




Verbal Behavior Analysis of Teaching Story Recall to Children with Autism: A Replication and Extension

Daniel E. Conine^{1,2}  · Lisa A. Guerrero³ · Erica Jones-Thomas⁴ · Sarah E. Frampton⁵ · Timothy R. Vollmer⁶ · Tina Smith-Bonahue⁷

Accepted: 6 March 2023 / Published online: 24 March 2023
© Association for Behavior Analysis International 2023

Abstract

Children with autism spectrum disorder (ASD) may struggle with verbal behavior related to recall in various contexts. However, relatively little research has evaluated methods for improving recall among this population, and even fewer from a verbal behavior perspective. One socially important set of skills that relies upon a behavioral repertoire of recall is applied reading skills, such as reading comprehension and story recall. Valentino et al. (2015) designed an intervention package to teach children with ASD to recall short stories and conceptualized the behavior as an intraverbal chain. The present study replicated and extended that study with three school-aged children with ASD using a multiple baseline design across stories. For some participants and some stories, story recall was mastered under less intensive intervention conditions than in the previous study. When it was necessary to implement the full intervention package, the effects largely replicated previous research. Improvements in recall were correlated with increases in correct answers to comprehension questions. These data have important implications for clinicians and educators providing reading and recall interventions to children with ASD. Results also have theoretical implications for verbal behavior accounts of memory and recall, and suggest several possible avenues for future research.

Keywords Autism · Behavior chain · Intraverbal · Reading · Recall

✉ Daniel E. Conine
dconine@gsu.edu

¹ Department of Learning Sciences, Georgia State University, Atlanta, GA, USA

² College of Education & Human Development, Georgia State University, P.O. Box 3980, Atlanta, GA 30302, USA

³ Little Leaves Behavioral Services, Weston, FL, USA

⁴ Florida Autism Center, Gainesville, FL, USA

⁵ Department of Psychology, University of Nebraska Omaha, Omaha, NE, USA

⁶ Department of Psychology, University of Florida, Gainesville, FL, USA

⁷ College of Education, University of Florida, Gainesville, FL, USA

Memory and recall are longstanding topics of interest in the psychological and behavioral sciences. As proposed by Palmer (1991), whenever the term “recall” or “memory” is used, either term may actually refer to one of two distinct behavioral phenomena¹. The first behavioral phenomenon that we might describe as “recall” is best analyzed as a matter of stimulus control: whether discriminative stimuli maintain their effects when they are not encountered for extended periods of time. For example, a pigeon may be given food for pecking a green key, and over time this behavior comes under stimulus control such that the pigeon only pecks the key when it is green, not when it is red or unlit. We may say the pigeon “recalls” this arrangement if it still pecks only the green key when we place it in the chamber after a year goes by during which the pigeon is not exposed to this arrangement or given any practice pecking green keys.

The second sort of behavior that we might describe as recall is considerably more complex. This type of recall is best approached as a matter of problem-solving, and requires responding under joint control of both antecedent verbal stimuli and some other stimuli that are not present at the time of recall (Lowenkron, 2006; Palmer, 1991). For example, when presented with a verbal stimulus such as “What color shirt did you wear yesterday?” there is no relevant nonverbal stimulus to occasion an accurate tact (i.e., yesterday’s shirt). Instead, a person might use problem-solving strategies to generate supplementary verbal or nonverbal stimuli that increase the likelihood of a correct response (Palmer, 1991). For example, you might use self-intraverbal statements such as, “Yesterday was Tuesday. On Tuesdays I meet with my supervisor, so I wore something nice. That’s right, I wore my new blue shirt.” A person may also engage in visual imagining (Aguirre & Rehfeldt, 2015; Kisamore et al., 2011; Skinner, 1953; Skinner, 1974) as a part of this process, such that one can “see” oneself sitting in a meeting room with their supervisor, wearing a blue shirt (Palmer, 1991).

Children with developmental delays, such as autism spectrum disorder (ASD), often exhibit persistent delays in recall behaviors that may require targeted intervention (e.g., Bordignon et al., 2015; Krantz et al., 1981; Shillingsburg et al., 2017). One specific type of recall behavior in which children with ASD are likely to experience deficits relative to their neurotypical peers is recalling stories read to them by teachers (Baixauli et al., 2016; Diehl et al., 2006; Williams et al., 2006). This particular behavior, often referred to as *story recall*, is considered an important foundational skill among the broader repertoires associated with listening and reading comprehension (Kim & Pilcher, 2016; Reed & Vaughn, 2012; Shapiro et al., 2014). Thus, it is possible that deficits in story recall skills might underlie the broader array of deficits in reading skills that children with ASD often experience as compared with neurotypical peers during their academic years (McIntyre et al., 2017; Ricketts et al., 2013).

Palmer’s (1991) behavioral analysis of recall has the potential to support the design of effective behavioral interventions for these skills. However, to date there have been limited empirical investigations of behavioral interventions for directly

¹ For concision, we will exclusively use the term “recall” rather than “memory” throughout the remainder of this paper.

teaching story recall to children with ASD (Bailey & Arciuli, 2020), and even fewer from a behavioral perspective. Most notably, Valentino et al. (2015) designed a behavioral intervention for improving the story recall behavior of three children with ASD. The experimenters created picture books containing short stories, read the stories to the participants, and later asked them to recall the story (e.g., “Tell me the story about [name of story]”). The intervention included textual prompts, error correction, backward chaining, and the delivery of social and tangible reinforcers for correct recall. Improvements in story recall were reported for all three children. Moreover, after intervention was completed with a few initial stories, participants began to correctly recall stories that remained in a modified baseline condition that consisted of repeated reading and tangible reinforcement for correct recall. This latter finding suggested that the intervention might have produced improvements in a generalized repertoire of story recall, although that notion remains speculative due to the multiple baseline across stories design used in that study.

Replications of Valentino et al. (2015) would be valuable, and several features of the study also suggest some extensions, modifications, or additional questions to be explored. First, the intervention used in Valentino et al. was complex, and required substantial idiosyncratic modifications for two out of three participants. Valentino et al. suggested that it might be possible to produce an intervention that is more consistently effective (i.e., fewer individualized modifications) if differential reinforcement of unprompted responses (e.g., Campanaro et al., 2020; Vladescu & Kodak, 2010) and alternative, a time-based termination criterion during probe trials, were incorporated (Valentino et al., 2015). However, these suggested changes to the intervention have yet to be studied.

Second, additional replications are needed to further investigate the potential that this sort of direct intervention for story recall might produce generalized improvements in a broader repertoire of story recall, such that continued intervention may no longer be necessary after implementation with a few initial stories. Such an outcome was suggested by the data in Valentino et al. (2015), but with limited experimental control and for only three participants. Additional data are needed to investigate the potential consistency of such an outcome.

Third, some secondary dependent variables merit exploration in a replication of Valentino et al. (2015). In that study’s baseline sessions, recall was tested 30 s after reading, but it was tested 24 h after reading in the intervention condition. Additional research is needed to identify what impact various delays might have on the accuracy of story recall during intervention for this population. Valentino et al. also posited that improvements in story recall might set the stage for subsequent or concurrent improvements in other reading skills, such as answering questions about the stories. However, no reading comprehension behaviors, aside from story recall, were measured in the Valentino et al. study. Additional research is needed to determine whether behavioral interventions for story recall might produce corresponding increases in other reading comprehension behaviors like answering questions.

The overall aim of the current study was to replicate the Valentino et al. (2015) study with three additional children with ASD. Three main areas of inquiry guided this replication. First, we explored the impact of two specific changes proposed by Valentino et al.: (a) differential reinforcement of unprompted responses and (b)

time-based termination criteria during recall probes. Second, we replicated the use of reading-and-reinforcement baseline probes before, during, or after intervention with each story to explore whether outcomes suggestive of generalization could be replicated. Finally, we explored the impact of intervention on two secondary variables: (a) responding at varied delays from reading and (b) answering comprehension questions about the stories.

Method

Participants and Setting

Three boys diagnosed with ASD participated in the study: Nick, Edgar, and Albert, who were 11, 5, and 8 years old, respectively. All three attended a school program for children with ASD. Each child's teacher referred them to the study based on deficits in story recall and reading comprehension, and identification of story recall as an educational goal. Informed consent to participate was obtained from each child's parent or legal guardian. Teachers sought participant assent at the start of every session by asking whether the participant would like to go work on reading; the session began if the participant responded affirmatively.

To verify appropriateness for inclusion, the second author conducted the reading cluster of the WJ Tests of Achievement-IV (Schrank et al., 2014) with each participant at the start of the study in a session room with minimal distractions. We conducted subtests to calculate every participant's Broad Reading and Basic Reading Skills scores, and the Reading Recall subtest to verify the participants' reported deficits in story recall. Although not standard for this testing protocol, we used non-contingent praise and breaks from testing to maintain responding. We used participants' Basic Reading Skills grade equivalent score to inform construction of the stories used in the study (described below). Participants were eligible for the study if they scored 89 or lower on the Reading Recall subtest, and their Basic Reading grade equivalent was at least first grade (1.0). Table 1 lists pseudonyms, ages, and

Table 1 Participant demographics and reading scores

Participant	Age	Reading grade level	Broad reading	Basic reading skills	Reading recall
Nick	11	2	90 (63–77) Average	91 (87–96) Average	47 (<40–76) Very low
Edgar	5	2.1	84 (77–91) Low average	104 (100–109) Average	66 (44–89) Very low
Albert	8	1.3	90 (85–94) Average	92 (88–97) Average	66 (44–89) Very low

Note. This table denotes the standard scores for each of the reading domains or subtests for each participant. Scores were calculated using a 95% confidence interval, and their scores in this 95% confidence band are denoted in the parentheses. Each participant's reading grade level was calculated with their reading grade equivalent score for their Basic Reading Skills score

assessment scores for all participants. All participants scored in the Low Average to Average range in Broad Reading, which factors in multiple skills (e.g., decoding, oral reading fluency, comprehension). However, all participants scored in the Very Low range for Story Recall, indicating this skill as a relative deficit.

For Nick and Edgar, sessions were conducted during the school day on every day that they were in attendance. Albert transitioned to a new school between recruitment and the start of the study; thus, his sessions took place one day per week on scheduled visits to the building after school. Sessions took place in a small classroom containing a table, two chairs, story books, and preferred items. Nick's and Edgar's lead classroom teachers conducted their sessions. The first author conducted sessions for Albert, and had no prior educational history with him. For brevity, all three instructors will be referred to as the "teacher" throughout this paper.

Materials

As in Valentino et al. (2015), we created five stories for the study, rather than using commercially available stories, to eliminate the possibility that participants had prior exposure to the stories. Each story contained eight pages, with five words centered at the top and one illustration centered at the bottom of each page. Volunteer artists created the illustrations to depict the main idea or event described on each page. Stories were printed in color on white backgrounds (8.5 in by 11 in) and bound in a folder with three-prong fasteners. To ensure that there were narrative elements to recall, all stories contained at least one character who experienced a series of events. We analyzed reading level of all stories using the Lexile Analyzer (MetaMetrics, 2017) to verify appropriateness for participants' reading levels. All stories had a Lexile Measure of 100L–200L (roughly corresponding with a Kindergarten or first-grade reading level, see MetaMetrics, 2023), and a mean log word frequency between 3.0 and 3.5. Text and pictures for all stories are available in the Supplementary Information.

We created eight comprehension questions for each story, one corresponding to each page. The questions were designed with two key features. First, each question began with either who, what, where, when, or how (not "why"). Second, we designed each question as a literal comprehension question based on the text of each page, so that the answer was supplied by the story text with no inferences or outside information required (Day & Park, 2005). Literal comprehension questions were identified as the most appropriate educational goal in this area for the participants in consultation with teachers and the pre-study assessment results (Table 1).

Response Measurement

Story recall was the primary dependent variable, and it was scored on a page-by-page basis. The definition for story recall was designed to capture whether participants accurately recalled all important features of a story, while allowing for flexibility in how participants summarized or retold the stories. At the start of the study we designated two to four words from the text of each page as *key words*, which

are shown along with the full text of each page in Table 2. We used the following rules to select key words: (a) at least one subject noun and verb were key words for each page, (b) articles (a, an, the) were never key words, and (c) other words (e.g., adjectives, prepositions) were key words on a case-by-case basis, based on relevance to the overall meaning of that page. Before the study began, the first and second authors independently selected key words for each page, and then discussed any disagreements until they reached consensus regarding a final list of key words.

Correct story recall for each page was defined as the participant vocally stating all key words for that page, or acceptable substitutions for those key words (Valentino et al., 2015) during recall tests. Acceptable substitutions included: (a) pronouns instead of character names (without considering gender), (b) generic nouns instead of character names (e.g., “car” instead of “Toby”), or (c) synonyms for key words. Grammatical features like pluralization, subject-verb agreement, and verb tense were not considered. With respect to synonyms, data collectors judged words as acceptable synonyms based on the overall context of the story’s text, pictures, or both. For example, “Frog go the water” was considered an acceptable substitution for “*Frog hops to the pond*” (Table 2, The Hungry Frog). “Kyle sweeps the floor” was an acceptable substitution for “*Kyle helped clean the house,*” because the illustration for that page shows Kyle sweeping (Supplementary Information, Roger’s New Cage). If participants made a statement that paraphrased key words for more than one page, we scored correct recall for multiple pages at once. For example, if a participant said “Phil the elephant likes eating bananas,” we scored correct recall for pages 1, 2, and 3 (see Table 2, Phil and Frank). Finally, pages were scored as correct regardless of the order in which they were recalled. We scored *incorrect story recall* for a page if the participant either did not make any statement that met the correct recall definition or met the correct recall definition but also substituted or added any words that contradicted the key words of the page. For example, if a participant said, “Phil does not like to eat bananas,” we scored incorrect recall for page 3 (Table 2, Phil and Frank).

A secondary dependent variable was *correct question-answering*. At the end of all sessions, except in Baseline (see below), teachers asked all eight questions associated with a given story (see Table 2). Correct question-answering was defined as any response initiated within 10 s of the question that accurately corresponded to the story’s text, pictures, or both. Data collectors were provided with examples and non-examples of correct question-answering for each story (Supplementary Information).

For purposes of evaluating mastery criteria, we derived another variable, *total correct*. Total correct represented the number of pages for which the participant engaged in correct recall, correctly answered the corresponding question, or both. For example, if a participant correctly recalled pages 1, 2, and 8, and correctly answered questions for pages 3, 4, and 8, total correct was equal to five (i.e., pages 1, 2, 3, 4, and 8). A story was considered mastered if: (a) the participant correctly recalled at least six of eight pages (75%) and (b) total correct was equal to eight (100%). Stories were mastered if these criteria were met for two (Edgar) or three (Nick and Albert) consecutive sessions. We chose these criteria and calculated *total correct* based on the notion that recall likely does not need to include all story events to be functional, particularly if respondents can answer questions about missing details when asked.

Table 2 Story text, key words, and comprehension questions for each story

Component	Page	Phil and Frank	Polly Panda	Roger's New Cage	The Hungry Frog	Toby the Car
Text	1	Phil is a big elephant.	Polly Panda likes to climb.	Kyle has a pet hamster.	Frog lives in a tank.	Toby is a fast car.
	2	He really likes to eat.	Polly climbs a tall tree.	The hamster's name is Roger.	Frog doesn't like his food.	He goes to fun places.
	3	Phil likes to eat bananas.	Polly's mom calls her name.	Roger wants a new cage.	Frog jumps over the tank!	He drives around the park.
	4	Phil's friend is named Frank.	Now Polly must come down.	But Kyle has no money!	Frog hops to the pond.	One day Toby fell apart.
	5	Frank is a silly monkey.	That makes Polly feel sad.	Kyle asked Mom for money.	Frog's tummy rumbles.	He lost his two wheels.
					He's hungry!	
	6	Frank likes eating bananas too.	Polly goes home with Mom.	Kyle helped clean the house.	Frog eats a fly! Buzz.	Toby went to the shop.
	7	Frank shares bananas with Phil.	Mom makes a yummy dinner.	Mom gave Kyle some money.	Frog thinks that is yummy!	Toby got two new wheels.
	8	This makes Phil very happy.	Going home makes Polly happy!	Kyle bought a new cage.	Frog hops back to home.	Toby can drive fast again!
Question	1	What kind of animal is Phil?	What did Polly like to do?	What kind of pet does Kyle have?	What did Frog live in?	What was Toby?
	2	What does Phil really like to do?	What did Polly climb?	What was the hamster's name?	What doesn't Frog like?	Where does Toby go?
	3	What does Phil like to eat?	Who called Polly's name?	What did Roger want?	What did Frog jump over?	What does Toby do at the park?
	4	What is Phil's friend's name?	Then what did she have to do?	What didn't Kyle have?	Then where did he go?	What happened to Toby?
	5	What kind of animal is Frank?	How did that make her feel?	Who did Kyle ask for money?	What did Frog's tummy do?	What did Toby lose?
	6	What does Frank like to eat?	Where did Polly go with her mom?	What did Kyle clean?	What did the frog eat?	Then, where did he go?
	7	Who shared bananas with Phil?	What did Polly's mom make?	What did Kyle's mom give him?	When he ate that, what did he think?	When he was there, what did he get?
	8	How did that make Phil feel?	How does going home make Polly feel?	What did Kyle buy?	Then where did he go?	Then what could he do?

Key words for each page of text are underlined

Experimental Design

We used a combination of multiple probe and multiple baseline designs, both concurrent. For Nick and Edgar, we used a multiple probe strategy to stagger the initial introduction of a reading-and-reinforcement (RAR) baseline from a no-reading baseline. With Edgar, we also staggered the introduction of intervention on the RAR baseline in keeping with a multiple baseline design. This mirrors the approach Valentino et al. (2015) used with one participant (Roger); we took this approach for the same reason offered by Valentino et al.: to increase experimental control by decreasing the likelihood that intervening with one story would cause changes across multiple ongoing baselines in which stories continued to be read. For Albert, we took the approach used by Valentino et al. with the other two participants: introducing all stories to the RAR baseline simultaneously at the start of the study, and then staggering intervention across stories using a multiple baseline design. We used this approach with Albert because he attended the clinic only once a week. On each day that a participant was present for the study, we conducted sessions for all stories that had been introduced to either the RAR baseline or intervention according to these experimental designs, and sessions continued until mastery criteria were met. Delay probes were conducted 3 weeks and 6 months following mastery with all stories with each participant.

Interobserver Agreement (IOA)

The teacher collected primary data for each session. To calculate interobserver agreement (IOA), a trained observer collected secondary data either in vivo or from video recordings for a subset of sessions across all experimental conditions. Secondary data collectors included the first and second authors, other teachers at the school, or undergraduate research assistants. For both of our dependent variables (story recall, question-answering) we calculated IOA for each session by dividing the number of pages for which both observers agreed (i.e., scored a response as correct or incorrect) by the total possible (eight). For recall, mean IOA was 98% for Nick (range, 88–100%, collected for 24% of sessions), 94% for Edgar (range, 63–100%, collected for 29% of sessions), and 94% for Albert (range, 75–100%, collected for 24% of sessions). For question-answering, mean IOA was 98% for Nick (range, 75–100%, collected for 25% of sessions), 94% for Edgar (range, 75–100%, collected for 31% of sessions), and 99% for Albert (range, 88–100%, collected for 10% of sessions).

Pre-Experimental Assessments

At the start of the study, we conducted a pre-test to determine whether participants could read the words contained in the stories. For this pre-test, we printed each word included across all five stories on a flash card (112 words total). The teacher showed the child each word once, asked “What is it?”, provided praise for correct responses, and provided a neutral response (e.g., “okay”) for incorrect responses. If participants scored 80% or greater on this pre-test, they were deemed appropriate for inclusion in

the study, because any unknown words would be modeled by the teacher during the study's reading component and prompted during intervention. All participants had histories of learning to read new words with minimal modeling and prompting. We also conducted multiple-stimulus-without-replacement preference assessments with four foods and four toys (Conine & Vollmer, 2019); the four highest-preferred items were used in the study as described below.

General Procedures

At the start of each session, the teacher and participant sat across from or next to one another at a table. Sessions contained one or more discrete components, which varied across the experimental conditions. Figure 1 outlines the components of each condition and their order. Procedural descriptions for each component are provided in the condition descriptions in which they first appear, below.

Conditions and Session Components

Baseline (No Reading)

As in Valentino et al. (2015), the baseline condition was conducted for experimental control purposes to assess whether the question and story name alone (i.e., "Tell me the story about [name of story]") would occasion any responses that met the definition for correct recall even though participants had not yet heard or read the story (e.g., due to exposure to similar stories, information provided by the title). In baseline, the teacher did not read the story, show the book to the participant, or ask any questions about the story (Fig. 1). Sessions in baseline contained only a *recall test*, described below (Fig. 1).

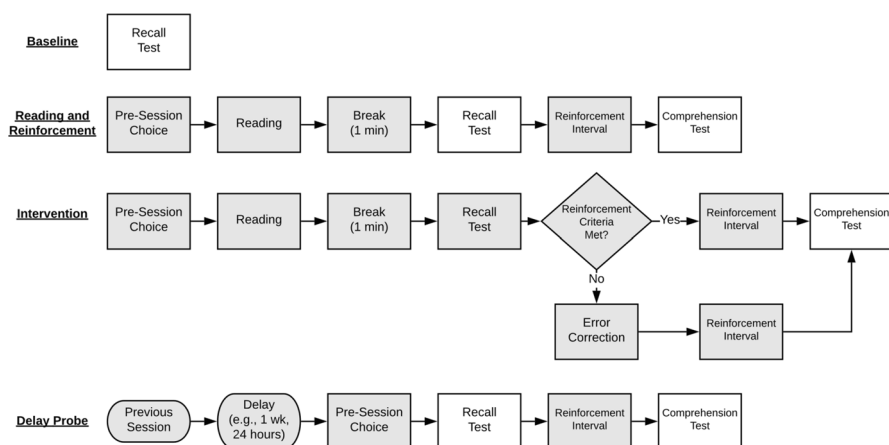


Fig. 1 Flow diagram of session components across experimental conditions. *Note.* Components in white indicate those in which data on participant's target behaviors were collected and reported

Recall Test The teacher began all recall tests with the instruction “Tell me the story about [name of story].” The teacher then allowed time for the participant to talk aloud until they met one of the *termination criteria*, which were: (a) the participant correctly recalled all eight pages of the story or (b) 30 s elapsed in which the participant did not correctly recall any *new* pages (i.e., pages that had not yet been correctly recalled in that same recall test). These termination criteria were used in all recall tests throughout the study. The teacher gave brief praise (e.g., “right”) the first time the participant correctly recalled each page of the story in a given recall test. The teacher did not respond to any repeated correct recall for the same page, to any other non-target vocalizations, or to incorrect recall. At the end of the recall test in baseline, the teacher provided non-specific praise regardless of participant responses (e.g., “okay, good work”). Different consequences were implemented at the end of the Recall test in other experimental conditions (described in those conditions under “Reinforcement Interval”, below).

Reading and Reinforcement (RAR) Baseline

This condition was modeled after the “Reading” condition in Valentino et al. (2015). The RAR baseline served as the first baseline measure of question-answering, and the first baseline of story recall given actual exposure to the stories. The RAR baseline contained the following components in order (Fig. 1): pre-session choice, reading, 1-min break, recall test (identical to baseline), reinforcement interval, and the comprehension test.

Pre-session Choice At the start of session, the teacher placed the four highest-preferred items from the preference assessment on the table and asked a question like “What do you want to work for?” The first item the participant selected (vocally or by touching) was designated as the reinforcer for the upcoming session. If the participant requested any item other than the four items on the table, and the requested item was available and appropriate to deliver (at the teacher’s discretion), that requested item was designated as the reinforcer. If the designated reinforcer was an iPad or computer, the teacher also asked the participant to select a specific video or application on the device.

Reading During the reading component, the teacher held the book open and upright with its text and pictures facing toward the participant. The teacher read the story out loud from start to finish (beginning with the title), turning each page after they finished reading it.

Post-reading Break (1 min) After the reading component, the teacher provided a statement such as, “Okay, you can take a break” and allowed the participant to engage in any activities of their choice for 1 min (e.g., sit alone, engage the teacher in conversation), except for leaving the table or engaging with the designated reinforcer. The teacher responded to any questions or conversations initiated by the participant during this break, except for any questions about the story.

Recall Test After the 1-min break, the teacher initiated a recall test, as described in *Baseline*.

Reinforcement Interval Once one of the termination criteria were met in the recall test, the teacher delivered reinforcement according to two rules. If the participant had not correctly recalled any pages, the teacher provided non-specific praise. If the participant had correctly recalled at least one page of the story during the recall test, the teacher provided praise (e.g., “great job!”) and the designated reinforcer. If the reinforcer was a food, the teacher provided one small portion (e.g., one skittle). If the reinforcer was a toy or video, the teacher provided access for 60 s.

Comprehension Test The second dependent variable, question-answering, was probed in the comprehension tests. Immediately after the reinforcement interval (above), the teacher began the comprehension test with a statement like “Now I’m going to ask you some questions about the story.” The teacher then asked each question for the story, one at a time and in order (i.e., one through eight; Table 2), waiting up to 10 s after each question. Regardless of the participant’s response (correct, incorrect, or no response), the teacher made a neutral comment (e.g., “all right”) and asked the next question. At the end of the comprehension test, the teacher provided non-specific praise (e.g., “nice work,” “thank you for answering those questions!”).

Intervention

Intervention contained all components in the RAR baseline with the addition of error correction and modified reinforcement criteria (Fig. 1). The pre-session choice, reading, 1-min break, recall test, and comprehension test were all conducted exactly as in the RAR baseline. One exception to this rule is that in the first session of Intervention with each story, there was no recall test (as in Valentino et al., 2015), because any effects of Intervention would not be detectable on a recall test until the participant had already encountered the modified reinforcement criteria and potential error correction during the first Intervention session.

Each Intervention session included a differential reinforcement criterion based on recalling a specific set of *targeted pages* (Valentino et al., 2015). Backward chaining with leaps ahead was used to set the targeted pages criterion. That is, the targeted pages for each session were selected from the end of the story with the initial criterion for each story requiring correct recall of at least page 8. However, if the participant had correctly recalled page 8 for the last two sessions of the RAR baseline, the initial criterion would require both pages 7 and 8 (and so on, if additional pages were mastered during the RAR baseline).

Throughout Intervention, one page, moving from the end to the beginning, was added to the targeted pages criterion after the participant met the reinforcement criterion during the Recall test for two consecutive sessions (e.g., page 8, then pages 7 and 8, then pages 6–8, and so on). Because it was possible for participants to correctly recall pages before they were required by the criterion, we used leaps ahead

(Spooner et al., 1986) when advancing the criterion. For example, if the participant correctly recalled pages 5, 6, 7, and 8 for two consecutive sessions when the criterion required only pages 7 and 8, the next reinforcement criterion would leap ahead to require pages 4 through 8. We continued advancing the reinforcement criterion in this manner until the participant met mastery criteria for the whole story. The reinforcement criterion was never regressed or decreased.

Reinforcement Interval If the participant met the reinforcement criterion during the recall test, the teacher immediately provided praise and a larger magnitude of the designated reinforcer. A larger magnitude was defined as four portions of a food or 4-min access to a toy. If the participant had not met the reinforcement criterion, the teacher did not deliver a reinforcer, and instead implemented a re-present-until-correct error correction procedure described below (Cariveau et al., 2019; labeled as a prompted trial and transfer trial in Valentino et al., 2015).

Error Correction At the start of all Intervention sessions, the teacher inserted one or more blank pages (i.e., plain white sheets of 8.5×11 in paper) into the story folder before each *targeted page*. The teacher skipped past these pages during the Reading component. These blank pages were used only during error correction as described below.

To begin error correction, the teacher repeated the recall instruction (“Tell me the story about [name of story]”) and opened the book to its first page, with text facing the participant. The presentation of the text was intended to prompt correct recall (i.e., reading the words on the page). If the participant did not respond or stopped responding in the presence of the text for 5 s at any point prior to engaging in correct recall for that page, the teacher provided supplemental prompts by pointing to the next word on the page and reading it out loud; the point remained and the vocal prompt was repeated every 5 s if needed until the participant said the word. Once the participant engaged in correct recall in the presence of the text for a given page (i.e., prompted recall), the teacher provided brief praise (e.g., “Good!”) and turned to the next page in the story, skipping past all of the previously inserted blank pages. The teacher continued this process until the participant had engaged in correct recall in the presence of the text (i.e., prompted) for all pages of the story.

Once the participant correctly recalled the story in the presence of all of the pages, the teacher began a second error correction trial by re-presenting the recall instruction once again (“Tell me the story about [name of story]”). For all of the non-targeted pages, a correct response was prompted exactly as described above (i.e., participants were immediately shown the text, given point and verbal prompts if needed). However, whenever the teacher turned to a targeted page during this second error correction trial, they: (1) showed only the blank page covering the targeted page, rather than the text, (2) said nothing, and (3) waited for up to 5 s for a response from the participant. If the participant engaged in correct recall for that page within 5 s of the teacher turning to the blank page, the teacher provided praise and turned to the next page of the story (if applicable, also covered by a blank page). If the participant did not respond correctly within 5 s of the teacher turning to a blank page, the teacher lifted the blank page to reveal the text, and prompted a response as done

on the first error correction trial. Then the teacher turned back to the blank page and waited another 5 s for an independent response. This process continued until either: (a) the participant engaged in correct recall for the targeted page in the presence of the blank page or (b) the blank page was re-presented five total times (this latter criterion was never met). Error correction continued in this manner until the participant correctly recalled the final page of the story (i.e., page 8) in the presence of a blank page. When this occurred, the teacher provided praise and the reinforcer, using the smaller magnitudes from the RAR baseline (i.e., 1 piece or 60-s access). This magnitude was used to arrange for differential reinforcement of unprompted recall at criterion on the recall test relative to correct responses during error correction (Vladescu & Kodak, 2010).

Intervention plus Question Interspersal (Albert only)

Near the end of the study, a modification was introduced to Intervention for two of Albert's stories when his correct responses remained below mastery criteria for an extended period of time. Sessions were identical to Intervention as previously described except that during the Reading component, the teacher asked the comprehension question corresponding to each page immediately after they finished reading the page. For example, after reading "Phil is a big elephant" the teacher asked "What kind of animal is Phil?" before moving on to the next page. After each of these questions, the teacher waited 10 s for Albert to respond, and provided praise for correct question-answering. For incorrect or non-responses to these questions, the teacher repeated the question and modeled a correct answer (e.g., "What kind of animal is Phil? Phil is an elephant."). If Albert responded correctly after this prompt, the teacher delivered praise and turned to the next page. If Albert did not respond correctly within 10 s of the model, the teacher continued to model the correct response every 10 s until Albert correctly answered the question, at which point the teacher resumed reading the story.

Delay Probes

We scheduled *delay probes* at pre-determined times throughout the study in order to evaluate whether increased delays between reading and the recall test would affect correct responding. These delay probes were scheduled to occur: (a) shortly after reading first began for each story and (b) after the mastery criterion was met with each story (some exceptions to this schedule occurred, see Figs. 2, 3 and 4). As shown in Fig. 1, delay probe procedures were identical to the RAR baseline, except that the teacher did not read the story at the start of delay probe sessions. Instead, delay probes occurred either 1 day (Nick and Edgar) or 1 week (Albert) after the most recent session, meaning that last time the story had been read to participants was in that previous session (i.e., 1 day ago for Nick and Edgar, 1 week ago for Albert). We also conducted delay probes using longer delays at the end of the study for all participants (i.e., a 6-week delay for Albert, and a 3-month delay for Nick and Edgar).

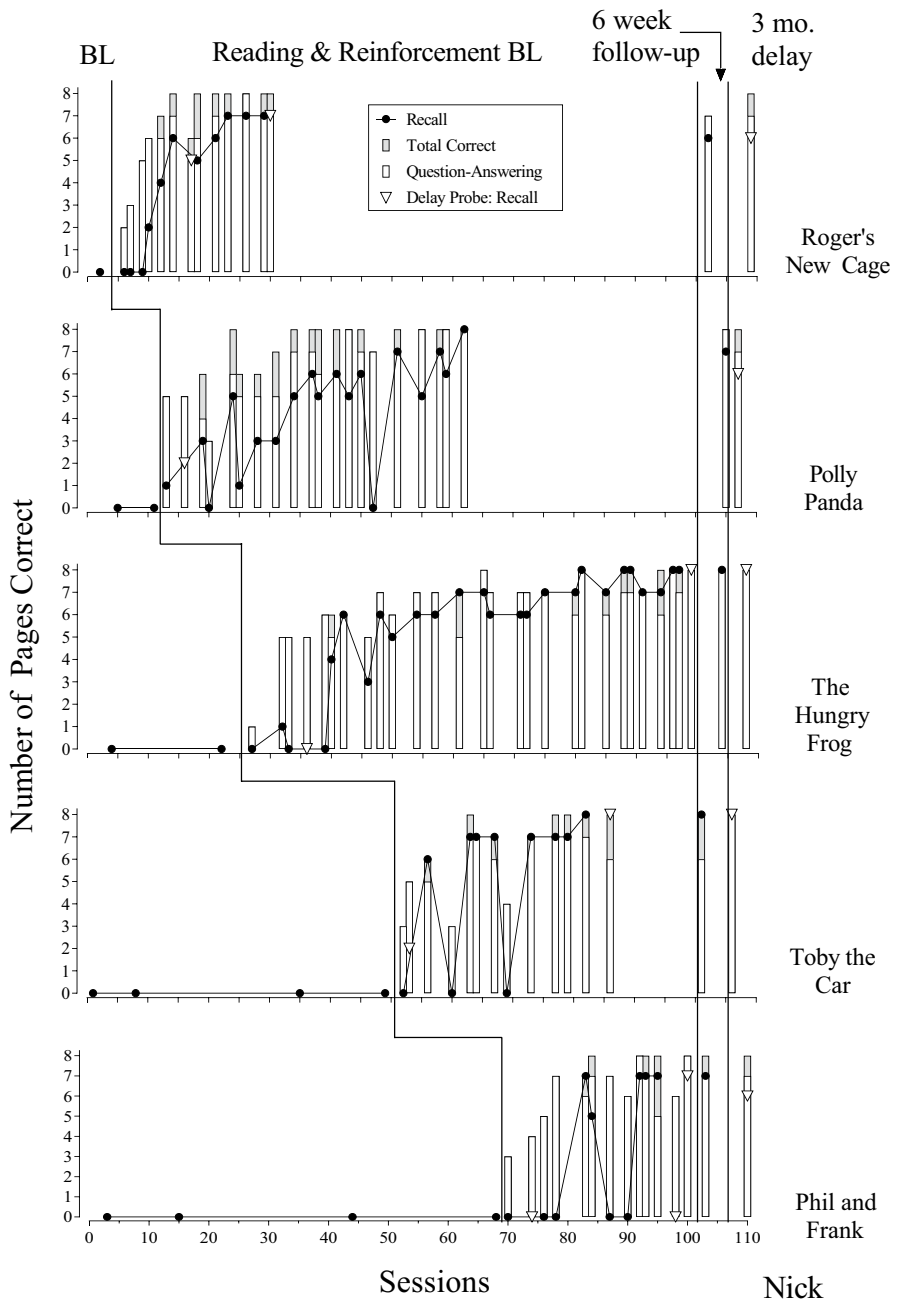


Fig. 2 Nick's correct recall, question-answering, and total correct for Nick during baseline (BL), reading and reinforcement baseline, and intervention

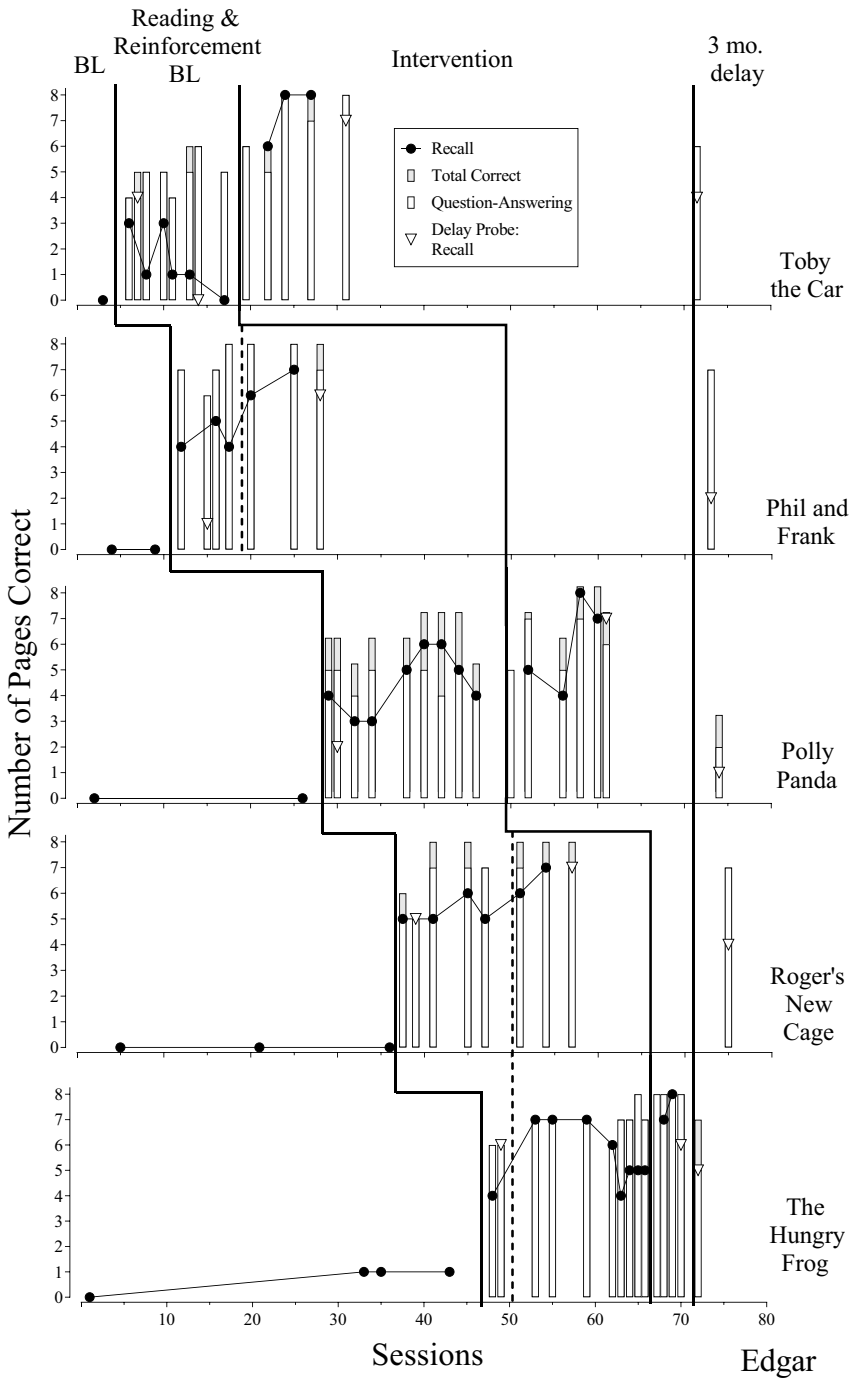
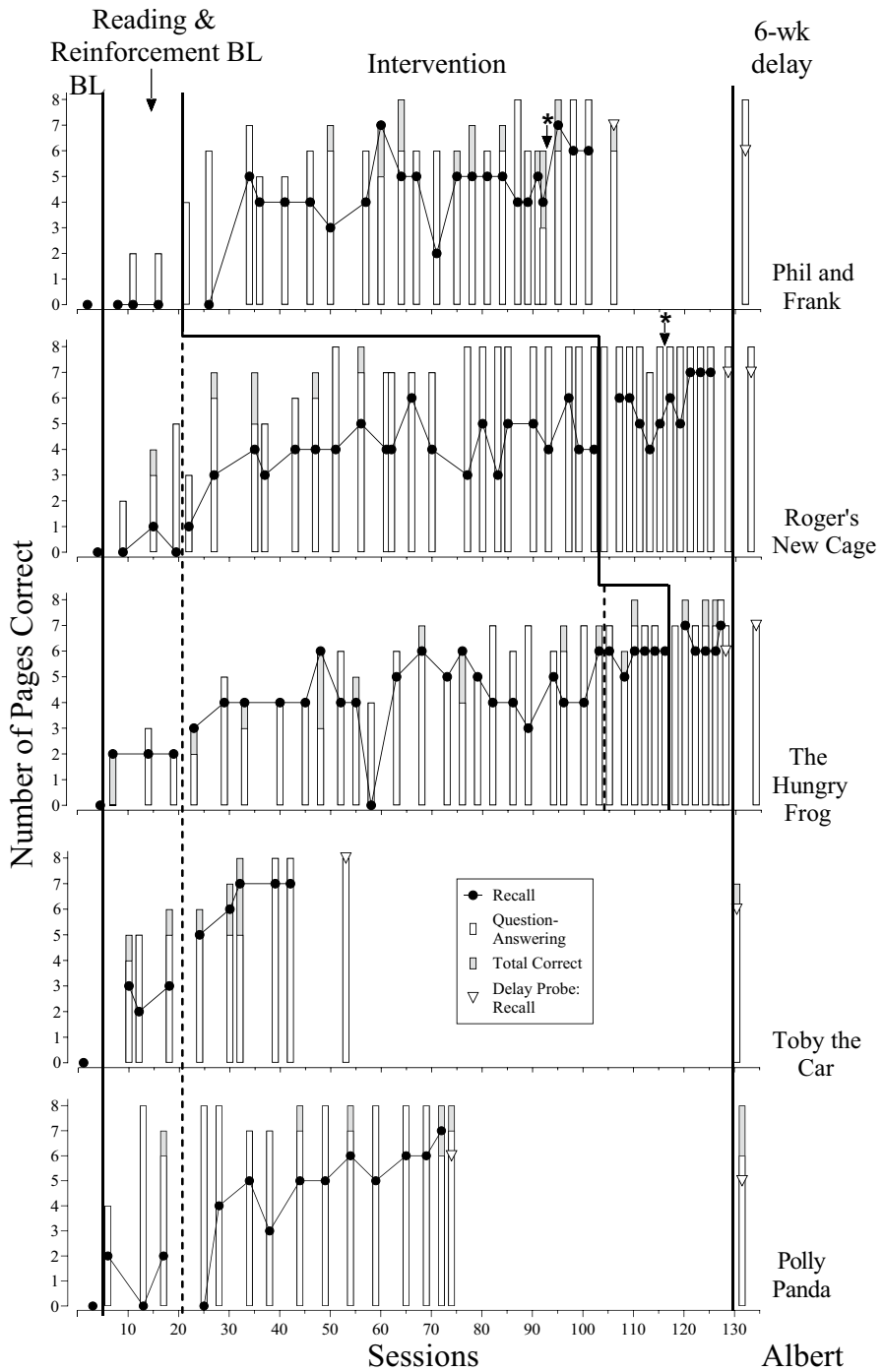


Fig. 3 Edgar’s correct recall, question-answering, and total correct for Edgar during baseline (BL), reading and reinforcement baseline, and intervention



- ◀ **Fig. 4** Albert's correct recall, question-answering, and total correct for Albert during baseline (BL), reading and reinforcement baseline, and intervention. Asterisks indicate the introduction of the question interspersal modification to intervention

Follow-up RAR Sessions

For Nick, we conducted an additional series of RAR baseline sessions for all stories 6 weeks after the last story was mastered.

Correlational Analysis

To address one of our secondary research questions regarding relationships between recall and question-answering, we calculated Pearson correlation coefficients between each participant's story recall and question-answering to supplement visual analysis regarding the extent to which improvements in recall were correlated with improvements in responses to comprehension questions. Because these data and the related research question are only correlational, we calculated correlations for each participant and each story using all sessions from across RAR and Intervention conditions, regardless of the condition that was in effect (excluding Baseline, in which no comprehension questions were asked).

Results

Scores on the word reading pre-assessment were 96% (107/112) for Nick, 96% (107/112) for Edgar, and 82% (92/112) for Albert. Figure 2 shows results for Nick, who mastered story recall for all stories in the RAR baseline. Nick did not correctly recall any pages for any stories in the initial no-reading baseline. RAR baseline was introduced first with Roger's New Cage. For the first three sessions, Nick engaged in zero correct recall but correctly answered multiple questions with an increasing trend, and began to engage in correct recall by the fourth RAR session. Nick then met mastery criteria for Roger's New Cage after 11 total sessions of RAR baseline. Nick mastered all other stories during RAR baseline, albeit after differing numbers of sessions and with slightly different patterns in recall and question-answering. Polly Panda, The Hungry Frog, Toby the Car, and Phil and Frank were mastered after 19, 26, 11, and 10 sessions of RAR baseline, respectively. For The Hungry Frog, Toby the Car, and Phil and Frank, Nick's responses followed a similar pattern as with Roger's New Cage (i.e., initial sessions with little to no recall, but correct question-answering). This pattern was slightly different with Polly Panda, for which Nick engaged in some correct recall from the beginning of the RAR baseline. In delay probes throughout the RAR baseline, Nick's correct recall and question-answering were generally equal to or within two pages of the number of correct responses in the most recent RAR baseline (Fig. 2). One exception is Session 98 (Phil and Frank), in which Nick only said "no" when asked to recall the story. The teacher initiated a second delay probe later in the same day (Session 100); Nick responded at mastery levels in

this probe. Nick also responded at mastery levels for all stories during 6-week follow-up RAR sessions and 3-month delay probes.

Figure 3 shows results for Edgar. Solid phase lines in Fig. 3 (and Fig. 4, described later) indicate condition changes, whereas dotted phase lines indicate when Intervention was introduced for a different story while a story was in the RAR baseline. These dotted phase lines are intended to enhance visual analysis of potential changes in recall across the multiple baselines. During the no-reading Baseline, Edgar engaged in zero correct recall for all stories except *The Hungry Frog*. In the last three baseline probes, Edgar said “frog eats flies,” which met the correct recall definition for page 6. Edgar had never read *The Hungry Frog*, so it is likely that these responses can be attributed to the title and prior learning history (i.e., learning that frogs eat flies). The RAR condition was introduced first with *Toby the Car*; Edgar engaged in some correct recall and question-answering, but correct recall eventually decreased to zero. During Intervention, recall immediately increased, and Edgar met mastery criteria after four sessions. Edgar mastered *Phil and Frank* after five RAR sessions; mastery-level responding for this story only occurred after Edgar first experienced the Intervention condition (with *Toby the Car*, session 19). Edgar engaged in some correct recall and question-answering from the start of the RAR condition with the story *Polly Panda*; these responses stabilized below mastery criteria with a downward trend. During Intervention with *Polly Panda*, Edgar’s responses did not increase immediately, but met mastery criteria after five sessions. Edgar mastered *Roger’s New Cage* after six total RAR sessions; however, as with *Phil and Frank*, Edgar’s responses were just below mastery levels until intervention was introduced with *Polly Panda* (session 50). Edgar’s responses for *The Hungry Frog* followed a similar pattern as for *Polly Panda*: approaching mastery criteria during RAR but eventually trending downward below mastery levels. Edgar then mastered *The Hungry Frog* after three Intervention sessions.

As with Nick, Edgar’s responses on delay probes throughout the study were usually within two pages of responses in the prior RAR or Intervention session. However, Edgar responded below mastery levels for all stories in the 3-month delay probes. Of note, a follow-up session that included reading was not conducted at the end of the study for Edgar, as was done with Nick. Thus, the final delay probes were conducted 3 months from the last reading of *The Hungry Frog*, but at greater and unequal delays from the last reading of the other four stories.

Albert (Fig. 4) recalled zero pages during initial no-reading baseline probes across all stories. When the RAR condition was introduced simultaneously across all stories, recall was low and stable across all for three sessions. Thus, Intervention was introduced first with *Phil and Frank*; Albert’s question-answering increased immediately, but correct recall did not increase until the third Intervention session. Across the other four stories, correct recall during RAR increased by two to three pages within one or two sessions of Intervention beginning with *Phil and Frank*. During the remainder of Intervention with *Phil and Frank*, correct responses increased but eventually stabilized just below mastery levels. Thus, we introduced the question-interspersal modification in Session 92, and Albert met mastery criteria after four additional sessions. Across the remaining stories, *Toby the Car* and *Polly Panda* were mastered in the RAR condition while Intervention was ongoing with *Phil and Frank*. Recall for *Roger’s New Cage* and the *Hungry Frog* stabilized just below mastery levels in RAR when *Phil*

and Frank were in Session 101. Thus, Intervention was introduced with Roger's New Cage; we saw no reliable improvements in recall, and question-answering remained at the ceiling (i.e., eight questions correct). We introduced question interspersal in Session 115 and Albert met mastery criteria after six additional sessions. Responses for The Hungry Frog remained stable throughout the first five sessions of Intervention with Roger's New Cage, and thus Intervention was introduced to The Hungry Frog in Session 118. Albert mastered The Hungry Frog after six Intervention sessions.

In Albert's 1-week delay probes conducted throughout the study, recall and question-answering were within one page of correct responses in the prior week's session. In 6-week delay probes at the end of the study, Albert responded at mastery levels for Phil and Frank Roger's New Cage, and near mastery levels for the other stories.

Correlations between Recall and Question-Answering (All Participants)

Before calculating correlation coefficients, we analyzed the distribution of our overall recall and question-answering data, which showed negative skew (-0.69 for recall, -0.92 for question-answering) and positive kurtosis (2.57 for recall, 3.55 for question-answering). The following correlation coefficients should be interpreted with this data distribution in mind. Table 3 and Fig. 5 show correlation coefficients between correct recall and correct question-answering across participants and across stories. Positive correlations between recall and question-answering were identified for all stories and all participants, many of which were statistically significant and greater than 0.5 (a large correlation; Cohen, 1988; Hemphill, 2003; Lovakov & Agadullina, 2021). The overall correlation coefficient between these variables for each participant was also strong and statistically significant (see Fig. 5 and Table 3 for all r coefficients and associated p values).

Discussion

The purpose of the present study was to replicate and extend the Valentino et al. (2015) study of a story recall intervention for children with ASD with three main questions. First, we implemented two procedural changes recommended by

Table 3 Correlation coefficients (r) between correct story recall and correct question-answering across participants and stories

Participant	The Hungry Frog	Phil and Frank	Polly Panda	Roger's New Cage	Toby the Car	All Stories
Nick	0.71 ($p < .001$)*	0.34 ($p = .33$)	0.69 ($p = .001$)*	0.88 ($p < .001$)*	0.94 ($p < .001$)*	0.69 ($p < .001$)*
Edgar	0.42 ($p = .19$)	0.49 ($p = .40$)	0.60 ($p = .03$)	0.40 ($p = .43$)	0.78 ($p < .001$)*	0.6 ($p < .001$)*
Albert	0.63 ($p < .001$)*	0.61 ($p < .001$)*	0.09 ($p = .77$)	0.78 ($p < .001$)*	0.64 ($p = .09$)	0.5 ($p < .001$)*

Note. p values are shown in parentheses after each correlation coefficient

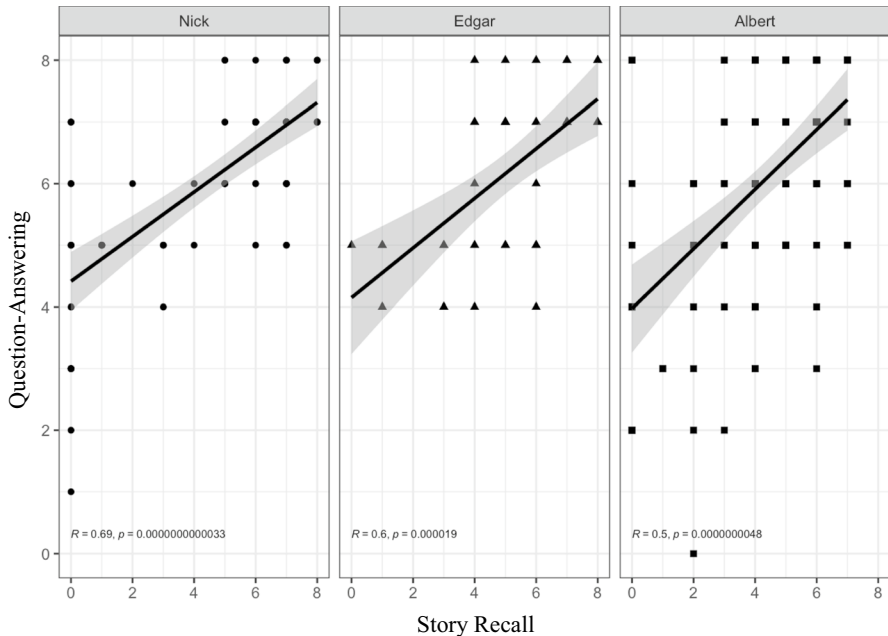


Fig. 5 Correlations between question-answering and recall for each participant (across all stories). *Note.* The shaded gray areas indicate the 95% confidence interval

Valentino et al.: (a) using differential reinforcement of unprompted responses during Intervention and (b) using time-based termination criteria in all recall tests. Second, we replicated the use of a reading-and-reinforcement condition, to explore whether patterns that suggested generative learning in Valentino et al. could be replicated with additional participants. Finally, we explored two additional variables: (a) whether correct recall varied at scheduled delays from reading and (b) to what extent improvements in story recall were associated with improvements in answering related, simple comprehension questions. We conducted a combined multiple probe and multiple baseline design across stories with three children with ASD, all of whom showed educationally significant deficits in story recall on standardized reading assessments. Two participants mastered story recall after exposure to Intervention (Edgar and Albert), with one unplanned intervention modification for one of these participants (Albert). The third participant (Nick) met mastery criteria in the RAR condition before Intervention was introduced with any story. Each of these findings has important implications for clinical and educational practice, as well as future research.

With respect to our first main research question, Edgar's and Albert's data indicate that the Intervention condition produced reliable improvements in story recall, especially with the first story for which Intervention was introduced (Figs. 3 and 4). Moreover, no additional modifications to Intervention were required for Edgar, and only one (question-interspersion) was required for Albert. Valentino et al. (2015) recommended that future studies modify their intervention to include differential

reinforcement and time-based termination criteria, specifically as a means of avoiding the number and complexity of idiosyncratic modifications that were required in that study. Those authors noted that a lack of time-based termination criteria may have artificially suppressed responding in all experimental conditions by arranging a contingency between repetitive statements and termination of the recall test; they also suggested that the nondifferential reinforcement they provided for independent recall versus recall after error correction may have similarly suppressed responding during intervention. The authors attributed the need for idiosyncratic intervention changes in that study to these specific response patterns. In the current study, we saw no such patterns of repetitive responding; therefore, the smaller number of idiosyncratic modifications required in the present study might have resulted from making these two recommended procedural changes.

However, there are several other important differences between the current study and Valentino et al. (2015) that make a direct comparison of the intervention's efficacy challenging. For example, we ensured that the stories we used were appropriately matched to the participant's reading levels, which was not done in Valentino et al.; we also used less stringent mastery criteria. Together, these changes may have made the target behavior "easier" to master. It is also possible that the participants in the current study presented with stronger reading skills at the start of the study (Table 1); there were no standardized reading assessments in Valentino et al. that could be used to compare participant skill sets across these two studies. Future research conducting component analyses, future studies that include standardized reading assessments, or a combination thereof, remains important to clarify the relative contributions of differential reinforcement and time-based termination criteria to the overall effects of this Intervention package.

Second, we replicated an effect noted by Valentino et al. (2015) with Edgar and Albert wherein changes in recall occurred across stories in the multiple baseline design after Intervention was introduced with other stories. This pattern suggests that improvements in a generalized repertoire of story recall may have emerged throughout the course of the study. In our study, this effect was most pronounced after Intervention was introduced with the first story for each participant (Toby the Car for Edward, Fig. 3; Phil and Frank for Albert, Fig. 4), but this effect was also replicated to a lesser extent for both participants on subsequent iterations of Intervention. However, it is important to note that in our study, as in the Valentino et al. study, the experimental control for this finding is substantially limited, because multiple baseline or multiple probe designs do not allow for a reliable empirical demonstration of such an effect. Improvements across tiers of a multiple baseline design constitute threats to experimental control at the same time as they are suggestive of generative responding (see Valentino et al., 2015 for further discussion). This challenge of capturing generalized repertoires in single-case designs has precedent in other studies of complex verbal behaviors such as recall that may be generalized in nature (e.g., Axe & Sainato, 2010; Frampton & Shillingsburg, 2020; Kohler & Malott, 2014; Shillingsburg et al., 2018). However, such outcomes are of great clinical and educational importance. If initial exposure to this intervention with a few stories leads to mastery-level recall under conditions of reading-and-reinforcement alone with other stories, a generalized

repertoire of story recall may have emerged and intensive intervention is no longer necessary.

Thus, it is important for future story recall studies to be conducted using alternative experimental designs such that this potential generative learning can be better captured while reducing threats to internal validity. For example, a large set of stories could be reserved for generalization tests while Intervention is provided for a smaller set of target stories, one at a time. If evaluated in a multiple baseline across participants design, elevated responding in the reserved generalization stories would not pose a threat to the demonstration of a functional relation.

Nick's results also present challenges for interpretation and application. Because mastery-level recall in the initial RAR condition was unanticipated, there was no baseline in which Nick actually read stories. Thus, there was no experimental control with respect to the effects of the RAR baseline on Nick's behavior, and a host of potential confounding variables such as practice effects and extraneous variables cannot be ruled out. Nevertheless, Nick's data are noteworthy and worth reporting because they suggest questions to be explored in future research. Namely, Nick's results suggest that at least some children with ASD who present with substantial deficits in story recall (Table 1) can show marked improvement with much simpler intervention packages. The RAR condition comprises components that are commonly used in shared book reading contexts with typically developing children, such as repeated reading (e.g., Shimono, 2018), praise and reinforcement for correct recall (e.g., Dolezal et al., 2007), and asking comprehension questions about stories after reading them (e.g., Fleury et al., 2014; Hindman et al., 2008; Lever & Senechal, 2011). The RAR condition may be of considerable social acceptability for use in classrooms.

Future research is warranted to explore the potential benefits of this RAR condition using different experimental baselines and component analyses. Moreover, as mentioned above, it will be important to continue to conduct standardized reading assessments in such future research to enhance the external validity of findings; more research is needed to identify the participant characteristics and skill sets that are associated with the various outcomes observed across participants in these studies to date. We collected standardized assessment data for our participants (Table 1), but these data alone do not seem sufficient to account for the major differences between Nick's results and those of Albert and Edgar. Additional types of reading pre-assessments should be explored in future studies. A related limitation of this study is that although we conducted assessments (Schrank et al., 2014) with each participant at the start of the study, we did not conduct parallel assessments at the end of the study. Doing so may be a useful supplemental measure to add in future research.

Our third purpose was to collect data regarding two secondary variables suggested by Valentino et al. (2015). Specifically, we examined whether various delays between reading and recall tests would affect correct recall and whether improvements in recall were correlated with improvements in answering comprehension questions. In terms of delay, recall did not vary substantially for our participants across delays of 1 min, 24 h, or 1 week (Figs. 2, 3 and 4). However, longer delays (e.g., 3 months) were associated with decreases in recall. The social importance of responding with correct recall at extreme delays such as 3 months is unclear. For example, previous studies have found that children with ASD do not show

significant differences from typically developing children when delays are added to memory tests, and rather that lower performance on memory tests in this population more often results from language, cognitive, or skill deficits rather than some core difference in responding at increased delays (e.g., Southwick et al., 2011). Thus, the decrements in recall that we observed at 3-month delays from the time of reading may be neither abnormal nor problematic.

The relationship between story recall and question-answering, and the notion that story recall may be a pre-requisite or aid for answering questions about stories, was a core rationale in the Valentino et al. (2015) study. We asked direct, literal comprehension questions after every session, and calculated Pearson correlation coefficients between recall and question-answering to supplement visual analysis. We identified strong, positive, and statistically significant correlations between these two variables for all three participants (Table 3, Fig. 5). However, because these data are correlational, the exact nature of this relationship remains unclear and requires further study. For example, visual analysis of Figs. 2, 3 and 4 suggests that although question-answering and recall often increased together, there were also many instances in which participants answered questions correctly despite engaging in no correct recall (e.g., all stories for Nick, initial RAR Baselines for some of Albert's stories). Those latter instances seem to contradict the notion that recall is a pre-requisite for answering questions about stories.

However, this does not preclude the possibility that subsequent exposure to Intervention was partially responsible for subsequent improvements in question-answering, or the possibility that participants engaged in covert recall during those early sessions. In keeping with Palmer's (1991) account, it seems reasonable that one could use recall as a problem-solving strategy to overtly or covertly generate verbal stimuli to answer questions in the absence of the text. Of course, such an account based on private events remains speculative and in need of further study. Future studies could attempt to collect data on overt by-products of covert problem-solving behaviors during recall tests, comprehension tests, or both (e.g., Kisamore et al., 2011; Sautter et al., 2011). Moreover, the correlations we found between these two variables could indicate a potential therapeutic effect in the opposite direction; some research suggests that intervening directly on listening comprehension can produce increases in narrative retell (Solari et al., 2020). Future research could attempt to test for this effect by probing story recall both before and after an intervention for question-answering. Overall, the exact nature of the relationship among these two behaviors remains an important topic for future research, with practical implications for curricular sequencing in reading interventions for children with and without ASD.

Several limitations of this study should be noted. As previously mentioned, mastery of story recall during the RAR condition (before intervention with Nick, after some initial intervention with Edgar and Albert) poses threats to experimental control. Future research should explore alternative experimental arrangements to answer these questions. Another important limitation is that we did not collect procedural integrity data. Procedural integrity data can be of special importance when interventions are implemented by teachers (McIntyre et al., 2007), as was the case in this study. Moreover, teachers' procedural integrity with different types of

interventions may be differentially affected by the many distractions of a classroom environment (Berdeaux et al., 2022), and various levels of procedural integrity may be required for certain interventions to be successful (e.g., Joslyn & Vollmer, 2020). The role of teacher procedural integrity in the success of this specific type of story recall intervention is not possible to evaluate in the current study because we did not collect these data, and remains important to explore in future research. Other limitations include the mastery criteria and response definition, which were not based on established norms. We set lower mastery criteria than Valentino et al. (2015), with the intention of mirroring the amount of recall necessary to be functional in an educational or social environment; it is likely socially acceptable to omit some details from recall, especially if one can answer questions about those details when asked (see Response Measurement). However, it is possible that our mastery criteria still required more detailed recall than is typical. We required participants to recall 75% of story content, but typically developing school-aged children often recall between 10–50% of story content (Reed & Vaughn, 2012). More research is needed to validate recall mastery criteria for children with and without ASD of various ages.

Finally, the current study and the Valentino et al. (2015) study together have implications for future research and practice in verbal behavior. Valentino et al. conceptualized story recall as a chain of complex intraverbal responses established primarily via textual and echoic prompting. This analysis may be suited for cases in which participants respond with every page of a story in consecutive order, but seems insufficient for instances in which participants recalled stories in the RAR baseline without intervention or recalled story pages out of order. (Though not presented here, a closer analysis of our data indicates that participants often did recall stories in different orders.)

When recall does not occur sequentially, it seems reasonable that it may occur under joint control of a number of different variables, of which intraverbal control is only one (Lowenkron, 2004; Palmer, 2016). One possibility is that recall responses may have initially occurred under control of the various stimuli that were present when teachers read to the participants: the auditory stimuli produced by the teacher's reading, the textual stimuli on the page, the visual stimuli (pictures) on the page, or some combination of these (Valentino et al., 2015). During the reading component, participants may have also behaved as speakers by overtly or covertly engaging in tact or textual response or reading along with the teacher (Miguel, 2016; Schlinger Jr, 1995). Anecdotally, participants sometimes read out loud with the teacher during the reading component. Transcripts of our sessions also show that participants sometimes recalled story elements that were represented only by story pictures and not by the text. For example, in the session 22 recall test for Roger's New Cage, Albert said, "Roger sweep them off." Although the text does not reference sweeping, Roger is shown sweeping the floor in the pictures (see Supplemental Information). We did not conduct a detailed analysis at this level, and so the degrees to which textual, auditory, or visual stimuli contributed to recall cannot be fully described, but could be explored in future studies.

We designed the stories and reading component of sessions to include auditory, visual, and textual stimuli all at once, because this also occurs when stories are typically read in preschool or elementary contexts; parents and teachers often read

picture books to their children while showing them the text and pictures. Future research could conduct component analyses to determine how these various stimuli interact to contribute to the development of reading and recall repertoires, and what such data mean for verbal behavior accounts of recall more broadly speaking. Similar statements can be made, and future analyses conducted, regarding the degree to which participants made paraphrases or omissions during recall as permitted by our response definition. Anecdotally, Nick provided nearly verbatim recall throughout the entire study, whereas Albert's and Edgar's recall initially contained more substitutions, omissions, and additions but became more verbatim throughout the study. While a full analysis of response variation is beyond the scope of this study, these observations suggest another area of inquiry for future research regarding recall behaviors.

The main implications of these findings for educational and clinical practice are that the overall intervention approach studied here (RAR baselines, followed by the Intervention condition if necessary, with ongoing RAR baselines to test for generalization) can produce mastery-level story recall and mastery-level answers to comprehension questions for some children with ASD. However, future research using alternative experimental designs is needed to evaluate the extent to which this intervention approach, broadly implemented, can produce generalized changes in reading and recall outcomes. Additionally, the fact that these interventions were conducted in a classroom setting by teachers suggests a general feasibility of implementation. The present findings also highlight the need for additional research in this area containing robust, standardized pre-assessments to more precisely characterize the external validity of these findings and help identify for whom optimal outcomes under various teaching arrangements are most likely.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40616-023-00183-2>.

Acknowledgements We would like to thank the Florida Autism Center for their collaboration and support of this project. We thank Yoko Fisher, Tim Greeley, Carlos Lopez, and Sharene Mullings for creating the artwork that was used in this study. We also thank Cynthia Dela Rosa, Lindsay Kalick, Abigail Petronelli, Naomi Edwards, and Jazzmique Lake for their assistance with data collection, and Cindy Cahill for assistance with sessions. We thank J. Stephanie Gonzalez for her helpful comments on an earlier draft of this manuscript, and Kathryn S. McCarthy for her assistance with statistical analysis.

Data Availability Additional data analyzed during this study are included in this article's supplementary information files. Additional datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Research Involving Human Participants All procedures performed in the study 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 Informed consent was obtained from all individual participants included in the study or their legal guardians.

References

- Aguirre, A. A., & Rehfeldt, R. A. (2015). An evaluation of instruction in visual imagining on the written spelling performance of adolescents with learning disabilities. *The Analysis of Verbal Behavior*, 31(1), 118–125. <https://doi.org/10.1007/s40616-015-0028-0>
- Axe, J. B., & Sainato, D. M. (2010). Matrix training of preliterate skills with preschoolers with autism. *Journal of Applied Behavior Analysis*, 43(4), 635–652. <https://doi.org/10.1901/jaba.2010.43-635>
- Bailey, B., & Arciuli, J. (2020). Reading instruction for children with autism spectrum disorders: A systematic review and quality analysis. *Review Journal of Autism and Developmental Disorders*, 7, 127–150. <https://doi.org/10.1007/s40489-019-00185-8>
- Baixauli, I., Colomer, C., Roselio, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, 59, 234–254. <https://doi.org/10.1016/j.ridd.2016.09.007>
- Berdeaux, K. L., Lerman, D. C., & Williams, S. D. (2022). Effects of environmental distractions on teachers' procedural integrity with three function-based treatments. *Journal of Applied Behavior Analysis*. <https://doi.org/10.1002/jaba.918>
- Bordignon, S., Endres, R. G., Trentini, C. M., & Bosa, C. A. (2015). Memory in children and adolescents with autism spectrum disorder: A systematic literature review. *Psychology & Neuroscience*, 8(2), 211. <https://doi.org/10.1037/h0101059>
- Campanaro, A. M., Vladescu, J. C., Kodak, T., DeBar, R. M., & Nippes, K. C. (2020). Comparing skill acquisition under varying onsets of differential reinforcement: A preliminary analysis. *Journal of Applied Behavior Analysis*, 53(2), 690–706. <https://doi.org/10.1002/jaba.615>
- Cariveau, T., La Cruz Montilla, A., Gonzalez, E., & Ball, S. (2019). A review of error correction procedures during instruction for children with developmental disabilities. *Journal of Applied Behavior Analysis*, 52(2), 574–579. <https://doi.org/10.1002/jaba.524>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Conine, D. E., & Vollmer, T. R. (2019). Relative preferences for edible and leisure stimuli in children with autism. *Journal of Applied Behavior Analysis*, 52(2), 557–573. <https://doi.org/10.1002/jaba.525>
- Day, R. R., & Park, J. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17(1), 60–73.
- Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, 34(1), 83–98. <https://doi.org/10.1007/s10802-005-9003-x>
- Dolezal, D. N., Weber, K. P., Evavold, J. J., Wylie, J., & McLaughlin, T. F. (2007). The effects of a reinforcement package for on-task and reading behavior with at-risk and middle school students with disabilities. *Child & Family Behavior Therapy*, 28(2), 9–25. https://doi.org/10.1300/J019v29n02_02
- Fleury, V. P., Miramontez, S. H., Hudson, R. F., & Schwartz, I. S. (2014). Promoting active participation in book reading for preschoolers with autism spectrum disorder: A preliminary study. *Child Language Teaching and Therapy*, 30(3), 273–288. <https://doi.org/10.1177/0265659013514069>
- Frampton, S. E., & Shillingsburg, M. A. (2020). Promoting the development of verbal responses using instructive feedback. *Journal of Applied Behavior Analysis*, 53(2), 1029–1041. <https://doi.org/10.1002/jaba.659>
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58(1), 78–79. <https://doi.org/10.1037/0003-066X.58.1.78>
- Hindman, A. H., Connor, C. M., Jewkes, A. M., & Morrison, F. J. (2008). Untangling the effects of shared book reading: Multiple factors and their associations with preschool literacy outcomes. *Early Childhood Research Quarterly*, 23(3), 330–350. <https://doi.org/10.1016/j.ecresq.2008.01.005>
- Joslyn, P. R., & Vollmer, T. R. (2020). Efficacy of teacher-implemented Good Behavior Game despite low treatment integrity. *Journal of Applied Behavior Analysis*, 53(1), 465–474. <https://doi.org/10.1002/jaba.614>
- Kim, Y.-S. G., & Pilcher, H. (2016). What is listening comprehension and what does it take to improve listening comprehension? *Interventions in Learning Disabilities*, 112(7), 1367–1387. <https://doi.org/10.1037/edu0000430>
- Kisamore, A. N., Carr, J. E., & LeBlanc, L. A. (2011). Training preschool children to use visual imagining as a problem-solving strategy for complex categorization tasks. *Journal of Applied Behavior Analysis*, 44(2), 255–278. <https://doi.org/10.1901/jaba.2011.44-255>

- Kohler, K. T., & Malott, R. W. (2014). Matrix training and verbal generativity in children with autism. *The Analysis of Verbal Behavior*, 30(2), 170–177. <https://doi.org/10.1007/s40616-014-0016-9>
- Krantz, P. J., Zalsenski, S., Hall, L. J., Fenske, E. C., & McClannahan, L. E. (1981). Teaching complex language to autistic children. *Analysis and Intervention in Developmental Disabilities*, 1(3–4), 259–297. [https://doi.org/10.1016/0270-4684\(81\)90003-3](https://doi.org/10.1016/0270-4684(81)90003-3)
- Lever, R., & Senechal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, 108(1), 1–24. <https://doi.org/10.1016/j.jecp.2010.07.002>
- Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485–504. <https://doi.org/10.1002/ejsp.2752>
- Lowenkron, B. (2004). Meaning: A verbal behavior account. *The Analysis of Verbal Behavior*, 20, 77–97. <https://doi.org/10.1007/BF03392996>
- Lowenkron, B. (2006). An introduction to joint control. *The Analysis of Verbal Behavior*, 22(1), 123–127. <https://doi.org/10.1007/BF03393034>
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659–672. <https://doi.org/10.1901/jaba.2007.659-672>
- McIntyre, N. S., Solari, E. J., Grimm, R. P., E Lerro, L., E Gonzales, J., & Mundy, P. C. (2017). A comprehensive examination of reading heterogeneity in students with high functioning autism: Distinct reading profiles and their relation to autism symptom severity. *Journal of Autism and Developmental Disorders*, 47(4), 1086–1101. <https://doi.org/10.1007/s10803-017-3029-0>
- MetaMetrics, Inc. (2017). *Lexile Analyzer*. Retrieved December, 2019, from <https://la-tools.lexile.com/free-analyze/>
- MetaMetrics, Inc. (2023). *Lexile Grade Level Charts*. Retrieved February, 2023, from <https://hub.lexile.com/lexile-grade-level-charts>
- Miguel, C. F. (2016). Common and intraverbal bidirectional naming. *The Analysis of Verbal Behavior*, 32(2), 125–138. <https://doi.org/10.1007/s40616-016-0066-2>
- Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on Verbal Behavior* (pp. 261–279). Context Press.
- Palmer, D. C. (2016). On intraverbal control and the definition of the intraverbal. *The Analysis of Verbal Behavior*, 32(2), 96–106. <https://doi.org/10.1007/s40616-016-0061-7>
- Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, 3, 187–217. <https://doi.org/10.1080/10888438.2010.538780>
- Ricketts, J., Jones, C. R. G., Happé, F., & Charman, T. (2013). Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders*, 43(4), 807–816. <https://doi.org/10.1007/s10803-012-1619-4>
- Sautter, R. A., LeBlanc, L. A., Jay, A. A., Goldsmith, T. R., & Carr, J. E. (2011). The role of problem solving in complex intraverbal repertoires. *Journal of Applied Behavior Analysis*, 44(2), 227–244. <https://doi.org/10.1901/jaba.2011.44-227>
- Schlinger, H. D., Jr. (1995). *A behavior analytic view of child development*. Springer Science & Business Media.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.
- Shapiro, E. S., Fritschmann, N. S., Thomas, L. B., Hughes, C. L., & McDougal, J. (2014). Concurrent and predictive validity reading retell as a brief measure of reading comprehension for narrative text. *Reading Psychology*, 35, 644–665. <https://doi.org/10.1080/02702711.2013.790328>
- Shillingsburg, M. A., Cariveau, T., Talmadge, B., & Frampton, S. (2017). A preliminary analysis of procedures to teach children with autism to report past behavior. *The Analysis of Verbal Behavior*, 33(2), 275–282.
- Shillingsburg, M. A., Frampton, S. E., Cleveland, S. A., & Cariveau, T. (2018). A clinical application of procedures to promote the emergence of untrained intraverbal relations with children with autism. *Learning and Motivation*, 62, 51–66. <https://doi.org/10.1016/j.lmot.2017.02.003>
- Shimono, T. R. (2018). L2 reading fluency progression using timed reading and repeated oral reading. *Reading in a Foreign Language*, 30(1), 152–179.
- Skinner, B. F. (1953). *Science and human behavior*. Macmillan.
- Skinner, B. F. (1974). *About behaviorism*. Vintage Books.

- Solari, E. J., Henry, A. R., McIntyre, N. S., Grimm, R. P., & Zajic, M. (2020). Testing the effects of a pilot listening comprehension and vocabulary intervention for individuals with autism. *Research in Autism Spectrum Disorders*, 71, 101501. <https://doi.org/10.1016/j.rasd.2019.101501>
- Southwick, J. S., Bigler, E. D., Froehlich, A., DuBray, M. B., Alexander, A. L., Lange, N., & Lainhart, J. E. (2011). Memory functioning in children and adolescents with autism. *Neuropsychology*, 25(6), 702–710. <https://doi.org/10.1037/a0024935>
- Spooner, F., Spooner, D., & Ulicny, G. (1986). Comparisons of modified backward chaining: Backward chaining with leap-aheads and reverse chaining with leap-aheads. *Education and Treatment of Children*, 9, 122–134.
- Valentino, A. L., Conine, D. E., Delfs, C. H., & Furlow, C. M. (2015). Use of a modified chaining procedure with textual prompts to establish intraverbal storytelling. *The Analysis of Verbal Behavior*, 31, 39–58. <https://doi.org/10.1007/s40616-014-0023-x>
- Vladescu, J. C., & Kodak, T. (2010). A review of recent studies on differential reinforcement during skill acquisition in early intervention. *Journal of Applied Behavior Analysis*, 43(2), 351–355. <https://doi.org/10.1901/jaba.2010.43-351>
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). The profile of memory in children with autism. *Neuropsychology*, 20(1), 21–20. <https://doi.org/10.1037/0894-4105.20.1.21>

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.