


# The Effects of Inference Instruction on the Reading Comprehension of English Learners With Reading Comprehension Difficulties

Remedial and Special Education  
2020, Vol. 41(5) 259–270  
© Hammill Institute on Disabilities 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0741932518824983  
rase.sagepub.com  
SAGE

Colby Hall, PhD<sup>1</sup> , Sharon Vaughn, PhD<sup>2</sup>,  
Marcia A. Barnes, PhD<sup>3</sup>, Alicia A. Stewart, MA<sup>2</sup>,  
Christy R. Austin, MEd<sup>2</sup>, and Greg Roberts, PhD<sup>2</sup>

## Abstract

Inference skill is one of the most important predictors of reading comprehension. Still, there is little rigorous research investigating the effects of inference instruction on reading comprehension. There is no research investigating the effects of inference instruction on reading comprehension for English learners with reading comprehension difficulties. The current study investigated the effects of small-group inference instruction on the inference generation and reading comprehension of sixth- and seventh-grade students who were below-average readers ( $M = 86.7$ ,  $SD = 8.1$ ). Seventy-seven percent of student participants were designated limited English proficient. Participants were randomly assigned to 24, 40-min sessions of the inference instruction intervention ( $n = 39$ ) or to business-as-usual English language arts instruction ( $n = 39$ ). Membership in the treatment condition statistically significantly predicted higher outcome score on the *Gates-MacGinitie Reading Test Reading Comprehension* subtest ( $d = 0.60$ , 95% confidence interval [CI] [0.16, 1.03]), but not on the other measures of inference skill.

## Keywords

reading comprehension, reading comprehension instruction, inferences, inference instruction, reading comprehension difficulties, English learners

Reading with comprehension involves building a coherent representation of a text in memory. Kintsch and van Dijk (1978) distinguished between three levels of text representation: the reader begins by accessing word meanings and syntactic knowledge (*surface-level representation*); next, the reader attends to information that is explicit in the text (*text-level representation*); finally, the reader retrieves general and topic-specific knowledge from memory and integrates this knowledge with information in text to create a more complete representation of the situation described (*situation model*). The coherence of the situation model reflects the degree to which appropriate, meaningful connections are established between (a) discrete pieces of information in text and (b) information in text and information in memory. To read with understanding, a student not only has to remember information in text but also has to generate inferences to discover implicit meanings.

Inference generation, then, is the process by which a reader integrates information within or across texts to create new understandings (Elleman, 2017; McNamara & Magliano, 2009). Researchers often distinguish between *text-connecting inferences* and *gap-filling inferences*.

*Text-connecting inferences*, sometimes called cohesive, bridging, close-to-the-text or inter-sentence inferences, rely on linguistic cues present in the text. Examples are anaphor resolution, lexical or “word-to-text integration” inference (e.g., Perfetti & Stafura, 2014; Yuill & Oakhill, 1988), and inference of word meanings from context clues (e.g., Cain, Oakhill, & Lemmon, 2004). *Gap-filling inferences*, sometimes called knowledge-based inferences, require the reader to go beyond the text and draw on background knowledge. Some researchers distinguish between gap-filling inferences that are necessary for maintaining text coherence and gap-filling inferences that are *not*

<sup>1</sup>The University of Texas Health Science Center at Houston, USA

<sup>2</sup>The University of Texas at Austin, USA

<sup>3</sup>Vanderbilt University, USA

## Corresponding Author:

Colby Hall, The Children’s Learning Institute, The University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 2441, Houston, TX 77030, USA.

Email: Colby.S.Hall@uth.tmc.edu

**Associate Editor:** Melissa Stormont

strictly necessary. Bowyer-Crane and Snowling (2005) provide the following example of a necessary gap-filling inference: “The campfire started to burn uncontrollably. Tom grabbed a bucket of water” (p. 192). To understand why Tom grabbed a bucket of water, it is necessary for the reader to activate the background knowledge that water puts out fire, and relate the second sentence to the first by generating the inference that Tom grabbed the bucket of water because he was trying to put out the fire.

Inference generation and other language processing skills may be more important for situation model construction as children progress from primary to secondary grades. Tighe and Schatschneider (2014) determined that while word reading fluency was the most influential predictor of reading comprehension for a diverse sample of third-graders, inferential reasoning had a greater influence on reading comprehension for students in Grades 7 and 10. In a sample of ninth-grade students, Cromley and Azevedo (2007) found that while vocabulary and background knowledge made the largest contributions to students’ comprehension of narrative and informational text, inference skill also predicted unique variance in students’ text comprehension; word reading made a smaller contribution, and the effects of reading strategy use were indirect, through inference. Ahmed et al. (2016) used multiple-indicator latent variables to measure the constructs in the Cromley and Azevedo (2007) direct and inferential mediation (DIME) model in a large sample of children ( $N = 1,196$ ) in Grades 7 through 12. When they controlled for measurement error and method bias, these component skills predicted virtually all of the systematic variance in reading comprehension, and inference making had the largest direct effect on reading comprehension. Longitudinal studies of reading comprehension further demonstrate the unique contribution of inference skill to growth in comprehension over time (e.g., Oakhill & Cain, 2012), and studies comparing skilled and less skilled adolescent comprehenders show reliable differences in inference-making groups even after taking into account reading and reading-related skills such as vocabulary and working memory (e.g., Barth, Barnes, Francis, York, & Vaughn, 2015; Denton et al., 2015).

## Reading Comprehension and English Learners (ELs)

Although all students with reading comprehension difficulties typically demonstrate difficulties with oral language comprehension (Spencer, Quinn, & Wagner, 2014), ELs with reading comprehension difficulties tend to have more pronounced oral language comprehension weaknesses than English-only (EO) students with reading comprehension difficulties (Spencer & Wagner, 2017). Cross-sectional studies conducted with monolingual and language-learning upper elementary students in Canada (Grant, Gottardo, &

Geva, 2011), the United Kingdom (Babayiğit, 2014), the Netherlands (e.g., Droop & Verhoeven, 2003), and the United States (Cho, Capin, Roberts, Roberts, & Vaughn, in press) suggest that vocabulary knowledge and other linguistic comprehension variables may make a greater contribution to reading comprehension for EL students than for EO students in the upper elementary and secondary grades. In the study conducted by Cho et al. (in press) with participants who were ELs with significant reading difficulties, linguistic comprehension variables made a greater contribution to reading comprehension than word reading variables, while the opposite pattern held true for EO students. These cross-sectional findings align with other research suggesting that specific reading comprehension difficulties may be more prevalent in populations of ELs than in populations of monolingual students (Lesaux, 2006; Lesaux & Kieffer, 2010; Nakamoto, Lindsey, & Manis, 2007; Spencer & Wagner, 2017).

Vocabulary and oral language comprehension instruction is thus a particularly important component of reading comprehension instruction for ELs with reading comprehension difficulties beyond the primary grades. In addition, because ELs with reading comprehension difficulties have reading comprehension weaknesses that are greater than their oral language weakness (Spencer & Wagner, 2017), it is especially important to address aspects of language comprehension that are unique to written text, and particularly to the complex, academic texts that students in Grades 4 and above encounter in their classrooms. Academic texts are more likely to contain unfamiliar topics, low-frequency vocabulary words, and syntactically complex sentences (Lee & Spratley, 2010). The comprehension of academic texts differs from the comprehension of oral language, relying more on (a) students’ background knowledge and ability to integrate background knowledge with information in text, as well as (b) students’ skill in connecting ideas within and across information-dense sentences (Wolfe & Woodwyk, 2010). All of these factors make the comprehension of textual language more cognitively and linguistically demanding than the comprehension of oral language, and explain the prevalence of “long-term English learners” (LTELs; Menken, Kley, & Chae, 2012; Olsen, 2010), who continue to be identified as lacking English proficiency after more than 6 years of education in United States schools. Olsen (2010) reported that 59% of all students identified as “ELs” in California secondary schools have been enrolled in U.S. schools for 6 years or more, and other reports published in New York and Texas estimate that similar numbers of secondary school students identified as lacking proficiency in English were educated in U.S. elementary schools (Capps & Fix, 2005; Olsen, 2014). Long-term ELs are often fluent in conversational English but lack academic language proficiency in both English and their L1. Olsen (2014) notes that “despite the fact that English tends to be

the language of preference for these students, the majority are ‘stuck’ at intermediate levels of English oral proficiency or below” (p. 5). Many ELs in the secondary grades thus need intensive, supplemental instructional supports to gain the academic discourse and text processing skills that they need to be successful in school.

There is a dearth of research on reading comprehension instructional interventions for ELs in the secondary grades (Hall et al., 2017). Richards-Tutor et al. (2016) meta-analyzed 12 studies published between 2000 and 2012 that evaluated the effects of reading interventions for ELs with or at risk for experiencing academic difficulties. While the seven included studies that were conducted in kindergarten and first grade produced statistically significant, positive effects for interventions targeting beginning reading skills (ES range, 0.58–0.91), there were mostly null effects for the three included interventions conducted with older participants (i.e., students in Grades 4 through 8). For these three studies, which investigated the effects of the *Reading Mastery*, *Corrective Reading*, and *Wilson Reading* programs, differences between groups were not statistically significant for 88%, or 23 out of 26, reading outcomes measured. Hall et al. (2017) meta-analyzed research on reading instruction across academic contexts for ELs in Grades 4 through 8 and determined that interventions for ELs at these grade levels yielded a negligible mean effect size of  $g = 0.01$  on standardized reading outcomes.

Rates of annual growth on measures of reading skill decreases as grade level increases (Bloom, Hill, Black, & Lipsey, 2008), suggesting that reading comprehension skill is less easily influenced in older students than it is in younger students. Because of the oral language difficulties reviewed above, reading comprehension may be particularly resistant to change for ELs in the secondary grades. Still, another explanation for these small or nonexistent effects in favor of reading interventions tested in previous research is that these reading comprehension interventions developed for older ELs have not adequately focused on the written language comprehension skills required when reading secondary-level academic texts.

### **Inference Instruction for Students With Reading Comprehension Difficulties**

Despite evidence that inference generation makes an important contribution to reading comprehension (e.g., Ahmed et al., 2016; Oakhill & Cain, 2012), relatively few studies have investigated the impact of explicit inference instruction on the inferential comprehension or general reading comprehension of either EL or EO struggling readers. Hall (2016) located only nine peer-reviewed studies published between the earliest indexed year of searched databases and 2013 that examined the effects of inference

instruction relative to a control or business-as-usual comparison condition on at least one reading comprehension or inference generation outcome for struggling readers in Grades 1 through 12. Among these studies, most reported comparatively positive effects of inference instruction on researcher-developed measures of inference skill (effect sizes for group design studies ranged from  $g = 0.72$  to  $g = 1.85$ ). Four studies reported positive effects in favor of treatment on norm-referenced, standardized measures of general reading comprehension (effect sizes ranged from  $g = -.03$  to  $g = 1.96$ ).

Elleman (2017) meta-analyzed 25 studies of inference instruction interventions for both struggling and proficient readers and found that inference instruction was beneficial for students’ general comprehension ( $d = 0.58$ ) and inferential comprehension ( $d = 0.68$ ). The overall effect on inference measures for less-skilled readers was larger ( $d = 0.80$ ) than for skilled readers ( $d = 0.55$ ). Again, most measures employed in included studies were researcher-developed and closely aligned with the instruction provided in studies; relatively few effects ( $k = 7$ ) across only five studies were derived from norm-referenced, standardized measures (overall effect,  $d = 0.53$ ). On average, inference instruction was effective in improving less-skilled readers’ literal comprehension of text, as well as their inferential comprehension. Studies showed positive results in relatively short periods of time (i.e., less than 10 hr). Students who received inference instruction in a small group of 10 students or fewer benefited more than students who received instruction in a larger group.

Since the publication of these reviews, Reed and Lynn (2016) determined that middle grade students with learning disabilities who received explicit inference instruction significantly improved their pre- to posttest performance on a multiple-choice test of reading comprehension. Barth and Elleman (2017) reported that middle grades struggling readers who received an intervention designed to simultaneously build content knowledge and teach multiple inference generation strategies made significant gains relative to a business-as-usual comparison group on a proximal measure of content knowledge and on a standardized measure of general reading comprehension. Denton et al. (2017) found that ninth-graders with reading comprehension difficulties randomly assigned to a nine-session inference instruction intervention that included multisyllable word study, explicit instruction in inference generation, and guided practice in thinking aloud about text made moderate but nonsignificant gains relative to students in a business-as-usual comparison condition on proximal measures of inference skill. There has been no study that has investigated the effects of an inference instruction intervention on the inferential and general reading comprehension of students with reading comprehension difficulties who are ELs.

**Table 1.** Participant Demographic Information.

Group	Ethnicity: Hispanic	FRPL	LEP	SED	Age in years
Control ( $n = 39$ )	38 (97.4%)	38 (97.4%)	28 (71.8%)	6 (15.4%)	12.603
Treatment ( $n = 39$ )	37 (94.9%)	38 (97.4%)	32 (82.1%)	3 (7.7%)	12.622
Total ( $N = 78$ )	75 (96.2%)	76 (97.4%)	60 (76.9%)	9 (11.5%)	12.612

Note. FRPL = free and reduced-price lunch status; LEP = with the designation limited English proficient currently or during the prior school year; SED = receiving special education services.

## Study Aims

The primary purpose of this study was to expand the research base on inference instruction for middle grades students with below-average reading comprehension by determining the relative effects of a small group, 14-week inference instruction intervention and a business-as-usual comparison condition on measures of inference generation and reading comprehension. Research questions were as follows:

**Research Question 1:** What is the effect of a small-group inference instruction intervention, compared with typical practice reading instruction, on the inference generation skill of middle grades students with below-average reading comprehension?

**Research Question 2:** What is the effect of a small-group inference instruction intervention, compared with typical practice reading instruction, on the general reading comprehension of middle grades students with below-average reading comprehension?

## Method

### Participants

**Students.** The study was conducted in an urban public charter school in Central Texas serving students in Grades 6 through 9, the vast majority of whom were Latino and grew up speaking Spanish as a first language. This school is similar to other “no-excuses” charter schools in Central Texas and throughout the United States: It articulates rigorous behavioral and academic expectations, enforces a strict disciplinary code, and delivers instruction during an extended school day and year. The school holds a random lottery to ensure that any student applicant has an equal opportunity to gain an offer of admission. Any sixth- or seventh-grade student at the study school who received a standard score of 97 or below on the *Gates-MacGinitie Reading Test* Reading Comprehension subtest (GMRT-RC; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000) was eligible for inclusion. Of the 109 students who returned consent forms providing permission to participate in the study, a total of 84 students met the screening criterion and were randomly assigned within grade to treatment

( $n = 43$ ) and comparison ( $n = 41$ ) conditions. After attrition, the sample comprised 78 students ( $n = 39$  in treatment,  $n = 39$  in comparison), which included twice as many seventh-graders ( $n = 46$ ) as sixth-graders ( $n = 16$ ). Students’ mean standard score ( $M = 86.7$ ,  $SD = 8.1$ ) at screening on the GMRT-RC was at the 19th percentile; scores ranged from the 1st percentile to the 42nd percentile. The demographic characteristics of the sample reflected the demographics of the school as a whole: Of the 78 participants in the final sample, 96.2% were Hispanic. Among all participants, 76.9% had the designation limited English proficient (LEP) currently or were within 2 years of exiting LEP status. Students were identified as LEP if they had a score indicating limited English proficiency as defined by the Texas Education Agency (TEA) on a test that was approved by the TEA. These tests included the *IDEA Proficiency Test—Oral English II* (IPT II; Ballard & Tighe, 2012), *Language Assessment Scales Links* (LAS Links; CTB/McGraw-Hill/Data Recognition Corporation, 2012), *Stanford English Language Proficiency Test* (SELP 2; Pearson, 2012), *Test of English Language Learning* (TELL; Pearson, 2015), and the *Woodcock-Muñoz Language Survey—Revised* (WMLS-R; Riverside Publishing/Houghton Mifflin Harcourt, 2010). Limited English proficiency was defined as a score (a) below Advanced on the IPT II; (b) below Level 4 on the LAS Links; (c) below Level 5 on the SELP 2; (d) below certain scale scores on TELL Listening or Speaking subtests (468 or 469, respectively, for Grade 6 students; 473 or 475, respectively, for Grade 7 students); or (e) at or below Emerging Proficiency on two or more WMLS-R Listening and Speaking subtests. After students were exited from LEP status, the study school continued to monitor students’ language acquisition progress for 2 years. If students scored in the “proficient” range on TEA-approved assessments for 2 years in a row, they no longer received monitoring. The vast majority of students in the final sample (97.4%) received free or reduced-price lunch; 11.5% received special education services. There were no significant differences between treatment and comparison conditions on demographic variables including ethnicity, LEP status, free and reduced-price lunch status, special education status, or age. Table 1 represents demographic characteristics of students in the treatment and comparison conditions.

**Tutors.** Students received reading instruction from the lead author (then a doctoral candidate with 5 years of experience teaching students in the upper elementary grades) and two other tutors (both doctoral students, with 5 and 7 years of teaching experience respectively). The author, who developed intervention materials and protocols, provided 4 hr of training to the other tutors prior to the start of the study.

### Treatment Condition

Inference instruction was delivered to small groups of three to six students during 40-min sessions, 2 to 3 times per week for a total of 24, 40-min sessions per student. The text used in the study was the novel *Wonder* (Palacio, 2012); it has a Lexile® level of 790L. At the beginning of each of the first 10 intervention sessions, students received explicit instruction in generating a specific type of inference. Students were taught to notice gaps and/or lack of coherence in text, to identify clue words or phrases, and to integrate information from knowledge with information in text. Simple graphic organizers scaffolded the process of knowledge-based inference generation, making visible the integration of information in text with information in background knowledge (Elbro & Buch-Iversen, 2013). The tutor modeled generating a particular type of inference while reading a passage from *Wonder* and then engaged students in guided practice using the same passage or a subsequent passage. Finally, the tutor directed students to continue reading *Wonder* independently with a partner or in a small group. For the first 10 sessions, students read aloud throughout the session; during the remaining 14 sessions, students were instructed to read at least every other page silently.

Each student's book was prepared with stopping points marked with sticky tabs. Tutors explained how students should stop at each sticky tab and refer to the next inference question. Stopping points were chosen deliberately; they were places where the text lacked coherence or where generating an inference would furnish a more complete situation model. The tutor taught and modeled for students how to discuss and find text evidence in support of potential answers to each inference question before choosing a final answer. Initially, student partnerships or small groups received feedback after they answered questions via a scratch-off answer sheet: If a partnership chose and scratched off the correct answer, a star was revealed. If the partnership chose and scratched off an incorrect answer, there was no star and the partnership knew that it would be necessary to discuss and select an alternate answer. After the first 10 sessions, students began discussing and writing down the answers to open-ended rather than multiple-choice inference questions.

### Comparison Condition

Students assigned to the comparison condition participated in their school's business-as-usual *Accelerated Reader*™ (AR) English language arts instruction. During AR instruction, students picked books independently and read at their own pace. When finished, each student took a short quiz on the computer. The purpose of the AR Reading Practice Quizzes was to determine whether students had read their books by evaluating their performance on literal comprehension questions. At the study school, AR instruction was delivered in 90-min blocks every other day (i.e., students in the comparison condition received an average of 225 min of AR instruction per week). Students in the treatment condition received 45 min of AR instruction and 40 min of inference instruction every other day (i.e., an average of 112.5 min of AR instruction and 100 min of inference instruction per week).

### Data Collection and Measures

We examined student outcomes on three measures of inference skill (the *Clinical Evaluation of Language Fundamentals, Fifth Edition Metalinguistics* [CELF-5 Metalinguistics] Making Inferences subtest; the researcher-developed *Making Inferences Reading Test*; the *Stanford Achievement Test, 10th Edition* [SAT-10], Reading Vocabulary subtest), and one measure of general reading comprehension (GMRT-RC). We administered each of these tests within 2 weeks of the start of the intervention and then again within 2 weeks of the last intervention day. A series of *t* tests for independent samples showed no statistically significant differences between the treatment group and the control group on any of these measures at pretest,  $p > .05$ .

**CELF-5 Metalinguistics, Making Inferences subtest.** This individually administered assessment evaluates a student's ability to generate gap-filling inferences on the basis of causal relationships or event chains presented in short narrative texts that are presented orally and in text form (Wiig & Secord, 2014). Students listen to the examiner describe a situation by its beginning and its ending; they then identify the best two out of four reasons given for the ending and provide an additional reason of their own invention. Internal consistency reliability was 0.81, test-retest reliability was 0.72, and interscorer reliability was 0.95.

**GMRT-RC.** This timed, group-administered assessment measures a student's ability to read and understand literary and informational passages (MacGinitie et al., 2000). About half of the 48 items on the GMRT-RC assess literal comprehension of information in text, while half require students to make within-text inferences (Kulesz,

Francis, Barnes, & Fletcher, 2016). Alternate-forms reliability coefficients range from .74 to .89 across Grades 6 to 12.

**Making Inferences Reading Test.** This researcher-developed measure of text-connecting and gap-filling inference skill closely resembles the passages and postreading inferential questions that students answered during the intervention. It consists of 11 passages between eight and 28 sentences in length from a book that has a 600L Lexile® level. Each passage is followed by two to four multiple-choice inference questions. There are 30 questions in all. The internal consistency of the test when administered with this sample was .76 (Cronbach's alpha). The test did not demonstrate high concurrent validity when correlated with scores on the Making Inferences subtest of the CELF-5 Metalinguistics at pretest ( $r = 0.53$ ) or at posttest ( $r = 0.31$ ). This lack of concurrent validity may be at least partly due to the fact that the *Making Inferences Reading Test* measured students' ability to make inferences while reading, in contrast to the CELF-5 Metalinguistics subtest, which measured oral language inference skill.

**SAT-10, Reading Vocabulary subtest.** This timed, group-administered assessment of reading vocabulary knowledge measures students' ability to infer word meanings based on context clues as well as the breadth of students' preexisting word knowledge. We investigated intervention effects on the test as a whole, as well as on the subset of eight items that, according to the test publisher, measure students' ability to derive word meaning based on context clues. Internal consistency reliability for the Reading section of the SAT-10 was reported as 0.87.

**Test of Word Reading Efficiency, Second Edition (TOWRE-2).** This individually administered measure of word reading accuracy and fluency includes two subtests (Torgesen, Wagner, & Rashotte, 2012). The Sight Word Efficiency (SWE) subtest assesses the number of real words printed in vertical lists that an individual can accurately identify within 45 s. The Phonemic Decoding Efficiency (PDE) subtest measures the number of pronounceable nonwords presented in vertical lists that an individual can accurately decode within 45 s. The average of alternate-forms and test-retest reliability coefficients for the TOWRE-2 each exceeded .90.

### Implementation Fidelity

All inference instruction lessons were audio-recorded. Two trained doctoral-level research assistant coders listened to audio recordings and scored fidelity of implementation of individual intervention components on a Likert-type scale addressing the quality of implementation (1 = *low*; 4 =

**Table 2.** Implementation Fidelity Data.

Aspect of implementation measured	N	M	SD	Range
<b>Type of instruction</b>				
Review	7	2.71	1.60	1–4
Explicit instruction	8	3.25	1.39	1–4
Modeling	8	3.63	1.06	1–4
Guided practice	8	3.63	1.06	1–4
Direction to read	23	4.00	0.00	4–4
Redirection to read	23	4.00	0.00	4–4
Direction to answer	23	4.00	0.00	4–4
Guidance to answer	23	4.00	0.00	4–4
Feedback	23	4.00	0.00	4–4
<b>Type of global observation</b>				
Global instruction	23	5.00	0.00	5G–5G
Global classroom management	23	5.00	0.00	5G–5G

Note. N = total number of observations. Some lessons did not include review, explicit instruction, modeling, and guided practice components; for this reason, these components were not observed during a number of observations. Thus, "N" for "Type of Instruction" varies. "1" = Low. "2" = Mid-low. "3" = Mid-high. "4" = High. "1G" = Low quality. "2G" = Mid-low quality. "3G" = Average quality. "4G" = Mid-high quality. "5G" = High quality.

*high*). Each coder had achieved 96% and 100% agreement, respectively, with a gold standard who was the author of the intervention and the director of the study. Ten percent of sixth-grade audiofiles ( $n = 10$ ) and 10% of seventh-grade audiofiles ( $n = 14$ ) were randomly selected for coding. Fidelity (see Table 2) was rated as high in a majority of observations across all intervention components.

### Results

All student groups received comparable amounts of instruction from each of the three tutors. Given this fact, clustering at the group level would be expected to be minimal. Accordingly, we fit single-level regression models to estimate the treatment's effect on the selected outcomes. Grand-mean centered pretest values for each outcome were included as covariates to improve statistical power. Scatter plots confirmed that each outcome's functional form was best described as linear. The data were multivariate normal based on the evaluation of histograms for each outcome and related Q-Q plots. No scores had standardized residuals with absolute values greater than 3. Finally, the assumption that residuals are uncorrelated was tested using Durbin-Watson's test. Values ( $d$ ) ranged from 1.89 to 2.24, suggesting that residuals are independent. Type I error was controlled using Benjamini-Hochberg's correction for false discovery rate (Benjamini & Hochberg, 1995). Effect sizes were calculated using mean gains for treatment and

**Table 3.** Group Comparison on All Measures of Main Effects.

Measures	Pretest <i>M</i> ( <i>SD</i> )		Posttest <i>M</i> ( <i>SD</i> )	
	<i>T</i>	<i>C</i>	<i>T</i>	<i>C</i>
CELF-5	7.35 (2.08)	7.82 (2.05)	9.03 (2.10)	9.15 (2.20)
SAT-10	635.9 (28.4)	632.2 (24.2)	641.2 (34.3)	633.8 (29.5)
MIRT	18.64 (4.94)	17.74 (5.29)	18.87 (5.35)	17.72 (5.33)
GMRT-RC	86.63 (8.06)	86.80 (8.22)	91.65 (9.33)	86.65 (9.00)

Note. *T* = Treatment (*n* = 39); *C* = Comparison (*n* = 39); CELF-5 = *Clinical Evaluation of Language Fundamentals, Fifth Edition Metalinguistics Making Inferences* subtest, for which scores are scaled scores (*M* = 10, *SD* = 3); SAT-10 = *Stanford Achievement Test, 10th Edition Reading Vocabulary* subtest, for which scores are scaled scores (*M* = 657.3; *SD* = 41.3). MIRT = *Making Inferences Reading Test*, for which scores are raw scores (minimum = 0; maximum = 30); GMRT-RC = *Gates MacGinitie Reading Test, Reading Comprehension* subtest, for which scores are standard scores (*M* = 100, *SD* = 15).

**Table 4.** Summary of Findings on All Measures of Main Effects.

Measure	$\beta$	<i>t</i>	<i>p</i>	<i>d</i>	95% CI	
CELF-5 Metalinguistics MI	.02	0.24	.81	0.17	-0.29	0.62
SAT-10 Reading Vocabulary	.07	0.81	.42	0.13	-0.25	0.50
Subset of inference items	.17	1.66	.10	0.45	0.00	0.91
<i>Making Inferences Reading Test</i>	.04	0.56	.58	0.05	-0.25	0.35
GMRT Reading Comprehension	.27	2.91	.01	0.60	0.16	1.03

Note. The group membership variable was dummy coded, with Treatment = 1 and Comparison = 0. Effect sizes were calculated using mean gains for treatment and comparison groups, standard deviations of pretest and posttest scores for each group, and the correlations between pretest and posttest scores for each group, using procedures described by Lipsey and Wilson (2001). CI = confidence interval; CELF-5 = *Clinical Evaluation of Language Fundamentals, Fifth Edition*; MI = *Making Inferences*; SAT-10 = *Stanford Achievement Test, 10th Edition*; GMRT = *Gates-MacGinitie Reading Test*.

comparison groups, standard deviations of pretest and posttest scores for each group, and the correlations between pretest and posttest scores for each group, using procedures described by Lipsey and Wilson (2001). Group comparisons for all measures are presented in Table 3. Table 4 represents a summary of findings on all outcome measures.

Group membership did not statistically significantly predict outcome scores on the CELF-5 Metalinguistics Making Inferences subtest,  $t(75) = 0.24, p = .81, \beta = 0.02; d = 0.17, 95\%$  confidence interval (CI) [-0.29, 0.62], or the researcher-developed *Making Inferences Reading Test*,  $t(75) = 0.56, p = .58, \beta = 0.04; d = 0.05, 95\%$  CI [-0.25, 0.35]. On the SAT-10 Reading Vocabulary subtest, there were also no statistically significant effects in favor of treatment,  $t(75) = 0.81, p = .42, \beta = 0.07, d = 0.13, 95\%$  CI [-0.25, 0.50]. Group membership also did not predict outcome scores on the subset of items on the SAT-10 Reading Vocabulary subtest that measured skill in inferring word

**Table 5.** Summary of Findings on All Measures of Main Effects for Students Designated LEP.

Measure	$\beta$	<i>t</i>	<i>p</i>	<i>d</i>	95% CI	
CELF-5 Metalinguistics MI	-.04	-0.36	.72	-0.02	-0.53	0.47
SAT-10 Reading Vocabulary	.09	0.92	.36	0.17	-0.27	0.61
<i>Making Inferences Reading Test</i>	.03	0.32	.75	0.02	-0.32	0.35
GMRT Reading Comprehension	.24	2.23	.03	0.49	0.02	0.97

Note. The group membership variable was dummy coded, with Treatment = 1 and Comparison = 0. Effect sizes were calculated using mean gains for treatment and comparison groups, standard deviations of pretest and posttest scores for each group, and the correlations between pretest and posttest scores for each group, using procedures described by Lipsey and Wilson (2001). LEP = limited English proficient; CI = confidence interval; CELF-5 = *Clinical Evaluation of Language Fundamentals, Fifth Edition*; MI = *Making Inferences*; SAT-10 = *Stanford Achievement Test, 10th Edition*; GMRT = *Gates-MacGinitie Reading Test*.

meanings from context at the  $p < .05$  level:  $t(75) = 1.66, p = .10, \beta = 0.17; d = 0.45, 95\%$  CI (0.00, 0.91). Group membership did statistically significantly predict outcome scores on the reading comprehension outcome, the GMRT-RC,  $t(75) = 2.91, p = .01, \beta = 0.27$ , with an effect size of  $d = 0.60, 95\%$  CI [0.16, 1.03]. There was a similar pattern of findings when data were disaggregated for students designated LEP ( $n = 60$ ). Table 5 summarizes findings for this subsample of students.

A *Making Inferences Student Interview* assessed students' perceptions of the intervention for the purposes of assessing the social validity of the instructional treatment. Students' responses were positive, ranging from 4.8 to 6.7 (using a scale for which 1 = *not at all*, 4 = *neutral*, and 7 = *very much*). When asked "What parts of inference instruction do you think helped you most?" students pointed to (a) the small group aspect of instruction, (b) the way in which they were encouraged to ask and answer inferential questions, (c) instruction that helped them infer word meanings from context and find text evidence to support inferences, (d) the opportunity to read aloud, and (e) encouragement to broadly "talk about the book" and "share our ideas." The last question on the survey asked students, "What suggestions do you have for changes to inference instruction that would make it more helpful to other students?" Most responses indicated that students did not suggest changing instruction. Some students requested more partner work instead of work in small groups (e.g., "Work with a partner more often," "Answer questions with a partner," "We should read with only one partner . . . to increase our reading").

## Discussion

This study investigated the effects of an inference instruction intervention on the inference generation and reading

comprehension of middle school struggling readers, the vast majority of whom were language minority students. Students who scored at or below a standard score of 97 ( $M = 86.7$ ,  $SD = 8.1$ ) on a standardized assessment of general reading comprehension were randomly assigned to the inference instruction treatment condition or a comparison condition, which consisted of business-as-usual computer-delivered *Accelerated Reader*<sup>TM</sup> instruction. The intervention consisted of 24 sessions (40 min, 2.5 times per week) of explicit text-connecting and gap-filling inference instruction and practice.

Group membership statistically significantly predicted outcome score on the standardized, norm-referenced measure of general reading comprehension,  $p = .01$ . On average, participation in inference instruction corresponded with a 0.27 standard deviation increase in students' reading comprehension scores at posttest compared with students in the comparison condition. Measured by the Cohen's  $d$  statistic, the size of the effect was  $d = 0.60$ . Group membership predicted outcome score on the GMRT-RC for the subsample of students designated LEP ( $n = 60$ ),  $p = .03$ . On average, participation in inference instruction corresponded with a 0.24 standard deviation increase in students' reading comprehension scores at posttest compared with students in the comparison condition ( $d = 0.49$ ). These effects are meaningful given what we know about the way in which gains on measures of standardized reading achievement decrease as students progress through the grades (Bloom et al., 2008; Scammacca, Roberts, Vaughn, & Stuebing, 2013). The effect is substantially larger than (a) mean effects of reading interventions reported between 1980 to 2004 on standardized reading outcomes for struggling readers in Grades 4 through 12 ( $g = 0.13$ ; Scammacca et al., 2013) and (b) mean effects of reading interventions reported between 1995 and 2015 on standardized reading outcomes for ELs in Grades 4 through 8 ( $g = 0.01$ , Hall et al., 2017).

Group membership did not predict posttest scores on any measure of inference skill at the  $p < .05$  level. However, for all three inference measures, adjusted posttest means were higher for students in the treatment group than for students in the comparison group. If we could be sure that the differences between groups represented true differences, then they would represent small but practically significant effects in favor of treatment, given the context of reading interventions for older students with learning difficulties. The effect size on the CELF-5 Metalinguistics Making Inferences subtest was  $d = 0.17$ , 95% CI [-0.29, 0.62], and on the SAT-10 Reading Vocabulary subtest, it was  $d = 0.13$ , 95% CI [-0.25, 0.50]. When considering only the subset of items on the SAT-10 Reading Vocabulary subtest that measured skill in inferring word meanings based on context clues, the effect size was  $d = 0.45$ , 95% CI [0.00, 0.91]. Future research

might consider investigating the statistical significance of inference instructional effects with a larger sample, in a study powered to discover effects of this size.

We had hypothesized that treatment students would demonstrate accelerated growth on our measures of inference skill, and particularly on the researcher-developed *Making Inferences Reading Test*, as more proximal measures typically yield higher intervention effects than distal measures (Swanson, 2000). One possible explanation lies in the fact that the inference measures were less proximal than we had intended, and all depended on students' breadth of knowledge as well as on their inference skill. The CELF-5 Metalinguistics Making Inferences subtest exclusively assessed a student's ability to generate gap-filling inferences on the basis of causal relationships or event chains presented in short narrative texts. The researcher-developed *Making Inferences Reading Test* consisted of 30% to 40% text-connecting inference questions and 60% to 70% gap-filling inference questions. Only the subset of SAT-10 Reading Vocabulary subtest items that measured ability to infer word meanings based on context clues were truly proximal to the inference instruction intervention; the remaining items measured the breadth of students' word knowledge, something that was not a target of this intervention.

In contrast, the GMRT-RC does *not* measure gap-filling inference skill. Kulesz et al. (2016) determined that the Grade 7 to 9 form of the GMRT-RC (Form S) consists of 50% literal or text memory items, 50% text-based inference items, and no gap-filling inference item. Therefore, statistically significant effects in favor of treatment on the GMRT-RC but not on the CELF-5 Metalinguistics Making Inferences subtest, the *Making Inferences Reading Test*, or the SAT-10 Reading Vocabulary subtest could indicate that inference instruction was effective in teaching ELs with reading comprehension difficulties to better remember literal information in the text and to make text-connecting inferences, but not effective in teaching students to make gap-filling inferences. While it may be possible to teach students to generate text-connecting and even gap-filling inferences during an instructional intervention of short duration, it is likely less possible to increase the breadth of word and world knowledge that is a necessary condition for gap-filling inference generation.

While it seems counterintuitive that inference instruction has the potential to improve students' literal comprehension of text, Elleman (2017) notes that the Construction-Integration (CI) model of text processing (Kintsch, 1988) provides a theoretical rationale for this finding. The authors of the CI model posit that the process of actively drawing connections between propositions or factual information in text during situation model construction strengthens the reader's memory for literal content. Elleman (2017) found that, on average, inference instruction had a positive effect



on literal comprehension (effect sizes [ $k = 18$ ] ranged from  $d = 0.46$  to  $d = 1.85$ , with an overall random weighted mean effect size of 0.28).

It is noteworthy that this inference instruction intervention had a statistically significant and substantial impact on the literal and text-connecting inferential comprehension of students with reading comprehension difficulties who were language minority students. Of the sample of participants in this study, 96.2% were Hispanic, and 76.9% were designated LEP currently or within the last 2 years. ELs constitute 9.8% of enrollment in U.S. public elementary and secondary schools (National Center for Education Statistics [NCES], 2016) and the population of EL students is growing faster than that of other populations nationally. But ELs in the United States demonstrate significantly lower academic achievement than their EO peers (NCES, 2015), with EL students at higher risk of grade retention and school dropout (Kena et al., 2015). Previous reading interventions for middle grades' ELs with reading comprehension difficulties have not demonstrated statistically or practically significant effects on students' reading comprehension (Hall et al., 2017; Richards-Tutor et al., 2016).

Elements of this intervention may be particularly beneficial for ELs who are struggling readers. Research demonstrates that ELs in the middle grades benefit from intervention that is focused on oral language/academic vocabulary development, including from instruction focused on the inference of word meanings from context clues (e.g., Hall et al., 2017; Lesaux, Kieffer, Kelley, & Harris, 2014; Spencer & Wagner, 2017). In addition, because students with limited English proficiency may expend more cognitive capacity on word-level processes (i.e., accessing word-level semantic information) than on sentence- or passage-level integration of information in text, this text integration-focused intervention may have had a particularly beneficial effect on reading comprehension for EL students.

### Limitations and Future Research

Several factors limit interpretation of study findings. First, our sample size ( $n = 78$ ) was limited by the number of consent forms obtained and by sample attrition. Six students dropped out of the study due to school transfers or scheduling issues ( $n = 4$  in Treatment,  $n = 2$  in Comparison). A larger sample size would have increased the power to levels that could have better detected effects and yielded more generalizable findings. It is also important to note that, because intervention students received instruction within small groups of three to six students while business-as-usual students read independently and took on-computer quizzes, group size may have been a confounding variable. It is not possible to distinguish the effects of participating in small group instruction from the effects of inference instruction. Likewise, because

intervention students received explicit instruction and comparison students did not, it is not possible to distinguish the effects of receiving explicit reading comprehension instruction in general from the effects of this particular approach to inference instruction.

Another limitation may have been the short duration of the intervention. Elleman (2017) found that inference instruction showed positive results in relatively short periods of time (i.e., less than 10 hr). Still, while few studies have examined the effects of interventions lasting longer than one school year (Vaughn & Wanzek, 2014), there is evidence to suggest that multiple-year intensive interventions are needed to support older students with significant reading difficulties (Vaughn & Fletcher, 2012). The current intervention provided students with 16 hr (24, 40-min sessions) of instruction. Perhaps larger intervention effects might be realized if instruction were of longer duration.

The wide range in reading comprehension ability of students included in this study (range = 1st percentile to 42nd percentile) may limit conclusions about the populations that are likely to benefit from this intervention. Still, it is important to note that results suggest that instruction was equally effective for students regardless of initial reading comprehension ability. Reading comprehension at screening was not a statistically significant moderator of intervention effects on the GMRT-RC ( $b = 0.05$ , 95% CI  $[-0.47, 0.58]$ ,  $t = 0.27$ ,  $p = .84$ ). In addition, when we restricted our analysis only to students who scored at or below a standard score of 90 (i.e., at or below the 25th percentile) on the GMRT-RC screener, students in the treatment group continued to statistically significantly outperform students in the comparison group on the GMRT-RC at posttest ( $p = .00$ ,  $d = .89$ ).

It would be useful to replicate this study with a larger sample that includes more EO students, in order to better understand if the effectiveness of this intervention depends on student levels of proficiency in the language of instruction. It would also be beneficial to collect more fine-grained information about student language proficiency to describe student proficiency in English as a continuous variable. Kieffer (2008, 2011) revealed noteworthy differences among subpopulations of ELs when contrasting reading growth trajectories of LEP, initially fluent English proficient (IFEP), and redesignated English proficient (RFEP) students from kindergarten until fifth or eighth grade. Other research suggests that the effectiveness of reading interventions may vary based on EL students' initial levels of English proficiency (Hwang, Lawrence, Mo, & Snow, 2015; Lawrence, Capotosto, Branum-Martin, White, & Snow, 2012). A related limitation of this study was the absence of Spanish-language measures of word reading, reading comprehension, and inference skill. It is unclear how accurately our English-language measures assessed

reading and inference skills for students who had limited English language proficiency, and it would have been informative to have more information about students' reading and inference generation proficiency in their L1. These factors may also have impacted students' responses to intervention. Finally, future research would benefit from the development and validation of measures of inference skill, including specific subtypes of inference skill, that are sensitive to incremental change in inference generation skill and measure gap-filling inference skill in a way not confounded by students' levels of knowledge. Such an instrument would allow researchers to better assess the degree to which inference instruction impacts gap-filling inferential reading comprehension as well as text-connecting inference generation and literal reading comprehension.

### *Implications for Practice*

Findings from the present study indicate that inference instruction, when delivered to small groups of three to six students during supplemental reading intervention blocks, is effective in improving the text-connecting inference generation and literal reading comprehension of middle school students with reading difficulties. Results suggest that EL students with reading comprehension difficulties benefit from receiving explicit instruction in generating specific inference types, including noticing gaps and/or lack of coherence in text, identifying clue words or phrases, integrating information within the text, activating background knowledge, and using background knowledge to fill a gap in the text. Teachers can use simple graphic organizers to scaffold the process of knowledge-based inference generation, making visible the integration of information in text with information in background knowledge. In this study, interventionists designated stopping points where the text lacked coherence or where generating an inference would furnish a more complete situation model. They then posed predetermined inference questions and explicitly taught and modeled for students how to discuss and find text evidence in support of potential answers to these questions.

### *Conclusion*

In addition to being a subject of research, inferential reading comprehension has been a focus of recent education policy. The National Governors Association Center for Best Practices, Council of Chief State School Officers (CCSSO; 2010) stresses the importance of building students' inferential comprehension to prepare students to enter a world in which colleges, businesses, and the obligations of citizenship demand ever more rigorous and critical comprehension of text. Students are expected not only to "read closely to

determine what the text says explicitly" but also "to make logical inferences from it; cite specific textual evidence when writing or speaking to support conclusions drawn from the text," "determine central ideas or themes," and "analyze how and why individuals, events, and ideas develop and interact" (CCSSO, 2010). Historically, reading instruction has often focused more on literal comprehension, or recall of information stated explicitly in text, than on inferential comprehension (e.g., Bintz & Williams, 2005; Graesser & Person, 1994; McKeown & Beck, 2003). Today, the *Accelerated Reader*<sup>TM</sup> educational software program, used in more than 37,000 schools worldwide and the most popular educational software for K-12 students in the United States (Renaissance Learning, 2015), encourages students to read books independently and answer exclusively literal postreading questions.

Given the importance of inference generation to reading comprehension (e.g., Ahmed et al., 2016), the efficacy of inference instruction in improving the inference generation skill of struggling readers (e.g., Elleman, 2017; Hall, 2016), and the value of inferential comprehension in today's schools and workplaces, inference instruction may represent a significant opportunity to ensure students' school and postsecondary academic achievement. The current study demonstrates the promise of a small-group inference instruction intervention. While there is a need for future research to determine that these findings are replicated with a larger sample, this study suggests that 16 hr (24, 40-min sessions) of inference instruction that incorporates teacher modeling via think-aloud, inference-eliciting questions during reading, and simple graphic organizers has the potential to improve literal and text-connecting inferential comprehension for ELs with reading comprehension difficulties.

### **Authors' Note**

Colby Hall, The Children's Learning Institute, The University of Texas Health Science Center at Houston.

### **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: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to The University of Texas at Austin as part of the Reading for Understanding Research Initiative. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors declare that they have no conflict of interest.

## ORCID iD

Colby Hall  <https://orcid.org/0000-0002-0779-1322>

## References

- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44*, 68–82.
- Babayigit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading, 37*, 22–47.
- Ballard, W. S. & Tighe, P. L. (2012). IDEA Proficiency Test. Brea, CA: Author.
- Barth, A. E., Barnes, M. A., Francis, D., York, M., & Vaughn, S. (2015). Bridging inferences among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing, 28*, 587–609.
- Barth, A. E., & Elleman, A. (2017). Evaluating the impact of a multistrategy inference intervention for middle-grade struggling readers. *Language, Speech, and Hearing Services in Schools, 48*, 31–41.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological, 57*, 289–300.
- Bintz, W. P., & Williams, L. (2005). Questioning techniques of fifth and sixth grade reading teachers. *Middle School Journal, 37*(1), 45–52.
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Educational Effectiveness, 1*, 289–328.
- Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*, 189–201.
- 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.
- Capps, R., & Fix, M. (2005). *Undocumented immigrants: Myths and reality*. Washington, DC: The Urban Institute.
- Cho, E., Capin, P., Roberts, G. J., Roberts, G., & Vaughn, S. (in press). Examining sources and mechanisms of reading comprehension difficulties: Comparing English learners and non-English learners within the simple view of reading. *Journal of Educational Psychology*.
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325.
- CTB/McGraw-Hill, Data Recognition Corporation (2012). Language Assessment Scales Links. Monterey, CA: Author.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., . . . Carter, S. (2015). Text-processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and informational text. *Reading Research Quarterly, 50*, 393–416.
- Denton, C. A., York, M. J., Francis, D. J., Haring, C., Ahmed, Y., & Bidulescu, A. (2017). An investigation of an intervention to promote inference generation by adolescent poor comprehenders. *Learning Disabilities Research & Practice, 32*, 85–98.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly, 38*, 78–103.
- Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading, 17*, 435–452.
- Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology, 109*, 761–781.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137.
- Grant, A., Gottardo, A., & Geva, E. (2011). Reading in English as a first or second language: The case of grade 3 Spanish, Portuguese, and English speakers. *Learning Disabilities Research & Practice, 26*, 67–83.
- Hall, C. (2016). Inference instruction for struggling readers: A synthesis of intervention research. *Educational Psychology Review, 28*, 1–22.
- Hall, C., Roberts, G. J., Cho, E., McCulley, L. V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review, 29*, 763–794.
- Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism, 19*(3), 314–332.
- Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., . . . Dunlop Velez, E. (2015). *The condition of education 2015* (NCES 2015-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*, 851–868.
- Kieffer, M. J. (2011). Converging trajectories reading growth in language minority learners and their classmates, kindergarten to grade 8. *American Educational Research Journal, 48*, 1187–1225.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163–182.
- Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*, 363–394.
- Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology, 108*, 1078–1097. doi:10.1037/edu0000126

- Lawrence, J. F., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. E. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition*, 15(3), 437–451.
- Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York: Carnegie Corporation of New York.
- Lesaux, N. K. (2006). Development of literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp. 75–122). Mahwah, NJ: Lawrence Erlbaum.
- Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, 47, 596–632.
- Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, 51, 1159–1194.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L., & Hughes, K. E. (2000). Gates-MacGinitie Reading Test. Rolling Meadows, IL: Riverside Publishing.
- McKeown, M. G., & Beck, I. L. (2003). Taking advantage of read-alouds to help children make sense of decontextualized language. In A. van Kleeck, A. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (6th ed., pp. 159–176). Mahwah, NJ: Lawrence Erlbaum.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, 51, 297–384.
- Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on “long-term English language learners”: Characteristics and prior schooling experiences of an invisible population. *International Multilingual Research Journal*, 6, 121–142.
- Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners’ word decoding and reading comprehension. *Reading and Writing*, 20, 691–719.
- National Center for Education Statistics. (2015). *The nation’s report card: 2015 mathematics and reading assessments*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- National Center for Education Statistics. (2016). *English language learner (ELL) students enrolled in public elementary and secondary schools, by grade, home language, and selected student characteristics: Selected years, 2008-09 through fall 2014 (Table 204.27)*. Washington, DC: Author. Retrieved from [https://nces.ed.gov/programs/digest/d16/tables/dt16\\_204.27.asp](https://nces.ed.gov/programs/digest/d16/tables/dt16_204.27.asp)
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Author.
- Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16, 91–121.
- Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for California’s long-term English learners*. Long Beach, CA: Californians Together.
- Olsen, L. (2014). *Meeting the unique needs of long term English language learners*. Washington, DC: National Education Association.
- Palacio, R. J. (2012). *Wonder*. New York, NY: Alfred A. Knopf.
- Pearson (2012). Stanford English Language Proficiency Test. New York, NY: Author.
- Pearson (2015). Test of English Language Learning. New York, NY: Author.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22–37.
- Reed, D. K., & Lynn, D. (2016). The effects of an inference-making strategy taught with and without goal setting. *Learning Disability Quarterly*, 39, 133–145.
- Renaissance Learning. (2015). *The research foundation for accelerated reader 360*. Wisconsin Rapids, WI: Author.
- Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children*, 82(2), 144–169.
- Riverside Publishing, Houghton Mifflin Harcourt (2010). Woodcock-Muñoz Language Survey–Revised. Rolling Meadows, IL: Author.
- Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2013). A meta-analysis of interventions for struggling readers in Grades 4-12: 1980-2011. *Journal of Learning Disabilities*, 48, 369–390.
- Spencer, M., Quinn, J. M., & Wagner, R. K. (2014). Specific reading comprehension disability: Major problem, myth, or misnomer? *Learning Disabilities Research & Practice*, 29, 3–9.
- Spencer, M., & Wagner, R. K. (2017). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: A meta-analysis. *Journal of Research in Reading*, 40, 199–217.
- Swanson, H. L. (2000). Issues facing the field of learning disabilities. *Learning Disability Quarterly*, 23(1), 37–50.
- Tighe, E. L., & Schatschneider, C. (2014). A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. *Reading and Writing*, 27, 101–127.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE: Test of word reading efficiency, 2nd edition*. Austin, TX: Pro-Ed.
- Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities*, 45, 244–256.
- Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice*, 29, 46–53.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals, fifth edition (CELF-5)*. Bloomington, MN: NCS Pearson.
- Wolfe, M. B., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology*, 80, 341–362.
- Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2, 33–45.