Building BRIDGES: A Design Experiment to Improve Reading and United States History Knowledge of Poor Readers in 8th Grade

Rollanda E O'Connor

Kristen D. Beach

Victoria Sanchez

Kathleen M. Bocian

Lindsay Flynn

O'Connor, R.E., Beach, K., Sanchez, V. Bocian, K., & Flynn, L. (2015). Building BRIDGES: A Design Experiment to Improve Reading and United States History Knowledge of Poor Readers in 8th Grade. *Exceptional Children, 81,* 399-425.

Funded by the Institute of Education Sciences, Grant number R324A120173, to the University of California, Riverside. Opinions expressed are those of the authors, and do not represent those of the Institute or U.S. Department of Education.

Abstract

We tested the effects of teaching reading skills through U.S. History content for 38 eighth-grade poor readers whose reading ability ranged from 2nd to 4th grade levels. Half of the students received special education services and half of the students were English Language Learners. Students were taught to decode multisyllabic words, learn meanings of academic words, and identify cause and effect relationships. They used easy levels of history text, and then bridged into more difficult text accounts of the same events. Results showed gains in performance across each cycle in each reading component. Comparisons between students in the intervention and their typically developing peers showed stronger gains for intervention students in vocabulary and comprehension strategies. Students with disabilities who received instruction in causes and effects of historical events scored similarly to typical readers in their general education history classes.

Keywords: reading, U.S. History, middle school, design-based experiment, intervention

Building BRIDGES: A Design Experiment to Improve Reading and United States History Knowledge of Poor Readers in 8th Grade

The Common Core State Standards (CCSS; National Governors' Association, 2010) stress higher order thinking skills across literacy and content areas, including identifying main ideas and cause/effect relationships, and comparing and contrasting features of ideas and points of view. For students with learning disabilities (LD) in secondary history classrooms, these stringent expectations may be compounded by low basic reading skills such as difficulty reading multisyllabic words and understanding what they mean (Bulgren, Graner, & Deshler, 2013). These reading difficulties inhibit students' likelihood of achieving the CCSS, as well as achieving positive life outcomes associated with satisfying employment and civic involvement (Wei, Blackorby, & Schiller, 2011).

On the one hand, the CCSS states that "Students must be able to read complex informational texts … with independence and confidence" on the other hand, many students with LD have difficulty reading any kind of text near grade level. As Swanson and Deshler (2003) concluded, this lack of basic skills, combined with the complexity of expository text in content areas, make it difficult for adolescents with LD to succeed in middle and high school. Teachers may find themselves caught between competing instructional goals as they weigh the importance of content acquisition against the reading and reasoning skills demanded by the CCSS.

Our research in this paper tackled multiple instructional dilemmas: (1) helping students with LD and other poor readers learn sufficient basic reading skills to engage in

reading history texts and historical documents; (2) teaching focused learning strategies to enable students with disabilities to grapple with important historical concepts; and (3) supporting teachers as they attempt to infuse instructional reading supports and strategies in their general and special education history classes. We concerned ourselves with student growth in reading and comprehension of history text, but equally with their teachers' endeavors to adjust instruction to include students with disabilities in the midst of addressing the CCSS. As Smith, Schmidt, Edelen-Smith, and Cook (2013) explained, "Practitioners want to know whether a practice will work for their students and whether it can be implemented in the realities of their classroom and school context (e.g., limited time and resources)." (p. 148) We reasoned that teachers might improve implementation of innovations for students with disabilities by participating in the development of instruction and small feasibility trials as these innovations were designed and integrated. This type of collaborative research focused on shared problems of practice is often referred to as design-based research (Smith et al.), and is the method we used in the current study.

The Structure of Design-based Research

Design-based research (DBR) evolves around collaborative, iterative, and systematic study focused on solving persistent problems of practice (Penuel, Fishman, Cheng, & Sabelli, 2011). Ideally, DBR includes multiple stakeholders' perspectives. In our case, teachers were concerned with a relatively high failure rate in history classes (near 20%), relatively large proportions of students who were English Language Learners (near 35%), mainstreaming of students with disabilities (10%), and looming implementation of the CCSS. Among failing students, most had difficulty accessing the

8th grade history text, which was used in class and for homework. Our university team was concerned with students' limited opportunities to improve their reading ability, given the congested middle school schedule and teachers' professed lack of training in teaching reading. Students were concerned with their own poor academic performance along with how their school failure impacted the perception of their peers. DBR brings these perspectives together in a commitment from all parties toward iterative, collaborative design, which has been advocated recently in professional journals such as *Exceptional Children* (Smith et al., 2013) and *Educational Researcher* (Anderson & Shattuck, 2012).

DBR consists of multiple iterations, as features of instructional routines are designed, tested, and integrated. Based on research for improving reading skills and history knowledge of students with disabilities (described later in this introduction), we established three 5-week instructional cycles. Each cycle involved three weeks of researcher-led instruction and ongoing revision (weeks 1-3), one week of revision following discussions with classroom teachers (week 4), and one week of implementation by classroom teachers in their intact classrooms (week 5). Each cycle began with direct teaching of 12 small groups of students (range 2 to 6 students each) who were poor readers in a pull-out setting. Daily observations of interactions in these instructional groups, frequent videotaping of instruction, and data on student learning led to ongoing improvement of each routine. After three weeks of researcher-led instruction, students provided their views (anonymously) on the most and least helpful instructional features during that cycle.

During each three weeks of each cycle that researchers instructed students, researchers and teachers met twice after school to share instructional samples, video

clips, and data on student learning, discuss feasibility and improvements from teachers' perspectives, and plan instructional routines for them to use in their own classes. We incorporated design improvements derived from observations of teacher implementation with their intact classes in the next 5-week cycle of small group pull-out instruction we developed with students who were poor readers, and so the process repeated across the year as each new feature was developed and evaluated in teachers' classes. Table 1 shows this iterative design, along with the reading strategies and history content addressed in each cycle. Our overall concern was developing capacity for sustaining change in instructional practice focused on easing the problem we all shared initially: Improving the ability of very poor readers to read history text with understanding. In the next section we discuss the research base for the specific strategies we included in the instructional package.

Research on Reading Improvement in Middle School

Teaching reading effectively to middle school poor readers is no small challenge, because poor readers are hampered by serious delays in word, vocabulary, and comprehension skills compared to peers, who mastered the reading basics years before (Swanson & Deshler, 2003). The research base for the three reading components we developed during the DBR is described below.

Decoding. For very poor readers, the first hurdle for comprehending text may be deciphering the words on the page (see Rack, Snowling, & Olson, 1992; Rosenthal & Ehri, 2008); however, effective interventions to address this problem differ from those designed for younger students. Lenz and Hughes (1990) developed a model for word recognition that was effective for students with LD in middle school that incorporated

elements of pattern recognition (word parts larger than the phoneme) and morphology (meaningful parts of words). In reviews of this topic (Bhattacharya & Ehri, 2004; Carlisle & Stone, 2005), interventions that applied larger-than-phoneme-unit decoding skills to words that were age and grade appropriate (multisyllabic words) improved word reading for middle schoolers. Key to the most effective interventions was analysis of words that contained known patterns and teaching patterns and morphemes that had immediate utility for reading long words. Therefore, our decoding strategy taught students to identify word patterns and morphemes across multisyllabic words.

Vocabulary. Although translating printed words to speech is the most fundamental act in reading, understanding meanings of words is also crucial for comprehension. Ebbers and Denton (2008) and Hairell et al. (2011) underscored the importance of developing meaning of words through practical and conversational connections in addition to reading text. By using words in appropriate conversational settings and discussions, studies have shown generalization of improved vocabulary to comprehension (Alishahi, Fazly, & Stevenson 2008; Beck, McKeown, & Kucan, 2013; Harniss, Caros, & Gersten, 2007; Mastropieri, Scruggs, & Fulk, 1990).

Deep understanding of words' meanings accumulates across different contexts, which provide the nuances that lead to decontextualized representation of meaning—one a reader can use to make sense of subsequent contexts containing the word (Baumann, Kameenui, & Ash, 2003; Cromley & Azevedo, 2007; Harniss et al., 2007). Therefore, word meanings were taught through student-friendly examples, rehearsed through conversation, and connected to other words in and across instructional sets.

Comprehension of Expository Text. Comprehension of text forms the next major hurdle for middle school poor readers (Denton, Wexler, Vaughn & Bryan, 2008; Roberts et al., 2008). Expository text (such as history) is particularly problematic for struggling readers, as it comprises a variety of text structures (Gajira, Jitendra, Sood, & Sacks, 2007); however, key structures such as cause and effect relationships have been taught successfully to students with LD in middle school (Williams, Stafford, Lauer, Hall, & Pollini (2009), especially when text difficulty is decreased (Wineberg & Martin, 2009). Additionally, we needed a reading comprehension strategy that history teachers would consider valuable. Due to the research base on cause and effect relationships in both reading and history instruction, along with its central role in the CCSS, we chose to focus on the cause and effect strategy.

Espin, Cevasco, van den Broek, Baker, and Gersten (2007) described the ability to make causal connections between events as the "cement of the universe" (p. 175), and noted that many students with LD have difficulty making these connections. In history texts, causal connections may be confused when events are not presented in a clear sequence that helps students determine which event preceded the other. Espin et al. demonstrated that when students with LD understand the causality among events, these events are more memorable than when they learn the same events as facts. When taught to identify and articulate causal connections, students organize events under an overarching thematic cause that framed the series of events, in essence turning the historical account into a narrative and making it memorable as a story.

De La Paz, Morales, and Winston (2007) demonstrated that cause/effect relationships could be taught successfully to students with LD, and that they can use this

understanding to improve written essays about historical events. Their collaboration (i.e., a researcher and two teachers in general education history and special education), in which the researcher developed instructional routines, teachers used the routines with their students, and modifications were made during and across instructional sessions, provides a model for our DBR. De La Paz et al. adjusted teachers' learning materials to make use of the strategy transparent for students' and teachers' first attempts, and then included a broader range of materials as students began to infer relationships more reliably.

An additional feature affecting results from strategy intervention for students with disabilities (Gajria et al., 2007) was practicing the strategy with text at easier reading levels than typical middle school text (see also Wineburg & Martin, 2009). When students can read the text, their comprehension is improved with consistent visual enhancements, such as diagrams to fill in or brief video segments, that help students anticipate the structure of what they read (Gersten, Baker, Smith-Johnson, Dimino, & Peterson, 2006) and organize information as they read it (Bulgren, Deshler, & Lenz, 2007; DiCecco & Gleason, 2002). Therefore, we taught students to apply vocabulary and cause and effect strategies first in easier (i.e., 4th grade readability) passages, and then read the same events in more difficult passages (i.e., portions extracted from their 8th grade textbook).

In all, studies of successful reading and history interventions for poor readers in middle school share features that include a focus on reading and understanding key vocabulary and teaching an overarching comprehension strategy such as cause and effect relationships. In a recent study of improving both reading comprehension and history

knowledge for typical learners, teachers of experimental classes implemented five distinct components (an overarching idea, essential words, new content, team-based learning, and team-based application) over 10-day cycles of instruction (Vaughn, Swanson et al., 2013). Participating students were relatively high achievers: 93% had passed the state reading assessment prior to beginning the study, 4.8% were eligible for special education services, and 1.2% were English Language Learners (EL). On the experimenter-designed test of history content, students in the treatment outperformed those in traditional instruction. Although the goals of Vaughn et al. and our study were similar, their results were not disaggregated for special education, EL, or reading ability, all of which could influence learning.

The Current Study

Struggling readers and their teachers were the particular focus of building our strategies to improve performance in reading and history, which we called BRIDGES: Building Reading Interventions Designed for General Education Subjects. Taking features of effective middle school intervention into account, we built BRIDGES successively over the course of a school year, using DBR to construct and integrate lesson components interfaced with feasibility trials by their teachers in classrooms. Each of the 5-week cycles included 3 weeks of small group pull-out instruction to develop and test an intervention component (i.e., decoding, vocabulary, or cause and effect relationships) with very poor readers, a week to analyze data and prepare for classroom teachers' implementation of the strategy, and a 1-week feasibility trial as general and special education teachers implemented the instruction with their intact classes that included the BRIDGES students.

Our specific research questions were: (1) Will students make gains in reading skills through reading lessons that focus on historical content? (2) Does improvement differ among students with disabilities, students who are EL, and other low achievers? (3) Does this reading instruction improve knowledge of historical content? (4) To what degree of fidelity do coached teachers implement reading instruction during history class? and (5) How do teachers view feasibility of these instructional routines in their roles as general and special education history teachers?

Method

Participants

The participating middle school, Hamilton, was located in a large urban district in the SW United States with over 40,000 students. Hamilton serves a 7th and 8th grade student population of 1080, with 54% Hispanic, 27% White, 8% African-American, and 8% Other. Over half the students receive free or reduced-price lunch.

Five of the 50 teachers at Hamilton participated; 4 taught 8th grade history in general education and one in special education. Years of teaching experience ranged from 1 to 30 years, three of the five teachers were credentialed in secondary history, one was credentialed in secondary English, and the special education teacher had a history endorsement. All agreed to participate during a spring meeting prior to implementation of this research.

Students. The school counselor helped to select student participants by sending permission letters to parents of 8th grade students in general education (GenEd) who as 7th graders scored below basic on the state test of English/Language Arts (ELA) in spring of 2012, and who received grades of D or F in 7th grade history, along with students who

received special education (SpEd) in the mild/moderate eligibility categories. This process generated a potential sample of 61 students. As this was a confidential recruitment handled by the school counselor, access to data for comparison of opt in/opt out samples is limited. Among the 22 students who chose not to participate in the study, 10 were receiving SpEd services and 9 were English Language Learners.

We obtained positive parent and student assent for 39 students. Of these, 22 received SpEd services, 25 were Hispanic, 9 were White, and 2 were African American. Three students (2 GenEd and 1 SpEd) were dropped from BRIDGES during our first two weeks of instruction: 1 due to parental request, 1 due to scheduling conflicts, and 1 due to excessive absences. Within the group of Hispanic students, 15 were classified as English Learners (EL): 9 from GenEd and 6 from SpEd. The demographics of the remaining 36 students are noted in Table 2.

Of the students with disabilities, 14 received SpEd services in self-contained and 7 in resource settings. Their primary eligibilities were Learning Disability (n = 14), Other Health Impairment (n = 4), and Autistic-Like Behaviors (n = 3). Six students had Speech and Language Disorder as a secondary eligibility. These students had several similar IEP goals that addressed reading fluency, reading comprehension, and compositional writing.

Measures

Students completed assessments that measured aspects of reading, including decoding, word identification, reading fluency, vocabulary, and comprehension, and also knowledge of history content and use of the cause/effect strategy. Our participants scored below the 25th percentile on standardized measures, and most students scored below the 15th percentile. These measures are described below and scores are shown in Table 3.

Descriptive Measures

Woodcock-Johnson Tests of Achievement III. To describe our BRIDGES participants, we administered four norm-referenced subtests (Word Identification, Word Attack, Picture Vocabulary and Passage Comprehension) from the Woodcock-Johnson Tests of Achievement III (WJ-III; Woodcock, McGrew, & Mather, 2001). The Word Identification subtest requires students to read aloud from a list words that increase in difficulty. The Word Attack subtest measures decoding of increasingly difficult nonwords. The Picture Vocabulary subtest requires students to point to pictures and name objects. The Passage Comprehension subtest requires students to read short sentences or passages (increasing in difficulty) and provide a contextually appropriate word for a deletion within the passage. Across subtests, reliabilities ranged from .81-.94, with validity ranging from .44 to .82.

Cognitive ability. The WASI (Wechsler, 2011) is a norm-referenced intelligence test comprising four subtests: Vocabulary, Block Design, Similarities, and Matrix Reasoning. Across subtests, reliabilities ranged from .81-.98, with validity of .66. Students without SpEd designations had no cognitive scores available in their files, and we administered the WASI to these students (n = 15).

Students with SpEd designations were administered an assessment of cognitive functioning as part of their school-based psychoeducational evaluation by the school psychologist at their school site (n = 21). At the SpEd program director's request, we used the school administered cognitive assessment as the estimate of cognitive functioning for these students. We report means and standard deviations across these cognitive measures for the students with disabilities in Table 3.

Test of Word Reading Efficiency. Two norm-referenced subtests (Sight Word Efficiency [SWE] and Phonetic Decoding Efficiency [PDE]) from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) were administered individually to students receiving intervention. SWE requires students to read a list of printed words; PDE requires students to read a list of pronounceable nonwords. We recorded the number of words or nonwords read accurately within 45 seconds. Reliability and validity on the SWE and PDE ranged from .83-.96 and .89-.92, respectively.

Measures to Monitor Progress

Curriculum-based oral reading fluency (CBM-ORF; Pearson, Inc., 2012) and maze passages (CBM-MAZE; Shinn & Shinn, 2002) were used to measure reading progress and assess the effects of instruction (Deno, 1985). CBM-ORF passages provide the number of words a student reads aloud correctly in one minute. Three CBM-ORF passages were individually administered to BRIDGES students, with the median score reported here. CBM-ORF reliability ranges from .82-.94 and validity from .66 to .83.

Eighth grade CBM-MAZE passages were group administered and require students to read a passage silently with every seventh word deleted and select a correct word to fill the deletion. Students read and make as many selections as possible in 3 minutes. We administered three passages and recorded the median score. Maze reliability ranges from .86-.99 and validity from .75-81.

Test of Silent Contextual Reading Fluency. The Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) is a norm-referenced test intended to measure general silent reading ability. The TOSCRF was group administered, alternating between the four equivalent forms. Students read short passages

arranged in rows of contextually related words without spaces or punctuation between the words and draw a line between the boundaries of as many words as possible in 3 minutes. Reliability ranges from .82-.99, and validity from .61-.89.

Tests of Learning During BRIDGES

We developed decoding, vocabulary, and cause and effect relationships quizzes that were individually (word) and group (vocabulary and cause/effect) administered to intervention students before and after each 3-week unit that addressed these instructional goals. Quizzes were used to assess learning and also to make adjustments in lesson design across instructional cycles.

Experimenter-designed Decoding Quiz. The decoding portion of each quiz consisted of a sample of specific words taught during the lessons (see Intervention section for a description of how words were selected for decoding instruction.). Cycle 1 introduced the decoding strategy, and so nearly half the taught words were tested. Students read a list of 48 words aloud prior to and within one week of completing Cycle 1. Words included –y endings, and the affixes -tion/-sion, con-, -ive. In Cycles 2 and 3 the decoding lists on quizzes were shortened to 11 and 14 words, respectively, to make room for the additional instructional targets of vocabulary and cause/effect relationships in later cycles, thereby keeping total testing length comparable across cycles. Cycles 2 and 3 included words with the patterns introduced in Cycle 1, and affixes –ish, -ive, dis-, and intra- were added. Across lists, 31% of words comprised two syllables, 28- 29%, respectively, comprised three syllables, 29-31% comprised four syllables, and 11% comprised 5 syllables. Student responses were tape recorded for accuracy of scoring. Although no time limit was established for reading the words, we generated two scores:

(1) the total time to read the words, and (2) an accuracy rating. Correctly reading the word earned two points; reading the word with inappropriate inflection or dropping a suffix earned one point. Taught and tested word lists are available from the author.

Experimenter-designed Vocabulary Quiz. Vocabulary words were selected from the thematic units found in the history texts (this selection process is described later). Three vocabulary scores were obtained for each student: a self-rating of word knowledge, scores for the definition of each target word and use of each target word in a sentence. Before raters asked students for definitions and sentences, students were asked to rate their own word knowledge using a 1 to 4 numbered scale adapted from Dale and O'Rourke's (1981) 0 – 3 scale: (1) I've never heard this word before; (2) I've heard of this word but I don't know what it means; (3) I've heard of this word and I think it has something to do with ___; and (4) I know this word and it means ___. If students responded with a 3 or 4 rating, the student was asked to define the word and supply a sentence. Student responses were audio taped and transcribed.

Similar to others (Cain, Oakhill, & Lemmon, 2004; van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001), we developed a 3 point scale (0 = incorrect, 1 = partially correct, 2 = correct) to describe the quality of student-generated definitions for target words. Scoring rubrics are available from the authors. See Table 5 for sample definition and sentence scores, beginning with Cycles 2 (for definitions) and 3 (for definitions *and* sentences) of instruction.

Experimenter designed Cause/effect quiz. These quizzes were group administered to intervention students before and after instruction on cause/effect relationships in Cycle 3. They assessed (1) whether students could apply the cause/effect

strategy independently to historical content that was addressed during the lessons (a test of learning), and (2) whether students could apply the strategy independently to extract and organize information from new text material at a similar level of difficulty (near transfer). Regarding instructional design, if students can apply strategies to previously taught material, but not to untaught material, more direct instruction, scaffolding, and repetition of the strategy may be necessary. We compared scores on these assessments with those collected from typical readers (i.e., students whose reading ability was too high to be selected for BRIDGES) from the same general and special education classrooms as the BRIDGES students.

Students were asked to identify the cause and effect relationships within 4 distinct passages that ranged in length between 50 and 81 words. Students indicated their responses by underlining and labeling causes and effects with a C or E, respectively, and circling the signal words. Two cause-effect formats were presented (see Williams, 2007, for test format). The first involved direct cause-effect relationships that included a linking signal word between each cause and effect clause (e.g., Children had to read books written for grown-ups <u>because</u> there were no storybooks for kids). The second cause-effect format was used in two passages, in which one cause was paired with two effects in separate sentences in the paragraph (one cause with multiple effects). Scoring is available from the authors.

Classroom history tests. Classroom history tests assessed retention of content following units of instruction, and although reliability could not be established, we monitored student progress prior to, during, and following cycles of intervention. Regarding instructional design of the DBR cycles, if students' scores significantly drop

from pre-intervention scores, declines would signal adjustment of content delivery in our lessons. Classroom history tests were group administered to BRIDGES students and typical readers in GenEd and SpEd classes by participating classroom teachers classes approximately every three weeks. Classroom history tests comprised multiple-choice questions related to the unit of instruction and were developed collaboratively by the GenEd history teachers at Hamilton to measure recall of facts, people, and events.

Intervention

We developed the intervention components sequentially and cumulatively, using our first instructional cycle to fine-tune a strategy for breaking apart and reading multi-syllabic words, our second cycle for teaching academic and history-focused meanings of these words (alongside decoding them) and our third cycle for analytic reading to identify cause/effect relations among historical events (alongside decoding and understanding meanings for academic and history-focused vocabulary). Because our research team delivered these lessons to small groups in a pull-out setting during the first three weeks of each cycle, we could make frequent adjustments during and after each cycle based on our documentation of student engagement and learning, and ongoing meetings with teachers during which we discussed the strategies and feasibility, showed videotapes of their students engaging in the strategies, shared data on student learning, and gathered teacher feedback. During week 4 of each cycle , we met again with teachers to design lesson segments they would field-test with their intact classes--including BRIDGES students—during the next week of upcoming content (week 5 of the cycle).

Cycle 1: Decoding Multisyllabic Words

Selecting words to decode in Cycle 1. We began by entering two lists of words into an excel database--the Cox Academic Word List (Coxhead, 2000) and the school's list of essential history words (Gr 7-8)--with columns by source. Next, we analyzed each chapter in the school's adopted history text that would be included in each cycle and generated a list of words that could be difficult for students with a 3rd grade reading ability because they contained multiple syllable and often were morphologically complex (e.g., constitutional, declaration, unavoidable). Last, we performed the same analysis on the Powerpoint slides used by classroom teachers during the unit and added any additional words. This file allowed us to cross reference sources for each potentially difficult word in the text, along with breadth of usage across sources. Specifics regarding initial and reduced lists are available from the authors.

Teaching students to decode multisyllabic words. Our procedure was based on a combination of the rule "Every syllable has at least one vowel (ESHALOV)" and BEST (O'Connor, 2007), which shares features with Lenz et al.'s (1990) DISSECT. ESHALOV required students to: (1) underline all of the vowels in a long word (e.g., unavoidable), (2) join any vowel teams into one vowel sound (i.e., oi), (3) identify known word parts (i.e., un-, -able), (4) count the number of word parts to expect (i.e., 5), (6) break the word into parts for decoding (i.e., un-a-void-able), and (7) try a pronunciation of the word. BEST included the steps: Break it apart, Examine the base word, Say each part, and Try the whole word, and follows researched recommendations for adolescents (Kamil et al., 2008).

During Cycle 1, about 15 minutes each day focused on teaching and rehearsing the strategy, and 30 minutes on reading and discussing history text at relatively easy levels (i.e., average readabillity Grade 3.8). By Week 3 about 10 minutes daily focused on the decoding strategy, because most students were using it independently, and we added short segments of 8th grade history text that contained these words. During Cycles 2 and 3, only 5 and 3 minutes per day respectively were spent on decoding words. When students encountered a word in text they could not read, they were encouraged to use ESHALOV or BEST to read the word.

Cycle 2: Learning Academic Vocabulary

Selecting vocabulary words to teach in Cycle 2. Similar to the process for selecting words for decoding, we selected vocabulary words after analysis of the chapter content to be taught during the cycle. We considered the reading level of our students, whether the word was critical to understanding the text, how frequently the word appeared in the text, and the contexts students would be able to access for variations in meaning and usage. As before, we cross-referenced our initial list with the Coxhead Academic Word List (Coxhead, 2000), the school's list of essential history words, and teacher unit slides and study guides. For teaching meanings, we also cross referenced our lists with Biemiller's *Words Worth Teaching* (Biemiller, 2010) and grade-level bands associated with those words.

Our goal was to select for instruction words that were: 1) used more than twice in the text, and 2) appeared on multiple lists, which ensured that selected words were central for understanding the history content of the unit. The reduction process resulted in a focused list of 18 words for Cycle 2 and 14 for Cycle 3.

Teaching the meanings of words. We followed the recommendations of Beck, McKeown, and Kucan (2013) and Lawrence, Capotosto, Branum-Martin, White, and Snow (2012) for teaching word meanings in iterative and interactive ways that involved student-friendly definitions and applications, multiple guided practice opportunities to use the words in varying contexts, and small group discussions where students used the words meaningfully with each other. Examples from Cycle 2 are available upon request from the authors. Cycle 2 lessons included about 5 minutes of decoding practice, 15 minutes of vocabulary development, and 20 minutes of reading and discussing history text at easier and more difficult levels.

Cycle 3: Identifying Cause and Effect Relationships

Students were introduced to the cause-effect comprehension strategy with a direct instruction lesson that oriented students to the purpose of cause and effect in reading and how to identify a cause-effect relationship using transitional signal words. All components were defined for students (e.g., cause, effect, signal word) and real-life events were used to illustrate cause-effect relationships (e.g., What was the effect of the tornados in Oklahoma? And what caused the damage?), with multiple opportunities for students to discuss their reasoning. The key steps of the cause-effect strategy were drawn from earlier instructional studies (see Gersten, Fuchs, Williams, & Baker, 2001; Solis et al., 2012; Williams et al., 2007) and included the following: (1) Read the passage, (2) Identify any signal words, (3) Reread phrases before and after signal words, (4) Identify cause and effect using signal words, (5) Check other phrases/sentences that could indicate cause or effect, and (6) Check your decisions (e.g., Which event happened first?)

Passages from the history textbook were modified to ensure cause-effect discourse markers were included within the passage, which focused on the Texas Revolution and Mexican-American War. Lessons were structured to discuss the causes and effects of the historical events within the passages the students read. In small groups or individually, students organized the events on graphic organizers under teacher direction. Cycle 3 lessons included about 2 minutes of decoding practice, 8 minutes of vocabulary development, 15 minutes of practice identifying cause-effect relationships among events, and 15 minutes of reading and discussing history text aside from cause-effect passages.

Observations and Fidelity

In a design experiment like this one, observations are used to determine fidelity to the current instructional script, but also to document areas in lesson content and pacing that may require adjustment for future lessons. We scripted each lesson and documented degree of adherence to that script; however, at the end of each day, the team (instructors and observers) discussed potential changes based on student engagement and learning, areas of student misunderstanding, and areas where instructional talk was unclear. More changes were made as we showed videos to teachers and gathered their concerns and suggestions for feasibility. During the first week of each cycle, each tutor was observed daily with the goals of improving student engagement and lesson delivery, pacing, and record keeping. Fidelity to implementation was documented in Weeks 2 and 3, and procedures are available from the authors.

GenEd and SpEd classroom teachers were observed twice as they implemented the target feature of the cycle (decoding in Cycle 1, vocabulary development in Cycle 2,

or cause/effect instruction in Cycle 3) during Weeks 4 or 5 of each cycle. Fidelity of teacher implementation in each cycle is reported in the results.

Results

We asked: (1) Will students make gains in key reading skills through reading lessons that focus on historical content? (2) Does improvement differ between students with disabilities, other low achievers, and ELs? (3) Does this reading instruction improve knowledge of historical content? (4) To what degree of fidelity do coached teachers implement reading instruction during history class? and (5) How do teachers view feasibility in their roles as general and special education history teachers? First, we address inter-rater agreement for our experimenter-designed measures, and then present data to answer the questions above.

Inter-Rater Agreement for Experimenter-designed Measures

A scoring rubric was developed and practiced on a random selection of 30 decoding and vocabulary assessments from pre and post testing periods. Details on calibration are available from the authors.

Decoding tests were scored immediately after administration by testers using the tape recording for accuracy, and 20% of all decoding tests were double-scored by the third author. Consistency estimates of interrater reliability were high across all cycles: intra-class correlations (ICC) were .995,,.978 and.990 for Cycles 1, 2, and 3, respectively.

Vocabulary tests were blinded so that student intervention status and time of testing (pre or post) was unknown to the raters. Agreement for definitions (Cycle 2. K = 0.78; Cycle 3, K = .79) and sentences (Cycle 3, K = .75) fell within the substantial agreement range according to Cohen's parameters (.41- .60 = moderate agreement; .61-

.80 = substantial agreement; .81-.99 = almost perfect agreement). Disagreements were discussed and resolved; final scores reflected consensus from the two raters.

Cause-effect tests were scored by the third author using the scoring rubric developed by all authors, with 15% of the assessments were double-scored by the second author. The consistency estimate of interrater reliability was high between the two raters, ICC = .99.

Student Growth in Reading Skills

To answer our research questions, we executed a series of repeated measures ANOVAs with time as the within- and SpEd status (SpEd or GenEd) and EL status (EL or English Only) as between-subjects factors (unless otherwise noted). Eta squared was used as the measure of effect size (ES). Where applicable, assumptions of sphericity were met unless noted, in which case tests with adjusted degrees of freedom were used. ORF and MAZE assessments were administered on four occasions (October, December, February, and May). TOSCRF and experimenter measures of decoding, vocabulary, and cause-effect relations were administered pre and post each intervention cycle.

ORF. Raw scores for BRIDGES students fell at or below the 15th percentile (based on national norms) at each time point. Because reading rate (words read correctly per minute) was low, we conducted separate analyses for rate and accuracy. The repeated measures ANOVA for rate revealed no significant interactions, all *p*s > .05. We found a significant main effect for time, $F(3,93) = 4.016$, $p = .01$, ES = .115, indicating significant change in wcpm across measurement occasions. The test of between subjects effects for SpEd (pooling over time) was significant, $F(1, 31) = 8.65$, $p = .006$, ES = .281.

Examination of univariate tests and plots revealed that GenEd students outscored SpEd students at each time point (all *p*s < .05).

We calculated accuracy by dividing the total words read correctly by the total words attempted. BRIDGES students read with an average of more than 90% accuracy at each time point. The assumption of sphericity for the repeated measures ANOVA was violated according to Mauchly's Test, $X^2(5) = 26.8$, $p < .001$. The Greenhouse-Geisser test of the main effects for time F(2.14, 70.6) = 8.84, $p < .001$, ES = .21, and time by SpEd interaction, F(2.14, 70.6) = 7.56, $p = .001$, ES = .19, were significant. Tests of within-subjects contrasts for the interaction were significant for Time 1 (T1) to Time 2 (T2), F(1, 33) = 6.5, $p = .015$, ES = .17 and Time 3 (T3) to Time 4 (T4), F(1, 33) = 11.57, $p = .002$, ES = .26, which mirrors the decoding effects reported later (i.e., SpEd students required more time to apply the strategy consistently). An examination of plots showed a steep decline in ORF accuracy scores for SpEd students between T1 and T2, but not for GenEd students. Additionally, accuracy scores increased sharply from T3 to T4 only for SpEd students. On average, general education students' accuracy scores remained consistent across time.

Maze. Average raw scores across intervention students for MAZE at each time point fell at approximately the 15th percentile compared to AIMSweb published norms. The repeated measures ANOVA showed a significant main effect for time, F(3, 84) = 19.142, $p < .001$, ES = .406. The interaction effects for time by SpEd, time by EL, and time by SpEd by EL were not significant (*p*s > .05), suggesting that all groups made progress.

TOSCRF. Standard scores on the TOSCRF for intervention students fell below national averages (M=100; SD=10) at each time point (all ps <.001). The repeated measures ANOVA indicated a main effect for time, F(3, 87) = 26.3, p<.001, ES = .476. The time by SpEd, time by EL, and time by SpEd by EL interaction effects were non-significant (all ps > .05). Tests of within subjects contrasts for time and for interactions showed significant change from each time point to the next (all ps<.001), with a significant time by SpEd interaction from T1 to T2 (p = .014). Examination of plots revealed steeper growth for GenEd students in BRIDGES from T1 to T2 than for SpEd students. The remaining tests of within subjects contrasts for the interaction were non-significant, ps >.05.

Student Growth on Experimenter Measures

Recall that decoding was taught in Cycle 1 and reviewed in Cycles 2 and 3, vocabulary meanings were taught in Cycles 2 and 3, and cause-effect relationships were taught in Cycle 3, which also included small amounts of decoding and vocabulary instruction. Table 4 shows scores on the experimental measures for intervention students, and where administered, for their typical reader classmates who did not participate in BRIDGES, but were taught the same history content. These students scored too high on the 7th grade state assessment to be included in BRIDGES pull-out instruction; however, they were in the same GenEd history classes as BRIDGES students.

Decoding, Cycle 1. The maximum possible score for our researcher developed decoding assessment was 96 points for Cycle 1. The repeated measures ANOVA showed a significant main effect of time, F(1, 33) = 123.03, ES = .79, and a significant time by SpEd interaction, F(1, 33) = 12.41, *p* = .001, ES = .27. Change in scores from pre- to

post-test was steeper for Sp.Ed than GenEd students. The time by EL and three way interaction were non-significant. Follow-up ANOVAs for pre- and post-test decoding scores by SpEd status revealed a significant main effect for SpEd status, $F(2, 32) = 7.82$, p=.002, where GenEd students outscored SpEd students at each time point, ps < .01.

Decoding, Cycle 2. The maximum score for the decoding assessment was 22 points in Cycle 2. The repeated measures ANOVA showed a significant main effect for time, $F(1, 21) = 29.74$, $p < .001$, ES = .59, and a significant time by SpEd interaction, $F(1, 21) = 4.68$, $p = .042$, ES = 0.18. Once again, change in scores from pre- to post-test was steeper for SpEd compared to GenEd students. The time by EL and three way interaction were non-significant. Follow-up ANOVAs using pre- and post-test decoding scores by SpEd revealed a significant main effect for SpEd status at pre-test, $W(1, 21.89) = 11.21$, $p = .003$, where GenEd BRIDGES students outscored SpEd students. Welch's test was used due to significant differences in error variances across groups at pre-test. No significant differences between groups were apparent at post-test.

Decoding, Cycle 3. The maximum score in Cycle 3 was 28 points. Unlike Cycles 1 and 2, the Cycle 3 repeated measures ANOVA did not indicate interaction effects; all students grew significantly and similarly with time, $F(1, 28) = 36.01$, $p < .001$, ES = .56. Tests of between subjects effects revealed a significant main effect for SpEd (pooling across time and EL), $F(1, 28) = 8.51$, $p = .007$, ES = .23. On average, GenEd students outscored SpEd students. Univariate ANOVAs using the Kruskal-Wallis test for post-test scores due to non-normality suggested the effect was significant at pre-, $p = .015$, and post-test, $p = .04$.

Vocabulary, Cycle 2. Recall that academic vocabulary was explicitly taught and assessed only in Cycles 2 and 3. The maximum score for vocabulary definitions was 22 points. The repeated measures ANOVA demonstrated a significant main effect for time, $F(1, 25) = 69.06$, $p < .001$, ES = 0.73. The time by SpEd and time by EL effects were non-significant as was the three way interaction. Therefore, SpEd, GenEd, EL and EOs experienced similar changes in their ability to define taught academic words, pre to post intervention.

Vocabulary, Cycle 3. The maximum score for vocabulary definitions and sentence production was 28 for Cycle 3. The repeated measures ANOVA for definitions revealed a significant main effect for time, $F(1, 30) = 95.17$, $p < .001$, ES = 0.76, and significant interaction effects for time by SpEd, $F(1, 30) = 5.78$, $p = .025$, ES = .16, and time by EL, $F(1,30) = 7.09$, $p = .012$, ES = .19. Change in scores from pre- to post-test was steeper for GenEd and EO students. The 3-way interaction was non-significant.

The repeated measures ANOVA for sentence scores revealed a significant main effect for time, $F(1, 30) = 96.46$, $p < .001$, ES = .76, and a significant time by EL interaction, $F(1, 30) = 20.178$, $p < .001$, ES = .40. Change in scores from pre to post-test was steeper for EO students. The time by SpEd and three way interactions were non-significant.

Cause and Effect Relationships, Cycle 3. The maximum score for identified causes, effects, and signal words was 32 points. The repeated-measures ANOVA for pre- to post-test total scores indicated a significant main effect for time, $F(1, 30) = 11.197$, $p = .002$, ES = 0.27. Pooling across groups, students identified more causes, effects, and signals at post-test ($M = 18.76$; SD = 6.99) than at pre-test ($M = 11.65$, SD = 7.52). The

between-subjects effect for SpEd was also significant, $F(1, 30) = 5.051$, $p = .032$, ES = 0.14). On average, SpEd students outscored GenEd students. Univariate ANOVAs indicated the effect was significant at post-test only, $p = .030$. The time by SpEd, time by EL, and three way interactions were not significant.

Comparisons with Classroom Peers

Because this design experiment did not have a typical control group, we compared pre- and posttest performance of BRIDGES students to their average reader classmates (e.g., students in the same GenEd classes that BRIDGES students attended when not receiving BRIDGES pull-out instruction). We executed repeated measures ANOVAs with time as the within subjects and intervention status (BRIDGES vs. typical reader peers) as the between subjects factor on measures of vocabulary sentences and cause-effect relationships in Cycle 3. BRIDGES students and typical reader classmates were exposed to the same history vocabulary and content during the same time frame; typical reader classmates received instruction as usual from their general education teachers.

Vocabulary, Cycle 3 Comparison. The repeated measures ANOVA indicated a significant time by intervention status interaction, $F(1, 61) = 50.32$, $p < .001$, ES = .45. Examination of a line plot suggested steeper change from pre to post for BRIDGES students (Figure 1), with virtually no change for typical peers. Follow-up univariate tests revealed significant differences for pretest sentence scores, $F(1,61) = 48.85$, $p < .001$, ES = .45 with typical readers outscoring BRIDGES students, and no significant differences for post-test scores, $p = 0.38$.

Cause-Effect Relationships, Cycle 3 Comparison. The repeated-measures ANOVA for pre- to post- total scores indicated a significant time by intervention status interaction, $F(1, 125) = 23.368$, $p < .001$, ES = 0.16. Examination of a line plot demonstrates steeper change from pre to post for BRIDGES students (Figure 2), with virtually no change for typical peers. Follow-up ANOVAs suggested that BRIDGES students scored fewer points (M = 11.42) at pre-test than their typical-reader peers (M = 19.70; $F(1,126) = 23.320$, p =0.000)). At post-test, we found no significant difference between intervention students (M = 19.33) and their typical-reader peers (M =18.15; $F(1,128) = .435$, p =0.511).

BRIDGES Student Achievement on Class U.S. History Tests

A major concern in pulling students out of their scheduled U.S. History classes to teach reading skills was whether their grades in history might suffer, and we frequently discussed this issue with the teachers and administrators at Hamilton. During instructional cycles, students participated in BRIDGES 4 days per week, with the 5th day back in class taking the same history quizzes and exams as their class peers who were average readers. We collected class quiz and test grades during the first 8 weeks of the year prior to BRIDGES participation, during the BRIDGES instructional cycles, and during the 4th quarter following the intervention. Grading mechanics differed considerably across teachers, and so we did not compare quarter grades across teachers or classrooms.

Determining differences for students in SpEd proved difficult, because test administration was highly accommodated and grade totals depended also on extra credit and service learning. These opportunities for earning extra points were diminished

during students' participation in BRIDGES, and we found no discernible difference in grades prior to and during BRIDGES participation.

Differences for BRIDGES students in the GenEd classes were more apparent; however, measurable differences on quizzes and tests did not necessarily translate into improved quarter grades. For 20% of the GenEd students, percentage correct on tests was so low prior to BRIDGES that large improvements still generated near-failing grades (i.e., scores ranged prior to BRIDGES between 37-43% for this group of students, and rose following the three-week BRIDGES cycles to 62-66%). Half of the GenEd participants improved their grades from failing (average 58%) to the C range (average 74%) on class tests. Another 30% of GenEd participants had some of their grades 'excused' due to BRIDGES participation; however, 3 of these students scored above 90% on the test of the Constitution and Bill of Rights (taught during BRIDGES), which was referred to by two participating teachers as the "Mother of All Constitution Tests." Unfortunately, during the 4th quarter when students worked full time in their GenEd history classes, they did not maintain their higher level of performance on new material taught in these classes.

Teacher Implementation and Fidelity of Instruction

Using teacher recommendations from discussions of feasibility of implementation in their classes, the research team provided each of the five teachers with materials to pilot 4 unique lessons for each of the three strategies, using the history content they suggested for the 5th week of each cycle. In almost all cases, teachers were observed across two pilot lessons in each cycle, and inter-rater reliability was calculated at 96.67% over six shared observations.

The observation instrument noted specific characteristics of the teaching that were integral to lesson fidelity. We did not rate the quality of the teacher implementation of these specific behaviors, but only the presence of the teaching behaviors, because we wanted to determine the feasibility of these lessons for teachers in real world settings: across GenEd and SpEd teachers with real and persistent content coverage pressure. The lessons developed for teachers to pilot were planned to take 10-15 minutes, targeted the strategy that was developed during the preceding cycle, and used the instructional content teachers had planned for that week. The teachers also requested that pilot materials be suited to whole-class presentation with a teacher-directed lecture style format, with accompanying visual aids (Powerpoint slides) and opportunities for students to write during the lecture (e.g., Cornell Notes, which was the method of note-taking students would be required to use in high school the following year). Table 6 notes the mean scores across the observed lessons for each of the strategies in categories that were common to two or more lesson strategies.

Decoding. The decoding lesson was the most scripted of the three strategies. Teachers were coached to model the process, use direct instruction to assist students in identifying the vowels, break the word apart mindfully, and then blend the word together as they monitored these activities for student accuracy. Because many 8th graders need little help with decoding, we used unusual words (e.g., alektorophobia) in the GenEd classrooms for teachers to try out the strategy with a range of readers. Across the observed lessons, we saw modeling of the process 28.57%; direct instruction on vowel teams 42.86%; direct instruction in mindfully breaking apart the word 14.29%, and checking for student accuracy 42.86%. GenEd teachers told us that decoding instruction

was not expected of history teachers, and only two of the five teachers had course work or professional development in word reading.

Vocabulary. The vocabulary instruction included student friendly definitions and sentences, and historical context sentences. Students rehearsed word meanings, applied them in new contexts, and extended the provided definition to new sentences. Teachers always presented these sentences as provided in the pilot lessons; however, they were unlikely to give students the opportunity to repeat the word alone or in a sentence (mean 2.26 times per lesson) or to provide opportunities for student discussion (whole class or small group) to discuss word meanings or create sentences using the word.

Cause and effect relationships. The Cause and Effect strategy taught students to recognize inferential discourse markers as tools to identify cause and effect. The specific steps in these lessons were followed frequently by the teachers: introducing and reviewing the strategy and specific words that signaled cause and effect relationships, and providing guided and independent practice to read and apply these steps. Teachers were less likely to encourage student discussion of passage meaning (44.4%) or to ask students to justify their selections (22.2%).

Teacher Comments on Feasibility

We collected data from teachers using a 7-item anonymous questionnaire regarding the feasibility of using BRIDGES strategies in the classroom. These data are summarized in Table 7. Teachers found decoding of least value of the three strategies, although the SpEd teacher mentioned continuing use of ESHALOV after BRIDGES cycles ended. Teachers perceived decoding as unnecessary for most of their students, and indicated little confidence in their ability to teach decoding successfully (3.0 on a 5

point scale). Conversely, teaching academic vocabulary and identifying cause and effect relationships were perceived as both useful for the majority of students and as valuable. The confidence and ease with which teachers used the cause and effect strategy was also high (4.0).

Teachers viewed strategies as moderately difficult for struggling students to use, despite their success in BRIDGES. Based on our observations of teacher implementation, this perception may be due to teacher inexperience with direct explicit instruction, a critical feature of BRIDGES. We noted that teachers asked students for individual or choral repetition of key terms, strategies, vocabulary less frequently than BRIDGES instructors; that teachers provided information without checking for understanding across students; and that teachers were unlikely to ask students to 'share out' to the class or with partners/small groups as a basis for corrective feedback.

Teacher Discussion of Implementation

The end-of-year focus group interview of participating teachers was conducted and recorded by a faculty colleague who had not been part of the daily operation of BRIDGES. Teachers responded positively to the improved motivation and confidence observed in students who participated in BRIDGES. As one teacher commented, " … the effects of this project may not be revealed to us immediately. Being able to decode and access text might just give them what they need to finish school." The renewed interest in learning carried forward to the GenEd classroom: teachers noted that students worked to complete assignments more frequently following BRIDGES cycles, and to participate in discussions. Negative peer pressure at times competed with the students' public display of renewed interest in learning; however, as teachers noted, "even when

their hands started going down in front of their peers, they still had now the confidence to come up and see us privately. They may have gotten the notion that approaching a teacher for a help or support at this level was almost unacceptable. They would not want to divulge that they didn't know how to do these things and they may even have not thought that the teacher would be willing to work with them. So I started seeing some of those students in the classroom when things were unclear, they started coming up to us for support and they were coming up to us with confidence. That was important."

Teachers acknowledged the gap between the amount of history content covered in BRIDGES and their own classrooms. Moving at a pace that was appropriate for the BRIDGES students, as well as incorporating vocabulary and decoding skills, resulted in fewer topics taught. However, these topics were taught to a mastery level for the majority of BRIDGES students, as demonstrated on the BRIDGES content assessments and also for the test on the US Constitution, which was designed by the classroom teachers. Nevertheless, even for a three-week, three-cycle pullout for students who were failing in their classrooms and also failed 7th grade history, teachers struggled to justify the differentiated instruction provided by BRIDGES. They were acutely aware of pacing pressures, state wide testing, the integrity of grade reports, and adherence to coverage of history content.

Conversely, teachers appreciated the BRIDGES strategies for vocabulary and cause and effect *because* they differentiated instruction. One teacher noted…" it absolutely fits (other teachers murmur yes), and in terms of the classroom, we have a huge spectrum, and this is a great opportunity for differentiation in our classroom. It

really is, you can have this whole huge range of words, and let kids learn as many as they can."

Discussion

Finding time to provide intensive reading instruction is more difficult in secondary than in elementary schools because course work required for graduation limits opportunities for teaching reading to poor readers. Moreover, reading instruction in middle school is often eliminated in favor of tutoring support for passing courses (Bulgren et al., 2013). For these reasons, along with the history teachers' concern for the upcoming transition toward the CCSS, we focused reading intervention in this DBR on history content, and supplanted history instruction during BRIDGES cycles. Key findings included growth during BRIDGES instruction on experimenter measures of decoding, vocabulary, and identifying cause and effect relationships, and on contextual reading (i.e., Maze and TOSCRF scores). Growth in academic vocabulary and cause and effect relationships was greater for BRIDGES participants than for typical readers in their history classes. Few interactions were found for disability or EL status in growth on experimenter or standardized measures. Students also improved their knowledge of history content; however, their failing grades prior to and in-between BRIDGES cycles made these gains difficult for their teachers to reward through quarter and semester grades.

Generalized Gains in Reading Skills

Students increased their silent reading rate across the three cycles, which is somewhat unusual for studies with disabilities in middle school (e.g., see Denton et al., 2008; Vaughn et al., 2013), although gains in oral reading rate for our participants were

minimal at best. We found a consistent pattern across progress monitoring measures in which students with disabilities required more time to consolidate reading skills sufficiently to apply them in general reading tasks such as Maze passages and TOSCRF.

Likewise, BRIDGES students without disabilities responded to the decoding instruction very quickly, and began to use the ESHALOV and BEST strategies within days of initial instruction, whereas students with disabilities took more time to internalize the strategies and made their strongest gains after a full cycle of instruction and practice. Examination of Table 4 hints at generalization of decoding skills for the students with disabilities. During Cycle 1, their scores improved from 41 to 65%. Their Cycle 2 pretest with new multisyllabic words began at 58% and rose to 93%, even though instruction in decoding decreased from 15 min in Cycle 1 to about 5 throughout Cycle 2. This pattern continued in Cycle 3 (from 61% to 88%), with a new collection of words and only 3 min of daily decoding instruction included in BRIDGES lessons. In Cycles 2 and 3, 20% of tested words included an affix taught in the previous cycle(s). Thus the successive gains on pretests of new words across cycles provide evidence of generalization of the decoding strategy to novel words.

On average, students with disabilities scored below students without disabilities on most measures at each time point. Nevertheless, few time by disability or EL status interactions were significant, which suggests that growth during intervention did not differ across the three subgroups of poor readers (GenEd, SpEd, and EL).

Knowledge of Vocabulary, Cause and Effect Identification, and History Content

Vocabulary. Unlike decoding strategies, which can be applied to any new word, vocabulary knowledge tends to be word specific. On experimenter measures of the

academic vocabulary central to the covered units of history, students grew significantly over time during each instructional cycle. To place these gains in perspective, the measures of vocabulary and cause/effect relationships were also administered to typical-reader class peers before and after instruction by their general class teachers of the same classes from which BRIDGES students were drawn. These teachers had taught the same content units we taught in BRIDGES, but with only incidental vocabulary instruction.

Vocabulary pretest scores for BRIDGES students were predictably low (below 16%); however, the academic vocabulary scores for their typical reader peers, while better, were also low (44% correct). During Cycle 2 when vocabulary became the prime target, BRIDGES students grew significantly on knowledge and use of these words. In Cycle 3, when focused instruction and practice was reduced to 5 minutes per session, BRIDGES students continued to grow in use of the new words; however, their typical reader classmates made no notable gain on vocabulary use even though their history teachers used these words during their instruction, and had recommended that we teach them. Typical readers made no perceptible improvement in vocabulary during routine instruction, remaining below 50% correct after teachers completed their unit. If 5 minutes of interactive instruction can improve understanding of key content words for most students with and without disabilities in their classes, this time might be well spent.

Despite this recommendation, the only differential effect for students who were EL was for vocabulary in Cycle 3, during which EL scored lower than native English speakers at pre- and posttest. Recall that vocabulary development was the major instructional focus during Cycle 2, with 15 minutes of daily vocabulary instruction and practice. As we introduced the cause and effect strategy in Cycle 3, we decreased the

time spent on vocabulary to about 5 minutes per session. The main effect for EL status in Cycle 3 suggests that EL students may need more than 5 minutes of daily practice to learn and retain meanings of academic and history content words, which suggests an area for differentiating instruction.

Cause/effect relationships. The pretest of cause/effect relationships suggested considerable guessing by BRIDGES students, with substantial reduction in guessing on the posttest (30 errors on the pretest, compared to 9 errors on average on the posttest). Although the main effect between students in BRIDGES and their typical-reader peers was significant at pretest for measures of both vocabulary and cause/effect relationships, by posttest the groups performed similarly (see Figures 1 and 2). Students receiving BRIDGES instruction showed strong gains in academic vocabulary and identifying cause/effect relationships. Although their typical reader peers began each unit with more knowledge, they grew only slightly in connecting that knowledge to cause effect relationships with GenEd instruction. That the same historical events and vocabulary were taught in both settings lends support to the conclusion that BRIDGES instruction was responsible for these improvements in academic language and use of cause/effect strategies.

During the BRIDGES cycles, students improved their knowledge of history content as demonstrated on tests administered by their classroom teachers, with average gains of over 20%, or two letter grades. Nevertheless, content is a moving target, and so large improvements on one set of vocabulary words, or knowledge of the constitution or of westward expansion, did not transfer to new vocabulary words or post Civil War reconstruction. To keep poor readers actively engaged in reading and studying history,

and to maintain their gains on classroom history tests, teachers would need to change their instructional approach.

Feasibility in General and Special Education History Classes

While teachers in general education readily admit that history content presented through grade level texts is not accessible to low achieving students, available history materials offer scarce support for differentiating instruction in presentation, student assignments, or the scaffolding for students of differing ability. Our overarching goal was to explore how history teachers could teach reading skills through U.S. History content. We conducted this exploration through iterative development with students and their teachers in design-based research. Observations in history classes prior to implementing BRIDGES revealed that 8th graders with high-incidence disabilities and other students reading at primary-grade levels failed to engage in instructional activities of reading text, discussing, or answering questions, most likely because they could neither read the words in history texts and Powerpoint slides nor understand what they meant.

Our theory of change posited that by integrating word analysis instruction with age-appropriate words, vocabulary instruction that develops conceptual understanding, reading easier texts that bridge to general education materials, and comprehension strategies that enable students to understand the basis behind historical events, we could improve student's reading ability and perhaps their understanding of history, as well. In this goal, we succeeded.

Nevertheless, the ultimate usefulness of this work rests not only on improving reading skills for students with disabilities and poor readers, but the suitability of the

instructional approach for teachers in special and general education classrooms. The realities of middle school contexts, the inaccessibility of grade level texts to poor readers, and the need to differentiate instruction seamlessly, all factor into the validity of this instruction for teachers. By incorporating teacher recommendations and small tests of feasibility along the way, we hoped to set a stage for building teacher capacity and the potential for sustaining changes that could support student engagement and learning. In this goal, our success was limited.

With teacher cooperation, we built activities we thought would encourage teachers' "take-up" of jointly constructed routines. GenEd and SpEd teachers met with BRIDGES staff at mutually agreeable times several times during each implementation cycle and observed and rehearsed using BRIDGES strategies through videotaped lessons, modeling of strategies during meetings, and modeling with their students in their history classes. Prior to their implementation, teachers shared their concerns and suggestions, which were incorporated in the lesson development for the feasibility pilot in their classrooms. Following each 3-week cycle with researchers and students, teachers in SpEd and GenEd classes field-tested the target lesson components for four consecutive days following our small-group teaching and revision phases. What we learned through our observations of whole-class implementation and feedback from the implementing teachers was incorporated in the next cycle of instruction, ensuring continual improvement in lesson feasibility. Despite these efforts and teachers' interest in improving the participation of their poorest readers, implementation of some key instructional elements (providing feedback to students, allowing multiple practice

opportunities prior to individual assignments, reviewing words and strategies, actively engaging most students) was moderate, at best.

Teachers acted as resources by helping us to identify key problems, meeting with us regularly, and hosting the research on site; but also as constraints as school schedules, small proportions of students with perceived difficulties (i.e., fewer than half in all but the SpEd classes), and competing instructional demands (mandated curricular pacing) limited their implementation of strategies. With most of their instruction aimed at middle- to high-level readers, their checks for student understanding and monitoring conceptual engagement were minimal. Although they mentioned that BRIDGES students increased interest in reading words and task persistence when they returned to their classes following each cycle, they were unlikely to adopt the BRIDGES strategies connected to this change. Teachers told us that decoding instruction was not expected of them, and that although the vocabulary instruction was effective, it took too much instructional time.

Implications

The ability to read well with comprehension may be the single most important educational skill for passing high school exit exams and meeting the Common Core State Standards, and 8th grade was the year in which the Comprehensive Exam in History was first administered in this district. Principals and administrators are judged on how well students succeed on these exams and competencies. By choosing history as reading content, we selected a "high stakes" arena where motivation to improve both content knowledge and reading ability was high for stakeholders. For history teachers, improvement of reading cannot come at the expense of content; thus it became critical for

feasibility that history knowledge not suffer as reading skills were taught and improved.

Although the grade levels may vary across states, the cyclical pattern of U.S. History classes, commonly taught in 5th, 8th, and 11th grade, suggested promise; If we could develop a method that improves reading overall through the vocabulary, content, and information of U.S. History, the method and lessons could be useful at three time points in students' schooling. U.S. History is also key for poor readers seeking a GED, as well as for adult immigrants seeking citizenship. Moreover, an overarching goal of our education system is to prepare students to assume a participatory role in government in the U.S. and the world. Understanding U.S. History, as well as concomitant improvement in reading, could thus contribute to lifelong outcomes.

Nevertheless, change in public schooling comes slowly. Mandates such as Common Core State Standards are insufficient if training is not forthcoming or teachers perceive these changes as counterproductive. As Gamson, Lu, and Eckert (2013) comment, "increasing text complexity without serious attention to concomitant instructional supports is likely to further broaden and exacerbate the achievement gap without addressing underlying causes." (p. 388-389)

Gamson et al.'s comment led us to consider how much support teachers are likely to need. Participating teachers met with project staff 6 times formally and many more times informally across the 15 weeks of the three cycles. They provided the units of study, read our instructional scripts, and discussed lesson feasibility and how they would use them prior to attempting implementation. At the teachers' request, project staff prepared all slides, graphic organizers, and notebook reminders for their feedback and adjustment before teachers implemented with their students. Project staff modeled

implementation to demonstrate how to engage interactively with students and encourage small group student interaction to use the words, vocabulary, and concepts teachers identified as most important. When teachers allowed small group student interactions in their classes, we observed the poor readers who had participated in BRIDGES lessons helping their groups with the strategies they had previously learned during BRIDGES cycles. Nevertheless, after these interactive lessons, teachers tended to revert to whole-class lectures, in which poor readers had few opportunities to participate meaningfully—a pattern observed in both GenEd and SpEd classes. It is clear that the level of support provided through this research was insufficient for most teachers to implement the lessons in a manner that enabled meaningful participation of their lowest skilled students.

This study demonstrates that integrating key reading strategies with history content helps students to bridge across both domains. The question remains: Within the new pressures of the Common Core, with complex text analysis and primary sources as key components, what kinds of supports will teachers need to simultaneously differentiate instruction and incorporate basic, as well as advanced reading instruction into their daily work?

References

Alishahi, A., Fazly, A., & Stevenson, S. (2008). *Fast mapping in word learning: What probabilities tell us.* Proceedings of the12th Annual Conference on Computational Natural Language Learning. UK: Manchester.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41,* 16-25.

Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (pp. 752-785). Mahwah NJ: Erlbaum & Associates.

Beck, I., McKeown, M., & Kucan, L. (2013). *Bringing Words to Life: Robust Vocabulary Instruction*. New York: Guilford.

Bhattacharya, A., & Ehri, L.C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities, 37,* 331-348.

Biemiller, A. (2010). *Words Worth Teaching*. Columbus, OH: SRA/McGraw-Hill.

Bulgren, J., Deshler, D.D., & Lenz, K. (2007). Engaging adolescents with LD in higher order thinking about history concepts using integrated content enhancement routines. *Journal of Learning Disabilities, 40,* 121-133.

Bulgren, J.A., Graner, P.S., & Deshler, D.D. (2013). Literacy challenges and opportunities for students with learning disabilities in social studies and history. *Learning Disabilities Research & Practice, 28*, 17-27.

Cain, K., Oakhill, J., & Lemon, 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*(4), 671-681.

Carlisle, J., and Stone, C.A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*, 4, 428-449.Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311-325.

Dale, E. & O'Rourke, J. (1981). *The living word vocabulary.* Chicago: World Book/Childcraft International.

De La Paz, S., Morales, P., & Marlowe, P.M. (2007). Source interpretation: Teaching students with and without LD to read and write historically. *Journal of Learning Disabilities, 40*, 134-144.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Denton, C.A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research & Practice, 23*, 79-89.

Deshler, D. & Novossel, L. (2010). Implementing RTI in high school content classes. Strand D: Response to Intervention: Middle School and High School. Council for Exceptional Children Annual Conference. April 2010, Nashville.

DiCecco, V. M., & Gleason, M. M. (2002). Using graphic organizers to attain relational knowledge from expository texts. *Journal of Learning Disabilities, 35,* 306–320.

Duke, N.K. & Carlisle, J.F. (2011). The development of comprehension. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pps 199-228). London: Routledge.

Ebbers, S.M. & Denton, C.A. (2008). A root awakening: vocabulary instruction for older students with reading difficulties. *Learning Disabilities, Research & Practice, 23*(2), 90-102.

Edmonds, M.S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K.K., et al. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79,* 262-300.

Espin, C.A., Cevasco, J., van den Broek, P., Baker, S., & Gersten, R. (2007). History as narrative: The nature and quality of historical understanding for students with LD. *Journal of Learning Disabilities, 40,* 174-182.

Espin, C., Wallace, T., Lembke, E., Campbell, H., & Long, J.D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice, 25*(2), 60-75.

Gajria, M., Jitendra, A.K., Sood, S. & Sacks, G. (2007). Improving comprehension of expository text in students with LD: a research synthesis. *Journal of Learning Disabilities, 40*, 210-225.

Gamson, D.A., Lu, X., & Eckert, S.A. (2013). Challenging the research base of the Common Core State Standards: A historical reanalysis of text complexity. *Educational researcher, 42,* 381-391.

Gersten, R., Baker, S., Smith-Johnson, J., Dimino, J., & Peterson, A. (2006). Eyes on the prize: Teaching complex historical content to middle school students with learning disabilities. *Exceptional Children, 72*, 264-280.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*, 279–320.

Hairrell, A., Simmons, D., Swanson, E., Edmonds, M., Vaughn, S., & Rupley, W.H. (2011). Translating vocabulary research to social studies instruction: Before, during, and after text-reading strategies. Intervention in School and Clinic, 46, 204-210.

Hammill, D.D., Wiederholt, J.L., & Allen, E.A. (2006). TOSCRF: Test of silent contextual reading fluency. Examiner's manual. Austin, TX: PRO-ED.

Harniss, M., Caros, J., & Gersten, R. (2007). Impact of the design of U.S. history textbooks on content acquisition and academic engagement of special education students: An experimental investigation. *Journal of Learning Disabilities, 40,* 100-110.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., and Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A Practice Guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional

Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc.

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,* 437-451.

Lenz, B.K., & Hughes, C.A. (1990). A word identification strategy for adolescents with learning disabilities. *Journal of Learning Disabilities, 23,* 149-163.

Mastropieri, M.A. Scruggs, T.E. & Fulk, J.M. (1990) Teaching abstract vocabulary with the keyword method: effects on recall and comprehension. *Journal of Learning Disabilities, 23*(2), 92-96.

Morris, A.K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher, 40,* 5-14.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). The Common Core State Standards. Washington D.C., Author.

Pearson, Inc. (2012). *Aimsweb reading curriculum-based measurement administration and scoring guide.* Retrieved March 6, 2014, from http://www.aimsweb.com/wp-content/uploads/R-CBM-Admin_Scoring-Guide_2.0.pdf.

Penuel, W.R., Fishman, B.J., Cheng, B.H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40,* 331-337.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading, 11* (4), pp357-383.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29-53.

Richgels, D.J., McGee, L.M., Lomax, R.G., & Sheard, C. (1987). Awareness of four text structures: effects on recall of expository text. *Reading Research Quarterly, 22*, 177-196.

Roberts, G., Vaughn, S., Fletcher, J., Stuebing, K., & Barth, A. (2013). Effects of a response-based, tiered framework for intervening with struggling readers in middle school. *Reading Research Quarterly, 48*, 237-254.

Rosenthal, J. & Ehri, L.C. (2008;). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*, 175-191.

Scammaca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J. Reutebuch, C.K., et al. (2007) Interventions for adolescent struggling reading: A meta-analysis with implications for practice. Portsmouth, NH: Center on Instruction. RMC Research.

Shinn, M.R. & Shinn, M.M. (2002). *AIMSWeb Training Workbook: Administration and scoring of reading MAZE for use in general outcome measurement.* Retrieved March 6, 2014, from http://www.cnyric.org/tfiles/folder1052/Administration%20and%20Scoring%20of%20Reading%20MAZE.pdf.

Smith, G.J., Schmidt, M.M., Edelen-Smith, P.J., & Cook, B.G. (2013). Pasteur's quadrant as the bridge linking rigor with relevance. *Exceptional Children, 79*, 147-161.

Solis, M., Ciullo, S., Vaughn, S., Pyle, N., Hassaram, B., & Leroux, A. (2012). Reading comprehension intervention for middle school students with learning

disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities, 45*, 327-340.

Swanson, H.L., & Deshler, D.D. (2003). Instructing adolescents with learning disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities, 36*, 124-135.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999).*Test of Word Reading Efficiency*. Austin,Texas: Pro-Ed, Inc.

Van Daalen-Kapteijns, M., Elshout-Mohr, M., & De Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning, 51*(1), 145-181.

Vaughn, S., Swanson, E.A., Roberts, G., Wanzek, J., Stillman-Spisak, S.J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly, 48*, 77-93.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition*. Bloomington, MN: NCS Pearson, Inc.

Wei, X., Blackorby, J., & Schiller, (2011). Growth in reading achievement of students with disabilities, ages 7-17. *Exceptional Children, 78*, 89-106.

Williams, J., Nubla-kung, A., Pollini, S., Stafford, K., Garcia, A., & Snyder, A. (2007). Teaching cause/effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities,40*, 111-120.

Wineburg, S. & Martin, D. (2009). Tampering with history: Adapting primary sources for struggling readers. *Social Education, 73*, 212-216.

Wolf , M. (2007). *Proust and the Squid.* New York: HarperCollins Publishers.

Woodcock, R. W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside.

Zigmond, N., Danielson, L., Harris, J., & Mellard, D. (2010) What we know (and don't know) about middle and high school response to intervention. Strand D: Response to Intervention: Middle School and High School. Council for Exceptional Children Annual Conference. April 2010, Nashville.

Table 1

Cycles of Instruction, Reading Component Content, Instructors, and Setting

	Cycle 1			Cycle 2			Cycle 3		
	3 weeks	Revise	1 week	3 weeks	Revise	1 week	3 weeks	Revise	1 week
Reading Focus:									
Decoding	X			X			X		
Vocabulary				X			X		
Cause/Effect							X		
Reading Easy Text	X			X			X		
Reading 8th Gr Text				X			X		
Instructor	BRIDGES		Gen/Sp Teacher	BRIDGES		Gen/Sp Teacher	BRIDGES		Gen/Sp Teacher
Settings	Pull-out		Whole Class	Pull-out		Whole Class	Pull-out		Whole Class

Gen/Sp = General Education and Special Education Teachers

Table 2

Student Sample Demographics by Education Services Status

Student Services Status	Male			Female		
	African American	Hispanic	White	African American	Hispanic	White
General Education	1	6	1	0	6	1
Special Education	0	7	4	1	7	2
Total	1	13	5	1	13	3

Table 3

Reading Results for BRIDGES Students (Descriptive and Progress Monitoring)

Assessment		GenEd Intervention	Sped Intervention	All Intervention	ES Difference GenEd vs. Sped
Descriptive Measures					
WJ					
Word Identification		88.31 (6.0)	78.0 (10.3)***	82.71 (10.0)	.27
Word Attack		89.88 (8.7)	84.05 (10.8)	86.71 (10.9)	.01
Passage Comprehension		73.25 (9.9)	74.84 (9.5)	74.11 (9.6)	.01
Picture Vocabulary		76.38 (14.7)	80.0 (14.6)	78.34 (14.5)	.02
Basic Reading Cluster		88.25 (7.4)	79.58 (10.8)**	83.54 (10.3)	.18
Cognitive Ability					
IQ- Full Scale		80.60 (8.7)	84.45 (10.7)	82.80 (9.96)	.04
TOWRE					
SWE		84.63 (10.5)	75.58 (12.0)*	79.71 (12.1)	.14
PDE		86.06 (11.0)	75.21 (13.9)*	80.17 (13.6)	.16
Progress Monitoring					
Measures					
ORF					
	T1	113.31 (18.3)	86.10 (30.8)**	98.54 (29.0)	.23

	T2	109.69 (21.9)	82.53 (36.1)*	94.94 (33.0)	.17
	T3	109.25 (23.9)	81.16 (36.1)*	94.54 (33.6)	.17
	T4	115.00 (25.1)	93.89 (31.2)*	103.54 (30.1)	.13
MAZE 8th					
	T1	10.88 (2.5)	10.94 4.9)	10.92 (3.9)	.00
	T2	17.97 (3.7)	13.11 (4.6)**	15.32 (4.8)	.26
	T3	18.33 (5.4)	16.95 (6.2)	17.56 (5.8)	.01
	T4	21.03 (5.6)	16.95 (6.7)	18.81 (6.5)	.08
TOSCRF					
	T1	85.88 (8.4)	84.6 (8.4)	85.19 (8.3)	.00
	T2	92.31 (9.2)	84.75 (9.0)*	88.11 (9.8)	.14
	T3	92.13 (6.8)	88.32 (9.9)	90.06 (8.8)	.04
	T4	94.44 (6.9)	89.00 (9.2)	91.49 (8.6)	.09

Note. ORF = Oral Reading Fluency; T1 = Time 1; T2 = Time 2; T3 = Time 3; T4 = Time 4; WJ = Woodcock-Johnson Tests of Achievement III. IQ = Intelligence Quotient. PDE= Phonemic Decoding Efficiency. SWE = Sight Word Efficiency.

*p<.05; **p<.01; ***p<.001

Table 4

Researcher Created Measures

Assessment	GenEd Intervention		SpEd Intervention		All Intervention		Non-Intervention Comparison	
Decoding [a]	M (SD)	% of total	M (SD)	% of total	M (SD)	% of total	M (SD)	% of total
Cycle 1 Pre	68.06 (16.44)	70.9	40.9 (23.8)	42.6	53.82 (24.8)	56.1	NA	NA
Cycle 1 Post	82.3 (10.0)	85.7	65.4 (20.0)	68.1	71.28 (18.28)	74.3	NA	NA
Cycle 2 Pre	18.0 (1.41)	81.1	12.73 (6.45)	57.9	14.56 (6.0)	66.2	NA	NA
Cycle 2 Post	21.5 (1.07)	97.7	20.05 (3.06)	93.2	19.75 (4.69)	89.8	NA	NA
Cycle 3 Pre	19.69 (6.44)	70.32	16.95 (6.5)	60.5	19.69 (6.44)	70.3	NA	NA
Cycle 3 Post	26.93 (1.59)	96.2	24.56 (3.7)	87.7	25.59 (3.7)	91.2	NA	NA
Vocabulary [b]								
Cycle 2 Pre Definition	2.14 (2.34)	9.7	3.43 (1.91)	15.6	3.0 (2.1)	13.6	NA	NA
Cycle 2 Post	13.29	60.4	10.71	48.7	11.6	52.7	NA	NA

Definition	(3.55)		(4.87)		(4.6)			
Cycle 3 Pre	4.94	17.6	5.56	19.9	5.26	18.9	NA	NA
Definition	(2.5)		(3.4)		(2.99)			
Cycle 3 Post	16.62	59.4	14.22	50.8	15.35	54.8	NA	NA
Definition	(5.54)		(7.36)		(6.58)			
Cycle 3 Pre	4.18	14.9	5.39	19.25	4.82	17.2	12.37	44.18
Sentence	(2.74)		(2.68)		(2.74)		(5.57)	
Cycle 3 Post	10.19	36.4	12.28	43.9	11.29	40.3	12.55	44.82
Sentence	(4.18)		(5.75)		(5.11)		(6.07)	

Cause-Effect				
Cycle 3	M (SD)	M (SD)	M (SD)	M (SD)
Pre Total [c]	5.15 (3.66)	9.15 (6.52)	11.65 (7.52)	19.39 (8.87)
Post Total	16.06 (5.14)	21.1 (7.57)	18.86 (6.99)	18.08 (11.15)
History	Percentage	Percentage	Percentage	
Content Tests	Correct (SD)	Correct (SD)	Correct (SD)	
Cycle 1	70.35 (9.96)	94.89 (13.62)	83.43 (16.85)	NA
Cycle 2	85.9 (10.7)	80.3 (15.5)	81.9 (14.3)	NA
Cycle 3	90.62 (7.15)	91.52 (12.06)	91.12 (10.06)	NA

Note: [a] Total points possible varied by cycle and chapter: Cycle 1 = 96 points; Cycle 2 Ch. 7 = 22 points; Cycle 3 = 28 points.

[b] Total points possible varied by cycle: Cycle 2 = 22 points; Cycle 3 = 28 points

[c] Score reflects total number of causes, effects, and signal words identified across 4 passages; total possible = 32 points.

Table 5

Scoring Rubric Examples for Definitions and Sentences for the Words: Annex and Discriminate

Sample definition scores			
	0	1	2
Target Word	*incorrect*	*partially correct*	*correct*
Annex	something to drink	to add something	to add on to a building or region of land
Discriminate	to make angry	to treat someone differently	to treat someone differently based on a characteristic about them
Sample sentence scores			
	0	1	2
Target word	*incorrect*	*partially correct*	*correct*
Annex	I annexed my bottle of water.	We had to annex the house.	The U.S. increased its territory by annexing Texas.
Discriminate	I discriminate my brother.	Discrimination in the ‘60’s was bad.	Shelly and her friends discriminated against Emma because she didn’t go to their church.

Table 6

Implementation Data for General and Special Education Teachers

Decoding		**Vocabulary**		**Cause and Effect**	
Introduced Rule	100.0%	Introduced Word	100%	Introduced CE Concept	88.89%
Teacher Rule Repetition			6.57		
	100.0%	Teacher Repetitions of Word(s)	(3.46)		
Student Rule Repetitions	0%	Student Repetitions of Word	2.3 (1.29)		
Modeled strategy use	28.6%			Modeled strategy use	44.4%
				Student input on passage	
		Student Input on definition	70%	meaning	44.4%
				Students justify selections of	
		Student discussion of context	60%	cause/effect	22.2%
		Historical Context discussion	100%	Historical Context discussion	77.8%
Guided Practice	100%	Guided Practice	50%	Guided practice	88.9%
Independent Practice	85.7%	Independent Practice	50%	Independent practice	100%

Scaffolded	0	Scaffolded	50%	Scaffolded	44.4%
Appropriate corrective feedback	0	Appropriate corrective feedback	40%	Appropriate corrective feedback	66.7%

Table 7

Feasibility of Strategies in the Classroom Context: Means and Standard Deviations

	Decoding	Vocabulary	Cause and Effect
Percentage of students for whom strategy is useful	51%	85%	85%
Can be used routinely in classroom	divided response	yes	yes
Ease of teaching (5 as easy; 1 difficult)	3.5 (.58)	3.75 (1.5)	4.0 (1.41)
Confidence in teaching (5 confident; 1 no confidence)	3.0 (.82)	4.0 (1.41)	4.0 (1.41)
Appropriate responsibility of middle school teacher (1 disagree; 5 agree)	2.75 (1.26)	4.0 (1.41)	3.75 (1.26)
Strategy ease for struggling students (1 being easy and 5 being difficult)	3.0 (0)	3.25 (.96)	3.00 (.82)
Value of strategy for student learning	2.75 (.96)	4.00 (.82)	4.25 (.50)

Figure 1

Intervention and Typical Reader Vocabulary Sentence Score Interaction

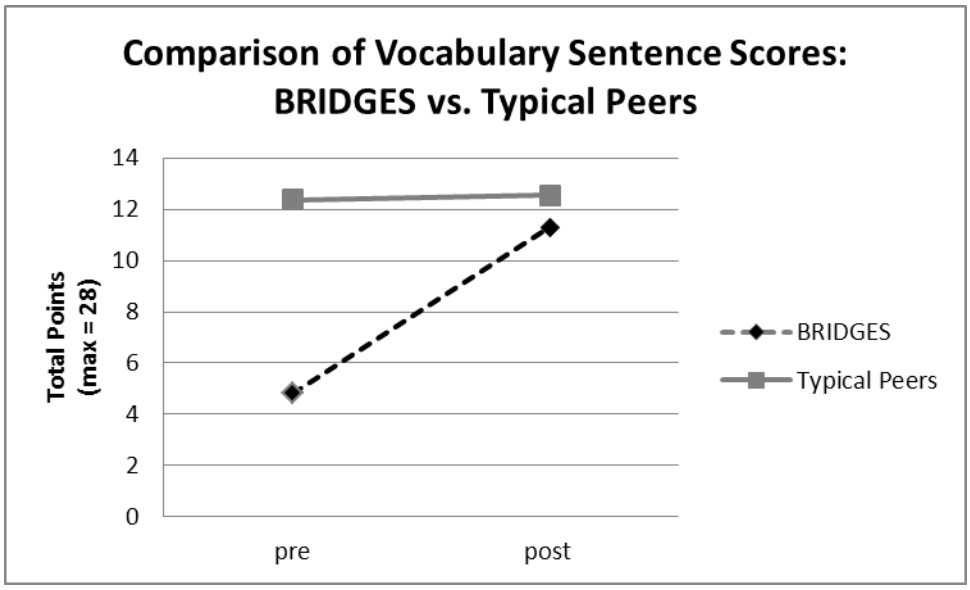

Figure 2

BRIDGES Intervention Students and Typical Reader Cause-effect Score Interaction