Study Protocol with Statistical Analysis Plan

Study Title: Retrieval Practice for Word Learning for Deaf and Hard of Hearing Children

NCT Number: NCT05512000

Document date: 02/07/2025

Increasing Word Learning Efficiency in Children who are Deaf and Hard of Hearing through Retrieval Practice

PI: Jena McDaniel
Jena.mcdaniel@vumc.org
Department of Hearing and Speech Sciences
1215 21st Avenue South, MCE South Tower Suite 8310, Nashville TN 37232

Table of Contents:

1.0 Background

Many children who are deaf and hard of hearing (DHH) are not achieving the language skills of their peers with typical hearing despite access to audiological, educational, and speech-language intervention (Antia et al., 2020; Boons et al., 2013; Geers et al., 2017; Geers & Sedey, 2011; Kyle & Harris, 2011; Lund, 2016; Nittrouer et al., 2019; Qi & Mitchell, 2012; Ruben, 2018; Werfel, 2017). Our **long-term goal is to increase the effectiveness and efficiency of language learning for children who are DHH through individualized interventions,** thereby improving their long-term educational and vocational success. Despite persistent deficits in language skills compared with hearing peers and the obligation to provide evidence-based services, few experimental language intervention studies for children who are DHH have been conducted (Brennan-Jones et al., 2014; Kaipa & Danser, 2016; Luckner & Cooke, 2010; Roberts, 2019). This paucity of evidence hampers practitioners' implementation of individualized, evidence-based interventions. As a result, students are at risk for receiving ineffective or even detrimental interventions.

Identifying active ingredients of language interventions for children who are DHH would build our theoretical understanding about what facilitates and impedes their language development. Once identified, active ingredients can be enhanced (e.g., modify intensity) through empirical investigations that consider individual differences. For children who are DHH, **retrieval practice** (testing effect) presents as a potential active ingredient for language intervention. Retrieving previously presented information (e.g., answering questions or free recall) is **hypothesized to enhance encoding and retention** because it entails the desirable level of effort (desirable difficulties hypothesis; Bjork & Bjork, 2011; Bjork, 1994), builds semantic elaborations and associations between words (elaborative retrieval account; Carpenter, 2009; Carpenter & Yeung, 2017), and/or updates representations to reflect multiple contexts (episodic context account; Karpicke et al., 2014). Replicated findings in children and adults with typical development (Fritz et al., 2007; Goossens et al., 2014; Jones et al., 2016; Karpicke et al., 2014, 2016; Lipowski et al., 2014; Marsh et al., 2012) and emerging results for individuals with disabilities (Coyne et al., 2015; Haebig et al., 2019; Knouse et al., 2016; Leonard, Deevy, et al., 2019; Leonard, Karpicke, et al., 2019; McGregor et al., 2017) have showed positive effects of retrieval practice for academically relevant tasks (e.g., word learning, textbook content, and recalling word lists or pairs) over short retention periods (e.g., one session or one week). For children who are DHH, the few experimental studies on language intervention have often used retrieval practice to some degree, though typically without isolating retrieval practice's effect and under another name (e.g., choral responses, cloze procedures, and receptive and expressive practice; Bobzien et al., 2015; Lund & Douglas, 2016; McDaniel et al., 2019; McDaniel et al., 2018; Reimer, 2019).

Antia, S. D., Lederberg, A. R., Easterbrooks, S., Schick, B., Branum-Martin, L., Connor, C. M., & Webb, M.-Y. (2020). Language and reading progress of young deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education* [advanced online release].

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 2,* 59-68.

Bjork, R. A. (1994). Memory and metamemory considerations in the. *Metacognition: Knowing About Knowing, 185*.

Bobzien, J. L., Richels, C., Schwartz, K., Raver, S. A., Hester, P., & Morin, L. (2015). Using repeated reading and explicit instruction to teach vocabulary to preschoolers with hearing loss. *Infants & Young Children, 28*(3), 262-280.

Boons, T., De Raeve, L., Langereis, M., Peeraer, L., Wouters, J., & Van Wieringen, A. (2013). Expressive vocabulary, morphology, syntax and narrative skills in profoundly deaf children after early cochlear implantation. *Research in Developmental Disabilities, 34*(6), 2008-2022.

Brennan-Jones, C. G., White, J., Rush, R. W., & Law, J. (2014). Auditory-verbal therapy for promoting spoken language development in children with permanent hearing impairments. *Cochrane Database of Systematic Reviews*(3).

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language, 92*, 128-141.

Coyne, J. H., Borg, J. M., DeLuca, J., Glass, L., & Sumowski, J. F. (2015). Retrieval practice as an effective memory strategy in children and adolescents with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 96*(4), 742-745.

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

Geers, A. E., Mitchell, C. M., Warner-Czyz, A., Wang, N.-Y., & Eisenberg, L. S. (2017). Early Sign Language Exposure and Cochlear Implantation Benefits. *Pediatrics, 140*(1). doi:10.1542/peds.2016-3489

Geers, A. E., & Sedey, A. L. (2011). Language and verbal reasoning skills in adolescents with 10 or more years of cochlear implant experience. *Ear and hearing, 32*(1 Suppl), 39S.

Goossens, N. A., Camp, G., Verkoeijen, P. P., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28*(1), 135-142.

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., . . . 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. doi:10.1044/2018_JSLHR-L-18-0071

Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the rainbow: Retrieval practice leads to better spelling than does rainbow writing. *Educational Psychology Review, 28*(2), 385-400.

Kaipa, R., & Danser, M. L. (2016). Efficacy of auditory-verbal therapy in children with hearing impairment: A systematic review from 1993 to 2015. *International Journal of Pediatric Otorhinolaryngology, 86*, 124-134.

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*, 350.

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition, 3*(3), 198-206.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation* (Vol. 61, pp. 237-284): Elsevier.

Knouse, L. E., Rawson, K. A., Vaughn, K. E., & Dunlosky, J. (2016). Does Testing Improve Learning for college students with attention-deficit/hyperactivity disorder? *Clinical Psychological Science, 4*(1), 136-143.
Kyle, F. E., & Harris, M. (2011). Longitudinal patterns of emerging literacy in beginning deaf and hearing readers. *Journal of Deaf Studies and Deaf Education, 16*(3), 289-304.
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. doi:doi:10.1044/2019_JSLHR-L-19-0221
Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., . . . 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. doi:10.1044/2018_JSLHR-L-18-0070
Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental psychology, 50*(4), 994.
Luckner, J. L., & Cooke, C. (2010). A summary of the vocabulary research with students who are deaf or hard of hearing. *American Annals of the Deaf, 155*(1), 38-67.
Lund, E. (2016). Vocabulary knowledge of children with cochlear implants: A meta-analysis. *The Journal of Deaf Studies and Deaf Education, 21*(2), 107-121. doi:10.1093/deafed/env060
Lund, E., & Douglas, W. M. (2016). Teaching vocabulary to preschool children with hearing loss. *Exceptional Children*, 0014402916651848.
Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory, 20*(8), 899-906.
McDaniel, J., Benítez-Barrera, C. R., Soares, A. C., Vargas, A., & Camarata, S. (2019). Bilingual versus monolingual vocabulary instruction for bilingual children with hearing loss. *The Journal of Deaf Studies and Deaf Education, 24*(2), 142-160.
McDaniel, J., Camarata, S., & Yoder, P. (2018). Comparing auditory-only and audiovisual word learning for children with hearing loss. *The Journal of Deaf Studies and Deaf Education, 23*(4), 382-398.
McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research, 60*(10), 2891-2905.
Nittrouer, S., Lowenstein, J. H., & Antonelli, J. (2019). Parental language input to children with hearing loss: Does it matter in the end? *Journal of Speech, Language, and Hearing Research*, 1-25.
Qi, S., & Mitchell, R. E. (2012). Large-scale academic achievement testing of deaf and hard-of-hearing students: Past, present, and future. *Journal of Deaf Studies and Deaf Education, 17*(1), 1-18.
Reimer, C. K. (2019). The effect of retrieval practice on vocabulary learning for children who are deaf or hard of hearing.
Roberts, M. Y. (2019). Parent-implemented communication treatment for infants and toddlers with hearing loss: A randomized pilot trial. *Journal of Speech, Language, and Hearing Research, 62*(1), 143-152. doi:10.1044/2018_JSLHR-L-18-0079
Ruben, R. J. (2018). Language development in the pediatric cochlear implant patient. *Laryngoscope investigative otolaryngology, 3*(3), 209-213.

Werfel, K. L. (2017). Emergent literacy skills in preschool children with hearing loss who use spoken language: Initial findings from the early language and literacy acquisition (ELLA) study. *Language, Speech, and Hearing Services in Schools, 48*(4), 249-259. doi:10.1044/2017_LSHSS-17-0023

2.0 Rationale and Specific Aims

This study is designed to test the effectiveness and efficiency of retrieval practice for word learning for children who are DHH.

Aim 1: Evaluate the effectiveness and efficiency of retrieval practice with and without feedback for the acquisition and retention for learning novel words. Both conditions are hypothesized to be effective with the feedback condition (i.e., explicit responses about accuracy given to child during teaching sessions) being more efficient than no feedback. Feedback is predicted to improve efficiency because it focuses the child's attention on unmastered material, prevents the illusion of success, and reduces repeated error.

Aim 2: Evaluate the effectiveness and efficiency of retrieval practice with spaced versus massed trials for the acquisition and retention for learning novel words. Both conditions are hypothesized to be effective with spaced trials (i.e., words are interspersed with one another) yielding more efficient learning than massed trials because spaced trials require more effort with fewer cues provided.

3.0 Inclusion/Exclusion Criteria

Children aged 5 to 9 years old who are DHH and use spoken English. Participants must have nonverbal cognitive abilities at least within the average range and receptive and expressive vocabulary standard scores of at least 70 to increase the likelihood that participants will exhibit word learning with a timeframe necessary for the study. Participants must use English as their only spoken language. The word sets must be carefully balanced by certain phonological features. When children are exposed to more than one spoken language, variation in their phonological systems can alter the balancing.

4.0 Enrollment/Randomization

Participants are enrolled via consent from a parent or legal guardian.

Each participant receives two active conditions and the no-teaching control condition. The specific active interventions are determined based on a random number generator.

This study is a clinical trial. Registration on ClinicalTrials.gov completed August 22, 2022.

5.0 Study Procedures

Eligibility Assessments

The assessment battery includes measures of vocabulary (Expressive Vocabulary Test, Third Edition; Peabody Picture Vocabulary Test, Fifth Edition; Dunn, 2018; Williams, 2018), and nonverbal cognition (Leiter International Performance Scale, Third Edition; Roid et al., 2013) skills. Participants must achieve standard scores of at least 70 for spoken receptive and expressive vocabulary skills to increase the likelihood that participants will exhibit word learning within a time frame necessary for the study. Individuals are excluded for below average nonverbal cognitive ability (i.e., standard score below 85 on the Leiter International Performance Scale; Roid et al., 2013). Medical, audiological, and school records supplement caregiver report (e.g., hearing and vision levels). Participants may not have uncorrected vision impairment (i.e., identified vision loss without use of corrective lenses that interferes with eligibility evaluation tasks), or evidence of severe motor impairment (i.e., insufficient motor skills to complete eligibility evaluation tasks independently). Such individuals are unlikely to benefit from the intervention and/or be able to participate in the study tasks.

Descriptive Assessments

Potential participants complete an eligibility evaluation. The assessment battery includes measures of language (Test of Language Development, Intermediate, Fifth Edition; Test of Language Development – Primary, Fifth Edition; Hammill & Newcomer, 2019; Newcomer & Hammill, 2019), speech production (Arizona Articulation Proficiency Scale, Fourth Edition, Fudala & Stegall, 2017), and speech perception (Test of Auditory Processing Skills; Bergeson & Pisoni, 2004; Martin & Brownell, 2005) skills. Caregivers provide demographic and background information and complete the Language Access Profile Tool (Hall & de Anda, 2021) to capture the diverse, dynamic nature of early language access of DHH children. Individual differences in these characteristics may influence word learning efficiency, but specifically how they do is not yet well understood (Carpenter et al., 2016; Jaeger et al., 2015; Karpicke et al., 2016; Moreira et al., 2019). These characteristics are expected to impact the study's external validity and to guide future research.

Procedures for Word Learning Sessions

Novel Word Set Development

Because the AATD requires sets of equal difficulty, the word sets are made of novel words balanced for number of phonemes and syllables, phonotactic probability, and lexical neighborhood density (Storkel, 2003; Storkel & Hoover, 2010a, 2010b). Novel words nearly eliminate the possibility of prior exposure and enable the creation of balanced word sets. Early developing consonants and tense vowels expected to be in the participants' phonemic repertoires (assessed prior to baseline) are primarily used, which aids acquisition receptively and expressively (Leonard et al., 1982). Words do not contain high frequency phonemes (based on acoustic frequency) because they influence word learning and are more likely to be less audible, inaudible, or distorted than lower frequency phonemes for DHH children (Niskar et al., 1998; Shargorodsky et al., 2010; Stelmachowicz et al., 2004b). The following words sets are used: (1) /zod/, /wam/, /get͡ʃ/, and /pib/; (2) /pem/, /vad/, /jot͡ʃ/, and /kub/; (3) /jum/, /tet͡ʃ/, /vob/, and /ked/.

Unfamiliar Objects
Each novel word is randomly paired to an unfamiliar object (e.g., building sets and hardware store items). Testing prior to baseline ensures that participants do not have a consistent form associated with any unfamiliar object or a meaning associated with any of the novel words.

Semantic Content
During intervention the examiner labels the novel object's location (e.g., toy box or bathtub) to increase semantic content. Including this content provides an early step towards using retrieval practice to teach rich vocabulary content that is necessary for achieving age expected language skills without threatening the study's internal validity. Locations are randomly assigned to avoid objects' perceptual features being associated with locations and pretested to ensure participants have the conceptual knowledge of each location to maintain balanced word sets across conditions.

Procedures
All sessions are in-person and videorecorded. Participants must pass a listening check to proceed with the session. Sessions take place in a quiet room (ambient noise $\leq$40 dB as measured by a sound level mobile application) in the child's home, educational setting, or clinic, depending on caregiver preference. Flexibility in location to match caregiver preferences is one way we guard against attrition. Having the examiner for all assessments and intervention sessions be a research team member, rather than school personnel, and conducting the intervention for a relatively short period of time are also designed to guard against attrition.

Dependent Variable Probe
The dependent variable probes (about 10 minutes) test the 12 target words in sets by condition. The primary dependent variable is the percent accuracy (i.e., correct responses / total trials x 100) for labeling target words expressively. The examiner asks the child to label the taught unfamiliar objects and one known object per set (i.e., "What's this?"). The participant is given up to 10 seconds to respond. The secondary dependent variables are the percent accuracy for (a) receptively identifying target words within 5 seconds in a field of six (i.e., four target objects, one foil, and a known object), (b) associated semantic information (i.e., location; "Where's this from?"), and (c) phonological precision for labeling target words expressively. The examiner will not provide any feedback regarding response accuracy for target words during the dependent variable probes. To maintain engagement, the examiner may provide positive reinforcement for non-target behaviors and for identifying the known word for each group, which is expected to be particularly important in the early sessions when accuracy for target words is expected to be low. The study's procedural manual specifies how the examiner should respond to specific child behaviors (e.g., self-corrections and articulatory errors). Primary and secondary coders double enter data into an electronic form using direct systematic observation of the video recordings (Ledford & Gast, 2018).

A research assistant codes at least 25% of sessions in each condition for each participant for interobserver agreement (IOA). The primary coder is blind to which sessions are coded for reliability. All coders are trained to .85 point-by-point agreement for independent coding of three consecutive sessions via recordings of non-participants

(Ledford & Gast, 2018). Training includes reading the coding manual, a didactic presentation, and group coding of non-participants with discrepancy discussions prior to independent coding. Discrepancy discussions occur regularly. If IOA falls below .85 for two sessions, coders must re-train to the initial training criteria.

Baseline Phase

Baseline sessions will occur three times per week and continue until the participant demonstrates a stable pattern of response, with a minimum of three sessions (Ledford & Gast, 2018). Participants complete a listening check then the dependent variable probe to test learning of the 12 target words.

Comparison Phase

Comparison phase sessions occur three times per week after the baseline phase is complete. They begin with a listening check, then the dependent variable probe, and then the two retrieval practice intervention conditions (i.e., feedback-massed, no feedback-massed, feedback-spaced, and no feedback-spaced) for the participant's specific, randomly assigned contrast. In each session, the second condition occurs immediately after the first. Participants are taught the meaning of novel nouns (Lund & Douglas, 2016; Reimer, 2019; Storkel et al., 2019). The independent variable that differentiates the conditions is either the presence of feedback or spaced versus massed trials. All other potentially influential variables are held constant. Table 1 provides a summary of the key teaching procedures for the massed trials conditions (with and without feedback). For the massed trials conditions, all exposures for one word are provided before presenting the next word in an order determined via a random number generator. The spaced trials conditions proceed in the same manner with the exception that the exposures for the four target words are interspersed with one another (e.g., an exposure for word 1, then one for word 3, then one for word 2, etc.). The interspersing of trials across words is especially apparent through the color coding of the steps by word in the Teaching Sessions section of the procedural manual.

Table 1

Summary of Teaching Procedures for Massed Trials Conditions

1. Examiner shows novel object, labels it, and states its location (e.g., "This is a *X*. The *X* is from the park.")
2. Examiner and participant color pictures of each novel object to provide shared context
3. Examiner says the label for the novel object ten more times and its location three more times
4. Examiner asks the participant to label the novel object four times and the location twice for retrieval practice

Feedback condition: Examiner provides direct feedback on accuracy and labels the target word (e.g., "Yes, that is a *X*!" or "Oops, that is a *X*!")	No-feedback condition: Examiner says statement without target word and then one with it to provide equal exposures to the feedback conditions (e.g., This looks neat. Here's the X!)

5. After the first word, examiner completes same procedure with the three remaining target words in the set
6. Examiner concludes set by asking participant to label and identify the picture for each word and state its location

Feedback condition: Examiner provides direct feedback without naming the target word	No-feedback condition: Examiner does not provide feedback and proceeds to next item

After a participant reaches the criterion of 75% accuracy for expressive labeling for three consecutive sessions for the four target words for one condition, instruction in the remainder of the comparison phase is provided only in the unmastered condition. The comparison phase ends when the unmastered condition reaches criterion. If the participant does not reach mastery after two times the number of sessions to criterion in the mastered condition, the maintenance phase begins then.

Maintenance Phase

Maintenance data are collected weekly via baseline procedures. Maintenance data are first collected a week after the participant achieves mastery for at least one condition. The same dependent variable probe used for the baseline and comparison phases is used to collect maintenance data. Sessions continue for 4 weeks after the comparison phase to permit sufficient time for a trend to emerge for visual analysis. If a consistent pattern is observed before 4 weeks have elapsed, the maintenance phase will end at that time. For example, the maintenance phase ends if a participant exhibits 0% accuracy for all conditions for two consecutive sessions.

Procedural Fidelity

Prior to study sessions, the examiner will demonstrate procedural fidelity (PF) of at least 85% accuracy across three consecutive sessions with non-participants. A research assistant codes at least 25% of study sessions across each phase and condition for each participant. The examiner is blind to which sessions are coded for PF. PF behaviors for probes are consistent across all phases. Only the independent variable varies across the active conditions' PF behaviors. If additional exposures occur, the total number is reported by participant and target word. If PF overall or for providing feedback or

spacing falls below 80% for two sessions, the examiner completes re-training. PF will be analyzed formatively at the behavior level.

6.0 Reporting of Adverse Events or Unanticipated Problems involving Risk to Participants or Others

All adverse events will be reported to the IRB and appropriate regulatory agencies within 7 business days. Any adverse event of a critical nature will be reported within 24 hours.

7.0 Study Withdrawal/Discontinuation

Participants may withdraw consent to participate in this study at any time. They also have the right to cancel permission to use and disclose further information collected about the child, in writing, at any time. If they cancel permission to use the child's information, the researchers will stop collecting additional information. However, the research team may use and disclose information that was gathered before they received the cancellation.

Assent is determined by demonstrating activities to the participants and asking if they want to participate. If a participant indicates he or she does not want to participate (though verbal and/or nonverbal means), the principal investigator or research assistant offers a break and then invites the participant to engage in the activity again. If the child continues to verbally and/or nonverbally (e.g., running away or crying) indicate he or she does not want to participate after three consecutive attempts to re-engage, the intervention is discontinued for that session. If a child declines to participate during three consecutive sessions, we will discontinue intervention with that participant and discuss that discontinuation with his or her caregivers.

8.0 Statistical Considerations

Visual Analysis

Visual analysis is the primary analysis method for single case research design (Ledford & Gast, 2018). Consistent with What Works Clearinghouse practices, we consider five features of the data pattern: level, trend, variability, overlap, and consistency (Kratochwill et al., 2010). We use visual analysis to determine whether there is functional relation between the independent variable (i.e., specific retrieval practice component) and the dependent variable. For each participant with the relevant conditions in their contrast, we examine whether there is evidence of a functional relation between the independent variable and dependent variable. We combine across participants based on the What Works Clearinghouse criteria for strong, moderate, and no evidence of an effect (What Works Clearinghouse, 2016).

Research questions related to effectiveness of the active interventions (i.e., RQs 1-4 and 7-10) will be judged based on accuracy for words taught in the active intervention condition versus those in the no-teaching condition. Research questions related to efficiency (i.e., RQs 5, 6, 11, and 12) will be judged based on accuracy for words taught in one active condition versus those taught in the other (e.g., retrieval practice with versus without feedback).

Feedback. Participants (n = 8) in Contrasts A (feedback vs. no feedback with massed trials) and B (feedback vs. no feedback with spaced trials) address the effect of feedback (see Figure 1). Table 2 displays how we will compare conditions within Contrasts A and B to evaluate both effectiveness and efficiency research questions. For all of these analyses, we use expressive labeling as the primary dependent variable. Therefore, performance for expressive labeling determines when phases change. Phonological precision, receptive performance, and knowledge of semantic content (i.e., location) are secondary variables that will be considered when interpreting the overall results. For example, one condition may exhibit greater accuracy for the primary dependent variable, but not for one or more of the secondary variables.

Figure 1
Study Contrasts

	Feedback	No feedback	Contrasts to Determine Effect of Feedback (RQs 1-6)
Massed	Feedback with massed trials	No feedback with massed trials	Contrast A: Feedback vs. no feedback (massed trials)
Spaced	Feedback with spaced trials	No feedback with spaced trials	Contrast B: Feedback vs. no feedback (spaced trials)
Contrasts to Determine Effect of Spacing (RQs 7-12)	Contrast D: Spaced vs. massed trials (with feedback)	Contrast C: Spaced vs. massed trials (no feedback)	

Table 2. *Comparisons of Relevant Conditions for the Effect of Feedback*

Research question						
Effectiveness or efficiency	Acquisition or retention	Contrast	Condition 1	Condition 2	Phase	Feedback RQs
Effectiveness	Acquisition	A	Feedback w/massed trials	No-teaching control	Comparison	1
		B	Feedback w/spaced trials	No-teaching control	Comparison	1
		A	No feedback w/massed trials	No-teaching control	Comparison	2
		B	No feedback w/spaced trials	No-teaching control	Comparison	2
Effectiveness	Retention	A	Feedback w/massed trials	No-teaching control	Maintenance	3
		B	Feedback w/spaced trials	No-teaching control	Maintenance	3
		A	No feedback w/massed trials	No-teaching control	Maintenance	4
		B	No feedback w/spaced trials	No-teaching control	Maintenance	4
Efficiency	Acquisition	A	Feedback w/massed trials	No feedback w/massed trials	Comparison	5
		B	Feedback w/spaced trials	No feedback w/spaced trials	Comparison	5
Efficiency	Retention	A	Feedback w/massed trials	No feedback w/massed trials	Maintenance	6
		B	Feedback w/spaced trials	No feedback w/spaced trials	Maintenance	6

Effectiveness. For research questions related to effectiveness, we determine whether there is an increase in accuracy for the active retrieval practice condition without an

accompanying increase, or with a smaller increase, for the no-teaching control condition. Given that the target words are novel words, participants are not expected to show increased accuracy for the dependent variables for the control condition.

For effectiveness of retrieval practice with feedback for *acquisition* (RQ1), we determine whether there is a separation of the data paths (as indicated by visual analysis and statistical analyses) for (a) the retrieval practice with feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with spaced trials condition and the no-teaching control condition in the comparison phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the comparison phase would provide evidence of the effectiveness of retrieval practice with feedback.

For effectiveness of retrieval practice without feedback for *acquisition* (RQ2), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice without feedback with spaced trials condition and the no-teaching control condition in the comparison phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the comparison phase would provide evidence of the effectiveness of retrieval practice without feedback.

For effectiveness of retrieval practice with feedback for *retention* (RQ3), we determine whether there is a separation of the data paths for (a) the retrieval practice with feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with spaced trials condition and the no-teaching control condition in the maintenance phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the maintenance phase would provide evidence of the effectiveness of retrieval practice with feedback for retention.

For effectiveness of retrieval practice without feedback for *retention* (RQ4), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice without feedback with spaced trials condition and the no-teaching control condition in the maintenance phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the maintenance phase would provide evidence of the effectiveness of retrieval practice without feedback for retention.

Efficiency. For research questions related to efficiency, we determine whether there is faster attainment of the mastery criterion for one retrieval practice condition versus another and whether mastery-level performance is better maintained.

For efficiency of retrieval practice with versus without feedback for *acquisition* (RQ5), we determine whether there is a separation of the data paths for (a) the retrieval practice with feedback with massed trials condition and the retrieval practice without feedback with massed trials condition (Contrast A) and (b) the retrieval practice with feedback with spaced trials condition and the retrieval practice without feedback with spaced trials condition (Contrast B) in the comparison phase. The number of sessions to mastery is

also considered when determining the presence of a functional relation related to efficiency. This separation of the data paths in the comparison phase would provide evidence of greater efficiency for the condition with the path that is displayed higher on the graph.

For efficiency of retrieval practice with versus without feedback for *retention* (RQ6), we determine whether there is a separation of the data paths for (a) the retrieval practice with feedback with massed trials condition and the retrieval practice without feedback with massed trials condition (Contrast A) and (b) the retrieval practice with feedback with spaced trials condition and the retrieval practice without feedback with spaced trials condition (Contrast B) in the maintenance phase. The number of sessions to fall below mastery is also considered when determining the presence of a functional relation related to efficiency. This separation of the data paths in the maintenance phase would provide evidence of greater efficiency for the condition with the path that is displayed higher on the graph.

Spacing. Participants (n = 8) in Contrasts C (spaced vs. massed trials without feedback) and D (spaced vs. massed trials with feedback) address the effect of spacing. Table 3 displays how we will compare conditions within Contrasts C and D to evaluate both effectiveness and efficiency research questions. As with the analyses for feedback, we use expressive labeling as the primary dependent variable. Phonological precision, receptive performance, and knowledge of semantic content are secondary variables that will be considered when interpreting the overall results.

Table 3. *Comparisons of Relevant Conditions for the Effect of Spacing*

Research question						
Effectiveness or efficiency	Acquisition or retention	Contrast	Condition 1	Condition 2	Phase	Spacing RQs
Effectiveness	Acquisition	C	No feedback w/spaced trials	No-teaching control	Comparison	7
		D	Feedback w/spaced trials	No-teaching control	Comparison	7
		C	No feedback w/massed trials	No-teaching control	Comparison	8
		D	Feedback with massed trials	No-teaching control	Comparison	8
Effectiveness	Retention	C	No feedback w/spaced trials	No-teaching control	Maintenance	9
		D	Feedback w/spaced trials	No-teaching control	Maintenance	9
		C	No feedback w/massed trials	No-teaching control	Maintenance	10
		D	Feedback with massed trials	No-teaching control	Maintenance	10
Efficiency	Acquisition	C	No feedback w/massed trials	No feedback w/spaced trials	Comparison	11
		D	Feedback w/massed trials	Feedback w/spaced trials	Comparison	11
Efficiency	Retention	C	No feedback w/massed trials	No feedback w/spaced trials	Maintenance	12
		D	Feedback w/massed trials	Feedback w/spaced trials	Maintenance	12

Effectiveness. For research questions related to effectiveness, we determine whether there is an increase in accuracy for the active retrieval practice condition without an accompanying increase, or with a smaller increase, for the no-teaching control condition. Given that the target words are novel words, participants are not expected to show increased accuracy for the dependent variables for the control condition.

For effectiveness of retrieval practice with spaced trials for *acquisition* (RQ7), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with spaced trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with spaced trials condition and the no-teaching control condition in the comparison phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the comparison phase would provide evidence of the effectiveness of retrieval practice with spaced trials.

For effectiveness of retrieval practice with massed trials for *acquisition* (RQ8), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with massed trials condition and the no-teaching control condition in the comparison phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the comparison phase would provide evidence of the effectiveness of retrieval practice with massed trials.

For effectiveness of retrieval practice with spaced trials for *retention* (RQ9), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with spaced trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with spaced trials condition and the no-teaching control condition in the maintenance phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the maintenance phase would provide evidence of the effectiveness of retrieval practice with spaced trials for retention.

For effectiveness of retrieval practice with massed trials for *retention* (RQ10), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the no-teaching control condition and (b) the retrieval practice with feedback with massed trials condition and the no-teaching control condition in the maintenance phase. This separation of the data paths with the retrieval practice condition above the no-teaching condition in the maintenance phase would provide evidence of the effectiveness of retrieval practice with massed trials for retention.

Efficiency. For research questions related to efficiency, we determine whether there is faster attainment of the mastery criterion for one retrieval practice condition versus another and whether mastery-level performance is better maintained.

For efficiency of retrieval practice with spaced versus massed trials for *acquisition* (RQ11), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the retrieval practice without feedback with spaced trials condition (Contrast C) and (b) the retrieval practice

with feedback with massed trials condition and the retrieval practice with feedback with spaced trials condition (Contrast D) in the comparison phase. The number of sessions to mastery is also considered when determining the presence of a functional relation related to efficiency. This separation of the data paths in the comparison phase would provide evidence of greater efficiency for the condition with the path that is displayed higher on the graph.

For efficiency of retrieval practice with spaced versus massed trials for *retention* (RQ12), we determine whether there is a separation of the data paths for (a) the retrieval practice without feedback with massed trials condition and the retrieval practice without feedback with spaced trials condition (Contrast C) and (b) the retrieval practice with feedback with massed trials condition and the retrieval practice with feedback with spaced trials condition (Contrast D) in the maintenance phase. The number of sessions to fall below mastery is also considered when determining the presence of a functional relation related to efficiency. This separation of the data paths in the maintenance would provide evidence of greater efficiency for the condition with the path that is displayed higher on the graph.

Statistical Analysis

There is no consensus on recommended statistical analyses for single case research designs, particularly for the adapted alternating treatments design, which is underrepresented in research on the use of statistical analyses and effect sizes in single case research. We will calculate Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009), an overlap-based effect size, and analyze differences in conditions analogous to those described for visual analysis. Because overlap is one feature used for determining a functional relation, NAP will provide guidance on the degree of confidence for inferring the amount of overlap between conditions. NAP is defined as the number of comparison pairs between phases A and B without overlap divided by the total comparisons. It is related to Tau, another overlap metric. Tau is a linear re-scaling of NAP (Zimmerman et al., 2018). NAP offers several advantages over other overlap-based effect sizes, including less human error in calculations (can be calculated by hand or via computerized methods), stronger validation by R^2, and narrower confidence intervals (Parker & Vannest, 2009). Parker and Vannest (2009) tentatively suggest the following guidelines for NAP effect sies: .93 – 1.00 = large, .66 – .92 = medium, and 0 – .65 = weak.

9.0 Privacy/Confidentiality Issues

All assessment data and medical/developmental history data will be de-identified and stored on a HIPAA-compliant database server, cloud storage, REDCap, and/or external hard drives. All storage locations will follow necessary confidentiality requirements. An ID and password are required to log into project computers and only key study personnel will have access. Each of these staff members will have participated in training regarding confidentiality. All data regarding participant performance will be de-identified by assigning a code for each participant. Identified participant information such as intake forms will be stored in centralized locked cabinets. Only the PI and trained key study personnel will have access to keys to the locked cabinets. Any forms or correspondences that contain identifiable material will be shredded with a secure

shredder. As indicated in our informed consent letter, deidentified data will be stored indefinitely. Contact information for families will be kept in the locked cabinets for up to 5 years past the end of the study. We will use videorecordings of some assessment sessions or portions of assessment sessions to measure fidelity and reliability. As part of our informed consent process, parents or guardians will indicate if they consent to sharing videos with other researchers and/or a larger audience for instructive purposes. Relevant audio and video recordings being used for research purposes will be stored indefinitely on the HIPAA-compliant, encrypted server, cloud storage, or hard drive.

CONSENT FORMS: Hard copies of consent forms are stored in a locked storage room (currently MCE 10222). All consent forms are maintained in a file folder labeled with IRB number. A log of consent forms is placed at the top of all consent forms and lists the date the consent was signed, who entered the consent on the log, the participant's name, and the participant's date of birth. This log is used for the IRB annual review. When the study is completed, the consent log and the consent documents will be destroyed.

DATABASE: A de-identified database will be created from study data. Excel or REDCap will be used for the database.

REDCap data may be downloaded into Excel files. Data may also be stored as SPSS data files. In all instances, database will be de-identified and will be stored as electronic files on password protected computer in a locked lab (currently MCE 10216) or on the password protected server, cloud storage, or hard drive.

10.0 Follow-up and Record Retention

This study is expected to be completed by November 2024.

Contact information for families will be kept in the locked cabinets for up to 5 years past the end of the study.

A repository will be created. Caregivers may select whether or not they agree with their children’s data being part of the repository for future studies on the informed consent form. Electronic data and recordings are stored on a HIPAA-compliant server, cloud-storage, and/or external hard drives. Participant video and audio recordings are labeled by participant ID, not their name. Access to the data repository is controlled by Jena McDaniel.