JSLHR

Research Article

The Neural Underpinnings of Processing Newly Taught Semantic Information: The Role of Retrieval Practice

Eileen Haebig,[a] Laurence B. Leonard,[b] Patricia Deevy,[b] Jennifer Schumaker,[b] Jeffrey D. Karpicke,[c] and Christine Weber[b]

Purpose: Recent behavioral studies have demonstrated the effectiveness of implementing retrieval practice into learning tasks for children. Such approaches have revealed that repeated spaced retrieval (RSR) is particularly effective in promoting children's learning of word form and meaning information. This study further examines how retrieval practice enhances learning of word meaning information at the behavioral and neural levels.

Method: Twenty typically developing preschool children were taught novel words using an RSR learning schedule for some words and an immediate retrieval (IR) learning schedule for other words. In addition to the label, children were taught two arbitrary semantic features for each item. Following the teaching phase, children's learning was tested using recall tests. In addition, during the 1-week follow-up, children were presented with pictures and an auditory sentence that correctly labeled the item but stated correct or incorrect semantic information. Event-related brain potentials (ERPs) were time locked to the onset of the words noting the semantic feature. Children provided verbal judgments of whether the semantic feature was correctly paired with the item.

Results: Children recalled more labels and semantic features for items that had been taught in the RSR learning schedule relative to the IR learning schedule. ERPs also differentiated the learning schedules. Mismatching label–meaning pairings elicited an N400 and late positive component (LPC) for both learning conditions; however, mismatching RSR pairs elicited an N400 with an earlier onset and an LPC with a longer duration, relative to IR mismatching label–meaning pairings. These ERP timing differences indicated that the children were more efficient in processing words that were taught in the RSR schedule relative to the IR learning schedule.

Conclusions: Spaced retrieval practice promotes learning of both word form and meaning information. The findings lay the necessary groundwork for better understanding of processing newly learned semantic information in preschool children.

Supplemental Material: https://doi.org/10.23641/asha.15063060

There has been a recent resurgence in research applying principles of retrieval practice to better understand learning. The core insight of this study is that, instead of retrieval serving solely as a means of assessing what has already been learned, the act of retrieval actually enhances learning. This property of retrieval makes it ideal for application to populations who face learning challenges. For example, we might be able to bolster learners' outcomes in an efficient way by adding a retrieval component to already-existing learning activities.

Children would seem to be an obvious beneficiary of retrieval practice, yet, to date, they have been the focus of relatively few studies (see Fritz et al., 2007; Karpicke et al., 2016; Leonard, Deevy, et al., 2019; and Leonard, Karpicke, et al., 2019, as recent exceptions). The aim of this study is to build on what is known about the process of retrieval practice in children's learning. We focus on word learning, as words and their meanings constitute an important part of the knowledge children must accrue during their development.

Karpicke and his colleagues have demonstrated in adult studies that retrieval attempts throughout a learning phase yield long-term retention improvements between 50% and 150% relative to conditions that have equivalent or greater

[a]Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge
[b]Department of Speech, Language, & Hearing Sciences, Purdue University, West Lafayette, IN
[c]Department of Psychological Sciences, Purdue University, West Lafayette, IN

Correspondence to Eileen Haebig: ehaebig1@lsu.edu

Editor-in-Chief: Stephen M. Camarata
Editor: Filip Smolik

Received August 17, 2020
Revision received February 18, 2021
Accepted April 13, 2021
https://doi.org/10.1044/2021_JSLHR-20-00485

Disclosure: The authors have declared that no competing interests existed at the time of publication.

study opportunities without retrieval prompts (Karpicke & Blunt, 2011; Karpicke & Roediger, 2007, 2008). These retrieval-based benefits appear to extend to children. For example, Karpicke et al. (2016) revealed that, relative to a repeated study condition, 9- to 12-year-old children were much more likely to recall and recognize words from a list that was taught if they participated in the retrieval practice condition. Karpicke et al. also found that the retrieval practice effect was broad and not influenced by the children's individual differences in reading comprehension skills or processing speed. In recent word learning studies in preschool children, Leonard, Karpicke, et al. (2019) found that both typically developing (TD) preschoolers and preschool children with developmental language disorder (DLD; often referred to as specific language impairment within the literature) also benefit from repeated retrieval practice. Leonard, Karpicke, et al. found that both groups of preschool children more successfully learned novel labels for exotic plants and animals when the labels were taught in a repeated retrieval condition relative to a repeated study condition. Although impressive, it is reasonable to ask whether similar benefits can be found under different retrieval conditions and whether certain retrieval schedules promote learning more than others.

Haebig et al. (2019) extended the Leonard, Karpicke, et al. (2019) study and demonstrated that preschool children benefitted from a particular retrieval schedule—one that spaced out retrieval opportunities. This schedule is often referred to as repeated spaced retrieval (RSR) practice or repeated retrieval with contextual reinstatement. Haebig et al. directly compared learning between words that were taught using either an RSR learning schedule or an immediate retrieval (IR) learning schedule. Importantly, the two conditions were equivalent in the number of exposures to each novel word label (i.e., word form) and the information about each item (i.e., meaning). Moreover, the conditions were also equivalent in the number of retrieval prompts. Thus, the only difference between the conditions was in the timing of the retrieval prompts. During the IR schedule, children were prompted to retrieve the word label and meaning immediately after teaching. In contrast, for words taught in the RSR condition, retrieval prompts were spaced out and appeared after exposure to other words (i.e., delayed retrieval). Haebig et al. found that preschool children with typical development and with DLD more accurately recalled the newly taught labels and meanings when words were taught using the RSR learning schedule. This effect was largest for recalling the novel labels. Word meaning recall was slightly higher when taught in the RSR condition; however, the word meaning recall accuracy approached ceiling, limiting the interpretation of the effects of RSR on learning semantic information.

Although the literature has yet to reach a consensus about the exact mechanism underlying retrieval-based effects, Karpicke et al. (2016) propose that the robust effects of retrieval practice are best explained by the "episodic context account." According to this account, repeated spaced retrieval practice prompts individuals to reinstate the prior learning context for a specific piece of information (i.e., contextual reinstatement; Lehman et al., 2014; Whiffen & Karpicke, 2017). This process thereby strengthens cues from the learning context that facilitated retrieval of the information. In addition, features from the current context may become integrated with the item representation to form an enhanced representation. Thus, the enhanced representation contains features from the past and current contexts that can serve as cues and thereby reduce the search set to further support success during subsequent retrieval attempts. In contrast to RSR schedules, IR schedules require little to no contextual reinstatement because the context has barely changed from the learning context that has just passed. Therefore, the representation of the learned item does not become significantly enhanced with the IR practice.

It is important to note that contextual reinstatement is believed to occur across various conditions. For instance, one can easily assume that contextual reinstatement occurs when the physical environment differs between the learning phase and the test phase. However, such an extreme change is not necessary to engage contextual reinstatement during a retrieval attempt. Whiffen and Karpicke (2017) present a particularly insightful example of retrieval benefits within a study with minimal contextual change. Within this study, college students were presented two lists of words that were separated by a distracter task. Following this, half of the participants were presented with all of the words from both lists together for a study phase. The other half of the participants were presented with all of the words from both lists but were asked to indicate for each word whether it had been presented in the first or the second list. Whiffen and Karpicke (2017) found during subsequent recall testing that accuracy was higher for the participants who had been asked to identify each word's list set membership. Importantly, the participants were not directed to attend to the order of the word presentation during the initial learning phase. Therefore, the participants had encoded details including the order of presentation without being directed to do so, and a prompt to recall this contextual detail positively influenced recall accuracy. Within the context of the Haebig et al. (2019) word learning study, contextual reinstatement is believed to be engaged when a word retrieval prompt occurs after the presentation of a word that differs from the target word. This slight change in context appears for words in the RSR learning condition but not for those in the IR learning condition.

Within the impressive retrieval practice studies that have been documented in the literature, it has become clear that the act of attempting to retrieve information facilitates learning. Adult studies have indicated that early success in recalling the target item leads to improved recall during a delayed test (Gordon et al., 2020). However, it has also been demonstrated that even if an individual is unsuccessful in retrieving the correct information, a benefit can still be experienced as long as feedback is provided. Butler et al. (2008) demonstrated the importance of feedback on both correct and incorrect retrieval attempts during the learning phase. For instance, adults experienced high levels of accuracy at the final test if they received feedback after providing an incorrect response during a retrieval attempt in the learning phase. Additionally, adults benefitted from feedback even when the retrieval attempt was accurate; this feedback benefit was particularly strong if the individual had low confidence in the accuracy of their response. In a word learning study conducted with sixth grade children, Metcalfe et al. (2009) demonstrated that children also benefit from feedback. As such, the recent retrieval practice studies that have been conducted with preschool children have included feedback (typically referred to as "study" or "re-study") after each retrieval prompt (e.g., Haebig et al., 2019; Leonard, Deevy, et al., 2019; Leonard, Karpicke, et al., 2019).

In addition to the behavioral results documenting recall performance, Haebig et al. (2019) presented neurophysiological data that revealed differences in the underlying neural correlates associated with processing the newly taught labels that were presented in the RSR and IR conditions. To do this, they examined an event-related brain potential (ERP) component that indexes semantic violations, that is, the N400 (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984). When the preschool children were shown a picture of one of the exotic plants or animals and presented with the incorrect label (i.e., mismatch trial), an N400 was apparent only if the label had been taught within the RSR learning condition. This finding provided further evidence that RSR schedules promote more effective learning of novel word forms.

The Haebig et al. (2019) study provided an important first step in using a multilevel approach to examine how some retrieval practice schedules promote word learning more than other schedules in preschool children. Importantly, successful word learning requires individuals to complete a complex process of processing the phonological information in the new words, associating it with the appropriate referent, mapping additional meaning onto the phonological form of the word, and relating the new meaning with previous conceptual knowledge (Alt et al., 2004; Hay et al., 2011; Nation, 2014). As such, the next logical step is to further examine how repeated retrieval practice influences learning of nonphonological aspects of words. This study extends the previous work conducted by Leonard, Karpicke, Haebig, and colleagues by applying a multilevel approach to examining how retrieval practice influences child learning of word meaning. This emphasis allows for a more thorough understanding of word learning that better captures depth of word knowledge.

It is important to carefully examine how young children learn semantic features of words because vocabulary depth is an important predictor of academic abilities. Specifically, depth of vocabulary knowledge predicts decoding skills for irregular words early in reading development (Ouellette & Beers, 2010), and it predicts reading comprehension skills (Lawrence et al., 2019; Ouellette, 2006). In addition, within the field of speech-language pathology, semantic learning is particularly important to understand because children with DLD have been found to have superficial word knowledge, evidenced by less information in word definitions (Marinellie & Johnson, 2002; McGregor et al., 2012, 2013) and less detailed definitions and drawings (McGregor et al., 2002). Children with DLD also recognize and report fewer semantic features of newly taught words (Alt et al., 2004). Lastly, compared to TD peers, children with DLD experience slower growth in lexical depth (even more so than growth in breadth of lexical knowledge; McGregor et al., 2013).

This Study

In this study, we present two experiments that expand upon Haebig et al. (2019) by shifting the focus to learning more semantically rich nouns using two retrieval-based learning schedules in TD preschool children. Haebig et al. demonstrated that retrieval practice that requires contextual reinstatement (i.e., RSR) promotes word learning relative to a retrieval practice that requires little to no contextual reinstatement (i.e., IR). As previously noted, this difference in learning condition was rather small when testing children's learning of word meaning, which potentially occurred because children attained near-ceiling levels during word meaning recall tests. Therefore, because true word knowledge incorporates much more than knowledge of the word's phonological form, in this study, we doubled the semantic information that was taught for each novel word. Therefore, in addition to learning novel labels for unfamiliar animals and plants, the children were explicitly taught two semantic facts: what each item likes and what happens to each item when it grows. We then examined whether learning of semantic information benefits more from an RSR learning schedule relative to an IR learning schedule.

In Experiment 1, we present behavioral data from a 5-min postlearning test and a 1-week postlearning test. The behavioral measures included a word form recall task, word meaning recall task, and a form–referent link recognition task. In addition, in Experiment 2, we expand upon the Haebig et al. (2019) study by incorporating a different electroencephalographic (EEG) task. Haebig et al. (2019) demonstrated in a picture–label match–mismatch task that children developed stronger referent–label (word form) pairings when they were learned in the RSR condition. In Experiment 2, we shift our focus to the newly taught semantic information, which is a novel contribution to the child ERP literature more broadly. The ERP correlates of semantic processing offer information about the depth of learning resulting from the two learning conditions. This study serves as an initial step in better understanding the role of repeated retrieval practice on semantic learning, which is important given that children with atypical language development (e.g., children with DLD and children with autism spectrum disorder) frequently demonstrate deficits in depth of word knowledge in addition to limitations in breadth of word knowledge (e.g., Anns et al., 2020; McGregor et al., 2012, 2013).

Across the two experiments that we present in this study, we ask the following questions: Does RSR practice enhance novel word learning to a greater degree than IR practice? Is this advantage seen for both word form and meaning? Based on previous findings, we hypothesize that word forms and meanings taught in the RSR condition will yield more successful learning than word forms and meanings taught in the IR condition. Furthermore, we predict that we will observe evidence of enhanced RSR-associated learning of word meaning information at both the behavioral and neural levels.

Experiment 1

Method

Participants

Participants included 20 TD children that were between the ages of 4;0 and 6;3 (years;months; *M* = 5.05 years, *SD* = 1.63). The sample included 11 girls and nine boys. All of the children were reported to be Caucasian, and two children were also reported to be Hispanic. The study was approved by the Institutional Review Board. All participants provided verbal assent, and a parent or legal guardian provided informed written consent.

Standardized Assessments

The children completed a battery of standardized assessments. Receptive vocabulary was assessed using the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007), and expressive vocabulary was assessed using the Expressive Vocabulary Test–Second Edition (Williams, 2007). Nonverbal cognitive scores were assessed using either the Primary Test of Nonverbal Intelligence (*n* = 5; Ehrler & McGhee, 2008) or the Kaufman Assessment Battery for Children–Second Edition (*n* = 10; Kaufmann & Kaufman, 2004); however, nonverbal cognitive scores were not obtained for five children because the visit protocol was shortened. Each child performed within or above 1 *SD* from the mean on all language and cognitive measures. In addition, parents confirmed that their children were not currently and did not have a history of receiving special education services. See Table 1 for participant characteristics.

Hearing was screened at 20 dB through headphones at 500, 1000, 2000, and 4000 Hz (American Speech-Language-Hearing Association, 1997). All children passed each frequency in at least one ear. Handedness was assessed using an abbreviated handedness assessment (Edinburg Handedness Inventory; Oldfield, 1971). All children were right-handed with the exception of one child who was left-handed.

Table 1. Participant characteristics.

Characteristic	*M*	*SD*
Chronological age (years)	5.05	0.63
Nonverbal cognition (standard score)[a]	117.13	12.64
Receptive vocabulary (standard score)[b]	118.40	10.06
Expressive vocabulary (standard score)[c]	114.65	10.07
Maternal years of education	16	2.18

[a]Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008) or Kaufman Assessment Battery for Children–Second Edition (Kaufmann & Kaufman, 2004). Nonverbal cognitive scores were not obtained for five children. [b]Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007). [c]Expressive Vocabulary Test–Second Edition (Williams, 2007).

Word Learning Task

Children were taught six novel words that corresponded to six unfamiliar pictures. Two words were monosyllabic, following a consonant–vowel–consonant (CVC) syllable shape, and four words were disyllabic with syllable–initial stress with an even number following a consonant–vowel–consonant–vowel (CVCV) and consonant–vowel–consonant–vowel–consonant syllable shape. Together, monosyllabic and disyllabic words make up 90% of the word tokens that are spoken to children between 2 and 6 years of age, based on the child-directed language transcripts in the CHILDES database (Roark & Demuth, 2000). The consonants within the novel words consisted of early-emerging sounds that can be easily produced by most preschoolers. None of the novel words contained the same word–initial phoneme. The six novel words were the following: /nɛp/, /jʌt/, /daɪbo/, /fumi/, /kodəm/, and /pobɪk/. The two CVC novel words also were used in the Leonard, Karpicke, et al. (2019) study, and all six novel words were used in Haebig et al. (2019); none of the participants in this study participated in the Haebig et al. or Leonard, Karpicke, et al. study.

Three novel words were taught in the IR learning condition, and three were taught in the RSR learning condition. Furthermore, each novel word was counterbalanced for the learning condition (IR vs. RSR) across children. The novel words were matched between the learning condition on syllable shape, phonotactic probability (average biphone frequency), and phonological neighborhood density, based on the Storkel and Hoover (2010) child language corpora database. The picture referents of exotic plants and animals consisted of colored photographs used by McGregor (2014); the real names of these items are typically unknown even to adults. Our linguistic and pictorial stimuli can be found in our Supplemental Material S1.

The word learning task was presented using a computer presentation program on a laptop. The novel words were taught using a block design, with words blocked within each learning condition. The children completed four blocks across two consecutive days. Two blocks (each block consisting of an IR and RSR sequence) were presented on each day with a 5-min break provided between each block; each block presentation was approximately 10 min in length. Lastly, we counterbalanced the order of the learning conditions across children.

In addition to presenting the word form, we provided semantic information about each novel item (i.e., what each item likes and what happens when it grows). The words that provided the semantic information were early-acquired words that are included on the MacArthur–Bates Communicative Development Inventory (Fenson et al., 2006), which is a parent-report checklist to document word knowledge for children between the ages of 8 and

30 months. Importantly, none of the semantic information could be inferred from the picture; therefore, associative learning was necessary to correctly connect the semantic information to each novel word and pictured referent. Each novel word was presented 4 times within each trial, and each piece of semantic information was presented once (e.g., “This is a /nɛp/. It’s a /nɛp/. A /nɛp/ likes worms.” “When a /nɛp/ grows, it gets round.”). Therefore, in total, each novel word was presented 32 times, and each word meaning (i.e., what it likes and what happens when it grows) was presented 8 times. Furthermore, each word form and each word meaning had six retrieval opportunities during the learning phase. The auditory stimuli were recorded by a young adult female from the Midwest. The sound stimuli were normalized to have an amplitude of approximately 65 dB using Praat software (Boersma & Weenink, 2006).

At the beginning of the word learning task, children were told that they were going to play a game with two cartoon characters. The examiner explained, “they went on a trip and they learned about some new animals and plants. They want to tell you about each one so you can learn about them too. They will show you a picture and tell you the name, and what it likes and what happens when it grows. Do you think you can remember?” The examiner then used a laptop computer to present the word learning experiment. As depicted in Figure 1, during the first sequence of presentations of words that were taught in the RSR condition, the word form and meaning were initially presented to the children (i.e., first “study”). Following the initial study presentation, the children were immediately prompted to retrieve the information that was taught. In this retrieval trial, the picture of the target animal or plant was displayed on the screen and prompts for the word label and meanings were presented (form: “What’s this called? What do we call this?”; meaning-like: “What does this one like? What does it like?”; and meaning-grow: “What happens when it grows? When it grows, it…”). After each retrieval trial, the label and semantic information were presented again (i.e., second “study,” serving as feedback). The subsequent retrieval trials for the same word occurred only after the two other RSR words had been presented. Thus, the first retrieval occurred with no other words intervening, but the second two retrieval trials occurred after two other words had intervened; this formed what we referred to as a 0–2–2/RSR schedule. In other words, the last two retrieval trials were spaced out by two intervening words (creating a change in the temporal context). In contrast, all retrieval trials that occurred for words in the IR condition appeared immediately after a study trial for that same word. Thus, no other words were presented before retrieval trials, leading to a 0–0–0/IR schedule. Because the 0–0–0/IR schedule did not involve a change in temporal context, limited to no contextual reinstatement was assumed. Across both conditions, retrieval trials were always followed by “study” trials that served as feedback regardless of the child’s response. Additionally, as previously noted, the number of exposures and retrieval prompts were equivalent across both learning conditions.

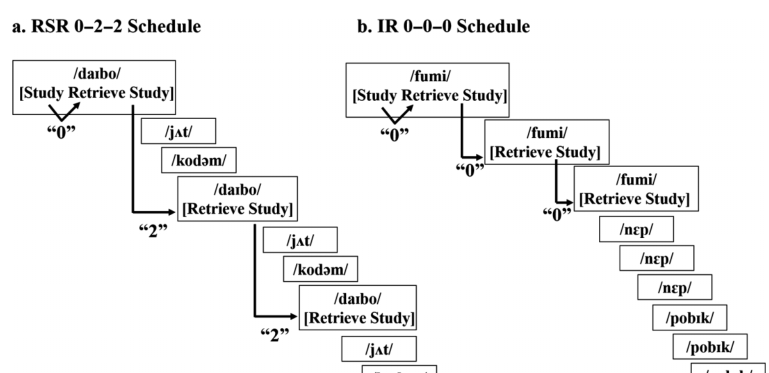
Behavioral Test Phase

Following the completion of the last block of the teaching phase on the second day, children were given a 5-min break and then participated in a test phase. Word form recall and word meaning recall were tested (e.g., form: “What’s this called? What do we call this?”; meaning-like: “What does this one like? What does it like?”; and meaning-grow: “What happens when it grows? When it grows, it…”). One week following the 5-min test, the children returned and repeated the word form and meaning recall tests; they also completed a form–referent link recognition test. In the form–referent link recognition test, they were presented with an array of four pictures and asked to point to the picture that corresponded to the label that was provided (e.g., “Where’s the /nɛp/?”).

Scoring and Reliability

We scored child responses to the word form and meaning recall tests according to accurate responses within each

Figure 1. Learning condition design. IR = immediate retrieval; RSR = repeated spaced retrieval.



learning condition. When scoring child attempts to produce the target, we first determined whether the child's production was a plausible or implausible attempt at the target. When doing so, we referenced each child's performance on a real word speech sound probe that we administered. The speech sound probe consisted of a list of phrases or short sentences that contained a word with a target phoneme to assess each phoneme and phoneme position that appeared in the novel words (e.g., /jʌt/ – "I like **y**ou" "Wear a ha**t**"). Next, we used an adapted version of the Edwards et al. (2004) scoring system, wherein we assigned up to 3 points to each consonant for correct place, manner, and voicing features and to each vowel for correct backness, height, and length features. An additional point was given if the child produced the accurate prosodic shape of the target (e.g., CVCV). With this combined score, we then determined whether the child's production had a higher score as an attempt at the target word compared with a score that would have been attained if the child's production had been an attempt at a different novel word that had been taught. For instance, the production of /pomɪg/ for the target word /pobɪk/ would earn 14 points (3 + 3 + 2 + 3 + 2 + 1), but it would only earn 10 points if it had been an attempt to produce /kodəm/ (2 + 3 + 1 + 2 + 1 + 1). If the child's production score aligned with the target word, the child was given credit for accurately producing the target (e.g., a child production of /pomɪg/ was credited as an acceptable attempt for /pobɪk/, resulting in 1 point contributing toward the word form accuracy score). Accuracy scores for the meaning responses were focused on the semantic content of the word; therefore, a point was granted even if the target word was mispronounced (e.g., "wocks" for "rocks"). When creating the word form and meaning accuracy scores, each word form response and meaning response resulted in either 0 points or 1 point, which were then summed within a learning condition. Each word form and meaning was tested twice, resulting in a maximum score of 6 points within each learning condition.

Scoring reliability was conducted by a second person who independently scored the 5-min and 1-week recall responses for five children (25%). Reliability was computed by comparing the accuracy scores. Word form recall agreement was 98.33%, and word meaning recall agreement was 100% (for both likes and grows meaning tests).

Analysis Plan

To address our research questions, we conducted a series of mixed-effect models. In these models, the dependent variable consisted of either the sum of the word form accuracy score or the sum of the word meaning accuracy score from each learning condition and test session. The independent variables consisted of learning condition (IR vs. RSR) and test session (5-min vs. 1-week). Random intercepts were set at the child level, and repeated measures (e.g., word form recall at 5-min test and 1-week test) were nested within the children.

Results

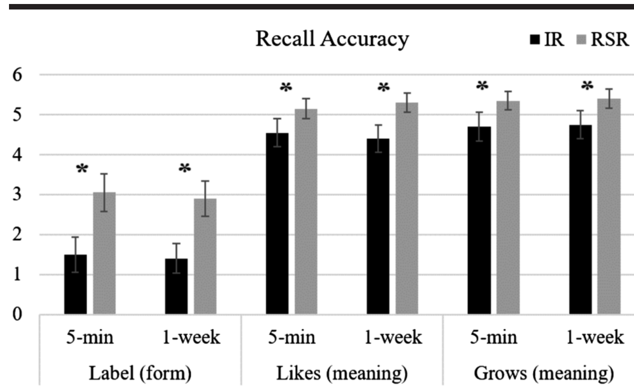
Word Form

In line with our previous RSR studies, the children more successfully recalled the novel words that were taught in the RSR condition relative to words that were taught in the IR condition (*Estimate* = 1.53, *SE* = 0.32, t = 4.71, p < .001). On average, the children scored 1.5 points higher in the RSR condition than in the IR condition (5-min: M_{IR} = 1.50, M_{RSR} = 3.05, Cohen's d = 0.75; 1-week: M_{IR} = 1.40, M_{RSR} = 2.90, Cohen's d = 0.83). There were no significant differences in accuracy between the testing sessions (time: *Estimate* = −0.13, *SE* = 0.32, t = −0.39, p = .70), and there was no interaction between learning condition and time (*Estimate* = −0.05, *SE* = 0.65, t = −0.08, p = .94). The RSR advantage during the 5-min and 1-week tests is notable given that recall accuracy on retrieval trials during the learning phase was higher for the words taught in the IR learning condition (average IR recall accuracy during learning = 16.41 and average RSR recall accuracy during learning = 8.47, with 18 recall prompts for each condition). See Figure 2 for the 5-min and 1-week recall accuracy.

Word Meaning

As also depicted in Figure 2, word meaning recall was higher overall relative to word form recall. Importantly, children more successfully recalled word meaning when the semantic information was taught within the RSR learning condition (*Estimate* = 0.70, *SE* = 0.15, t = 4.60, p < .001). On average, the children scored 0.75 points higher when recalling what each item likes (5-min: M_{IR} = 4.55, M_{RSR} = 5.15, Cohen's d = 0.44; 1-week: M_{IR} = 4.40, M_{RSR} = 5.30, Cohen's d = 0.69) and 0.65 points higher when recalling what happens when the item grows when taught in the RSR condition relative to the IR condition (5-min: M_{IR} = 4.70, M_{RSR} = 5.35, Cohen's d = 0.48; 1-week: M_{IR} = 4.75, M_{RSR} = 5.40, Cohen's d = 0.48). Additionally, the recall

Figure 2. Novel word recall—Form and meaning. Error bars represent standard errors. Each word form and meaning was tested twice, resulting in a maximum score of 6 within each learning condition. IR = immediate retrieval; RSR = repeated spaced retrieval.



accuracy of the semantic information did not significantly differ according to the type of semantic information (*Estimate* = −0.20, *SE* = 0.15, $t = -1.31$, $p = .19$). Also, there was no significant effect of time (*Estimate* = 0.03, *SE* = 0.15, $t = 0.16$, $p = .87$). Lastly, there were no significant interactions across learning condition, time, and semantic information (*p*s > .60).

Form–Referent Link Recognition

We also examined pointing accuracy for the form–referent link recognition test that was administered at the 1-week test. There was no significant difference in pointing accuracy between the learning conditions (*Estimate* = −0.45, *SE* = 0.25, $t = -1.77$, $p = .076$). Overall, pointing accuracy was high ($M_{IR} = 5.80$, $SD_{IR} = 0.52$; $M_{RSR} = 5.35$, $SD_{RSR} = 1.04$). Nine of the 20 children performed at ceiling for both learning conditions during the form–referent link recognition test; seven additional children only produced one error during the test.

Discussion

The behavioral results gathered from Experiment 1 provide additional support that RSR enhances word learning and retention more than a repeated IR schedule. Importantly, these results demonstrate that RSR facilitates learning of both word form information and word meaning information. This finding is also notable because, relative to the Haebig et al. (2019) study, we have doubled the semantic information that was associated with each novel word. Despite this increased semantic load, children were able to continue to demonstrate learning of word form and meaning. The facilitation of multiple components of word learning is important because effective word knowledge extends beyond having a phonological representation of a word and mapping it to a referent. Holding a deeper understanding of the word by associating it with other known concepts allows the word to be more fully integrated within the child's mental lexicon and world knowledge. Furthermore, as shown in previous work, both breadth and depth of word knowledge are important for concurrent and future reading skills (Ouellette, 2006; Ouellette & Beers, 2010).

The current findings also provide further support that spacing out retrieval is important. As previously noted, it is believed that retrieval attempts that are spaced apart from the presentation of the target information promote contextual reinstatement. During spaced retrieval, the child is likely to reinstate the prior learning context for a specific piece of information and strengthen cues from the learning context that facilitated retrieval of the information. Additionally, enhanced representations are believed to develop during retrieval attempts because features from the context during retrieval may become integrated with the item representation. Lastly, the feedback that was provided (in the form of a study trial that directly followed the retrieval trial) also likely supported performance during the 5-min and 1-week tests. The importance of spaced retrieval is especially notable because, during the learning phase, the children successfully responded more frequently to the retrieval prompts that occurred within the IR schedule relative to the RSR schedule. Therefore, the production of the word form and meaning during the learning phase alone is not sufficient to yield successful recall during the 5-min test or the 1-week test. Although repeated productions of targets have been associated with more stable motor articulatory movements when producing a newly learned word (Heisler et al., 2010), our findings indicate that even the attempt to retrieve word form and meaning information provides a larger boost in learning and long-term (1-week) retention.

Across the behavioral tasks, the only test that did not yield a significant learning condition effect was the form–referent link recognition task. In this task, children were required to only demonstrate fairly crude knowledge of the word form and its association with a pictured referent. In our previous word learning studies (Haebig et al., 2019; Leonard, Karpicke, et al., 2019), the TD children also had very high accuracy overall on the form–referent link recognition test, which did not differentiate according to the learning condition. However, to assess the underlying processes associated with a receptive measure of word form, Haebig et al. also included a task that measured ERPs when children heard matching and mismatching picture–label pairings. Haebig et al. identified differences in ERP components when TD children were exposed to mismatching picture–label pairings only if the labels had been taught in the RSR learning condition. In Experiment 2, we extend this work by examining the underlying neural patterns associated with processing the newly taught word meaning information.

Experiment 2

In Experiment 2, we sought to provide more in-depth information about the effects of retrieval practice on the semantic aspects of word learning. To do this, we assessed ERPs to compare the real-time neural indices underlying semantic processing for word meanings that were taught in the two learning conditions. Specifically, we assessed children's learning of what each novel item "likes" (e.g., "A /nɛp/ likes worms."). Online measures like ERPs provide complementary insight about the information processing that precedes a behavioral response.

ERP components reflect synchronized neural activity from a population of neurons that are elicited from a stimulus, such as a visual light, or reflect a cognitive process, such as access to semantic information (Luck, 2014). Although spatial resolution is poor, ERPs have high temporal resolution, providing valuable information about processing abilities (Luck, 2014). In Experiment 2, we focus on two ERP components that are associated with lexical access and integration: the N400 and the late positive component (LPC).

The N400 component is believed to index lexical-semantic access and the semantic fit of an item within a certain context (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984). Following a semantic violation, such as a semantically anomalous word within a sentence (e.g., “Grass is *purple*.”), the N400 appears as an increase in negative polarity that peaks between 200 and 600 ms. Years of research has documented the N400 in individuals ranging in age from infants to the geriatric population (Ford et al., 1996; Fox et al., 2010; Parise & Csibra, 2012). Furthermore, a variety of experiment designs have been used to elicit the N400 in young children, including auditory-only tasks with semantic congruity and incongruity and picture–label match–mismatch tasks (e.g., Gerwin et al., 2021; Haebig et al., 2018; Kuipers & Thierry, 2013; Lindau et al., 2017; Weber-Fox et al., 2013). The N400 is most prominent in the centroparietal electrodes. Furthermore, in contrast to young adults, the mean amplitude of the N400 is typically larger and peaks later in children (Hahne et al., 2004; Holcomb et al., 1992). In addition, the N400 reduces in mean amplitude with repeating incongruent stimuli (Batterink & Neville, 2011; Besson et al., 1992); however, even with repetition, the N400 remains detectible in preschool children (Haebig et al., 2018). The ability to detect the N400 with nonsequentially repeating stimuli is particularly important for learning studies because many experimental child studies must limit the number of items that are taught to allow for a reasonable degree of successful learning.

The LPC is a post-N400 ERP component that is associated with semantic processing (Juottonen et al., 1996; Sabisch et al., 2006). Although the LPC has been associated with several cognitive processes, one semantically related process includes effortful postlexical integration of verbal meaning following a semantic anomaly (Hahne et al., 2004; Van Petten & Luka, 2012; Weber-Fox et al., 2013). It is believed that the LPC may index additional processing resources that are necessary to support extended processing or semantic reanalysis (DeLong, Troyer, & Kutas, 2014; Kolk et al., 2003). Furthermore, it has been suggested that the N400 may primarily index automatic semantic processing, whereas the LPC may index more controlled/postlexical processing that is associated with conflict monitoring and semantic reanalysis and integration (DeLong, Quante, & Kutas, 2014; Kutas et al., 2011; Van Petten & Luka, 2012; although, see Kutas & Federmeier, 2011, for a more thorough discussion of automatic and controlled semantic processing associated with the N400). As with the N400, developmental changes have been noted with the LPC. Juottonen et al. (1996) suggest that the LPC may be more evident when children develop more efficient semantic memory processing abilities. Lastly, within individuals, the LPC amplitude increases with nonsequentially repeated stimuli (Renoult et al., 2010), and the LPC can be detected when TD preschool children view nonsequentially repeated stimuli (Haebig et al., 2018).

Previous studies have used matching and mismatching picture–label pairings of familiar words to elicit the N400 in young children (e.g., Haebig et al., 2018; Kuipers & Thierry, 2013). In the word learning study presented by Haebig et al. (2019), semantic processing was examined by presenting pairs of matching and mismatching picture–label pairings of the newly taught words. Haebig et al. found that, relative to matching (i.e., correct) picture–label trials, mismatching trials elicited an N400. However, the N400 was only elicited when the label had been taught in the RSR learning condition. Haebig et al. suggested that the children were more successful in learning the novel words that were taught in the RSR condition, and therefore, mismatching picture–label pairings elicited a more pronounced N400. In contrast, the novel words were less successfully learned when they were taught in the IR condition (that involved little to no contextual reinstatement), and thus, mismatching picture–label pairings elicited a nondetectable N400. In this study, we extend the previous experiment by using ERPs to examine a different aspect of lexical-semantic knowledge acquired in the word learning task, that is, word meaning information. Because we demonstrated that children were successful in learning what each item likes in both learning conditions in Haebig et al. (2019), in this study, we anticipated that trials that presented incorrect label–meaning pairings would elicit an N400 and LPC. However, we predicted that children would demonstrate more efficient semantic processing when the semantic information had been taught within the RSR learning condition relative to the IR condition. Efficiency of processing would be demonstrated by earlier-emerging ERP components (Haebig et al., 2017; Pijnacker et al., 2017). Previous work has examined differences in the timing of ERP components such as the N400 in children that differ in age (e.g., 3-year-olds vs. 4-year-olds; Silva-Pereyra et al., 2005) and children that have typical and impaired language (e.g., TD vs. DLD; Pijnacker et al., 2017). Across these studies, differences in the early and late ERP component windows have been interpreted as indexing differences in processing rate or processing efficiency (Holcomb et al., 1992; Pijnacker et al., 2017; Silva-Pereyra et al., 2005).

Method

Participants

All of the children from Experiment 1 also participated in Experiment 2 during the 1-week test visit; however, data from three of the children were not included in Experiment 2 because they did not yield enough usable EEG data to be retained in the analyses. This subset of 17 children had a mean chronological age of (5.10 years (*SD* = 0.64). The Experiment 2 study also was approved by the Institutional Review Board. All participants provided verbal assent, and a parent or legal guardian provided informed written consent.

EEG/ERP Test Phase

As previously noted, long-term retention of knowledge of the novel words and their meanings was tested a week following the completion of the word learning task. During the 1-week test session, children completed a semantic processing task during which online EEG data were collected. As a first step in examining the neural correlates of semantic processing of newly taught semantic features, we assessed only one of the semantic features that was taught: what each item likes. Data from Haebig et al. (2019) demonstrated that children were able to learn what each item likes in a similar word learning task, justifying the a priori selection of this semantic feature in this study. Furthermore, as reported in Experiment 1, recall of the two different semantic features did not statistically differ; therefore, we would not expect the ERP patterns to be different if we had used the other semantic feature. The semantic processing task followed a match–mismatch paradigm. During match trials, a picture of one of the novel words was displayed on a screen and an auditory recording of the correct label and correct semantic information was played (e.g., visual stimuli: picture of a nep; auditory stimuli: "A /nɛp/ likes worms.") via sound field at approximately 65 dB. In mismatch trials, the semantic information did not correspond to the target but instead corresponded to one of the other five targets (e.g., visual stimuli: picture of a nep; auditory stimuli: "A /nɛp/ likes rain."). At the end of each trial, children judged whether the semantic information was correct or incorrect. Each semantic target was presented 20 times, that is, 10 in the match condition and 10 in the mismatch condition, resulting in a total of 120 test trials. Match and mismatch trials were pseudorandomized so that there were never more than three match or mismatch trials in a row and no more than three consecutive IR or RSR stimuli presented. In addition, the same label never appeared sequentially more than twice, and the same label–meaning pairing never repeated in sequential trials.

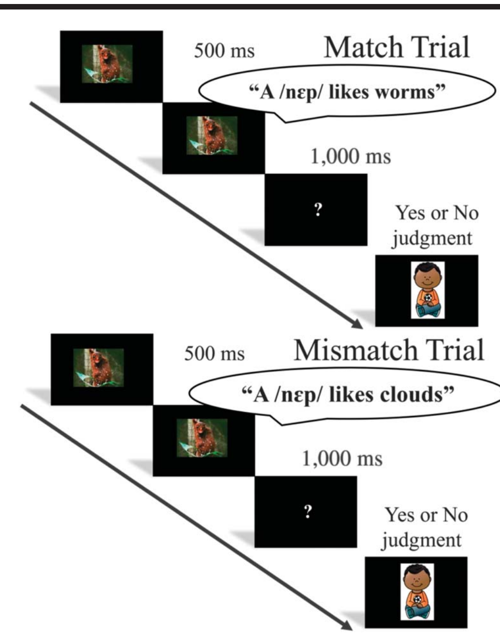
Visual task stimuli consisted of two-dimensional pictures that depicted each referent; these were the same pictures that were used during the teaching phase of the word learning study. The images were approximately 23.5 cm wide and 18 cm tall and were presented on a 47.5-cm monitor that was 164 cm in front of the child. Auditory stimuli were recorded from a young female adult with a Midwestern dialect. Event codes were inserted into the audio files to mark the acoustic onset of the semantic feature. The recorded sentences were an average of 2,107 ms in length and ranged in duration between 1,803 and 2,334 ms, and the length of the final word in the sentence (semantic feature) had a mean duration of 813 ms (range: 662–946 ms). Sound stimuli were normalized to have an amplitude of approximately 65 dB using Praat software (Boersma & Weenink, 2006).

At the beginning of the task, the examiner explained to the child that he/she was going to see pictures and hear sentences. The examiner also explained that some of the sentences were right and others were not right and that the child was going to tell her whether each one was right. Children first completed eight practice trials during which different familiar pictures (e.g., baby and fish) appeared on the screen and an auditory sentence was presented with semantically congruent/matching or incongruent/mismatching information (e.g., matching: "A baby likes a bottle."; mismatching: "A fish likes a bottle."). Children were prompted to judge whether or not the sentence made sense by verbally saying "yes/no" or "good/not good." The examiner provided feedback to the children about their judgments. Following the practice trials, the children completed 120 test trials. At the beginning of each trial, one of the six images appeared in the center of the screen at a height visual angle of 3.14° and width visual angle of 4.19°. The picture remained on the screen in silence for 500 ms before the auditory stimuli were presented via a speaker that was mounted above the display screen. Following the completion of the audio file, the picture remained on the screen for an additional 1,000 ms (the total time of picture on display was approximately 3,000 ms). Afterwards, a question mark "?" appeared in the center of the screen to prompt the child to judge whether the label and semantic information matched or mismatched. Once the child made a verbal judgment, the examiner recorded the child's response by pressing one of two buttons on a response pad. Once the child's response was recorded, the question mark was replaced by a picture of a smiling child who was sitting until the examiner advanced the task to the next trial. See Figure 3 for a depiction of the trial progression. Throughout the task, 14 breaks were provided, that is, one every 8–12 trials. Half of the breaks consisted of short video clips (4–5 s) of nature scenes with music or other engaging stimuli; the other half consisted of an engaging picture. During a break with a picture, the child was allowed to place a sticker on a visual schedule to chart progress through the task.

EEG Recordings

While the children completed the task, EEG data were also being recorded. We measured electrical activity at the scalp using a 32-electrode array, secured in an elastic cap (ActiveTwo head cap, Cortech Solutions). The electrodes were positioned over homologous hemisphere

Figure 3. Event-related brain potential task trial.



locations according to the International 10–10 system (Jurcak et al., 2007). Locations were as follows: lateral sites F7/F8, FC5/FC6, T7/T8, CP5/CP6, and P7/P8; midlateral sites FP1/FP2, AF3/AF4, FC1/FC2, F3/F4, CP1/CP2, P3/P4, PO3/PO4, and O1/O2; and midline sites FZ, CZ, PZ, and OZ. The electrical recordings were referenced to the average of the electrodes on the left and right mastoids during the data-processing procedures. Electrodes placed over the left and right outer canthi recorded horizontal eye movements. Vertical eye movement was monitored through recordings from electrodes placed over the left inferior and superior orbital ridge. The continuous electroencephalogram data were recorded using the Biosemi ActiveTwo system.

ERP Measures

The EEG data were processed using EEGLAB and ERPLAB (Lopez-Calderon & Luck, 2010), which are MATLAB toolboxes (MathWorks). The EEG signals were downsampled at a rate of 256 Hz and were bandpass filtered from 0.1 to 30 Hz with a 12-dB roll-off to remove high-frequency noise and to minimize offsets and drift. Eye artifact was removed through independent component analysis (ICA; EEGLAB). ICA identifies independent sources of EEG signals and yields components that represent patterns from the EEG signal. Components that represent artifact, such as blinks, horizontal eye movements, and voltage drifts, were identified by two independent trained research assistants, and discrepancies were resolved by a third research assistant. The data were epoched from 200 ms prior to the onset of the semantic feature to 2,000 ms poststimulus to allow for averaging and ERP component measures. Epochs were baseline corrected from −200 ms to the onset of the semantic feature (0 ms). The EEG channels underwent automatic voltage-dependent thresholds to remove any trials that still contained artifact. Each participant was required to contribute at least 18 artifact-free trials within each condition, similar to other preschool child ERP studies (e.g., Gerwin et al., 2021; Haebig et al., 2019; Pijnacker et al., 2017; Silva-Pereyra et al., 2005). Within the IR condition, the average number of artifact-free trials within the match condition was 24.35 and that within the mismatch condition was 24.94. Within the RSR condition, the average number of artifact-free trials within the match condition was 24.24 and that within the mismatch condition was 23.82. Finally, the EEG epochs from artifact-free trials were averaged within task conditions for each individual, and analyses were conducted to examine the N400 and LPC ERP components.

For each ERP component, we determined the region of interest (ROI) by consulting the literature and examining the grand average waveforms. We measured the N400 from centroparietal electrodes (CP1, CP2, CZ, and PZ) and we measured the LPC from parietal-occipital electrodes (PO3, PO4, O1, OZ, and O2). For both the N400 and the LPC, we captured the temporal aspects by selecting two windows for each component: one representing an early window and another representing a late window (Haebig et al., 2019; Pijnacker et al., 2017). The early N400 temporal window was set to 300–500 ms, and the late N400 window was set to 500–700 ms, aligning with the N400 child literature (Haebig et al., 2018, 2019; Juottonen et al., 1996; Pijnacker et al., 2017; Sabisch et al., 2006). The early and late LPC windows were 1,000–1,200 ms and 1,200–1,500 ms, respectively. Our LPC temporal windows also aligned with the LPC child literature (Haebig et al., 2018; Juottonen et al., 1996; Sabisch et al., 2006; Weber-Fox et al., 2013). In addition to referring to the literature and the grand average waveforms to select our temporal windows, we examined each child's waveforms to ensure that the windows captured the components of interest.

Analysis Procedures

We used mixed-effect models to examine whether there were differences in behavioral judgments and mean amplitude for ERP components according to learning condition. As before, random intercepts were set at the child level, and repeated measures (e.g., match/mismatch, ROI electrodes) were nested within the children. When analyzing the behavioral judgments, we controlled for response bias by calculating A' scores and using them as the dependent variable (Grier, 1971; Rice et al., 1999). Briefly, A' scores serve as a measure of the proportion of correct responses in a two-alternative forced-choice task. The A' value consists of scores from a control condition and an experimental condition (e.g., matching sentences and mismatching sentences). The formula is as follows: $A' = 0.5 + (y - x)(1 + y - x) / 4y(1 - x)$, where y represents correct identifications (hits) and x represents incorrect identifications (false alarms; Linebarger et al., 1983). An A' value of 1.00 represents perfect discrimination of correct and incorrect sentences. An A' value of 0.50 indicates chance performance, for example, a "yes" response to 50% of the correct sentences and to 50% of the semantically anomalous sentences.

Results

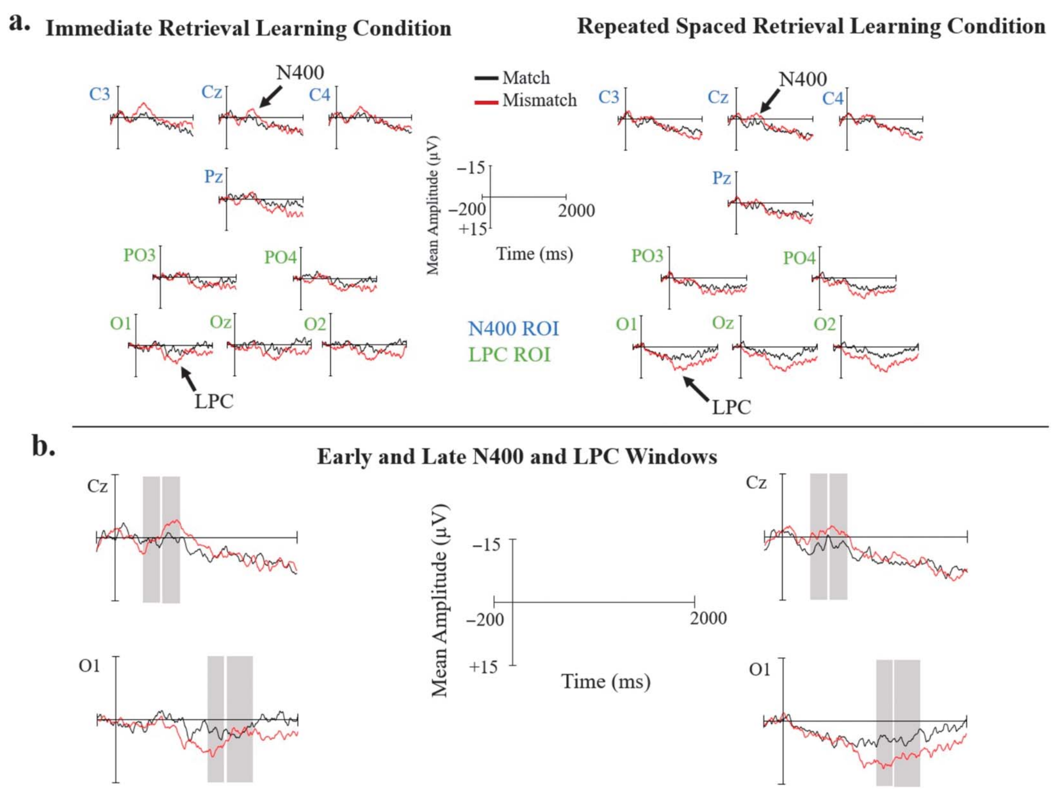
Behavioral Performance

We first assessed children's behavioral judgments made during the ERP task. Descriptively, the children had a mean A′ score of 0.81 (SD = 0.17) for the IR condition and a mean A′ score of 0.84 (SD = 0.17) for the RSR condition. Judgment accuracy did not statistically differ according to learning condition (*Estimate* = −0.45, SE = 0.25, t = −1.77, p = .08).

N400 Mean Amplitude

Next, we examined the ERPs to better understand the neural indices associated with processing the semantic information that was taught using the IR and RSR schedules. Figure 4 depicts the waveforms for the IR and RSR conditions. When comparing match and mismatch trials within the IR learning condition, we found that there was no significant difference in mean amplitude between the match and mismatch trials during the early N400 window (*Estimate* = 0.26, SE = 0.75, t = 0.34, p = .73). However,

Figure 4. Waveforms within the N400 region of interest (ROI) and late positive component (LPC) ROI. The waveforms in panel "a" of the figure include all electrode cites that comprised each ROI. Panel "b" of the figure provides an enlarged image of one electrode cite for the N400 and the LPC. Within each, the gray rectangles depict the early and late windows that were examined to better capture the temporal differences of the event-related brain potential components.



during the late window, an N400 was elicited by mismatching label–meaning pairings (i.e., mismatch trials; *Estimate* = −2.95, *SE* = 0.88, *t* = −3.34, *p* = .001).

In contrast to the IR condition, when comparing the mean amplitude values between the match and mismatch trials within the RSR condition, we found that there was a significant difference between match and mismatch trials during both the early and late N400 windows (*Estimate* = −2.25, *SE* = 0.74, *t* = −3.04, *p* = .003; *Estimate* = −2.37, *SE* = 0.81, *t* = −2.94, *p* = .004, respectively). The mismatch trials elicited a robust N400 during both the early and late windows.

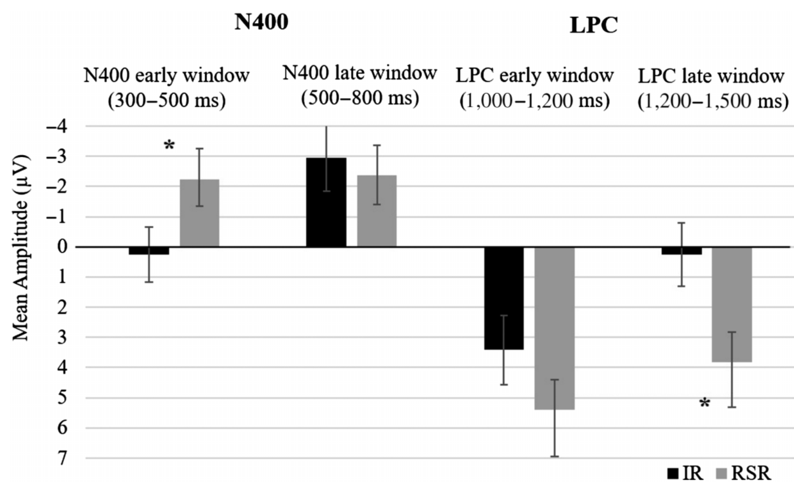
To directly compare the IR and RSR conditions, we examined the difference waves. Within the early N400 window, the difference between match and mismatch trials was significantly larger when the word meanings were taught within the RSR learning condition (*Estimate* = −2.50, *SE* = 1.10, *t* = −2.28, *p* = .024). However, because the N400 elicited from mismatch trials was robust within the late N400 window across both learning conditions, there was no significant difference in the difference-wave mean amplitudes between the learning conditions in the later window (*Estimate* = 0.58, *SE* = 1.06, *t* = 0.55, *p* = .584). Figure 5 depicts the mean amplitude from the difference waves for each learning condition and ERP window.

LPC Mean Amplitude

We also compared the LPC as an additional measure of semantic processing. When examining the difference between match and mismatch trials for word labels and meanings that were taught in the IR condition, we found a significant difference during the early LPC window (*Estimate* = 3.42, *SE* = 0.97, *t* = 3.52, *p* < .001). Mismatching label–meaning pairs were associated with a positive shift in polarity following the N400 component. However, there was no significant difference in mean amplitude between match and mismatch trials during the late LPC window (*Estimate* = 0.25, *SE* = 0.92, *t* = 0.28, *p* = .783). Next, we examined the difference between match and mismatch trials for items that were taught within the RSR learning condition. RSR mismatching label–meaning pairs elicited a significant positive polarity shift following the N400 component during both the early and late LPC windows (*Estimate* = 5.41, *SE* = 1.22, *t* = 4.45, *p* < .001; *Estimate* = 3.82, *SE* = 1.22, *t* = 3.12, *p* = .002, respectively).

Lastly, we directly compared the LPC associated with the difference between match and mismatch trials for each learning condition. There was no significant difference in the mean amplitude of the LPC during the early LPC window (*Estimate* = 1.99, *SE* = 1.41, *t* = 1.42, *p* = .159). However,

Figure 5. Mean amplitude difference between match and mismatch trials within learning condition. Error bars represent standard error. IR = immediate retrieval; LPC = late positive component; RSR = repeated spaced retrieval.



as depicted in Figure 5, there was a significant difference during the late LPC window (*Estimate* = 3.57, *SE* = 1.31, $t = 2.72$, $p = .007$), indexing a continued positive polarity elicited by mismatching label–meaning pairings for items taught within the RSR condition.[1]

Discussion

The ERP data provide further evidence that RSR promotes word learning. In particular, our ERP data highlight deeper aspects of meaning and semantic processing compared to the relatively more superficial label–referent word learning associations that we documented in a previous study (Haebig et al., 2019). To our knowledge, our ERP data provide the first demonstration of examining newly associated word meanings (i.e., labels and semantic information) by testing these learned associations using the N400 and LPC in preschool children. In the previous literature, the N400 and LPC have been used to examine processing of already-known semantic associates (e.g., shoe–sock, "eat a blanket"; Pijnacker et al., 2017; Sirri & Rämä, 2015; Torkildsen et al., 2007). In order to elicit an N400 and LPC in this study, the children needed to first learn that one semantic feature was linked to a specific label. Once this connection was learned, a semantically incongruent label–meaning pairing elicited semantic reanalysis, resulting in an N400 and LPC. The differences in timing of the ERP components indicate that the children attained different levels of learning when novel word labels and meanings were taught using the IR learning schedule versus the RSR learning schedule.

As with the behavioral data, the ERP data indicate that the children learned the word meaning information better when it was taught in the RSR learning condition. Mismatching label–meaning pairings that were taught in the RSR learning condition resulted in an earlier onset of the N400 (with significant differences between match and mismatch trials for both the early and late N400 windows). Similarly, mismatch trials corresponding to the RSR learning condition elicited an LPC with a longer duration, relative to the LPC associated with the IR learning condition. These findings indicate that the novel word labels were more effective in priming the corresponding semantic information if the words had been taught in the RSR learning condition. Our ERP findings aligned with our predictions. We interpret the earlier occurring N400 associated with the RSR learning condition as an indication that the children were more quickly able to detect the error in the mismatching label–meaning pairings, demonstrating more efficient processing.

Additionally, we interpret the LPC with a longer duration associated with items taught in the RSR learning condition as an indication that the children more strongly associated the label–meaning pairings, given previous findings that suggest the LPC is more evident when children are more efficient in processing such semantic information (Juottonen et al., 1996). A recent study that applied RSR in an adjective word learning task has also reported differences in the LPC (Gerwin et al., 2021). Gerwin et al. found that mismatch trials elicited a more mature LPC profile for adjectives that were taught in an RSR learning condition relative to a repeated study learning condition. They also suggested that the preschool children likely formed a richer representation for the words (adjectives) that were taught in the RSR condition and were therefore more efficient in the semantic reanalysis associated with the LPC.

[1]These patterns of findings hold even if the N400 and LPC data are analyzed differently, with models that contain data from both the early and late windows. As would be expected, there were significant interactions between trial type (match–mismatch) and window (early–late) for the IR learning condition. Additionally, models testing the difference-wave mean amplitudes also yielded significant interactions between learning condition and window.

It is notable that the ERP data differentiated the learning conditions while the child judgments did not. This highlights the benefit of incorporating ERPs because they offered information that was more sensitive and revealed learning differences during a receptive task that the behavioral judgments were not able to provide (see also Malins et al., 2013; Shafer et al., 2005, 2011). These findings also complement the ERP findings presented in Haebig et al. (2019), wherein a different sample of TD preschool children demonstrated a more robust N400 when presented with mismatching label–referent mappings when the novel word labels had been taught in the RSR relative to the IR learning condition. Therefore, across the two studies, we have demonstrated that RSR promotes multiple aspects of word learning that can be detected when examining the neural indices of lexical-semantic processing.

General Discussion

The behavioral and neural data presented in this study provide complementary support to the assertion that RSR promotes multiple aspects of word learning in preschool children. Our results indicate that RSR learning schedules promote the learning of both word form information and word meaning information. In this study, we extended the word learning results presented in Haebig et al. (2019) and shifted our focus to word meaning. This was done by doubling the semantic information that was taught about each novel word (e.g., what each item likes and what happens to each item when it grows). In addition, we carefully examined word meaning by measuring ERP components to assess the neural underpinnings of processing deeper semantic information than mere label–referent mappings. In this unique contribution to the literature, we first demonstrated that preschool children develop semantically rich representations for novel words and that evidence of lexical-semantic learning can be demonstrated at the behavioral and neural levels. The second contribution that this study provides is that rich lexical-semantic learning is promoted when learning experiences incorporate RSR over and above learning experiences that incorporate repeated IR.

Word Form

Our word label recall data replicate the Haebig et al. (2019) finding that RSR promotes the learning of word form in preschool children relative to IR. In Experiment 1, we found that, on average, children accurately recalled 1.5 more novel labels for words that were taught in the RSR learning condition. We interpret this large effect in learning condition as children develop enhanced representations of the word form, which was facilitated by prompts to retrieve the novel words in slightly changing contexts, that is, by retrieving words following an exposure to a different novel word. The contextual reinstatement required during the spaced retrieval is believed to strengthen recall not only by developing the enhanced representation during each spaced retrieval attempt but also by restricting the memory search set during retrieval (Karpicke et al., 2014). Furthermore, as previously noted, the RSR benefit at the 5-min and 1-week tests was impressive given that the opposite pattern of recall accuracy was observed during the learning phase. During learning, children were more accurate in recalling the word forms for items taught in the IR condition relative to the RSR condition. Therefore, it is believed that the act of attempting to retrieve the word following a delay from the study trial (RSR) promotes learning.

The current findings are also notable because word form learning does not appear to have been negatively impacted by the increase in semantic information that was taught. Descriptively, the children in this study demonstrated similar word form recall accuracy relative to the TD children in the Haebig et al. (2019) study, wherein only one semantic feature was taught. However, it is important to note that the frequency of word form exposure was not equivalent across the two studies; this study design included additional exposure to the word form (32 total exposures) relative to the Haebig et al. (2019) study (24 total exposures to word form).

Additionally, although the RSR learning condition effect was strong in the word recall task, this was not the case for the form–referent link recognition test. Given that the receptive pointing task only required children to demonstrate a minimum of superficial knowledge of word form information, it is not surprising that the children had near-ceiling performance for the words that were taught in both learning conditions. The more stringent and sensitive test incorporated in this study was the recall test. Our word form recall test findings not only replicate the behavioral findings presented in Haebig et al. (2019) but also align with their ERP evidence of enhanced learning of word form information when words were taught in the RSR learning condition. Lastly, it is notable that the children retained their knowledge of word form at both the 5-min and 1-week test points.

Word Meaning

Despite teaching double the amount of word meaning information in this study, the children continued to demonstrate high levels of successful learning of word meaning. The children also demonstrated improved learning of semantic information when the word meanings were taught within the RSR learning schedule relative to the IR learning schedule. On average, the children recalled almost one additional word meaning that corresponded to what each item likes (0.75 points higher word meaning scores) and word meanings corresponding to what happens when each item grows (0.65 points higher) if taught within the RSR learning condition. This pattern aligns with the word meaning recall findings presented in the studies of Haebig et al. (2019) and Leonard, Karpicke, et al. (2019) when word meaning only consisted of information about what each item likes. Importantly, the pattern of strong recall of word meaning information and enhanced recall for RSR items was apparent at the 5-min and 1-week test points.

In contrast to the expressive measure of word meaning, during the receptive match–mismatch label–meaning pairing judgment task, children did not demonstrate significantly better behavioral accuracy when judging pairings for items taught within one learning condition or the other. As with the receptive form–referent link recognition task, our receptive measure of word meaning was not particularly taxing. Given this, we benefitted from including an online measure of processing during this task. The N400 and LPC ERP components provided meaningful insight into more subtle differences in semantic learning attained through the two learning conditions. Although mismatching label–meaning pairings elicited the N400 and LPC for both learning conditions, we identified differences in the temporal aspects of each ERP component. The onset of the N400 occurred earlier, within the early N400 window, if the word meaning had been taught in the RSR learning condition. This indicates that the children may have been more efficient in processing and identifying the label–meaning mismatch. Furthermore, the mismatch trials resulted in a more persistent LPC, significant for both the early and late LPC windows when meaning had been taught in the RSR learning condition, indexing that the mismatching label–meaning pairings caused greater disruptions in processing when feature–label pairings were more strongly associated to one another.

Although we are not aware of a previous study using ERPs to measure the strength of associative learning in preschool children, other studies have used eye-gaze measures to assess online processing of newly associated familiar words in preschool children. For instance, Borovsky et al. (2014) briefly taught associations between familiar words to preschool and school-age children (e.g., "The **mouse** is eating an **apple**"; Borovsky et al., 2014). Following the learning phase, Borovsky et al. used online eye tracking to assess children's anticipatory looks between the agent (e.g., mouse) and the target (e.g., apple). Borovsky et al. found that, following this brief teaching phase, the preschool children were less efficient in associating the two familiar words, resulting in a lack of anticipatory looks to the target, despite being able to associate the words in an offline measure. School-age children, however, demonstrated anticipatory looks. Borovsky et al. suggested that the preschool children may have been less efficient in reactivating the newly taught associative pairs. This pattern of results and interpretation aligns with our ERP findings. In this study, it is likely that the preschool children developed a stronger associative pairing between the novel word form and the word meaning when they had been taught in the RSR learning condition relative to the IR learning condition. As a result, the picture of the referent and presentation of the novel label served as stronger primes to the word meaning for RSR meanings. This led to more efficient processing, with RSR mismatching trials eliciting an earlier N400 and a more persistent LPC in our preschool sample. In contrast, the priming between the label and meaning may have been weaker for IR items, and mismatched pairings between the two were less efficiently processed.

Study Limitations

Although this study contributes novel information about word learning to the retrieval practice literature, we must acknowledge some limitations. First, this study only examined novel noun learning. Although this reflects the broader word learning literature, future work should have a more concerted effort to examining other word classes. Two recent studies have reported adjective word learning while employing retrieval practice techniques (Gerwin et al., 2021; Leonard, Deevy, et al., 2019). Additional word classes will provide insight into children's ability to learn new words and to generalize their knowledge in different ways (e.g., by flexibly changing features of tense and agreement for verbs). Second, this study would have benefited from a larger sample size and different age groups to develop a more in-depth understanding of word learning and age-related differences in the neural underpinnings of word processing. Third, the ERP task design included differences in the number of times stimuli were repeated. Specifically, during the match trials, each word and its corresponding meaning (e.g., "A /nɛp/ likes worms.") repeated 10 times throughout the task. In contrast, mismatching stimuli (e.g., "A /nɛp/ likes clouds.") were presented only twice, and different semantic information was paired with each label across the mismatch trials. Therefore, although each semantic feature (e.g., worms and clouds) appeared 10 times in a match condition and 10 times in a mismatch condition, the specific sentence frame in the mismatch condition was only presented twice. Previous studies have demonstrated that stimuli repetition can influence ERP components (Besson & Kutas, 1993; Besson et al., 1992). It is possible that the frequent repetitions in the match condition may have strengthened the matching label–meaning associations and, as a result, enhanced the N400 and LPC for the mismatch trials. Importantly though, the numbers of repetitions were exactly the same across both learning conditions. Therefore, any influence the repetitions may have had would have equally influenced both learning conditions.

Clinical Implications and Future Studies

This study provides support to the use of RSR to promote learning in children (Gordon, 2020; Karpicke et al., 2016; Leonard, Deevy, et al., 2019). The emphasis on word meaning in this study is particularly relevant because depth of word knowledge is predictive of important literacy and academic outcomes in children (Ouellette, 2006). In addition, previous word learning studies of children from at-risk populations have found that emphasizing word meaning through elaboration during teaching or providing explicit definitions or linguistic contexts to support word meaning supports word learning in children (Justice et al., 2005; Ralph et al., 2020; Wilkinson & Houston-Price, 2013). Although these studies have incorporated learning strategies such as providing frequent repetitions, elaborations, and explicit definitions, they have not incorporated prompts for retrieval practice. Applying RSR to such word learning experiences would appear to be beneficial.

Although this study provides meaningful evidence supporting the utility of RSR, future work could further advance our understanding of RSR. Such work also could facilitate the development of learning tasks that optimize learning while also fitting into realistic educational and therapeutic curricula. For instance, although RSR was found to be superior to IR in this study, the TD children only learned a portion of the novel word forms that were taught. It is possible that future studies may find a more effective learning schedule that incorporates RSR to facilitate word learning to a larger extent. Additionally, although recent work has begun to explore the effects of RSR on word learning in preschool children with DLD (Haebig et al., 2019; Leonard et al., 2020; Leonard, Deevy, et al., 2019; Leonard, Karpicke, et al., 2019), other populations with known language learning difficulties would benefit from similar research (e.g., children with dyslexia, autism spectrum disorder, and Down syndrome). Lastly, future studies should examine whether RSR promotes word learning for other word classes and if word learning is sufficiently enhanced to promote different types of generalization.

In conclusion, we have demonstrated that at both the behavioral and neural levels, RSR enhances multiple aspects of word learning (i.e., word form and meaning). These findings highlight the importance of spaced retrieval relative to IR. Given the importance of word knowledge for child outcomes, it will be important to extend this research to different populations and tasks that more closely match educational materials and activities.

Acknowledgments

This research was funded by a grant from the National Institute on Deafness and Other Communication Disorder (R01 DC014708; PI: L. B. Leonard). Eileen Haebig was supported by a postdoctoral fellowship Training Grant T32 DC00030 (PI: L. B. Leonard). The authors would like to thank the families who participated in this study. Also, they thank Evan Usler, Ranjini Mohan, Gina Catania, Connor Slavich, and Katie Gerwin for their help with data collection and processing.

References

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping. *Journal of Speech, Language, and Hearing Research, 47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*. http://www.asha.org/policy

Anns, S., Gaigg, S. B., Hampton, J. A., Bowler, D. M., & Boucher, J. (2020). Declarative memory and structural language impairment in autistic children and adolescents. *Autism Research, 13*(11), 1947–1958. https://doi.org/10.1002/aur.2282

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience, 23*(11), 3181–3196. https://doi.org/10.1162/jocn_a_00013

Besson, M., & Kutas, M. (1993). The many facets of repetition: A cued-recall and event-related potential analysis of repeated words in same versus different sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1115–1133. https://doi.org/10.1037/0278-7393.19.5.1115

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4*(2), 132–149. https://doi.org/10.1162/jocn.1992.4.2.132

Boersma, M., & Weenink, D. (2006). *Praat: Doing phonetics by computer*. University of Amsterdam.

Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language, 73,* 1–14. https://doi.org/10.1016/j.jml.2014.02.001

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning Memory and Cognition, 34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61,* 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass, 8*(12), 631–645. https://doi.org/10.1111/lnc3.12093

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. NCS Pearson.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence*. Pro-Ed.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J. S., & Bates, E. (2006). *The MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Brookes.

Ford, J. M., Woodward, S. H., Sullivan, E. V., Isaacks, B. G., Tinklenberg, J. R., Yesavage, J. A., & Roth, W. T. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography and Clinical Neurophysiology, 99,* 235–246. https://doi.org/10.1016/S0921-884X(96)95049-3

Fox, A. M., Anderson, M., Reid, C., Smith, T., & Bishop, D. V. M. (2010). Maturation of auditory temporal integration and inhibition assessed with event-related potentials (ERPs). *BMC Neuroscience, 11,* Article 49. https://doi.org/10.1186/1471-2202-11-49

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology, 60*(7), 991–1004. https://doi.org/10.1080/17470210600823595

Gerwin, K. L., Leonard, L. B., Schumaker, J., Deevy, P., Haebig, E., & Weber, C. (2021). Novel adjective processing in preschool children: Evidence from event-related brain potentials. *Journal of Speech, Language, and Hearing Research, 64*(2), 542–560. https://doi.org/10.1044/2020_JSLHR-20-00332

Gordon, K. R. (2020). The advantages of retrieval-based and spaced practice: Implications for word learning in clinical and educational contexts. *Language, Speech, and Hearing Services in Schools, 51*(4), 955–965. https://doi.org/10.1044/2020_LSHSS-19i-00001

Gordon, K. R., McGregor, K. K., & Arbisi-Kelm, T. (2020). Optimising word learning in post-secondary students with developmental language disorder: The roles of retrieval difficulty and retrieval success during training. *International Journal of Speech-Language Pathology,* 1–14. https://doi.org/10.1080/17549507.2020.1812719

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*(6), 424–429. https://doi.org/10.1037/h0031246

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with development Language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research, 62*(4), 944–964. https://doi.org/10.1044/2018_JSLHR-L-18-0071

Haebig, E., Leonard, L. B., Usler, E., Deevy, P., & Weber, C. (2018). An initial investigation of the neural correlates of word processing in preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research, 61*(3), 729–739. https://doi.org/10.1044/2017_JSLHR-L-17-0249

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., & Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *Journal of Neurodevelopmental Disorders, 9,* 1–21. https://doi.org/10.1186/s11689-017-9201-1

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience, 16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Developmental Science, 13*(5), 722–730. https://doi.org/10.1111/j.1467-7687.2009.00930.x

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8*(2–3), 203–241. https://doi.org/10.1080/87565649209540525

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research, 4*(2), 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage, 34,* 1600–1611. https://doi.org/10.1016/j.neuroimage.2006.09.024

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks. *Language Speech and Hearing Services in Schools, 36*(1), 17–32. https://doi.org/10.1044/0161-1461(2005/003)

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology, 7,* 1–9.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 60). Elsevier Academic Press.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kaufmann, A., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children*. AGS.

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85,* 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex, 49,* 2853–2860. https://doi.org/10.1016/j.cortex.2013.01.012

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163. https://doi.org/10.1038/307161a0

Lawrence, J. F., Hagen, A. M., Hwang, J. K., Lin, G., & Lervåg, A. (2019). Academic vocabulary and reading comprehension: Exploring the relationships across measures of vocabulary knowledge. *Reading and Writing, 32,* 285–306. https://doi.org/10.1007/s11145-018-9865-2

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S. L., & Kueser, J. B. (2020). After initial retrieval practice, more retrieval produces better retention than more study in word learning of children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(8), 2763–2776. https://doi.org/10.1044/2020_JSLHR-20-00105

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S., Weber, C., Kueser, J. B., & Haebig, E. (2019). Adjective learning in young typically developing children and children with developmental language disorder: A retrieval-based approach. *Journal of Speech, Language, and Hearing Research, 62*(12), 4433–4449. https://doi.org/10.1044/2019_JSLHR-L-19-0221

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J. B., & Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder I: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research, 62*(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Lindau, T. A., Giacheti, C. M., da Silva, I. B., & de Souza, D. d. G. (2017). Semantic processing in children 0 to 6 years of age: An N400 analysis. *Revista CEFAC: Atualizacao Cientifica Em Fonoaudiologia e Educacao, 19*(5), 690–701. https://doi.org/10.1590/1982-0216201719513517

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361–392. https://doi.org/10.1016/0010-0277(83)90015-X

Lopez-Calderon, J., & Luck, S. J. (2010). *ERPLAB Toolbox (1.1.0)*. http://erpinfo.org/erplab

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Malins, J. G., Desroches, A. S., Robertson, E. K., Newman, R. L., Archibald, L. M. D., & Joanisse, M. F. (2013). ERPs reveal the temporal dynamics of auditory word recognition in specific language impairment. *Developmental Cognitive Neuroscience, 5,* 134–148. https://doi.org/10.1016/j.dcn.2013.02.005

Marinellie, S. A., & Johnson, C. J. (2002). Definitional skill in school-age children with specific language impairment. *Journal of Communication Disorders, 35,* 241–259. https://doi.org/10.1016/s0021-9924(02)00056-4

McGregor, K. K. (2014). What a difference a day makes: Change in memory for newly learned word forms over 24 hours. *Journal of Speech, Language, and Hearing Research, 57*(4), 1842–1850. https://doi.org/10.1044/2014_JSLHR-L-13-0273

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders, 42,* 35–47. https://doi.org/10.1007/s10803-011-1210-4

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45*(5), 998–1014. https://doi.org/10.1044/1092-4388(2002/081)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48,* 307–319. https://doi.org/10.1111/1460-6984.12008

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37,* 1077–1087. https://doi.org/10.3758/MC.37.8.1077

Nation, K. (2014). Lexical learning and lexical processing in children with developmental language impairments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369,* 1–10. https://doi.org/10.1098/rstb.2012.0387

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*(7), 189–208. https://doi.org/10.1007/s11145-008-9159-1

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science, 23,* 728–733. https://doi.org/10.1177/0956797612438734

Pijnacker, J., Davids, N., Van Weerdenburg, M., Verhoeven, L., Knoors, H., & Van Alphen, P. (2017). Semantic processing of sentences in preschoolers with specific language impairment: Evidence from the N400 effect. *Journal of Speech, Language, and Hearing Research, 60*(3), 627–639. https://doi.org/10.1044/2016_JSLHR-L-15-0299

Ralph, Y. K., Schneider, J. M., Abel, A. D., & Maguire, M. J. (2020). Using the N400 event-related potential to study word learning from context in children from low- and higher-socioeconomic status homes. *Journal of Experimental Child Psychology, 191,* 104758. https://doi.org/10.1016/j.jecp.2019.104758

Renoult, L., Brodeur, M. B., & Debruille, J. B. (2010). Semantic processing of highly repeated concepts presented in single-word trials: Electrophysiological and behavioral correlates. *Biological Psychology, 84,* 206–220. https://doi.org/10.1016/j.biopsycho.2010.01.014

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judgments of an extended optional infinitive grammar. *Journal of Speech, Language, and Hearing Research, 42*(4), 943–961. https://doi.org/10.1044/jslhr.4204.943

Roark, B., & Demuth, K. (2000). Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell, S. A. Fish, & T. Keith-Luca (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 597–608). Cascadilla Press.

Sabisch, B., Hahne, A., Glass, E., von Suchodoletz, W., & Friederici, A. D. (2006). Lexical-semantic processes in children with specific language impairment. *NeuroReport, 17*(14), 1511–1514. https://doi.org/10.1097/01.wnr.0000236850.61306.91

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience, 17*(7), 1168–1180. https://doi.org/10.1162/0898929054475217

Shafer, V. L., Schwartz, R. G., & Martin, B. (2011). Evidence of deficient central speech processing in children with specific language impairment: The T-complex. *Clinical Neurophysiology, 122,* 1137–1155. https://doi.org/10.1016/j.clinph.2010.10.046

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cognitive Brain Research, 23,* 247–258. https://doi.org/10.1016/j.cogbrainres.2004.10.015

Sirri, L., & Rämä, P. (2015). Cognitive and neural mechanisms underlying semantic priming during language acquisition. *Journal of Neurolinguistics, 35,* 1–12. https://doi.org/10.1016/j.jneuroling.2015.01.003

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods, 42,* 497–506. https://doi.org/10.3758/BRM.42.2.497

Torkildsen, J. V. K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics, 20,* 332–351. https://doi.org/10.1016/j.jneuroling.2007.02.003

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83,* 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Weber-Fox, C., Hampton Wray, A., & Arnold, H. (2013). Early childhood stuttering and electrophysiological indices of language processing. *Journal of Fluency Disorders, 38,* 206–221. https://doi.org/10.1016/j.jfludis.2013.01.001

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1036–1046. https://doi.org/10.1037/xlm0000379

Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess …: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*(3), 591–613. https://doi.org/10.1017/S0142716411000889

Williams, K. T. (2007). *Expressive vocabulary test* (2nd ed.). Pearson Assessments.