *a*), which is more consistent with letter forms commonly used in other early reading materials in educational settings. Picture-text compound prompts appeared in a row at the bottom of the SIR–WL page. A separate word list was used during daily probes which included the eight target words from both target sets randomly interspersed across two rows of four words each. A 22.9 cm by 31.8 cm whiteboard and 17.8 cm by 24.1 cm iPad© were used during generalization assessments (described below).

Dependent Variables

Independent observers recorded unprompted correct and incorrect responses. Data collectors recorded an unprompted correct response if the participant emitted the predefined response within 5 s of the textual stimulus being revealed on the word

list. The participant was allowed 5 s to respond to each textual stimulus in a row. Self-corrections that occurred before the presentation of the next target or feedback were also scored as correct. Data collectors recorded an unprompted incorrect response if the participant did not respond within 5 s or any response other than the target response was emitted. We calculated the percentage of unprompted correct responses by dividing the number of unprompted correct responses by the total number of unprompted correct and incorrect responses and multiplying by 100.

Observers also recorded the number of exposures to each target stimulus during instructional sessions by tallying the number of presentations during picture-text compound prompting sessions. We calculated the total number of exposures to mastery for each target by summing the number of exposures during all picture-text compound prompting sessions until responding met the mastery criterion.

Design

We evaluated the effects of picture-text compound prompting during SIR–WL on the acquisition, generalization, and maintenance of sight words using a concurrent multiple-baseline design across target sets. We selected a concurrent multiple-baseline design as we hypothesized that history represented the greatest threat to internal validity. Specifically, because the participants received instruction outside of the research protocol, it was possible that the participant would be exposed to the targets outside of research sessions. The staggered panel of the multiple-baseline design allowed for the detection of historical and repeated testing as threats to internal validity. The staggered panel also allowed for replication. Specifically, once responding met the mastery criterion to the targets in the top panel, the experimenter introduced picture-text compound prompting to the targets in the staggered panel.

Procedure

Pre-assessment

All target sight words were nouns. The experimenter prepared a pool of targets for each participant based on their performance on prior assessments (i.e., easyCBM and mastered Fry word list; Fry, 1980). Specifically, the experimenter selected targets that had a similar number of letters to words correctly read during these assessments. The experimenter first conducted a textual pre-assessment. Each trial of the pre-assessment included the presentation of a textual stimulus and the instruction to “read it.” Following a correct response, the experimenter delivered praise and removed the word from the list of potential targets. Incorrect responses were unsequated. Targets were assigned to a target set only if the participant emitted an incorrect response during the textual pre-assessment. Next, the experimenter conducted a tact pre-assessment using graphics from www.canva.com. The experimenter typed the target word (e.g., *boat*) into the search bar of the Canva application on an iPad, selected *graphics*, and instructed the participant to tact the visual stimuli. If the participant emitted a correct tact, the experimenter delivered

praise and the participant was allowed to select one of the visual stimuli to be used as the tact prompt in the compound. If the participant selected a visual stimulus that could be confused when presented alone as a prompt, the experimenter instructed the participant to select a different stimulus. For example, if the target textual stimulus was *boat* and the participant selected a picture that included a character fishing in a boat, the experimenter would instruct the participant to select a picture that only included a boat.

The experimenter attempted to equate targets across instructional sets using logical analysis procedures outlined by Cariveau et al., (2021, 2022). Each target set included textual stimuli with a similar number of letters and syllables. The experimenter also required that all targets in a set began with the same first letter (e.g., fire, fork, frog, fish) except for Silas' first evaluation due to experimenter oversight.

Daily Probe

The experimenter conducted a single daily probe before instruction began on the target words each day. All targets from Sets 1 and 2 were randomly presented across two rows at the top of the page. The experimenter instructed the participant to "read the words." Correct responses produced praise and incorrect responses resulted in no differential consequences. If the participant's responses met the mastery criterion of 100% correct responses during the daily probe, then no additional instruction was presented for those targets. If the participant responded incorrectly to any target in the instructional set during the daily probe, a picture-text compound prompting session was immediately conducted.

Baseline

We included an initial baseline phase for all target sets. During this phase, daily probes were conducted without any other intervention. The experimenter also conducted generalization probes during the baseline phase (described below). Specifically, the experimenter presented each target using the whiteboard and tablet modalities. All participants emitted no correct responses during baseline generalization probes conducted before instruction.

Picture-Text Compound Prompting and SIR-WL

Four picture-text compound stimuli which represented each of the targets in the set were presented at the bottom of the word list. All picture-text compounds included the picture appearing above the textual element. During prompted trials, the experimenter removed the blank sheet to reveal the picture-text compound array. To respond correctly, the participant was required to differentially respond to the picture-text compound that included the matching textual element to the target textual stimulus. The participant could then respond by naming the picture element of the compound, which was scored as a prompted correct response. Prompted correct responses produced praise and presentation of the next target.

Figure 1 shows the SIR–WL instruction sequence. During SIR–WL, the first appearance of a target in the word list resulted in a 0-s prompt delay trial. Thus, instruction began with prompted trials for both targets in Row 1. Next, the same targets were presented in a new order and the participant was provided with an unprompted opportunity to both targets (Row 2). If the participant emitted an incorrect response to either target in Row 2, the experimenter presented the prompt and, following a correct prompted response, instruction continued for the same targets (i.e., Rows 1 or 2) until the participant emitted unprompted correct responses to both targets in any row (i.e., Row 1 or 2). The experimenter then revealed a new block of words that included a new target, which appeared in the first position in the row (i.e., Row 3, Block 2; see Fig. 1). The experimenter immediately presented a prompt (i.e., revealed the picture-text compound array) for the new target. Thereafter, unprompted opportunities were presented for all targets in the block. Instruction continued until the participant emitted an unprompted correct response to all targets from the instructional set (i.e., Block 3) in a row or until 3 min elapsed.

Generalization

After responding met the mastery criterion on the daily probe, the experimenter assessed generalization to whiteboard and tablet-based modalities. Correct responses produced praise and incorrect responses produced no differential consequences. During whiteboard (i.e., handwritten) probes, the four targets were handwritten by the experimenter on a 22.9 cm by 31.8 cm whiteboard. The targets appeared in no particular order. Generalization to the tablet was arranged as an analog to a flashcard presentation method. Specifically, a single target appeared in the middle of the screen using the same font as training. All targets were individually presented using PowerPoint© on an iPad©.

Maintenance

The experimenter conducted maintenance probes 7–14 days following mastery. All procedures were identical to those used during daily probes.

Remedial Instruction

If the participant emitted less than 100% unprompted correct responses during the maintenance probe, the experimenter reintroduced picture-text compound prompting for all targets from that instructional set until responding again met the mastery criterion. Another maintenance probe was conducted seven days following mastery.

Interobserver Agreement and Procedural Fidelity

A second independent observer was present during at least 42.8% ($M=67.8\%$) of daily probes and 16.7% ($M=46.3\%$) of teaching sessions across participants. Each presentation of a target word (i.e., exposure) was recorded as a trial. As such,

the number of trials during teaching sessions varied based on the participant's performance. Trial-by-trial interobserver agreement (IOA) was calculated by dividing the total number of trials with an agreement by the total number of trials and multiplying by 100. Mean IOA was 98.8% (range, 75.0–100%) during daily probes and 100% during teaching sessions.

An observer recorded procedural fidelity during at least 42.9% ($M=71.2\%$) of daily probes and 100% of picture-text compound prompt sessions across participants. The observer recorded fidelity separately for each target presented during daily probe and instructional sessions. We calculated percent of targets presented with procedural fidelity by dividing the number of targets presented with fidelity by the total number of targets presented and multiplying by 100. During daily probes, the observer recorded a trial as being implemented with fidelity if the experimenter presented praise or no differential consequences following unprompted correct or incorrect responses, respectively. During teaching (i.e., SIR–WL) sessions, the observer scored procedural fidelity for each target presented if the experimenter adhered to all procedural components for that target. The experimenter must have (a) correctly presented an unprompted or prompted opportunity, (b) delivered praise following an unprompted correct response, (c) introduced a new target after 100% unprompted correct responses were emitted in one row, and (d) terminated instruction after 100% unprompted correct responses were emitted in one row of the final block or 3 min elapsed. Procedural fidelity during daily probes was 100%. Mean procedural fidelity during teaching was 99.9% (range, 94.4–100%).

Results

Figures 2, 3, 4, 5, 6, and 7 show participants' performance during instructional sessions and across daily, generalization, and maintenance probes. The findings from Juliette's evaluation are shown in Fig. 2. Juliette emitted no errors during instruction for Set 1 targets and three errors during instruction for Set 2 targets. Her responding met the mastery criterion in two and three daily probes for Sets 1 and 2, respectively. Juliette emitted 100% correct responses during all generalization and maintenance probes across both target sets.

Figure 3 shows Silas' first evaluation. Silas emitted a single incorrect response during instruction of Set 1 targets and no incorrect responses during instruction of Set 2 targets. His responding met the mastery criterion following a single daily probe for both target sets. Generalization to handwritten and tablet-based modalities was observed for all targets in Set 1. For Set 2, Silas emitted a single error during the whiteboard (i.e., handwritten) generalization probe. Nevertheless, his performance maintained at 100% correct responding for all targets in both sets. Figure 4 shows Silas' second evaluation (i.e., Sets 3 and 4). Silas emitted a total of five incorrect responses during instruction across both Sets 3 and 4. His responding met the mastery criterion in one and two daily probes for Sets 3 and 4, respectively. During generalization and maintenance probes, Silas emitted 100% correct responses across both sets.

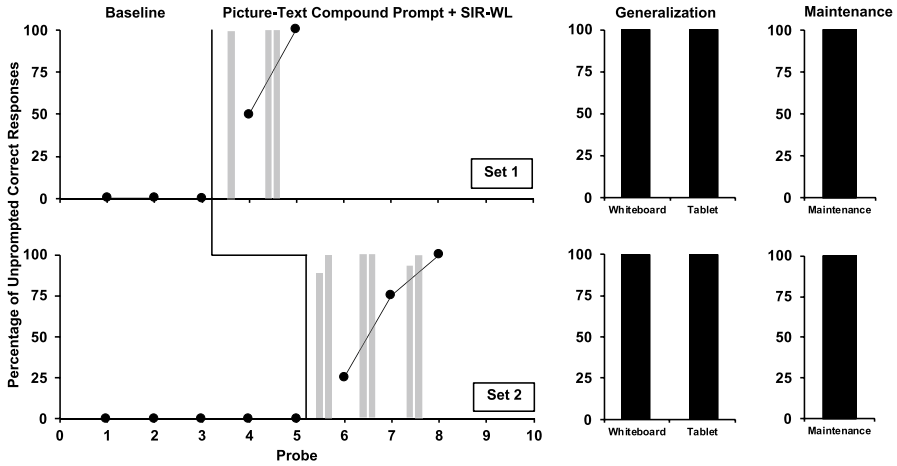


Fig. 2 Percentage of Unprompted Correct Responses during Juliette's Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list

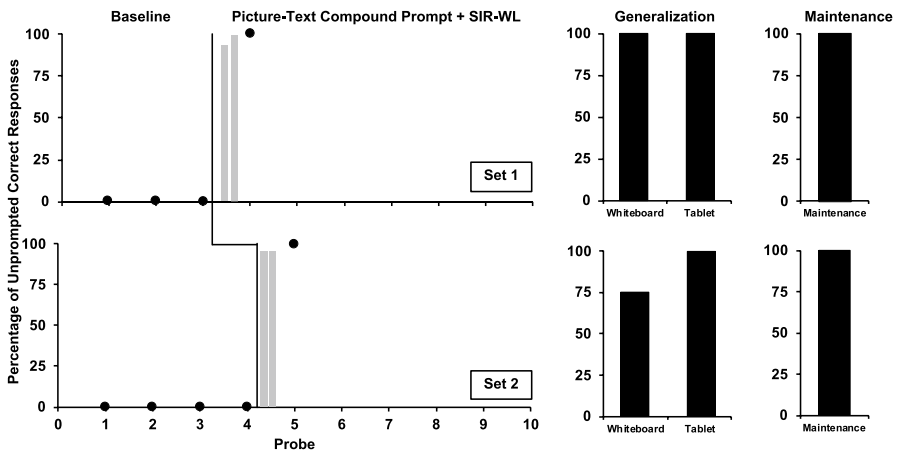


Fig. 3 Percentage of Unprompted Correct Responses during Silas' First Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list

The results from Piper's evaluation are shown in Fig. 5. Piper's percentage of unprompted correct responses showed an increasing trend across instructional sessions for both target sets. Responding met the mastery criterion in two daily probes for both target sets. For Set 1, generalization to the tablet-based modality was observed for all four targets; however, two errors were emitted during probes using the whiteboard modality. During the subsequent maintenance assessment, Piper responded with 100% correct responses. As a result, we conducted a second generalization probe to the whiteboard modality, and she emitted correct responses

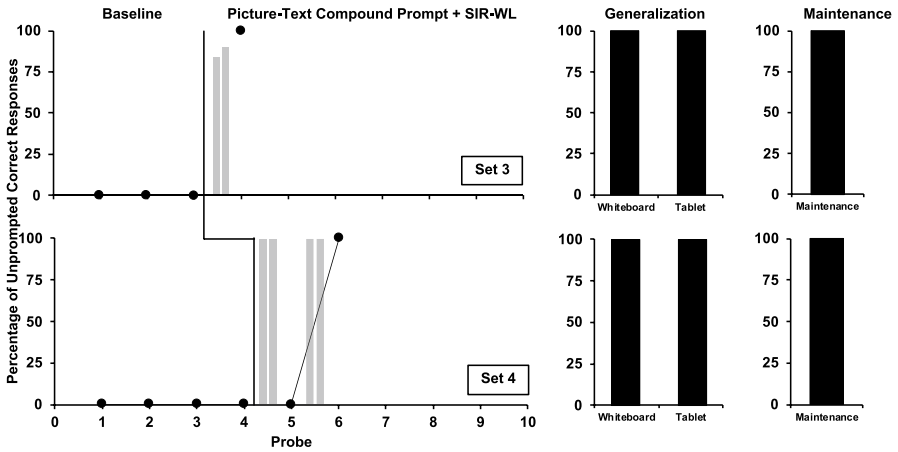


Fig. 4 Percentage of Unprompted Correct Responses during Silas' Second Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list

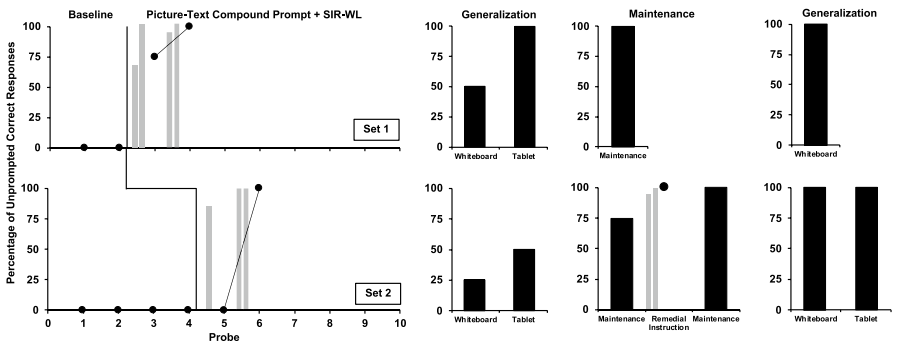


Fig. 5 Percentage of Unprompted Correct Responses during Piper's Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list

to all targets. For Set 2, Piper responded correctly to two of the targets during generalization probes to the tablet modality and a single target on the whiteboard modality. During the subsequent maintenance probe, she emitted correct responses to three of the four targets. We reintroduced picture-text compound prompting as a remedial instruction procedure and observed high levels of accurate responding during instruction and her responding met the mastery criterion following a single probe. She continued to respond with 100% accuracy during subsequent generalization and maintenance probes.

Figure 6 shows the results of Marlie's first evaluation (i.e., Sets 1 and 2). Similar to Piper, Marlie's correct responding showed an increasing trend across instructional sessions for both Sets 1 and 2. Two daily probe sessions were required

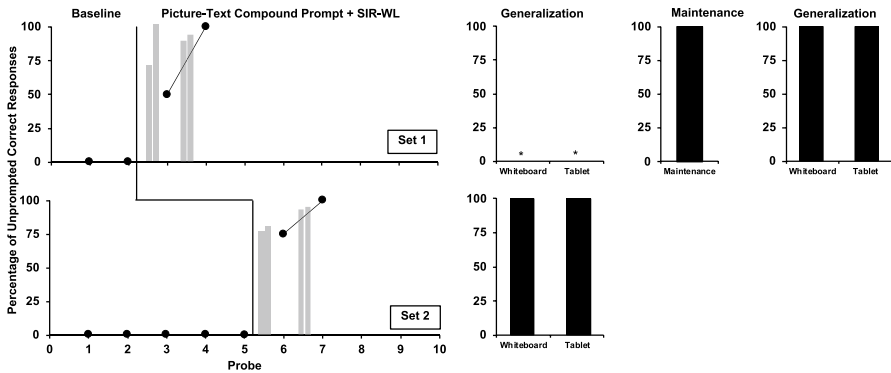


Fig. 6 Percentage of Unprompted Correct Responses during Marlie's First Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list. An asterisk (*) indicates that no correct

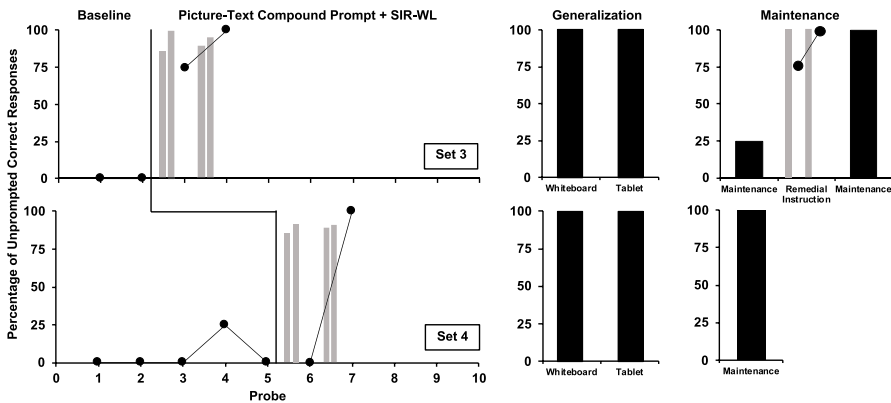


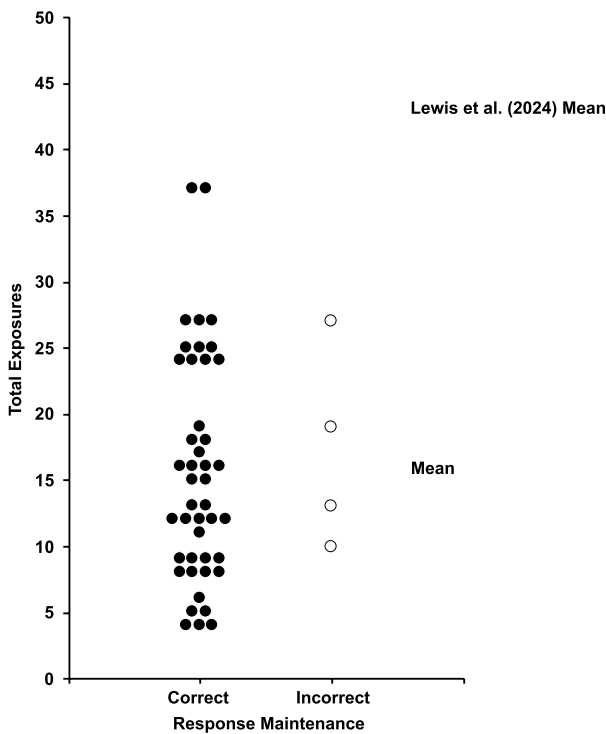
Fig. 7 Percentage of Unprompted Correct Responses during Marlie's Second Evaluation. Closed circles represent daily probe performance; gray bars represent responding during Picture-Text Compound Prompt and SIR-WL instructional sessions. SIR-WL=Strategic incremental rehearsal-word list

for responding to meet the mastery criterion for both target sets. Responding did not generalize to tablet-based or whiteboard modalities for any target in Set 1. Nevertheless, she responded with 100% accuracy during the maintenance probe. Given this performance, we repeated the generalization probes following the maintenance assessment and she exhibited perfect accuracy for all targets during both generalization assessments. For Set 2, Marlie responded accurately during all generalization probes. Maintenance probes for Set 2 could not be conducted due to an extended school break.

The results from Marlie's second evaluation are shown in Fig. 7. Marlie's correct responding during instruction increased across sessions for both target sets. Like Sets 1 and 2, her responding met the mastery criterion in two daily probes for both Sets 3 and 4. Marlie emitted 100% correct responses during

generalization probes to the whiteboard and tablet modalities across both sets and responding maintained for the Set 4 targets. Nevertheless, Marlie emitted a single correct response during the Set 3 maintenance probe. Picture-text compound prompting was reintroduced and responding met the mastery criterion in two daily probes and maintained during a second maintenance probe.

The total exposures to produce responding at the mastery criterion and performance on the initial maintenance probes for individual targets across all participants are shown in Fig. 8. Only four errors were emitted to targets during maintenance probes. An average of 15.7 exposures per target were required to produce responses at mastery levels. Lewis et al. (2024) also recently reported similar outcomes from their study regarding the number of per target exposures required to produce responding at the mastery criterion (i.e., 43 exposures). The mean reported by Lewis et al. (2024) is included in Fig. 8 as a point of comparison when interpreting the current findings compared to previous research on SIR–WL. No targets in the current study exceeded the mean number of exposures to mastery reported by Lewis et al. (2024).



Note. Each data point represents a single target.

Fig. 8 Total exposures and maintenance outcomes of individual targets. Each data point represents a single target

Discussion

The current findings demonstrate the efficacy of picture-text compound prompts embedded in SIR–WL during sight-word instruction for four children exhibiting reading challenges. All participants exhibited mastery level performances following three or fewer daily probes and 40 of 44 targets maintained during one-week probes. Performance on generalization probes was generally high, although errors were emitted during whiteboard generalization probes for ten targets. The current study provides further evidence supporting the efficacy of SIR (e.g., January et al., 2017; Kupzyk et al., 2011; Lozy & Donaldson, 2019). This study also replicated the work of Lewis et al. (2024) by embedding SIR in a word list. Prior research on SIR has commonly made reference to the procedure as a *flashcard* method (e.g., January et al., 2017; Kupzyk et al., 2011; Lozy & Donaldson, 2019); however, Lewis et al. (2024) suggested that the efficacy of SIR was likely due to the systematic introduction of targets and not the presentation modality. Our findings support the contention of Lewis et al. (2024) and extend prior research by embedding a novel prompt type during instruction.

The inclusion of picture-text compound prompts extends prior research in several ways. First, all previous studies on SIR have used echoic prompts (e.g., January et al., 2017; Kupzyk et al., 2011; Lozy & Donaldson, 2019). Similar to the arguments presented by Lewis et al. (2024), we believed that the efficacy of SIR was not restricted to flashcard modalities or echoic prompting procedures. Thus, the current modification contributes to our understanding of the requisite procedural components of SIR and suggests that modality and prompt type are not among them. We were further interested in studying picture-text compound prompts as they may allow for reduced demands on the instructor at least relative to procedures that rely on echoic prompts. Specifically, although the current procedures necessitated the presence of an instructor, future research might evaluate whether similar picture-text compound prompts can be embedded in reading-related tasks and used in lieu of instructor supports. For example, providing the learner with picture-text compounds that correspond to concrete nouns in a story might allow the learner to practice reading the passage without instructor oversight. The use of picture-text compound prompts further extends prior research by demonstrating an effective method by which pictures may be included during reading tasks. Other researchers have used more intrusive methods (i.e., effortful for the instructor or learner; see definition by McGhan & Lerman, 2013) to ensure that pictures do not overshadow or block control by the textual stimuli including stimulus fading (Richardson et al., 2017; Wu & Solman, 1993), directly teaching the participant to match the picture and textual stimuli (Richardson et al., 2017), or altering the size of the picture or text (Singh & Solman, 1990; Solman et al., 1992) with varying degrees of success. The current findings suggest that systematically arranging pictures and text in a manner that requires that the learner differentially respond to the text may engender subsequent control by those stimuli. This finding is promising, although additional research is needed to compare picture-text compound prompting to

other arrangements (e.g., text only) to determine whether differences exist in efficacy, efficiency, or learner preference for these arrangements.

Although the current study did not include a comparison condition, the total number of exposures required to produce mastery was considerably lower than those reported by Lewis et al. (2024). Specifically, an average of 15.7 (range, 4–37) exposures per target were required in the current study, whereas Lewis et al. (2024) reported a mean of 43 exposures per target. Although we might be tempted to conclude that picture-text compound prompting was responsible for these discrepant findings, it is also possible that the word types may have contributed to the rapid development of stimulus control. Specifically, the current study included concrete nouns as targets, whereas Lewis et al. (2024) included a range of word types (e.g., nouns, verbs, and adjectives). Previous research suggests that word imageability defined as “the degree to which the referent of the word evokes a mental image” (de Groot, 1989, p. 824), and word frequency (i.e., frequency of word in print; see Fry, 1980) may affect recall (Coltheart et al., 1988) and may be similarly implicated in the rapid acquisition of targets reported here. Of course, behavior analysts may be dissatisfied with an experimenter’s ability to measure the *mental imagery* evoked by a particular word. As an alternative, they might instead consider whether a learner’s history suggests that the same response is controlled by some nonverbal stimulus (e.g., under tact control). The possibility that certain word types or stimulus–response relations may influence rates of acquisition is relevant to research utilizing comparison designs such as the adapted alternating treatments design (Sindelar et al., 1985) given that experimenters are tasked with ensuring target sets are of equal difficulty (see Cariveau et al., 2021, 2022). Cariveau et al. (2021) recommended that researchers measure participants’ emission of the target response under other possible conditions, such as assessing the tact relation (e.g., saying “cat” when presented with a picture) when targeting the textual relation (e.g., saying “cat” when presented with the word *cat*) as a measure of target novelty. Cariveau et al. (2022) further suggested that word types might be considered when equating targets in the adapted alternating treatments design, although only three of the 49 reviewed articles reported equating targets based on word type. Regardless, a limitation of picture-text compound prompting is that its utility is constrained to certain word types and also requires that the response be under tact control. Given that these requirements might also suggest that targets in the current study would be more rapidly acquired, at least relative to those targets used by Lewis et al. (2024), conclusions about the superiority of picture-text compound prompts are not possible. Instead, future research might compare picture-text compound prompts to other methods (e.g., echoic prompts), while ensuring that targets are equated such that word type and pre-existing relations are not confounded across conditions.

The current study also found that the participants’ performance maintained during maintenance probes for 40 out of 44 targets. In contrast, Lewis et al. (2024) reported maintenance of 32 out of 44 targets, although, as noted above, the reason for differences in the percentage of targets maintained remains speculative. Participants’ responses also remained high during generalization probes to the tablet; however, generalization failures to the whiteboard modality were reported for ten targets. Importantly, Piper (Set 1) responded incorrectly to half and Marlie (Set

1) responded incorrectly to all targets during whiteboard generalization probes, yet both participants exhibited accurate responding during maintenance probes. Because performance maintained, we repeated the generalization probes and observed accurate responding despite no additional instruction for both participants. This finding suggests that some uncontrolled feature (e.g., the instructor's handwriting) of the whiteboard modality may have impacted the participants' performance during the first generalization probe. Although it would be possible to arrange for greater experimental control using handwriting-like fonts, we chose to assess performance using experimenter-generated stimuli given that the participants completed several other instructional activities using the whiteboard. Future research seeking to identify the culprit for similar generalization failures to experimenter-written stimuli might require that the participant name each letter in the word as a type of differential observing response (see Farber & Dickson, 2023). Doing so might allow for the instructor to correct any errors that may result from illegible writing before conducting the generalization assessment.

Future research is needed to address certain limitations of the current study. One such limitation is the omission of a comparison condition. As noted above, conclusions regarding the possible superiority of picture-text compound prompts during sight-word instruction cannot be made given the current experimental design. Instead, the current study serves as an initial demonstration of the efficacy of picture-text compound prompts. Future research might extend this work by comparing picture-text compound prompts to other common procedures (e.g., echoic prompts) and include other dependent measures of interest such as instructor and participant preference. The current design also does not allow for conclusions to be made regarding the efficacy of picture-text compound prompts in overcoming the picture-text problem. Specifically, it is unclear whether the current participants' sight-word acquisition would have been hindered by the simultaneous presentation of pictures and text or whether the current procedures may have engendered greater control by the textual stimuli than alternative strategies. Future research should compare picture-text compound prompts to other methods aimed to overcome the picture-text problem such as text-only or stimulus fading procedures (e.g., Richardson et al., 2017). The current study is also limited in that it included a small number of participants. Future research might include more participants or participants exhibiting a range of reading performances to identify the conditions under which picture-text compound prompts may be effective. A final limitation of the current study is the arrangement of picture-text compound prompts during a highly structured and instructor-delivered intervention. Future research might evaluate the efficacy of picture-text compound prompts when arranged during learner-driven or group-based activities. Such an arrangement might allow for differentiated instruction to be delivered simultaneously to multiple learners while reducing the need for additional oversight by the instructor.

The current study replicated the finding of Lewis et al. (2024) suggesting that SIR is efficacious when presented in a word list. We also extended previous work on SIR by embedding a novel prompt type, picture-text compound prompts, to teach sight words to participants exhibiting reading challenges. Additional work is needed to determine the efficacy and utility of picture-text compound prompts when embedded

in other instructional procedures, such as those intended to remediate the picture-text problem. Nevertheless, the current findings suggest that picture-text compound prompts embedded in SIR–WL resulted in rapid transfer of stimulus control and maintenance, which may represent a viable method for teaching important academic repertoires to children in need of targeted interventions.

Funding This study was not funded.

Declarations

Conflict of interest Tom Cariveau currently serves as an Associate Editor for the *Journal of Behavioral Education*. All other authors declare no conflicts of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Caregiver permission and daily assent were obtained from all individual participants included in the study.

References

- Alonzo, J., & Tindal, G. (2010). *Teacher's manual for regular easyCBM: Getting the most out of the system*. University of Oregon.
- Broomfield, L., McHugh, L., & Reed, P. (2008). The effect of observing response procedures on the reduction of over-selectivity in a match to sample task: Immediate but not long term benefits. *Research in Developmental Disabilities, 29*(3), 217–234. <https://doi.org/10.1016/j.ridd.2007.04.001>
- Broomfield, L., McHugh, L., & Reed, P. (2010). Factors impacting emergence of behavioral control by underselected stimuli in humans after reduction of control by overselected stimuli. *Journal of the Experimental Analysis of Behavior, 94*(2), 125–133. <https://doi.org/10.1901/jeab.2010.94-125>
- Cariveau, T., Batchelder, S., Ball, S., & La Cruz Montilla, A. (2021). Review of methods to equate target sets in the adapted alternating treatments design. *Behavior Modification, 45*(5), 695–714. <https://doi.org/10.1177/0145445520903049>
- Cariveau, T., Helvey, C. I., Moseley, T. K., & Hester, J. (2022). Equating and assigning targets in the adapted alternating treatments design: Review of special education journals. *Remedial and Special Education, 43*(1), 58–71. <https://doi.org/10.1177/0741932521996071>
- Carnine, D. W., Silbert, J., Kame'enui, E., Slocum, T. A., & Travers, P. A. (2017). *Direct instruction reading* (6th ed.). Pearson.
- Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*(1), 1–12. <https://doi.org/10.1111/j.2044-8295.1988.tb02270.x>
- de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 824–845. <https://doi.org/10.1037/0278-7393.15.5.824>
- Deitz, S. M., & Malone, L. W. (1985). Stimulus control terminology. *The Behavior Analyst, 8*(2), 259–264. <https://doi.org/10.1007/BF03393157>
- Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis, 32*(1), 25–33. <https://doi.org/10.1901/jaba.1999.32-25>
- Farber, R. S., & Dickson, C. A. (2023). The classification and utility of the differential observing response. *Behavior Analysis: Research and Practice, 23*(3), 179–194. <https://doi.org/10.1037/bar0000272>

- Feng, H., Chou, W., & Lee, G. T. (2015). Effects of tact prompts on acquisition and maintenance of divergent intraverbal responses by a child with autism. *Focus on Autism and Other Developmental Disabilities*, 32(2), 133–141. <https://doi.org/10.1177/1088357615610540>
- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, 45(3), 195–203. <https://doi.org/10.1177/0022219412442150>
- Gomes-Ng, S., Kim, P. B., Cowie, S., & Elliffe, D. (2023). Reevaluation of overselected stimuli: Emergence of control by underselected stimuli depends on degree of overselectivity. *Journal of the Experimental Analysis of Behavior*, 120(1), 155–170. <https://doi.org/10.1002/jeab.850>
- Hathaway, K. L., Schietz, K. M., & Detrick, J. (2021). Evaluating the effects of instructional prompts and strategic incremental rehearsal on the letter identification mastery of two typically developing kindergarteners. *Behavior Analysis in Practice*, 14(1), 20–35. <https://doi.org/10.1007/s40617-020-00456-5>
- Hodkinson, A. (2017). Constructing impairment and disability in school reading schemes. *Education*, 45(5), 572–585. <https://doi.org/10.1080/03004279.2016.1143520>
- January, S. A., Lovelace, M. E., Foster, T. E., & Ardoin, S. P. (2017). A comparison of two flashcard interventions for teaching sight words to early readers. *Journal of Behavioral Education*, 26(1), 151–168. <https://doi.org/10.1007/s10864-016-9263-2>
- Kennedy, T., & Cariveau, T. (2023). *Picture-text compounds in early reading: A descriptive review*. Advance online publication. <https://doi.org/10.1007/s42822-023-00139-0>
- 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. <https://doi.org/10.1901/jaba.2011.44-781>
- Lewis, T. K., Cariveau, T., Brown, A., Ellington, P., & Stocker, J. (2024). *Efficacy of strategic incremental rehearsal in a word list [Manuscript submitted for publication]*. University of North Carolina Wilmington.
- 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/dbb0000089>
- McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Cataldi, E. F., Mann, F. B., Barmar, A., Nachazel, T., Barnett, M., & Purcell, S. (2019). *The Condition of Education 2019 (NCES 2019–144)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics. <https://files.eric.ed.gov/fulltext/ED594978.pdf>
- McGhan, A. C., & Lerman, D. C. (2013). An assessment of error-correction procedures for learners with autism. *Journal of Applied Behavior Analysis*, 46(3), 626–639. <https://doi.org/10.1002/jaba.65>
- McIntosh, R., Vaughn, S., Schumm, J. S., Haager, D., & Lee, O. (1993). Observations of students with learning disabilities in general education classrooms. *Exceptional Children*, 60(3), 249–261. <https://doi.org/10.1177/001440299406000306>
- McKenna, J. W., Shin, M., & Ciullo, S. (2015). Evaluating reading and mathematics instruction for students with learning disabilities: A synthesis of observation research. *Learning Disability Quarterly*, 38(4), 195–207. <https://doi.org/10.1177/0731948714564576>
- National Assessment of Educational Progress [NAEP]. (2022). *NAEP report card: Reading*. National Center for Education Statistics. <https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4>
- National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. (NIH Publication No. 00-4754)*. U.S. Government Printing Office.
- O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities*, 38(6), 532–538. <https://doi.org/10.1177/00222194050380060901>
- Phipps, L., Robinson, E. L., & Grebe, S. (2022). An evaluation of strategic incremental rehearsal on sight word acquisition among students with specific learning disabilities in reading. *Journal of Behavioral Education*, 31(1), 281–297. <https://doi.org/10.1007/s10864-020-09398-y>
- Rahn, N. L., Wilson, J., Egan, A., Brandes, D., Kunkel, A., Peterson, M., & McComas, J. (2015). Using incremental rehearsal to teach letter sounds to English language learners. *Education and Treatment of Children*, 38(1), 71–91.

- Reed, P., Reynolds, G., & Fermandel, L. (2012). Revaluation manipulations produce emergence of underselected stimuli following simultaneous discrimination in humans. *The Quarterly Journal of Experimental Psychology*, *65*(7), 1345–1360. <https://doi.org/10.1080/17470218.2012.656663>
- Richardson, A. R., Lerman, D. C., Nissen, M. A., Luck, K. M., Neal, A. E., Bao, S., & Tsami, L. (2017). Can pictures promote the acquisition of sight-word reading? An evaluation of two potential instructional strategies. *Journal of Applied Behavior Analysis*, *50*(1), 67–86. <https://doi.org/10.1002/jaba.354>
- Samuels, S. J. (1967). Attentional process in reading: The effect of pictures on the acquisition of reading responses. *Journal of Educational Psychology*, *58*(6), 337–342. <https://doi.org/10.1037/h0020045>
- Sindelar, P. T., Rosenberg, M. S., & Wilson, R. J. (1985). An adapted alternating treatments design for instructional research. *Education & Treatment of Children*, *8*(1), 67–76.
- 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. <https://doi.org/10.1901/jaba.1990.23-525>
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press.
- Solman, R. T., Singh, N. N., & Kehoe, E. J. (1992). Pictures block the learning of sightwords. *Educational Psychology*, *12*(2), 143–153. <https://doi.org/10.1080/0144341920120205>
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, *18*(3), 137–146. <https://doi.org/10.1111/1540-5826.00070>
- Vaughn, S., Wanzek, J., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention examining higher and lower responders. *Exceptional Children*, *75*(2), 165–183. <https://doi.org/10.1177/001440290907500203>
- Vedora, J., & Conant, E. (2015). A comparison of prompting tactics for teaching intraverbals to young adults with autism. *The Analysis of Verbal Behavior*, *31*(1), 267–276. <https://doi.org/10.1007/s40616-015-0030-6>
- Walpole, C. W., Roscoe, E. M., & Dube, W. V. (2007). Use of a differential observing response to expand restricted stimulus control. *Journal of Applied Behavior Analysis*, *40*(4), 707–712. <https://doi.org/10.1901/jaba.2007.707-712>
- Wu, H., & Solman, R. T. (1993). Effective use of pictures as extra stimulus prompts. *British Journal of Educational Psychology*, *63*(1), 144–160. <https://doi.org/10.1111/j.2044-8279.1993.tb01047.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.