Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
1–14

DOI: 10.1177/0022219419862266
journaloflearningdisabilities.sagepub.com
SAGE

Learning Lexical Information Depends Upon Task, Learning Approach, and Reader Subtype

Neena M. Saha, MS[1], Stephanie N. Del Tufo, PhD[1], and Laurie E. Cutting, PhD[1]

Abstract

Learning to read relies upon the integration of phonological, orthographic, and semantic information. However, no studies have investigated how children with varying reading abilities learn phonological-orthographic (PO) and semantic aspects of novel words as a function of both learning approach (LA; e.g., learning new words in isolation or context) and outcome (fluency or comprehension). In this study, 45 children participated in three tasks that differentially tested PO and semantic attributes of novel pseudo-words learned through two learning approaches. Children were classified into groups as having dyslexia (DYS), having specific reading comprehension deficits (S-RCDs), or being typically developing readers (TD). Differences were found between groups, with S-RCD poorer than TD on semantic but not PO components of learning. Children with DYS displayed impaired results on both semantic and PO learning but showed an interaction on task by LA performance. Specifically, in the DYS group, isolation LA yielded better performance on PO learning, while context LA was better for semantic learning. These results indicate that (a) children with S-RCDs have a unique learning profile that is dissociable from DYS and TD and (b) reading impairments are not static but rather influence acquisition of reading skill in different ways, depending on reading profile.

Keywords

dyslexia, specific reading comprehension deficits, vocabulary

Given the abundant positive individual and societal outcomes associated with acquiring successful reading skills (e.g., Kern & Friedman, 2009) it is alarming that upwards of 60% of children in the United States are not reading at grade level (U.S. Department of Education, 2017). While there is a number of reasons underlying poor performance on national testing, one key feature of reading success is adequate word knowledge (Carroll, 1993). *Word knowledge* (also sometimes broadly referred to as *lexical information;* Perfetti & Stafura, 2014) is defined here as semantic, phonological, or orthographic information that aids word recognition (Seidenberg & McClelland, 1989). Multiple studies have found that word knowledge aids both fluency and comprehension. For example, Nagy and Scott (2000) found that to comprehend a text, 90% of words included in the text need to be known, and as decoding skills are mastered, vocabulary (i.e., the quality of lexical representations) becomes a better predictor of reading comprehension than are decoding skills (Protopapas, Sideridis, Mouzaki, & Simos, 2007). Therefore, knowledge of both semantic and phonological-orthographic information in a word is an important facet of skilled reading.

The Reading Systems Framework is a comprehensive model of reading that integrates both word reading and reading comprehension and has explanatory power across reader subtypes (Perfetti & Stafura, 2014). The Reading Systems Framework model designates the *lexicon* (a person's knowledge of phonological, orthographic, and semantic attributes of words) as the bridge between word recognition and comprehension. Children with well-specified (specific and redundant) phonological-orthographic knowledge and flexible semantic knowledge will possess a strong mental lexicon (Perfetti, 2007) that can be called upon to aid in online word recognition and comprehension of words during reading. This in turn enables fluent reading, and fluent reading allows cognitive resources to redirect from decoding to comprehension—the ultimate goal of reading.

Perfetti and Stafura's (2014) "mental lexicon" is roughly synonymous with a given child's vocabulary size, and research has shown that vocabulary size is an important predictor of reading ability (Protopapas et al., 2007). Typically developing young children acquire several new words a day

[1]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Laurie E. Cutting, PhD, Peabody College, Vanderbilt University, 416C One Magnolia Circle, Nashville, TN 37230, USA.
Email: laurie.cutting@vanderbilt.edu

(Oakhill, Cain, & Elbro, 2014); however, children who enter school with poor vocabularies, such as those with learning disabilities (LD), are often at a disadvantage (Oakhill et al., 2014) and have difficulty catching up to their peers (Stanovich, 1986). Research on how children with LD learn new words is critical in order to design effective interventions and help them catch up to their typically developing peers.

How Children Learn New Words

Impact of Learning Approach and Word Information Type

Short-term word-learning paradigms have been used to examine how children learn various attributes of new words across different types of conditions (referred to throughout this article as "learning approach"). Some early research using word-learning paradigms investigated the role of learning words in context. Ehri and Roberts (1979) found that children who read words embedded in sentences learned more about the semantic aspects of those words, whereas children who learned words in isolation could read the words faster and knew more about the orthographic form of the words. Recent work found a similar interaction of learning approach by task demand (or the particular aspect of word knowledge being recruited in the task) in skilled readers: Undergraduate students remembered semantic information better from words read in context and remembered orthographic information better from words that were presented in lists (Martin-Chang & Levesque, 2015). In addition, Martin-Chang and Levy (2006) found that (a) the type of task influenced performance and (b) the method of reading instruction was directly linked with the intended reading outcomes. More specifically, they found that across two reader groups (skilled and unskilled), words learned in isolation were read faster when tested in list form, whereas words learned in context were read faster when tested in context. However, they used real words, rather than novel pseudo-words, and therefore were not able to account for word familiarity.

This research shows that different learning approaches implicitly foster different types of word knowledge (lexical information). While this research was conducted in typically developing individuals, short-term word-learning paradigms have been used to examine differences in learning in individuals with varying reading ability.

Differences Among Reader Groups

While previous investigations have focused on information learned in context versus isolation, short-term word-learning paradigms can also investigate how children with various reading profiles respond to tasks that require different forms of word knowledge. Arguably, the differential effects of word information and learning approach are most critical when examining children with reading disabilities because it can inform intervention practices.

Developmental dyslexia. In this article, we also refer to developmental dyslexia as *dyslexia* or *reading disability*, and it is abbreviated to the acronym DYS when referring to our sample of children with dyslexia. Developmental dyslexia is perhaps the most widely known of the specific LD, with an estimated prevalence of 5% to 17% among school-age children (Shaywitz, Morris, & Shaywitz, 2008). Dyslexia is often characterized by deficits in word-level processing, which compromises the speed and accuracy of reading (Vellutino, Fletcher, Snowling, & Scanlon, 2004). Studies have demonstrated that children with dyslexia exhibit pervasive and persistent deficits in phonological processing that are compounded in the presence of increasing lexical information (such as long words and difficult text). In terms of short-term word-learning paradigms, children with dyslexia have been shown to have difficulties learning phonological and orthographic information (e.g., Bailey, Manis, Pedersen, & Seidenberg, 2004; Castles & Holmes, 1996).

Specific reading comprehension deficits (S-RCDs). While much inquiry during the past few decades has focused on the phonological factors of impaired reading, research has clearly documented that there is a subset of children who have a selective deficit in comprehension skill with intact word-reading skills (e.g., Cain & Oakhill, 2006). This impairment is referred to as *specific reading comprehension deficit*, or S-RCD. Despite estimates suggesting that 10% of children have S-RCD (Nation, 2005), these children are much less studied than children with phonological deficits (Hulme & Snowling, 2009). S-RCD is characterized by intact decoding skill with accompanying deficits in semantic processing and/or broader oral language skills (e.g., Colenbrander, Kohnen, Smith-Lock, & Nickels, 2016; Nation, Snowling, & Clarke, 2007).

Several researchers have utilized word-learning paradigms to better characterize the nature of S-RCDs. Ricketts, Nation, and Bishop (2007) employed a short-term word-learning task to test orthographic, phonological, and semantic aspects of nonwords, and they found that children with poor comprehension showed learning equivalent to that of control group members across tasks but poor retention of semantic information over time. On a task of semantic learning, Cain, Oakhill, and Elbro (2003) found that children with poor comprehension skills were less able to integrate the information from context than were control group children when the defining context for the word increased in distance from the target word. Last, Nation et al. (2007) taught children with poor comprehension and their typically developing peers semantic and phonological attributes of invented words and examined their performance on a series of tasks that required them to differentially recruit decoding and comprehension skill. They found that there was no difference in the number of trials it took children with poor comprehension to learn the phonological attributes of the words relative to their peers, but they were significantly impaired at recalling the semantic attributes. However, Nation et al. (2007) did not change the context of the words presented; all words were presented orally alongside a colored flashcard with a picture representing the meaning. While the above studies suggest that current reading ability interacts with task demands when predicting performance, the role of learning approach remains an open question.

Examining Differences Among Reader Groups as a Function of Task and Learning Approach

While it is clear that word knowledge is critical for vocabulary development, and vocabulary contributes to reading comprehension, little research exists on how readers with specific LD actively acquire new word knowledge as a function of task demands (what is being tested) and learning approach (isolation vs. context). Indeed, very few studies (if any) have examined how both phonological-orthographic and semantic word information are learned differentially across reader groups and how learning approach (isolation vs. context) interacts with type of word information and reader group. Therefore, we addressed these gaps in the research by teaching and testing both phonological-orthographic and semantic word information across two learning approaches (isolation vs. context) for every participant (i.e., it was a fully crossed design). Assessing both types of word information (phonological-orthographic and semantic) in a meaningful context is, arguably, more ecologically valid than using isolation-only learning approaches. Specifically, in classroom environments—whether independent reading or vocabulary learning—children are almost always provided a word embedded in a meaningful context. Furthermore, our sample consisted of children with varied reading skills so we also were able to examine the effect of existing reading impairments on new word acquisition. If existing reading impairments are not static, and thus affect acquisition of new word knowledge, then examining performance in a word-learning paradigm across different types of readers could yield information that would be useful for developing targeted interventions for students who have specific LD.

Hypotheses

Broadly, we were interested in examining differences among reader groups as a function of task (which differentially recruits the type of word information: semantic vs. phonological-orthographic) as well as learning approach (context vs. isolation). Specifically, we investigated the following questions:

1. Across various tasks that differentially recruit phonological or semantic information, how do children with selective reading deficits (phonological-orthographic for DYS; semantic for S-RCD) acquire phonological-orthographic and semantic attributes of new words?
2. How does a word-learning approach (isolation vs. context) interact with current reading ability across tasks?

We hypothesized that compared to the DYS group, the S-RCD group would experience difficulty learning new semantic information yet have intact word phonological-orthographic learning, while the DYS group would experience difficulty learning both types of information. We also examined whether the manner in which new words were learned (isolation vs. context, or *learning approach*) facilitates performance across the different subtypes of reading ability and/or different aspects of word learning (phonological-orthographic vs. semantic). More specifically, we hypothesized that children with dyslexia would benefit from the isolated learning approach for tasks with outcome measures that involved fluency (therefore reflecting phonological-orthographic learning), whereas the contextual learning approach would be more beneficial for tasks with semantic outcome measures.

Method

Participants

Study participants were 87 children in Grades 4 through 8 (43 boys; mean age = 11.6 years, age range = 9.14–14.83 years). Approximately 64% of the sample listed their race as White, 30% Black/African American, 2% more than one race, 2% Asian; 2% preferred not to answer. Participants were recruited through flyers posted at the Johns Hopkins School of Medicine and in other areas around the Johns Hopkins community. All participants were native American English speakers with normal hearing, and all had normal or corrected-to-normal vision. In addition, participants had no history of developmental disability (other than dyslexia and/or attention-deficit/hyperactivity disorder), major psychiatric illness, or neurological disorder. In accordance with the Johns Hopkins Institutional Review Board policy, all human participant procedures were preapproved, and participants properly consented to the study and were compensated for their time.

Measures

All participants completed a behavioral battery of tests, including several measures of reading ability (see Table 1). We used the fourth edition of the *Gates-MacGinitie Reading Test* (GMRT-4; MacGinitie, MacGinitie, Maria, & Dreyer,

Table 1. Demographic Information Across Groups.

Characteristic/Test	TD Group $n = 19$	DYS Group $n = 14$	S-RCD Group $n = 12$
Age (years)	11.6 (1.4)	11.5 (1.5)	11.6 (1.5)
Grade	5.8 (1.5)	5.6 (1. 5)	5.7 (1.6)
TOWRE-2 percentile rank	53.3 (13.6)[a]	12.7 (7.3)[a,b]	49.9 (14.4)[b]
GMRT-4 percentile rank	69.9 (19.3)[a,b]	12.4 (7.1)[a]	17.1 (7.0)[b]
WISC-IV Perceptual Reasoning Index Standard Score	107.6 (13.0)[a,b]	91.4 (13.0)[a]	93.8 (13.4)[b]

Note. TD = typically developing readers; DYS = dyslexia; S-RCD = specific reading comprehension deficit; TOWRE-2 = *Test of Word Reading Efficiency*, second edition; GMRT-4 = *Gates-MacGinitie Reading Test*, fourth edition; WISC-IV = *Wechsler Intelligence Scale for Children*, fourth edition. TD group members were selected from a larger sample and matched as closely as possible to the range of behavioral scores. Groups were formed as follows: TD: > 35th percentile on decoding (TOWRE-2) and > 25th percentile on comprehension (GMRT-4); DYS: < 25th percentile on decoding (TOWRE-2) and comprehension (GMRT-4); S-RCD: > 35th percentile on decoding (TOWRE-2) and < 25th percentile on comprehension (GMRT-4).
[a]Means with the same superscript differ significantly at $p < .01$.
[b]Means with the same superscript differ significantly at $p < .01$.

2000) and the second edition of the *Test of Word Reading Efficiency* (TOWRE-2; Torgesen, Wagner, & Rashotte, 1999) to categorize participants into groups based on their comprehension and word reading fluency abilities, respectively. Both assessments have high psychometric properties. Brief descriptions of these two tests are provided below, with more detailed descriptions in Appendix A.

GMRT-4. GMRT-4 is frequently used to assess reading comprehension ability in children (Cain & Oakhill, 1999; MacGinitie, MacGinitie, Maria, & Dreyer, 2008). The GMRT-4 comprehension test consists of 48 multiple-choice questions pertaining to 11 passages covering various subjects and have different lengths and prose types. Students are tested on their ability to understand both explicit and implicit information in the passages.

TOWRE-2. TOWRE-2 was selected to evaluate participants' single-word reading and pseudo-word reading fluency. TOWRE-2 consists of two subtests: Sight Word Efficiency and Phonetic Decoding Efficiency, both of which are timed. The Sight Word Efficiency subtest assesses the number of real printed words that are read correctly in 45 s, and the Phonetic Decoding Efficiency subtest measures the number of decodable pseudo-words that can be read correctly in 45 s.

In addition, given prior research suggesting that cognitive factors are related to word-learning differences (e.g., Cain, Oakhill, & Lemmon, 2004), we decided to assess performance IQ (PIQ) through use of the perceptual reasoning index subtest of the fourth edition of the *Wechsler Intelligence Scale for Children* (WISC-IV; Wechsler, 2004).

Participant Groups

The 87 participants were classified into groups based on a combination of their TOWRE-2 and GMRT-4 scores (see Table 1). Typically developing readers (TD group) were required to score above the 35th percentile on the TOWRE-2 and above the 25th percentile on the GMRT-4. Readers with DYS were required to score below the 25th percentile on both the TOWRE-2 and GMRT-4. Finally, readers with a S-RCD were required to score above the 35th percentile on the TOWRE-2 but below the 25th percentile on the GMRT-4. This yielded 60 participants in the TD group, 14 in the DYS group, and 12 in the S-RCD group. To create comparable sample sizes across groups and similar levels of decoding ability for the TD and S-RCD groups, participants were selected from the TD group who had the same range of TOWRE-2 scores as the S-RCD group, using a procedure similar to that used in Nation et al. (2007). This resulted in a final sample of 19 children in the TD group.

Procedure

Word-learning paradigm. The word-learning paradigm required participants to learn monosyllabic pseudo-words that were created by changing one letter of real monosyllabic words (see Tables A1 and A2 in Appendix A). For example, the pseudo-word "quobe" was created by changing the letter "t" to "b" from the real word "quote." Across two learning approaches (context and isolation) the target pseudo-words to be learned were matched on length, and their new assigned meanings (the new semantic information to be learned) were matched on part of speech (half nouns, half verbs) and imageability, which was calculated via the MRC Psycholinguistics Database (Coltheart, 1981; see Appendix A: Table A2). *Imageability* is defined as "the ease with which a word gives rise to a sensory mental image" (Bird, Franklin, & Howard, 2001, p. 73), and details about how these were calculated can be found in Paivio, Yuille, and Madigan (1968). The pseudo-words were matched across the learning approach of interest (context vs. isolation) but not across learning and test, fluency, and comprehension phases. Rather, all words that were

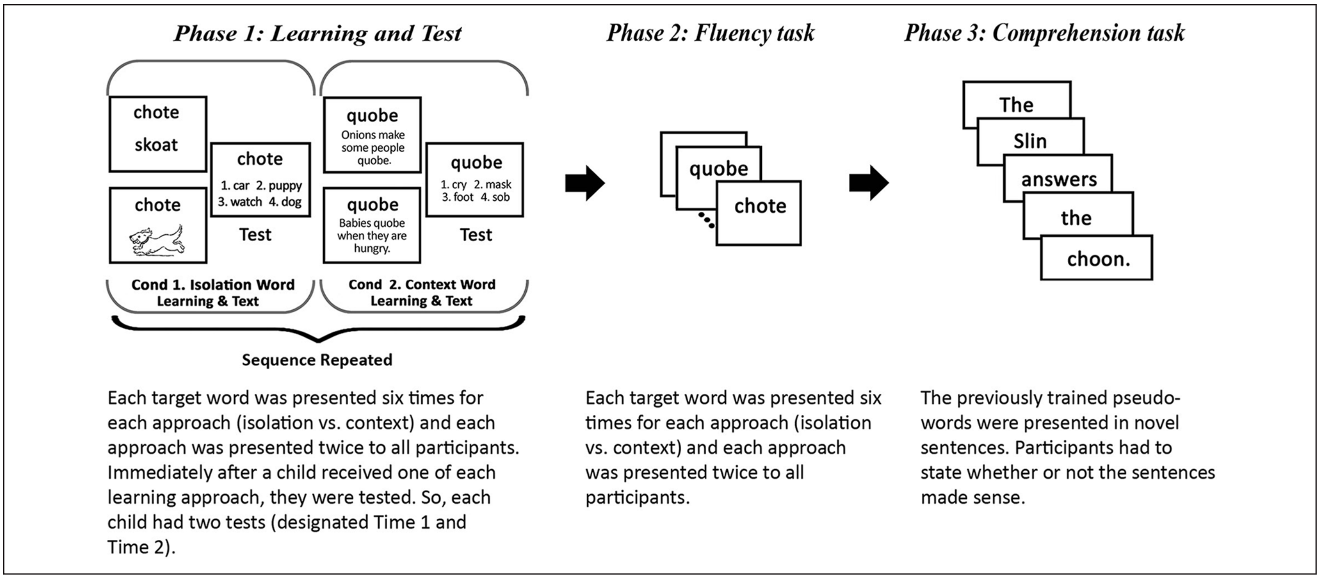

Figure 1. Sequence and description of the word-learning paradigm. All testing took place over the course of one day.

presented in the learning and test phases were presented again in the fluency and comprehension phases.

An overview of the sequence is presented visually in Figure 1; further details can be read in Clements-Stephens et al. (2012), who used the same stimuli and paradigm in adult readers. Note that Clements-Stephens et al. (2012) additionally reported neuroimaging findings; although neuroimaging data were collected, findings are not reported here.

Phase 1: Word learning and test. Forty monosyllabic pseudo-words that followed orthographic and phonological rules of English were split into *two learning approaches: isolation* (words presented on a computer screen with alternating trials of either a rhyming word or a meaningful picture) and *context* (words presented in one of two meaningful sentences). In the isolation approach, participants were shown either a rhyming pseudo-word or a picture that represented the meaning of the word, situated directly below the pseudo-word that was to be learned (see Figure 1). Before the learning sessions started, children were told they would see new words and that their task was to try to learn them. During the individual trials, children were instructed to state both pseudo-words ("chote," "skoat") on the rhyming trials, or the target pseudo-word and the picture ("chote," "dog") on the meaningful picture trials. If either was incorrect, the participant was immediately corrected. The child thus heard only his or her own vocalization of the word, unless it was mispronounced, in which case the experimenter pronounced it once for the participant (e.g., if the child read the word "chote" incorrectly, the experimenter would state the correct version and then let the child continue).

In the context approach, the target word was presented at the top of the computer screen as well as embedded in one of two meaningful sentences (i.e., the sentence defined the word), and children were instructed to read the target word and the sentence out loud. They also were instructed to state the meaning of the target word and were corrected if either their pronunciation or the meaning was incorrect. Each target word was presented six times for each approach, and each approach was presented twice to all participants. Words were presented randomly within each learning approach, and the order was counterbalanced across participants.

Immediately after a child received one of each learning approach (isolation and context), he or she was *tested.* Since the sequence was repeated, each child had *two tests (Designated Time 1 and Designated Time 2).* At test, each word was presented with four options below. For example, if they were trained to learn that "chote" meant "dog," the four options would include "dog" (correct), "puppy" (semantically related), "car" (previously trained meaning from the same set), and "watch" (randomly selected word not used in the experiment). Participants were instructed to read the word unassisted, read the answer choices, and pick the correct choice. If they chose the wrong answer (but not if they mispronounced the word or answer choices), the trial was marked as incorrect. This task assessed how well the children learned the meaning of the words presented in the two learning approaches.

Phase 2: Fluency task. In the fluency task, the previously trained pseudo-words were presented one at a time on a computer screen, and children were instructed to read them out loud as fast as possible. Accuracy, as well as reaction time, was recorded. Each of the 40 words was randomly presented on a computer screen 20 times (see Figure 1). This task assessed how well children learned the pronunciations of the words (phonological-orthographic word knowledge).

Phase 3: Comprehension task. The previously trained pseudo-words were presented in novel sentences. Some of the sentences contained two of the trained pseudo-words (from the same learning approach), and some sentences contained untrained pseudo-words that the child had never encountered in the learning paradigm. Participants had to state whether the sentences made sense (see Figure 1). This task required that the children had learned the original meaning of the target word, had retained it, and could comprehend it in a novel context. In other words, the task assessed knowledge of semantic learning.

Data Analysis

Data were formatted in SPSS Version 24 (IBM Corporation, 2015). Missing values were replaced with the group mean, and outliers were replaced with a 3.0 standard deviation cutoff by using the Improve Data Quality feature under the Transform function. The data were then reformatted so that repeated measures (time and/or learning approach, depending on the task) were nested within participant. As in Clements-Stephens et al. (2012), individual linear mixed-model regressions were performed to predict performance on each task since the data across learning paradigm tasks were noncommensurate. Therefore, mixed-effects models were assessed for each task individually using the lme4 package in *R* (Bates, Maechler, Bolker, & Walker, 2015). For all models, the method of estimation was maximum likelihood. Across all task analyses, participants were specified as a random intercept and served to control for associated intraclass correlation (Pinheiro & Bates, 2000). For only Model 1 of the fluency task we included time bins as a random slope (see Fluency Task section for time bin description). We employed a forward-fitting procedure to determine the model of best fit using likelihood ratio tests.

For all task analyses, age and PIQ were added to the models described below as fixed effects to see if they improved fit. If inclusion of age or PIQ improved model fit, follow-up mediation analyses were conducted.

Word learning and test task. In the test of word learning, the dependent variable was percentage correct, and the fixed effects were reader group (DYS, TD, or S-RCD), learning approach (isolation or context), and time (Test 1 or Test 2; see Figure 1).

Fluency task. Two linear mixed models were used to analyze the fluency data. In the first model, the dependent variable was mean reaction time (MRT; in milliseconds [ms]). This was the only best fitting model that included a random slope. Therefore, the random effect included in this model was time bin (random slope) by participant (random intercept). The fixed effects were reader group (DYS, TD, or S-RCD), learning approach (isolation or context), and bin number (the 20 trials were grouped into four time bins). Time Bin 1 represented Trials 1 through 5, Time Bin 2 represented Trials 6 through 10, Time Bin 3 represented Trials 11 through 15, and Time Bin 4 represented Trials 16 through 20. In the second model, the dependent variable was fluency MRT (which was transformed to normalize the distribution), and number of word-reading errors was entered as a predictor. Time bin number was removed because analyses were conducted only on Time Bin 1 (Trials 1–5). In addition to group and learning approach, number of fluency errors—as well as the interactions between fluency errors, reader group, and learning approach—was entered as fixed effects.

Comprehension task. Two linear mixed models were used to analyze the sentences task (which tested knowledge of word meaning [semantic knowledge]). In the first model, the dependent variable was percentage correct, with learning approach (isolation or context) and reader group (DYS, TD, or S-RCD) as fixed effects. In the second model, the dependent variable was reaction time (ms).

Results

Word-Learning Paradigm

Task 1: Learning at test. When predicting the percentage correct of word learning at test all three primary independent variables (reader group, learning approach, and test time) were significant. For the effect of time, $F(1, 135) = 64.048$, $p < .001$, percentage correct at Time 2 was higher than at Time 1: $b = .098$, $SE = .038$, $t(1, 135) = 2.590$, $p < .01$, suggesting children improved across time. For reader group, $F(2, 45) = 6.952$, $p < .01$, TD performed better than both S-RCD: $b = .140$, $SE = .056$, $t(2, 91.46) = 2.499$, $p = .014$, and DYS: $b = .126$, $SE = .054$, $t(2, 91.46) = 2.353$, $p = .021$, with no significant difference between DYS and S-RCD ($p = .813$). This suggests that both the DYS and S-RCD groups did not learn as well as the children in the TD group. For learning approach, $F(1, 135) = 10.828$, $p < .01$, children performed better on words learned in context than in isolation: $b = .070$, $SE = .041$, $t(1, 135) = 1.711$, $p = .089$ (see Table 2 for all ANOVA results and Appendix B: Table B1 for all means). There were no significant two-way interactions among any of the main effects, nor was there a significant three-way interaction between time, learning approach, and reader group. In addition, including age as a fixed effect did not result in a significant improvement in model fit ($p = .696$). This confirmed that learning at test results were not the result of age differences. We also investigated the impact of PIQ on learning. The inclusion of PIQ did improve the model

Table 2. Analysis of Variance Results for Learning Paradigm Tasks as a Function of Reader Group, Time, and Learning Approach.

	ANOVA *F*							
Learning Paradigm Task	Reader Group	Time	Learning Approach	G × T	G × L	T × L	G × T × L	G × E × L
Test (% correct)	6.952**	64.048***	10.828***	0.562	0.427	1.354	0.660	—
Fluency MRT	3.954*	2.770*	0.868	1.230	2.197	1.179	1.920	—
Fluency MRT (with errors)	3.103	—	0.117	—	2.222	—	—	3.694*
Sentences (% correct)	3.165*	—	10.288***	—	1.217	—	—	—
Sentences MRT	4.867*	—	5.142*	—	1.183	—	—	—

Note. Learning paradigm tasks are test, fluency, and sentences. G = reader group; TD = typically developing readers; DYS = dyslexia; S-RCD = specific reading comprehension deficit; T = time (if several time points were included); L = learning approach (isolation or context); E = fluency errors; MRT = mean reaction time.
*$p < .05$. **$p < .01$. ***$p < .001$.

fit ($p = .003$); however, when PIQ was added to the previous model, there was no longer an effect of reader group on learning. This suggested that PIQ might mediate the effect of reader group on learning at test (see Appendix A: Mediation Analyses and Appendix B: Figure B1).

Task 2: Fluency. Two linear mixed models were used to analyze the fluency data. The first model entered MRT as the dependent variable and reader group, learning approach, and time bin number as independent variables. There was a significant effect of time bin number, $F(3, 171) = 2.77$, $p = .024$, with MRT significantly slower for Time Bin 1 as compared to Time Bin 4: $b = −2.696$, $SE = .812$, $t(3, 132.970) = -3.320$, $p = .001$. There were no significant differences between (a) Time Bin 1 and Time Bin 2: $b = 5.703$, $SE = .726$, $t(3, 174.24) = –.785$, $p = .433$; (b) Time Bin 1 and Time Bin 3: $b = –.992$, $SE = .792$, $t(3, 141.060) = −1.254$, $p = .212$; or (c) Time Bin 2 and Time Bin 3: $b = –.422$, $SE = .680$, $t(3, 202.080) = -.620$, $p = .536$. There was a significant difference between Time Bin 2 and Time Bin 4: $b = −2.13$, $SE = .740$, $t(3, 166.180) = −2.871$, $p = .005$, and between Time Bin 3 and Time Bin 4: $b = −1.703$, $SE = .671$, $t(3, 207.350) = −2.537$, $p = .012$. In each case, word reading performance was slower at the earlier time point (see Appendix B: Table B2 for all means across time bins).

There was also an effect of reader group, $F(2, 55) = 3.954$, $p = .043$. The DYS group was slower than the S-RCD group: $b = −2.78$, $SE = 1.23$, $t(2, 72.56) = −2.260$, $p = .03$, and the TD group: $b = −2.82$, $SE = 1.10$, $t(2, 72.56) = −2.561$, $p = .01$, but there was no statistically significant difference of fluency MRT between the TD and S-RCD groups: $b = –.041$, $SE = 1.15$, $t(2, 72.56) = –.036$, $p = .971$. There was no significant effect of learning approach ($p = .353$). There were no significant two-way interactions among any of the independent variables, nor was there a significant three-way interaction among time, learning approach, and reader group (see Table 2 for all ANOVA results and Appendix B: Table B1 for all means).

In the second model, the number of errors was included as a predictor variable. There was a significant effect of number of errors, $F(1, 55.757) = 4.931$, $p = .03$, and there was also a significant three-way interaction of reader group, learning approach, and fluency errors, $F(2, 39.485) = 3.694$, $p = .03$ (see Appendix B: Figure B2).

Follow-up analyses were run to further determine what drove the three-way interaction. A significant difference in total errors was found across groups, $F(2, 45) = 23.312$, $p < .001$; the DYS group made more errors than the S-RCD group ($p = .049$) and the TD group ($p < .001$). The difference between the TD and S-RCD groups was not significant ($p = .121$; see Appendix B: Table B1). The difference in number of errors made by learning approach was significant for the DYS group, $F(1, 14) = 7.143$, $p = .008$, with fewer errors made on words learned in isolation compared to context. This was true only for the DYS group: Errors were not significantly different across learning approach for the S-RCD, $F(1,12) = .818$, $p = .366$, or TD, $F(1, 19) = 2.882$, $p = .090$, groups.

Task 3: Comprehension. In the first model, the dependent variable was percentage correct, and the independent variables were learning approach (isolation or context) and reader group (DYS, TD, or S-RCD). There were significant effects for reader group, $F(2, 45) = 3.615$, $p = .035$, and learning approach, $F(1, 45) = 10.288$, $p = .003$, but no interaction (see Table 2 for all ANOVA results). Post hoc tests of reader group showed that the TD group performed significantly better than the S-RCD group: $b = .112$, $SE = .054$, $t(2, 76.49) = 2.088$, $p = .040$, but that there was no difference in performance between the TD and DYS groups ($p = .199$) or S-RCD and DYS groups ($p = .428$). Beta coefficients indicated that group members performed better for words learned in context than in isolation: $b = .115$, $SE = .045$, $t(1, 45) = 2.540$, $p = .015$ (see Appendix B: Table B1 for means).

Since there were only two options for this task ("made sense" vs. "did not make sense"), it is possible that results could have been due to chance effects. However, we believe this is unlikely because performance was above chance for all groups in Phase 1 learning and was also above chance for performance on the comprehension task (see Appendix B: Table B1). Moreover, the participants were significantly faster in making judgments about meaning of words they were actually trained on versus untrained pseudo-words (see "Additional Comprehension Results" in Appendix B for details).

In the second model, the dependent variable was reaction time (ms), and the independent variables were learning approach (isolation or context) and reader group (DYS, TD, or S-RCD). There were significant effects of reader group, $F(2, 45) = 4.687$, $p = .014$, and learning approach $F(1, 45) = 5.142$, $p = .028$, but no interaction (see Table 2 for all ANOVA results).

Post hoc tests of reader group showed that the TD group performed better than the DYS group: $b = -125.026$, $SE = 39.180$, $t(3, 48.270) = -3.191$, $p = .002$, but there was no difference between the TD and S-RCD groups ($p = .136$) or the DYS and S-RCD groups ($p = .157$). Beta coefficients indicated that participants were faster for words learned in context rather than in isolation (see Appendix B: Table B1 for means).

The inclusion of age and PIQ as covariates did not result in a significant change in model improvement for the two fluency models ($p = .098$, $p = .528$; $p = .831$, $p = .072$) or for the two comprehension task models ($p = .986$, $p = .458$; $p = .167$, $p = .481$, age and PIQ, respectively), confirming that our effects are not due to confounding variables.

Discussion

The results of this study confirmed our hypothesis that task demands, learning approach, and reader subtype shape learning of new words. Across all outcome measures, typically developing children performed better than their peers with reading disabilities, but performance within the two groups with reading disabilities differed by whether the outcome measure emphasized phonological-orthographic connections or semantics. This suggests that existing impairments are not static but rather influence how children with disabilities actively acquire lexical information.

Performance Across Reader Groups and Task: Unique Learning Profile for the S-RCD Group

These results provide evidence that children with S-RCD show a unique learning profile compared to their TD peers and their peers with combined deficits in both decoding and comprehension (DYS). Specifically, this study demonstrated that children with S-RCD, like children with DYS, are characterized by generally poor performance on a word-learning paradigm. However, unlike children with DYS, they exhibit similar reaction times and accuracy on fluency trials as compared to children in the TD group, suggesting that their ability to learn the phonological-orthographic components of words is intact. In contrast, however, and similar to children with DYS, children with S-RCD were less accurate on the sentence comprehension task, which required applying semantic attributes of words. This unique pattern of results across outcome measures from different word-learning approaches mirrors static deficits described in the literature, and it contributes to the growing literature showing that these impairments exist—and perhaps contribute to—ongoing word-knowledge acquisition.

While Nation et al. (2007) did not include a fluency task in their paradigm, they did test knowledge of both phonological and semantic attributes of the words. Our results are consistent with their findings: Children with S-RCD have intact phonological abilities but exhibit a selective impairment in the semantic system.

Performance Across Reader Groups, Task, and Learning Approach

Test. At test, the TD group performed better than either of the groups with reading impairments. While the model failed to show a significant interaction of reader group by learning approach, performance was better for the S-RCD and DYS groups in the context learning approach, which is consistent with results from prior work (e.g., Nation & Snowling, 1998).

Fluency task and comprehension (sentences) task. Unlike the test task, there was a significant three-way interaction among errors, reader group, and learning approach in the fluency task. When the interaction was explored further, we found that the effect of learning approach was different across reader groups. Specifically, children in the DYS group made fewer word reading errors on words that were learned in isolation, while the differences in errors by learning approach were not significant for the S-RCD or TD groups. This suggests that when children with phonological-orthographic impairments (DYS) will be tested on oral word reading fluency, learning words in isolation will enhance performance. This is consistent with results of previous research showing that isolation is beneficial for learning phonological-orthographic lexical information (Ehri & Roberts, 1979; Martin-Chang & Levesque, 2015). Martin-Chang and Levesque (2015) have posited that the division of labor model (Harm & Seidenberg, 2004) provides a theoretical framework for understanding why training words in isolation might benefit phonological-orthographic learning. Specifically, this framework suggests that cognitive resources are differentially recruited, depending on the nature of the task. Perhaps, in our case, the lack of distractor word-forms in the isolation condition meant that cognitive resources could be fully allocated to learning the phonological-orthographic form, thus enabling better learning.

The same pattern was not seen in the sentences task, which recruited knowledge of semantic features of words. For all reader groups (TD, S-RCD, and DYS), the context learning approach led to higher accuracy and faster reaction time. These results are consistent with those of previous research showing that contextual information is beneficial for learning the meaning of words (Ehri & Roberts, 1979; Martin-Chang & Levesque, 2015). This suggests that regardless of type of reading impairment, context facilitates learning word meanings (semantic knowledge).

Study Limitations

This study had several limitations. First, as others have noted (e.g., Nation & Snowling, 1998) comparisons between children with good and poor reading abilities often confound decoding and comprehension. In an attempt to separate these two constructs, we created groups of readers based on standardized tests of reading using cut-off scores. While other researchers also have employed this method (e.g., Nation et al., 2007), an alternative approach would be to use continuous measures (Genard, Mousty, Alegria, Leybaert, & Morais, 1998). However, the continuous measures approach addresses slightly different research questions than those posed in this study.

Second, this study utilized a relatively small sample size per reader group. Future studies could oversample children with specific LD to increase the size of the groups.

Another limitation of our word-learning paradigm was the fact that children saw the word to be learned twice as many times in the context condition as in the isolation condition. Specifically, in the context condition the target word was presented at the top of each individual stimuli screen as well as in an embedded sentence. This greater exposure of the orthographic form of the pseudo-word could have led to increased learning in the context condition versus the isolation condition. However, this boost, if real, seems to have influenced only the DYS and S-RCD groups because for the percentage correct at Time 2 at test, the TD group exhibited the same percentage correct (.82) across conditions (see Appendix B: Table B1). This could be an interesting area of future investigation.

Conclusion

Understanding the factors that contribute to skilled reading is imperative for improving individual child reading outcomes. Word knowledge (lexical information) is a key component of fluent reading with comprehension. These results show that children with S-RCD have a unique word-learning profile, with impairment unique to tasks that rely on semantic processing. S-RCD is not a “static” impairment; children with this profile have deficits in learning new lexical-semantic information in addition to known deficits in processing of lexical-semantic information. Furthermore, our results show that depending on the task goals, learning approach interacts with reader subtype. Learning words in isolation is more beneficial for children with decoding deficits when being tested on oral reading, while learning words in context is more beneficial for learning semantic information across all reader groups. This has practical implications for the design of educational curricula and can help inform what works, for whom, and when.

Appendix A: Additional Method Information

Table A1. List of Pseudo-Words Used in the Word-Learning Paradigm.

List A				List B			
Isolation		Context		Isolation		Context	
Barp	girl	brip	key	vint	Girl	swoy	key
Quobe	cry	chote	dog	glant	Cry	meest	dog
Choon	phone	glant	climb	plute	Phone	quobe	climb
Meest	cat	nade	monkey	chote	Cat	slin	monkey
Fope	write	heke	read	jate	Write	zale	read
Gleeb	fly	kay	bear	troil	Fly	nur	bear
hain	cut	jate	ski	maip	Cut	fope	ski
jord	wash	grite	swim	lorp	Wash	preet	swim
krile	sleep	dirn	kick	smope	Sleep	teef	kick
lut	game	sneal	hand	fim	Game	bufty	hand
nisk	bird	lorp	drive	gilk	Bird	jord	drive
preet	throw	fim	car	grite	Throw	lut	car
bufty	foot	plute	apple	sneal	Foot	choon	apple
slin	hair	reen	duck	nade	Hair	trun	duck
swoy	book	maip	run	brip	Book	hain	run
teef	eat	smope	cheer	dirn	Eat	krile	cheer
trun	mouse	troil	dance	reen	Mouse	gleeb	dance
nur	cow	gilk	tree	kay	Cow	nisk	tree
whum	sing	yab	build	yab	Sing	wum	build
zale	drink	vint	boy	heke	Drink	barp	boy

Note. All pseudo-word stimuli were created by changing one letter of a real monosyllabic word not used elsewhere in the study. The new definition of the pseudo-word was matched across isolation and context on length, part of speech (half nouns, half verbs), and imageability, calculated via the MRC Psycholinguistics Database (Coltheart, 1981).

Table A2. Means and Standard Deviations for Word Length and Imageability.

Word Type	Word Length	Imageability
Set 1 trained pseudo-words	4.0 ± 0.8	561.4 ± 52.6
Set 2 trained pseudo-words	4.1 ± 0.9	553.5 ± 131.7
High-frequency words[a]	4.2 ± 0.7	463.2 ± 123.3
Low-frequency words	4.2 ± 0.7	523.8 ± 82.8
Untrained pseudo-words	4.2 ± 0.7	

[a]High-frequency words are significantly different from trained pseudo-words. The new definition of the pseudo-word was matched across isolation and context on length, part of speech (half nouns, half verbs), and imageability, calculated via the MRC Psycholinguistics Database (Coltheart, 1981).

Standardized Assessment Descriptions

The *Gates-MacGinitie Reading Test* (GMRT-4) is frequently used to assess reading comprehension ability in children (Cain & Oakhill, 1999; MacGinitie et al., 2008). Participants completed the appropriate level (4–7) of GMRT-4 based on their current grade level. Levels 4 through 7 consist of a vocabulary test and a comprehension test; however, only the comprehension test was administered as part of this study. The comprehension test consists of 48 questions pertaining to 11 passages that cover various subjects and have different lengths and prose types. Students are tested on their ability to understand both explicit and implicit information in the passages. The GMRT-4 has high reliability scores for internal consistency, alternate forms, and total test stability, ranging from .88 to .90 (MacGinitie et al., 2008). Construct validity was established through high correlations with the third edition of GMRT-4, other reading tests, and the PSAT and SAT (MacGinitie et al., 2008).

The second edition of the *Test of Word Reading Efficiency* (TOWRE-2; Torgesen Wagner, & Rashotte, 1999) was selected to evaluate participants' single-word-reading and pseudo-word-reading fluency and has high psychometric attributes

(Torgesen, Wagner, & Rashotte, 2012). TOWRE-2 consists of two subtests: Sight Word Efficiency and Phonetic Decoding Efficiency, both of which are timed. The Sight Word Efficiency subtest assesses the number of real printed words that are read correctly in 45 s, while the Phonetic Decoding Efficiency subtest measures the number of decodable pseudo-words that can be read correctly in 45 s. Internal consistency, test-retest, and interrater reliability metrics were all high, ranging from .86 to .97 (Hayward, Stewart, Phillips, Norris, & Lovell, 2008). Content, criterion, and construct validity were also assessed, with the correlation coefficients well above the minimum standard of .80 (Hayward et al., 2008).

The Perceptual Reasoning Index of the *Wechsler Intelligence Scale for Children*, fourth edition (WISC-IV; Wechsler, 2004) is a composite score including performance on the following tasks: Block Design, Picture Concepts, Matrix Reasoning, and Picture Completion. These tasks provide a performance IQ (PIQ) that is not influenced by language abilities, which is important given that our sample had known language impairments. The WISC-IV Perceptual Reasoning Index has high internal consistency (.92) and test-retest reliability (.85). Evidence of validity was established through high correlations with other standardized tests of intelligence.

Mediation analyses. To address the potential confound of age and PIQ differences across reader groups, both covariates were entered into the models. There were no significant differences between groups on age, and the models with age entered as a covariate did not indicate better fit for any of the tasks. This indicates that age was not a potential confound. However, there were significant group differences on PIQ across reader groups (see Table 1 in the main article), and when entered as a covariate, the model predicting word learning at test was significantly improved, thus prompting a test of mediation of PIQ on word learning at test.

PIQ scores were transformed to *z* scores before Structural Equation Modeling Mediation analysis was used to determine if the addition of PIQ provided a link between our independent and dependent task variables. The model was fitted using the lavaan package (Rosseel, 2012) in *R* using the robust maximum likelihood estimation. Standard errors were calculated using bootstrapping procedures (see Preacher & Hayes, 2004, for bootstrapping as an alternative approach to Sobel's test). Finally, the Yuan-Bentler estimation scaling correction factor was included as a method for smaller samples (Bentler & Yuan, 1999).

Appendix B: Additional Results Information

Table B1. Means and Standard Deviations Across the Three Tasks.

Test Task							Fluency Task				Comprehension (Sentences) Task							
	Time 1: % Correct			Time 2: % Correct				Time Bin 1: Number Word Reading Errors				% Correct				MRT (ms)		
Group	Iso	Con	Total	Iso	Con	Total	Group	Iso	Con	Total	Group	Iso	Con	Total	Group	Iso	Con	Total
TD[a,b]	.66	.73	.69	.82	.82	.82	TD[a]	4.7	4.3	9.0	TD[a]	.63	.72	.69	TD[a]	770	769	770
	(.14)	(.17)	(.13)	(.19)	(.15)	(.16)		(3.7)	(4.1)	(8.5)		(.16)	(.13)	(.13)		(114)	(129)	(120)
S-RCD[a]	.52	.59	.55	.67	.71	.68	S-RCD[b]	11.8	12.4	24.0	S-RCD[a]	.52	.64	.58	S-RCD	833	808	820
	(.16)	(.14)	(.13)	(.19)	(.19)	(.19)		(8.9)	(9.9)	(20.0)		(.16)	(.18)	(.16)		(116)	(115)	(114)
DYS[b]	.53	.59	.56	.63	.69	.65	DYS[a,b]	37.3	42.2	79.0	DYS	.57	.59(	.58	DYS[a]	896	879	887
	(.15)	(.15)	(.10)	(.12)	(.12)	(.13)		(21.1)	(23.2)	(56.0)		(.12)	.15)	(.10)		(104)	(106)	(103)

Note. Standard deviations are in parentheses. MRT (ms) = mean reaction time (in milliseconds); Iso = isolation learning approach; Con = context learning approach; TD = typically developing readers; S-RCD = readers with specific reading comprehension deficits; DYS = readers with dyslexia.
[a]Row labels with the same superscript, within a given task, differ significantly at the $p < .05$ level.
[b]Row labels with the same superscript, within a given task, differ significantly at the $p < .05$ level.

Table B2. Means and Standard Deviations for Fluency Mean Reaction Times (in Milliseconds) During the Fluency Learning Paradigm Task.

	Time Bin 1[a,d]			Time Bin 2[b]			Time Bin 3[a,c]			Time Bin 4[b,c,d]		
Group	Iso	Con	Total	Iso	Con	Total	Iso	Con	Total	Iso	Con	Total
TD[e]	798 (182)	797 (176)	797 (178)	786 (154)	801 (156)	750 (241)	757 (138)	763 (133)	718 (218)	751 (141)	757 (153)	713 (226)
S-RCD[f]	794 (136)	789 (117)	791 (124)	779 (143)	794 (152)	787 (124)	787 (170)	775 (147)	780 (163)	762 (115)	744 (112)	750 (115)
DYS[e,f]	949 (178)	935 (187)	942 (180)	902 (156)	876 (141)	896 (149)	857 (168)	812 (145)	874 (196)	829 (150)	860 (213)	921 (300)

Note. Standard deviations are in parentheses. Iso = isolation condition; Con = context condition; TD = typically developing readers; S-RCD = readers with specific reading comprehension deficits; DYS = readers with dyslexia. Significance values were not examined by group and time because there were no significant two- or three-way interactions.
[a]Column headings (Time Bins 1–4) with the same superscript differ significantly at the $p < .05$ level.
[b]Column headings (Time Bins 1–4) with the same superscript differ significantly at the $p < .05$ level.
[c]Column headings (Time Bins 1–4) with the same superscript differ significantly at the $p < .05$ level.
[d]Column headings (Time Bins 1–4) with the same superscript differ significantly at the $p < .05$ level.
[e]Row labels with the same superscript differ significantly at the $p < .05$ level.
[f]Row labels with the same superscript differ significantly at the $p < .05$ level.

Task 1: Learning at Test

Structural equation mediation analysis was used to determine whether performance IQ (PIQ) was an underlying variable influencing the relationship between reader group and word learning at test. The latent dependent variable Word Learning at Test significantly predicted Word Learning ($p < .05$) across all four measures: Time 1 Isolation, Time 1 Context, Time 2 Isolation, and Time 2 Context. The specified model fit was good: $\chi^2_{(13)} = 23.62$, comparative fit index = 0.93, Tucker-Lewis Index = 0.90, and standardized root-mean square residual = 0.046. There was a significant negative effect of reader group on PIQ (Path a: $b = -1.02$, $SE = 0.26$, $z = 3.95$, $p = .0001$). Since children with dyslexia (DYS) and children with a specific reading comprehension deficit (S-RCD) were coded as greater than typically developing readers (TD), this indicates that readers in the TD group had significantly higher PIQs than children with DYS and S-RCD. There was a significant positive effect of PIQ on word learning at test (Path b: $b = 0.046$, $SE = 0.017$, $z = 2.75$, $p = .006$), indicating that higher PIQ results in greater word learning at test. Prior to the inclusion of PIQ there was a significant effect of reader group on word learning at test (Path c: $b = -0.103$, $SE = 0.033$, $z = 3.14$, $p = .002$); however, after the inclusion of PIQ the effect became only marginally significant (Path c-prime: $b = -0.057$, $SE = 0.032$, $z = 1.78$, $p = .075$). In addition, there was a significant indirect effect of reader group on word learning at test (Path AB: $b = -0.047$, $SE = 0.02$, $z = 2.26$, $p = .024$), suggesting that PIQ partially mediates the effect that reader group has on test performance (see Figure B1).

Additional Comprehension Results

As an additional gauge on learning, we examined whether children were significantly faster when making judgments about meaning of words they were actually trained on versus untrained pseudo-words. This was done for words learned in both the isolation and the context learning approaches; both were statistically significant: $t(1, 40) = -4.813$, $p < .001$ and $t(1, 40) = -3.112$, $p = .003$, for untrained pseudo-words versus words trained in isolation and context learning approaches, respectively. In other words, children's reaction time was fastest for sentences composed entirely of real words (that were not trained), followed by trained pseudo-words, and finally untrained pseudo-words (slowest). These findings thus confirm that children were responsive to the learning trials and were not treating all words as pseudo-words.

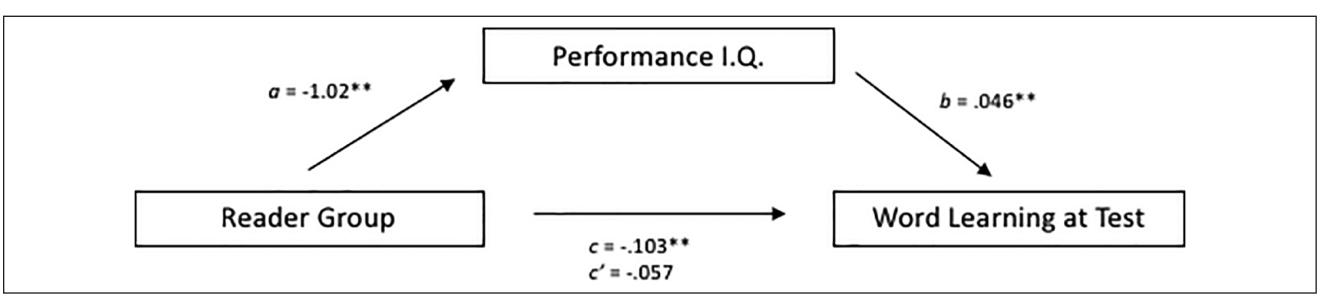

Figure B1. Structural Equation Modeling Mediation analysis was used to determine whether the addition of performance IQ provided a link between our independent and dependent task variable. The results showed that performance IQ partially mediates the relationship between reader group and word learning at test.

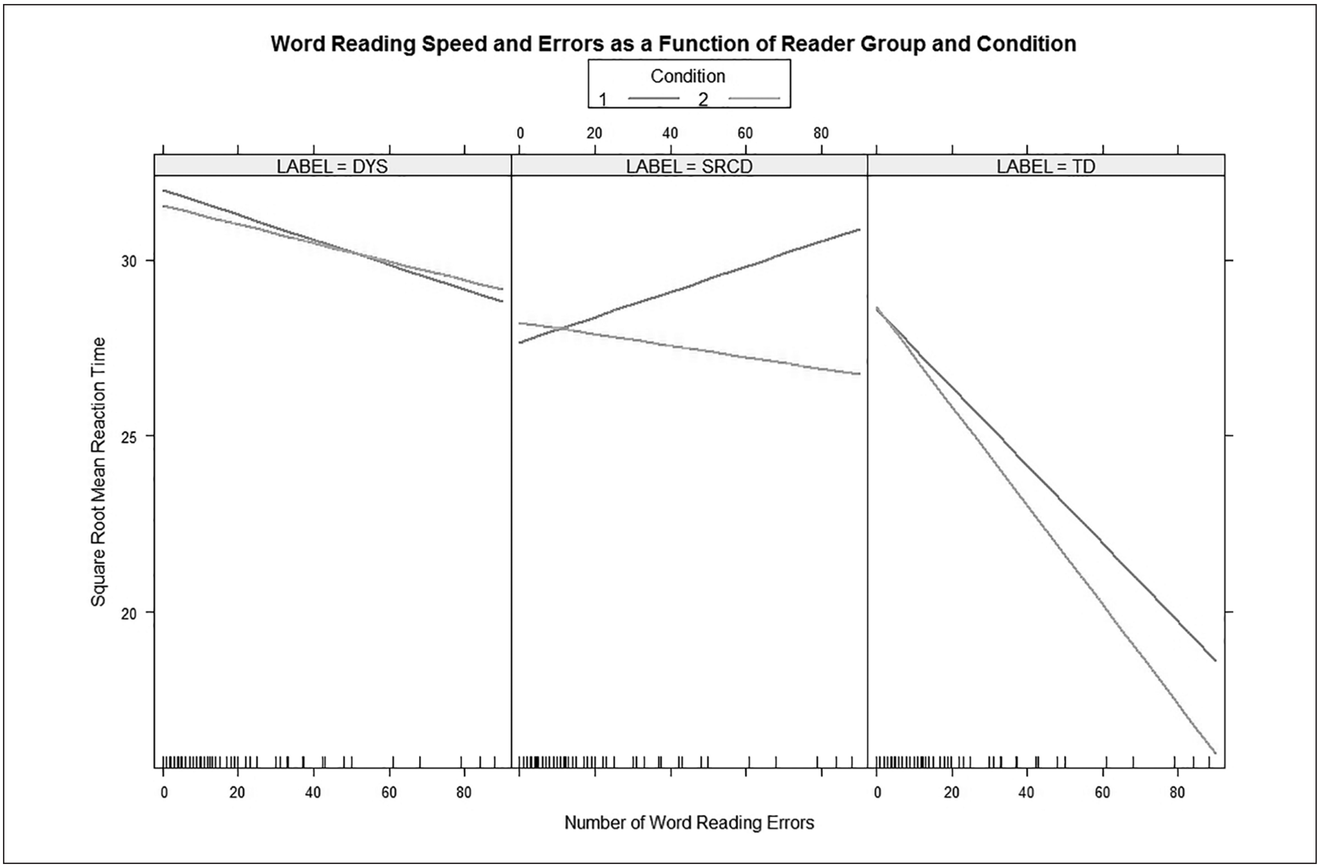

Figure B2. Predicted probabilities of mean reaction time on the fluency task as a function of reader group and condition. Number of errors are represented by the tick marks along the *x*-axes. Learning approach is represented by line saturation; the darker lines represent the isolation learning approach, whereas the lighter lines represent the context learning approach.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported in part by NIH grant R01 HD44073, the Learning Disabilities Research Center (NIH grant P50HD052121), Johns Hopkins School of Medicine General Clinical Research Center (NIH grant M01- RR00052), the F.M. Kirby Research Center (NIH/NCRR grant P41RR15241), and OSEP training grant H325D140073.

References

Bailey, C. E., Manis, F. R., Pedersen, W. C., & Seidenberg, M. S. (2004). Variation among developmental dyslexics: Evidence from a printed-word-learning task. *Journal of Experimental Child Psychology*, *87*, 125–154.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using "Eigen" and S4*. R package version 1.1-8. Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bentler, P. M., & Yuan, K. H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*, 181–197.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*, 73–79.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489–503.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683–696.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671–681.

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, *30*, 681–694.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MA: Cambridge University Press.

Castles, A., & Holmes, V. M. (1996). Subtypes of developmental dyslexia and lexical acquisition. *Australian Journal of Psychology*, *48*, 130–135.

Clements-Stephens, A. M., Materek, A. D., Eason, S. H., Scarborough, H. S., Pugh, K. R., Rimrodt, S., . . . Cutting, L. E. (2012). Neural circuitry associated with two different approaches to novel word learning. *Developmental Cognitive Neuroscience*, *2*, S99–S113.

Colenbrander, D., Kohnen, S., Smith-Lock, K., & Nickels, L. (2016). Individual differences in the vocabulary skills of children with poor reading comprehension. *Learning and Individual Differences*, *50*, 210–220.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33*, 497–505.

Ehri, L. C., & Roberts, K. T. (1979). Do beginners learn printed words better in contexts or in isolation? *Child Development*, *50*, 675–685.

Genard, N., Mousty, P., Alegria, J., Leybaert, J., & Morais, J. (1998). Methods to establish subtypes of developmental dyslexia. In P. Reitsma & L. Verhoeven (Eds.), *Problems and interventions in literacy development* (pp. 163–176). Dordrecht, the Netherlands: Springer.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.

Hayward, D. V., Stewart, G. E., Phillips, L. M., Norris, S. P., & Lovell, M. A. (2008). At-a-glance test review: Test of word reading efficiency (TOWRE). In *Language, phonological awareness, and reading test directory* (pp. 1–6). Edmonton, Alberta: Canadian Centre for Research on Literacy.

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and cognition*. Chichester, UK: John Wiley & Sons.

IBM Corporation. (2015). *IBM SPSS statistics for Macintosh, Version 23.0*. Armonk, NY: Author.

Kern, M. L., & Friedman, H. S. (2009). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology*, *30*, 419–430.

MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2000). *Gates-MacGinitie reading tests manual, Levels 4–6*. Itasca, IL: Riverside.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2008). *Gates-MacGinitie reading tests, fourth edition, bundled technical and supplemental report*. Rolling Meadows, IL: Riverside.

Martin-Chang, S., & Levesque, K. (2015). Reading words in and out of connected text: The impact of context on semantic and orthographic processing. *Scientific Studies of Reading*, *19*, 392–408.

Martin-Chang, S. L., & Levy, B. A. (2006). Word reading fluency: A transfer appropriate processing account of fluency transfer. *Reading and Writing*, *19*, 517–542.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah, NJ: Erlbaum.

Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 248–265). Oxford, UK: Blackwell.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, *69*, 996–1011.

Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech Language Pathology*, *9*, 131–139.

Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension: A handbook*. Abingdon-on-Thames, UK: Routledge.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York, NY: Springer-Verlag.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2007). Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading*, *11*, 165–197.

Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*, 235–257.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more: Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*, 1–36.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review Psychology*, *59*, 451–475.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *22*, 360–407.

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *TOWRE–2 test of word reading efficiency*. Austin, TX: Pro-Ed.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 examiner's manual*. Austin, TX: Pro-Ed.

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress. (2017). *National assessment of educational progress reading assessment*. Retrieved from https://www.nationsreportcard.gov/reading_2017/nation/scores?grade=4

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

Wechsler, D. (2004). *Wechsler intelligence scale for children* (4th ed.). London, UK: Pearson Assessment.