Embedding Explicit Instruction of Transfer to Improve At-Risk Students' Reading Comprehension in Informational Texts

By

Samuel Allen Patton III

Dissertation

Submitted to the Faculty of the

Peabody College of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

September 30, 2019

Nashville, Tennessee

Approved:

Douglas Fuchs, Ph.D.

Lynn S Fuchs, Ph.D.

Jeanne Wanzek, Ph.D.

Nathan Clemens, Ph.D.

DEDICATION

To my wife Kelly, whose love and support encouraged me, through good times and bad, to keep striving. I would not be where I am today without you. Thank you for everything.

To my parents, Sam and Suzanne; to my brother, Seth; to Bruce, Corinne, Lynn, and Forrest; to Emma, Caroline, Rhodes, and Gabby; to all of you, your love and positivity have helped me beyond measure.

and

To Lucille and Joey Pancakes, the two best writing buddies a guy could ever ask for.

ACKNOWLEDGMENTS

This work would not be possible without the enduring support of my committee chair, mentor and advisor, Dr. Doug Fuchs. In working with him both prior to and upon becoming his student, he has motivated me to become both a better researcher and a better person. His expertise, curiosity, and drive are an inspiration.

I am especially grateful to Dr. Fuchs and to the other members of my committee, Drs. Lynn Fuchs, Jeanne Wanzek, and Nathan Clemens, for their guidance and feedback during this journey. I have profited enormously from their thoughtful questioning and insight. I feel thrilled and humbled to have had the chance to learn so much from them.

I am tremendously thankful for the friends and colleagues I've worked with over the years: project coordinators, research assistants, cohort-mates, and instructors. I especially want to thank Loulee Yen Haga, Wen Zhang Tracy, Dr. Devin Kearns, Dr. Peng Peng, Meagan Walsh, Emma Hendricks, and Wooliya Kim, each of whom played a critical part in helping me reach this milestone. Many thanks are also due to the teachers and students who participated in this study.

Finally, thank you to my wife Kelly and all of my family. Your love and support have made this all possible.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Introduction

Reading informational texts with understanding is a necessary skill throughout one's academic career and beyond. Evidence indicates, however, that many students' comprehension of nonfiction texts is inadequate. Only 36% of 4th-grade students assessed on the National Assessment of Educational Progress (NAEP) scored at grade level or higher; only 12% of 4th graders with disabilities did so (U.S. Department of Education, 2017). And this likely *overestimates* the proportion of American students who read competently from informational texts because the NAEP assesses comprehension of both nonfiction and fiction texts, and the former tends to be more challenging (Best, Floyd, & McNamara, 2008; Goldman & Rakestraw, 2000; Hall, Sabey, & McClellan, 2005; McNamara, Ozuru, & Floyd, 2011).

Recently, there has been a stronger instructional focus on informational texts, as reflected in the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). With this increased emphasis, there has been greater recognition of the negative consequences of failing to understand informational texts (Meneghetti, Caretti, & De Beni, 2006; Miller, McCardle, & Hernandez, 2010). This may be especially true for students with disabilities, who are likely significantly behind their typically developing peers in this regard (e.g., Gilmour, Fuchs, & Wehby, 2018). There is an obvious need for programs that foster adequate comprehension of informational texts.

That said, weak reading comprehension can be difficult to ameliorate because of its multifaceted nature (Catts & Kamhi, 2017; Snow, 2002). Comprehension is a product of a complex process of student, text, and task factors interacting with each other, and there are many opportunities in that process for breakdowns and failure (Cain, Oakhill, & Bryant, 2000).

Strategy Training and Skills Instruction

Two approaches to help students improve reading comprehension are strategy training and skills instruction. Explicit strategy training typically involves a process by which students are taught and practice step-by-step cognitive routines. It has been shown to help remediate students' deficits in mathematics (e.g., Powell & Fuchs, 2010), writing (e.g., Hebert, Bohaty, Nelson, & Roehling, 2018), and reading comprehension (Gajria, Jitendra, Sood & Sacks, 2007; Gersten, Fuchs, Williams, & Baker, 2001; National Reading Panel & National Institute of Child Health and Human Development, 2000; Shanahan et al., 2010; Wanzek & Vaughn, 2009). With respect to reading comprehension, strategies can be taught for retelling (e.g., Koskinen, Gambrell, Kapinus, & Heathington, 1988), paraphrasing (e.g., Hagaman & Reid, 2008), and inference-making (e.g., Barth & Elleman, 2017), among others. For students with reading disabilities, explicitly teaching comprehension strategies may be more beneficial than teaching them strategies implicitly (Manset-Williamson & Nelson, 2005).

Strategy training encourages students to make deliberate use of a process or processes prior to solving a problem (such as comprehending a passage). Direct skills instruction, by contrast, targets processes that operate below conscious thought (Afflerbach, Pearson, & Paris, 2008). In improving nonfiction reading comprehension, skills instruction may make use of the same activity as a strategy-based approach, such as retell (e.g., Gambrell, Koskinen, & Kapinus, 1991). It may also function more as "drill and practice," such as structured repeated reading (e.g., Vadasy & Sanders, 2008) or learning academic vocabulary words (McKeown, Crosson, Moore, & Beck, 2018), in which repeated exposure to passages and words promotes familiarity and automaticity.

Strategy training and skills instruction need not be mutually exclusive; in fact, they can work well hand-in-hand. For example, students may improve their skill at selecting the appropriate comprehension strategy through repetition and practice (Polya, 2014). Or, with enough practice, learned strategies can become automatic skills that require little or no premeditated effort to use. Interventions combining strategy training and skills instruction have improved students' comprehension, both in cases where relatively few strategies and activities (e.g., Williams et al., 2014) and many strategies and activities (e.g., Fuchs, Kearns, et al., 2018) were included.

Whereas strategy training and skills instruction can be beneficial for many students, they are not sufficient for all (Compton, Miller, Elleman, & Steacy, 2014; Gilbert et al. 2013). Students may demonstrate difficulty in generalizing acquired strategies and skills to distal measures (as in Fuchs, Hendricks, et al., 2018, and Williams et al., 2014), which may raise questions about their effect on global comprehension. Other students may not improve even after a prolonged series of consecutive intensive interventions (Wanzek & Vaughn, 2009). For such students with serious and persistent comprehension deficits, something more or something else is still needed.

A Case for Transfer

That something else may be to combine strategy training and skills instruction with activities meant to strengthen specific cognitive processes. Such cognitive processes, related to reading comprehension, include executive functioning (EF; e.g., Eason, Goldberg, Young, Geist, & Cutting, 2012), working memory (WM; e.g., Peng et al., 2018), and metacognition (e.g., Pressley, 2002). Research indicates that they all play important roles in reading comprehension. Students struggling with comprehension may have deficits in one or more of these processes

(Cain & Oakhill, 2006), in addition to possible weaknesses in vocabulary, decoding, fluency, background knowledge, and so on. So which processes should be targeted?

A case may be made for providing explicit instruction and practice in *transfer* (Bransford, Brown, & Cocking, 1999), facilitating the generalization of taught strategies and skills to new contexts. Transfer learning has been the subject of debate among scholars across disciplines for more than a century (Barnett & Ceci, 2002), but its importance and complexity have rarely been questioned. With respect to reading comprehension, successful transfer from teaching to performance would likely subsume the aforementioned cognitive processes. As students are introduced to new strategies and begin to develop their skills, transfer instruction could serve a metacognitive, self-regulatory function (Eason et al., 2012). Later, as students become more facile and automatic with the strategies, passage-level factors could dictate the function of transfer instruction. So, providing explicit transfer instruction may be more efficient than targeting one isolated process.

Instruction alone is not sufficient, however. As with strategy training and skills instruction, transfer instruction should be scaffolded and paired with opportunities for independent practice (Archer & Hughes, 2011; Rupley, Blair, & Nichols, 2009). Transfer exists on a spectrum, ranging from transfer to highly similar to highly dissimilar contexts. So, students may demonstrate independence in some circumstances and require much assistance in others. When students practice their reading comprehension on passages contextually similar to ones read previously, requiring relatively little transfer of learning, the (presumed) increased efficiency could help compensate for deficits in EF or WM (e.g., Alloway, Bibile, & Lau, 2013). When encountering dissimilar passages, transfer instruction could resume its role as a self-regulatory check as students expend greater effort applying their learned strategies. Motivation

and temperament are also important for fostering transfer of learning (Perkins & Salomon, 2012); students may transfer learning further and more easily when they are engaged and have experienced success in doing so previously.

Measuring Reading Comprehension… Or Something Else?

Successful transfer may be considered evidence of "true" learning, or mastery. If a student with weak reading comprehension receives effective instruction and subsequently performs well on a reading comprehension test, one would likely infer that some learning has occurred. However, this belief rests on an important assumption: namely, the test is an accurate representation of the construct. In the case of reading comprehension, this is not necessarily true.

Standardized, commercially-available, norm-referenced tests tend to be viewed as the "gold standard" when evaluating reading comprehension program efficacy (Slavin, Lake, Chambers, Cheung, & Davis, 2010). Moreover, their use has increased with a recent push for more rigorous reading research (Scammacca, Roberts, Vaughn, & Stuebing, 2015). Commercial comprehension tests are not generally aligned with reading comprehension programs and, because of this, are seen as distant or "distal" measures of far-transfer. These tests tend to require students to use multiple strategies on passages that vary considerably in terms of fiction/non-fiction, genres, and reading difficulty, in an effort to broadly measure reading comprehension. This separation between the comprehension tests and the reading programs they are meant to evaluate is a positive feature, according to guidelines from organizations such as What Works Clearinghouse (2017) and in the Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], & Joint Committee on Standards for Educational and Psychological Testing, 2014).

But there is reason to question the validity of such tests, despite the evidence typically presented in their respective technical manuals. As stated, reading comprehension is a complex and implicit process. For the sake of efficiency and practicality, these tests must measure the observable artifacts of the comprehension process (Pearson & Hamm, 2005). As a result, the choices that test developers make in operationalizing reading comprehension – from the genres of the included passages to the response modalities – can have an appreciable impact beyond student ability on test performance (Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008), including for students with comprehension deficits (Collins, Lindstrom, & Compton, 2018). This was observed in a recent study by Keenan and Meenan (2014), who found a median correlation of performance of .54 among four standardized reading comprehension tests.

A more comprehensive and meaningful approach to measuring the effects of a reading comprehension program might include more proximal measures of near- and mid-transfer learning to accompany and complement the commercial far-transfer tests (Gersten et al., 2005; Institute of Education Sciences & National Science Foundation, 2013). This would necessarily require researchers to become skilled test developers, as only they know their interventions well enough to develop measures appropriately distanced from them. Recent evidence suggests that such expertise in test development is not widespread (Patton & Fuchs, 2018). Nevertheless, by manipulating dimensions of near- and mid-transfer tests *a priori*, we may mitigate problems stemming from commercial tests' weak alignment with both a reading program's intent and the construct itself (Paris & Stahl, 2005). Using tests aligned in varying degrees to reading programs can also promote theory-building by illuminating (via mediator analyses) processes or mechanisms by which change in performance on standardized tests is occurring (Scammacca Lewis, Clemens, & Roberts, 2019).

The Current Study

As an extension of our recent work to strengthen 3rd- and 5th-graders' comprehension of informational texts (see Fuchs, Hendricks, et al., 2018), we developed two versions of an extant reading comprehension program. The first version was a slight variation of the extant program, hereafter referred to as the "base" program (or "Comp"). The second version of the program combined the base Comp program with instruction and activities designed to promote transfer of taught strategies to novel contexts ("Comp+Transfer"). Our research questions were as follows:

1. What is the efficacy of a Comp program for informational texts compared to BAU classroom instruction, as measured by commercially-available, distal comprehension tests and researcher-developed, proximal comprehension tests?
2. What is the efficacy of a Comp+Transfer program for informational texts compared to BAU classroom instruction, as measured by distal and proximal reading comprehension tests?
3. How much value does explicit transfer instruction and practice add to the base Comp program?

By implementing these two versions of the reading program (Comp and Comp+Transfer), we hoped to explore the possible added value of the explicit transfer instruction. We strengthened this exploration by using a control group and standardized reading comprehension tests and more proximal measures of our own design at varying levels of transfer. This would allow us to assess the potential value of both the comprehension program and the additional transfer instruction along a gradient. In other words, we would not simply assess whether or not students *could* transfer their learning but rather *how far* they could transfer it.

Method

Participants

Teachers. Student participants came from 97 4th and 5th grade classrooms in 20 schools in a large school district in the Southeastern United States. Most of the teachers at both grade levels were female and Caucasian and nearly half held graduate degrees. Approximately a quarter of the teachers were certified to teach English Language Learners (ELLs); 16% were certified in reading; and only six were certified in special education (see Table 1).

Student recruitment and selection. We wanted to identify students scoring no lower than the low-average range on word reading and substantially below average on reading comprehension because our program was designed to improve comprehension. It had little emphasis on word reading or reading fluency. After explaining to participating teachers the purpose of our study, they nominated a total of 1316 students for study inclusion. Teachers gave parental consent forms to the nominated students and obtained written affirmative responses from 547 of them.

A gated screening process was used to maximize efficiency and resources. In the first test session, 531 students were individually tested for 15 minutes on the Sight Word Efficiency (SWE) subtest of the Test of Word Reading Efficiency-2 (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012) and Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence-2 (WASI-2; Wechsler, 2011). This number of students reflected the fact that 16 were not tested because of changes in their availability for testing, or because they transferred to another school, or because they refused to participate. Following this first test session, 112 more students were excluded because they scored below our minimum TOWRE-2 SWE criterion of the 14th percentile. Fifteen more students were lost due to new changes in testing availability.

The remaining 404 students were then screened on the Reading Comprehension subtest of the Gates-MacGinitie Reading Tests-4 (Gates-MacGinitie; MacGinitie, MacGinitie, Maria, & Dreyer, 2000), which was given in small groups of 2-4 students. It took 45 minutes. An additional 133 students were excluded after this session because they scored above our maximum Gates-MacGinitie criterion of the 50th percentile. An additional 47 students were eliminated due to new changes in availability, school transfers, or participation in an earlier iteration of the program we were evaluating.

Following this, 224 students were then individually administered the Vocabulary subtest of the WASI-2. To remain eligible for study participation, they were required to score at or above our minimum criterion of a T-score of 37 on either the WASI-2 Matrix Reasoning or Vocabulary subtests. Six students with T-scores below 37 on both subtests were dropped from the study. Finally, 14 students were excluded due to transferring schools, parents changing their minds about their children's participation, frequent absences during the testing period, or repeated demonstrations of defiant behavior during testing. Thus, the final selection pool consisted of 204 students who were given the remaining tests. Table 2 provides demographic and screening data information for the students in the final sample.

Attrition. Of the 204 students randomly assigned to Comp, Comp+Transfer, and Control, complete data were collected on 189, representing an overall attrition rate of 7.4%. Within study groups, attrition rates were 11.8%, 4.4%, and 5.9% for Comp, Comp+Transfer, and Control, respectively. Thus, the largest difference in attrition was 7.4%, occurring between the two active treatment groups. According to What Works Clearinghouse (2017), using a liberal attrition standard (appropriate for the current study), the combination of the overall and differential attrition rates would likely represent *low* attrition bias. A CONSORT (Consolidated Standards of

Reporting Trials; Schulz, Altman, & Moher, 2010) flow diagram of student participant numbers from initial screening to data analyses is shown in Figure 1.

Procedures

Random assignment. Eligible students were assigned to the study groups according to the following procedure. First, they were grouped by school. Because those assigned to receive tutoring would be working with another student, each pair needed to attend the same school. Within each school, pairs were grouped by the times during the school day given by their teachers as permissible for tutoring. For each time block, students were then randomly assigned to one of the three study groups. Group comparability was examined by grade for TOWRE-2 SWE (raw and standard scores), Gates-MacGinitie, (raw and normal curve equivalent scores) and WASI-2 Matrix Reasoning T-score. No significant between-group differences were obtained.

Finally, for students assigned to one of the two active treatment groups, those in the same group were assigned to tutoring pairs. At some schools, there was only a single option for forming pairs (i.e., only two students were assigned to one treatment group in a school). But when four or more students in a school were assigned to the same group, pairs were formed by matching students with similar TOWRE-2 SWE scores. This was to increase the likelihood that paired students would be reading at about the same level. Between the two active treatments, students were tutored in 67 pairs.

Project staff and timeline. Testing and tutoring sessions were conducted by 28 graduate students hired as research assistants (RAs) and two project coordinators. Two of the RAs were doctoral students in special education and the remaining RAs were in master's programs in educational and counseling fields. The project coordinators and doctoral students trained and

supervised the RAs, but their responsibilities also included tutoring, testing, and program development. Pre-treatment testing occurred between late August and early October. Tutoring began in late October. Post-treatment testing was conducted from late February and late March.

Reading Comprehension Program: The Base Program

For this study, we evaluated two versions of our reading comprehension program. One version consisted of a "base" treatment (Comp), which involved the teaching of strategies and skills for use before, during and after reading texts. The other (Comp+Transfer) was an additive modification of the base program designed to explicitly promote transfer of the taught strategies and skills to different contexts. A description of the features common to both treatments follows.

Setting and pacing. As indicated, tutoring was provided to student pairs. Tutoring occurred 3 times weekly for 14 weeks, with each session lasting 43 minutes on average. Thus, tutoring was conducted for a total of about 30 hours (42 sessions x 43 minutes per session). Our intent was to conduct tutoring in a quiet place in the child's school (e.g., the library or an empty classroom). Sometimes, however, it occurred in hallways outside the children's classrooms. During each session, tutored pairs completed as much of a lesson as possible within the allotted time. Tutors were provided with written guidelines to ensure that no lesson component lasted for an inordinate amount of time.

Excluding four training lessons that introduced the comprehension strategies, lessons were grouped by five themed units (Explorers, Animals, Olympics, Change Your World, and Ancient Egypt). These units ranged from 5-11 nonfiction texts of 4-6 paragraphs in length. Lessons were not designed to be completed in one session; instead, student pairs worked at their own pace. However, the RAs were instructed to spend no more than two sessions on one lesson. For student pairs working slowly, the RAs were directed to increase pacing by limiting the

amount of student-guided error correction or by skipping portions of lessons. As a result, while all tutored students received approximately the same *amount* of tutoring, they completed between 29-34 lessons across the 42 sessions.

Materials. For each unit (Explorers, Animals, Olympics, etc.), students were given an attractive workbook that contained the texts, accompanying text features (e.g., pictures, maps, etc.), and a glossary for uncommon words. They were also given separate lesson worksheets containing comprehension questions to complete after reading the texts.

The texts consisted of original stories and those adapted from other sources. Members of the research team created or modified the texts to maintain an appropriately challenging reading level for the students, with relatively easier texts more frequently featured in the early themed units and harder texts appearing later. Each text's readability was estimated via Flesch-Kincaid grade level and Lexile scores; text coherence was measured by the Coh-Metrix Text Easability Assessor (Graesser, McNamara, & Kulikowich, 2011). Across texts, Flesch-Kincaid grade level scores ranged from 3.7-6.4, and Lexile scores ranged from 500-830. Coh-Metrix indices guided construction of texts to facilitate use of certain strategies (see below). For example, texts designed to promote the use of inference-making were constructed to have a low score on Causal Cohesion.

RAs conducted lessons using scripts mounted on a double-sided easel so one side, showing the lesson script, faced the RA and the other side faced the student pair. RAs followed the lesson scripts but did not read them verbatim. This was to encourage them to interact naturally with the students. The printed material on the student-facing side varied by page but typically provided reminders about the comprehension strategies or other instructions (e.g.,

"Keep your eyes on your workbook") appropriate for that specific point in the lesson. A second easel displayed posters outlining the steps of the strategy that was currently being implemented.

Before reading strategies. Students learned five strategies to use prior to reading a new text: (1) learn bolded vocabulary words, (2) look for text features, (3) preview text structure, (4) check background knowledge, and (5) make a prediction. In each non-training lesson, students first identified the bolded vocabulary words in the text and used the glossary to define them. Tutors led discussions about the meanings of some bolded words.

Next, students looked for text features, such as titles, headings, pictures, captions, maps, and graphs, to find clues about the most important information in the text. They then previewed the passage to identify its structure by reading the first sentence of each paragraph and looking for text structure words (TSWs). They learned that these words were associated with one of four text structures: Descriptive, Sequence, Compare-Contrast and Problem-Solution. Compare-Contrast texts, for example, would have TSWs such as "compare," "different," and "similar." Students were given graphic organizers, one for each text structure.

They then discussed their background knowledge, or what they already knew about a topic. They were told that some background knowledge could help them better understand a passage. They took turns asking each other what they already knew and what they wanted to learn about the passage topic. Using a curated media library, students then attempted to add to their existing knowledge by watching a video about the day's topic. Last, students made a prediction about the "big idea," or the most important idea, in the passage. Students were taught that the big idea is like the main idea of the whole passage.

During reading strategies. As students read, they were taught how to (1) clarify confusing concepts in the text and (2) make connections between what they were reading and

what they had previously read in the passage. Throughout each lesson, RAs prompted the students to identify concepts that were confusing as they read (or, if they were not confused, to share something that might be confusing to a younger reader). Students practiced clarifying those concepts by rereading the confusing portion of text, remembering information from a previous text, using background knowledge, checking the glossary, or asking for help. RAs were also encouraged to help students clarify confusing concepts that originated spontaneously, rather than adhering to the prescribed prompts in the script.

Following this, RAs prompted students to make a connection using one of three sentence stems: (1) "I wonder…"; (2) "I'd like to know more about…"; or (3) "____ made me think of…." Students could make connections from the current text to previously read texts, from the videos they had watched from the media library, or from their background knowledge.

After reading strategies. After reading the text, students were encouraged to use two more strategies to promote reading comprehension. The first was finding the main idea of a paragraph, a strategy adapted from Peer Assisted Learning Strategies (PALS; Fuchs, Fuchs, Mathes, & Simmons, 1997). The Main Idea strategy had three steps: (1) name the most important who or what, (2) tell the most important thing about the who or what, and (3) say the main idea. After reading each paragraph, the students answered a multiple-choice or open-ended question about the main idea. If the answer was incorrect, the tutors used a standard correction procedure that was meant to distinguish sentences that provided important information (the main idea) from sentences that provided details. The Main Idea strategy was also used to find the "big idea," the most important idea in the entire passage. Students learned that the big idea was related to the particular text structure (e.g., in a Problem-Solution passage, the big idea should mention the

problem and its solution). Students answered a multiple-choice question about the big idea after reading the passage.

The second after-reading strategy was "In or Out" to help answer factual and inferential comprehension questions. The In or Out strategy had five steps: (1) Find the Key Words in the question, (2) Go back – find the Key Words in the passage, (3) Read around, (4) Answer the question and prove it with evidence, and (5) If you can't prove it…Brainstorm. Students were taught to identify "key words" in comprehension questions. These key words were clues that helped students in two ways. First, the key words told what the question was asking about. Second, they would help students find the answer in the passage. For example, the key words in the following factual question about astronaut Guy Bluford are underlined: "What did Bluford do after college?" Students used the In or Out strategy to identify the key words, look for them in the passage, and then read nearby sentences to find answers to factual questions.

Students were told that complete answers to inference questions could not be found in a passage. After ensuring that the answer was not in the passage, they brainstormed what they already knew about the key words. Then, they combined their background knowledge with clues from the passage to answer inferential questions.

Release of control. As students progressed through the lessons, the RAs gradually reduced their support for the use of strategies. The goal was for the student pairs to independently use the strategies, with the RA providing guidance and corrections as needed. That said, RAs were told to meet the students at their level of need on a strategy-by-strategy basis. Students were encouraged to lead the routine for the before-reading strategies, the Clarify strategy starting in Lesson 11, and for the remaining strategies in Lesson 13.

Comp Versus Comp+Transfer

Comp treatment. The strategies described above were implemented in both active treatment groups. Students in the Comp treatment engaged in one after-reading activity that was unique: Main Idea Recall. Students took turns reciting from memory the main ideas they had created for each paragraph as they read the text. The text was removed from view, but the students were shown the text's accompanying structural graphic organizer as a visual aid.

Comp+Transfer treatment. Students in the Comp+Transfer group engaged in four activities unique to the treatment that were designed to facilitate generalization of strategies and skills taught and practiced in the tutoring sessions. First, the students were given checklists outlining the before-, during-, and after-reading strategies. They were required to check each strategy as they used it during the lesson, which served as a self-regulatory prompt when they implemented the strategies with greater independence. Second, Comp+Transfer students were encouraged to silently use the Main Idea strategy, as opposed to the Comp students who created main ideas aloud as a pair. Third, when answering In or Out questions after reading, they were instructed to identify the question type as "factual," "inferential," or "main idea." They marked each question with an F, I, or M on their lesson-specific worksheets.

Finally, starting in Lesson 18, Comp+Transfer students practiced strategy use independently by completing a Reading Challenge. The Reading Challenges were administered every third lesson and consisted of short passages from varying genres and had accompanying comprehension questions. Students were given their strategy bookmarks and encouraged to use them to understand the passage and answer the questions. Following completion of the Reading Challenge, the RAs engaged students in brief discussion of their strategy use, inquiring how helpful it was.

These activities were not intended to function simply as "test taking strategies." In each lesson, prior to engaging in the activities above, RAs (via scripts) explicitly explained to students the importance of transferring their learning in an effort to forge a stronger connection between the strategies used within the tutoring context and their applicability to other contexts. This effort reflected the cultivation of a *learning culture of demand* (Perkins & Salomon, 2012), with the hope that students would recognize patterns across contexts and demonstrate their learned strategies appropriately. To illustrate, the question type identification activity did *not* require students to, for example, cross out obviously incorrect answers for multiple-choice questions (although students were not discouraged from doing so if they did this spontaneously). Instead, the activity's focus was on identifying key words of the question and having students explain the reasoning behind their choice afterward.

Measures

Commercially-available reading comprehension measures. Reading comprehension was assessed using a variety of standardized and experimenter-created measures. Two standardized tests of reading comprehension were administered: the Reading Comprehension subtest of the Wechsler Individual Achievement Tests-III (WIAT; Wechsler, 2009) and the Gates-MacGinitie. On the WIAT, students read a selection of texts (typically three) and answer open-ended factual and inferential questions about them. Questions are read aloud by the tester and students are allowed to view the texts as they answer them. Sample-based reliabilities were not obtained because, within grade levels, not all fourth and fifth graders were assessed on identical items due to the passage reversal rule. The manual provides internal reliabilities of .85 and .91 for students in grades 4 and 5, respectively. On the Gates-MacGinitie, students read 11 short passages and provide written answers to multiple-choice questions about them. Students

are given 35 minutes for the test. Sample-based Cronbach's alpha for the Gates-MacGinitie at pre- and posttreatment for students in grade 4 was .34 and .79, respectively; for students in grade 5, Cronbach's alpha was .62 and .82, respectively.

Experimenter-made reading comprehension measures. Three additional "near transfer" tests of reading comprehension were created and administered by the authors. First, a Near-Transfer Knowledge Acquisition test required students to answer 20 multiple-choice questions about vocabulary, facts, and ideas in the instructional passages used during treatment. The tester read aloud questions and answer choices, proceeding one question at a time so that all students could mark an answer in their test booklets. Sample-based Cronbach's alpha for Near-Transfer Knowledge Acquisition at pre- and posttreatment was .40 and .73, respectively.

Second, a Near-Transfer Reading Comprehension test required students to read four passages and provide written answers to multiple-choice questions about each. The passages and questions were similar in presentation (e.g., layout and design) to those used during treatment. None of the test passages had been seen previously by the students, but their content was drawn from topics that had been explored (inventors and the Civil Rights Movement). Sample-based Cronbach's alpha for Near-Transfer Reading Comprehension at pre- and posttreatment was .69 and .73, respectively.

Third, a Near Transfer Main Idea and Recall required students to read two passages aloud, each consisting of four paragraphs. Students orally provided a main idea statement following each paragraph. After reading the full passage, the tester removed it from view and the student was asked to recall the main ideas that he or she had created. Cronbach's alpha for the sample on the Near-Transfer Main Idea *generation* items at pre- and posttreatment was .71 and

.79, respectively. For the Near-Transfer Main Idea *recall* items, Cronbach's alpha for the sample at pre- and posttreatment was identical, .50.

The authors also created and administered separate "mid-transfer" and "far transfer" tests of reading comprehension. Both required students to read passages and answer questions about them. The Mid-Transfer Reading Comprehension test consisted of two nonfiction passages about topics not covered in tutoring. Students provided written responses to multiple-choice questions (some of which required multiple answers) and fill-in-the-blank questions. Additional questions required students to mark answers directly in the passage. Like the Near-Transfer test, the presentation of Mid-Transfer passages and questions was similar to those used in tutoring. Sample-based Cronbach's alpha for Mid-Transfer Reading Comprehension at pre- and posttreatment was .65 and .66, respectively.

The Far-Transfer Reading Comprehension test included four passages of varying genres (e.g., fictional narrative, persuasive essay, etc.) written about content not covered in treatment. Students provided written and oral responses to questions in a variety of formats, including multiple-choice, fill-in-the-blank, short answer, and sequencing. The presentation of the passages and questions were deliberately designed *not* to resemble those encountered by students in the treatment groups. Sample-based Cronbach's alpha for Far-Transfer Reading Comprehension at pre- and posttreatment was .44 and .46, respectively.

Word reading. Word reading was assessed with the TOWRE-2 SWE and Oral Reading Fluency (ORF) subtest of the Woodcock Reading Mastery Tests-3 (WRMT; Woodcock, 2011). Sight Word Efficiency requires students to read as many sight words as possible in 45 seconds from a list of words that increases in difficulty. Oral Reading Fluency requires students to read (typically two) passages of connected text as quickly as possible. Sample-based Cronbach's

alpha is not reported for either measure because it is not appropriate for speeded tests. For TOWRE-2 SWE, the examiner's manual reports test-reliability of .90 for students ages 8-12 and alternate form reliabilities of .89 and .83 for students ages 9 and 10, respectively. For WRMT ORF, the examiner's manual reports split-half reliability of .96 for students in grades 4 and 5, alternate form reliability of .84 for students in grades 3-8, and test-retest reliability of .80 for students in grades 3-8.

IQ. We used the Matrix Reasoning and Vocabulary subtests from the WASI-2. Matrix Reasoning assesses nonverbal reasoning with pattern completion, classification, analogy, and serial reasoning tasks. For each item, students select one of five options that best completes a visual pattern. Sample-based Cronbach's alpha was .52. The Vocabulary subtest evaluates expressive vocabulary, verbal knowledge, and foundational information. For each item, students identify a picture or provide a definition for a word read aloud by the tester. Sample-based Cronbach's alpha was .66.

Fidelity

Test training and testing fidelity. Prior to administering pre-treatment testing, RAs were trained to administer and score all assessments in standardized fashion. Over 4 weeks, RAs received approximately 9 hours of pre-treatment training and were required to practice administering and scoring the tests with a partner for a minimum of 3 hours. For each test, RAs were required to demonstrate at least 90% adherence to the standard administration and scoring rules during a fidelity check before administering the test to a student. If RAs failed a fidelity check, they were had to retake it for that measure until 90% adherence was achieved. Fidelity checks were conducted by PCs and doctoral students, using a checklist. Before post-treatment testing, RAs received 3 more hours of training and were required to pass another round of

fidelity checks with the same criteria in place (i.e., minimum of 90% adherence to standard administration and scoring procedures).

Tutor training and tutoring fidelity. Before tutoring students, RAs were trained to administer lessons in standard fashion. They received 8 hours of training in 2 days and were required to practice lesson administration with a partner for at least 2 hours. Before tutoring began, each RA was required to earn a fidelity score of 90% or higher. During tutoring implementation in the schools, three additional fidelity checks were conducted for each RA, two in person and one by audio recording. For the live checks, one PC or doctoral student observed the RA with a pair of students at a school. For the audio check, one PC listened to an audio recording of a tutoring session. All tutoring fidelity checks were scored with checklist items that reflected all program components. Adherence was determined as a percentage of the correctly-implemented behaviors out of all observed behaviors.

For the two in-school checks, overall observed fidelity across the two treatments was 94.4% and 95.2%. For the audio check, fidelity was 92.9%. Overall fidelity data for each observation by treatment group are presented in Table 3. In addition, Table 4 presents separate fidelity data for the Comp components (common across treatments) and the Transfer components (only in Comp+Transfer). Because a portion of the checklists were destroyed before they could be entered item-by-item into a database, percentages in Table 4 reflect data collected on only 4th-grade students during the second and third observation periods.

Inter-rater agreement for the tutoring fidelity checklists was calculated by having a second rater listen to audio files and complete a separate checklist for 16.4% of all sessions originally observed. The second rater's checklist was then compared to the original rater's checklist on an item-by-item basis. Agreement was calculated as a percentage of matched scores

from all recorded scores. Again, due to the loss of a portion of the checklists, the checklists used for calculating inter-rater agreement overrepresented 4th grade students in the middle and end of the treatment period. Overall inter-rater agreement was 89.7%.

Data Analysis

Data entry and scoring reliability. Children's performances on tests were double scored by two separate scorers and the data were double entered by two separate RAs into identical databases. Scoring discrepancies were resolved by one or more PCs or doctoral students by referring to the original test packet.

Inter-scorer reliability was calculated by having a third RA (i.e., not the original or double-scorer) listen to testing audio files and score a blank test protocol. This second protocol was then compared to the original on an item-by-item basis; agreement was calculated as a percentage of matched scores out of all recorded scores. Reliability was only calculated on tests for which students provided oral responses (e.g., WIAT) and not for written tests (e.g., Gates-MacGinitie). Thus, excluding the written tests, inter-scorer reliability was conducted on 22.4% of all test sessions. The reliability data are presented in Table 5.

Analytic plan. To analyze the outcomes of interest, we created three factor scores: Word Reading, by combining TOWRE-2 SWE and WRMT ORF; Near Transfer Reading Comprehension, by combining the Near-Transfer Reading Comprehension measure and Near-Transfer Main Idea generation items; and Far Transfer Reading Comprehension, by combining the Gates-MacGinitie, WIAT Reading Comprehension subtest, and experimenter-created Far-Transfer Reading Comprehension measure. (The Word Reading and Near-Transfer Reading Comprehension factors contain two equally-weighted measures, so they are more precisely composite scores.) The Mid-Transfer Reading Comprehension and Near-Transfer Knowledge

Acquisition measures were analyzed separately. The Near-Transfer Knowledge Acquisition test was not included in the Near-Transfer Reading Comprehension factor as it was a test of factual recall rather than passage comprehension.

We accounted for potential clustering effects in our data by running multilevel models, one for each outcome of interest: Near Transfer Knowledge Acquisition, Near-Transfer Reading Comp, Mid-Transfer Reading Comp, Far-Transfer Reading Comp, and Word Reading. Unconditional models for all outcomes were run to determine which higher levels of clustering (level 3 = school, level 2 = teacher, and level 2 = pair) were necessary to retain in each final model for accurate standard error estimation. Teachers and pairs were cross-classified at level 2 for students in the two active treatment groups. Students in the control group were nested in teachers and had no tutoring pair membership. Only random effects with values of 0.00 were omitted from the final model. In addition, we estimated group-specific random level-1 variance components to allow for heteroscedasticity but a single level-1 variance component was estimated in the final model in the case that all groups had level-1 variances within one point of each other (Sterba, 2017). Because of the relatively small number of schools in the sample, all multilevel models were estimated with restricted maximum likelihood estimation procedures and Kenward-Roger standard error adjustment (McNeish & Wentzel, 2017).

Once variance components were established for each outcome in the unconditional models, we added a number of fixed effects to the final models. Our interest was in the value-added academic benefit of the two versions of our program beyond typical ("business-as-usual") schooling. Therefore, each treatment group was compared to control by use of dummy variables. Comp was compared to control with d_Comp (Comp = 1; Trans = 0; Control = 0) and Comp+Transfer was compared to control with d_Trans (Comp = 0; Trans = 1; Control = 0). In

addition, two control variables were included to reduce error variance in the outcome: grade (grade 4 = -0.5, grade 5 = 0.5) and the pre-treatment score associated with the outcome variable.

Before obtaining results from final models, we ran each model in blinded fashion (using Stata's *quietly* command) to check normality and homoscedasticity assumptions of level-1 residuals. Only after remediating the models as necessary were final estimates obtained. Measures of effect size (Hedges' *g*, corrected for small samples) were calculated from the model coefficients (What Works Clearinghouse, 2017).

Results

Descriptive Results

Means and standard deviations (*SDs*) at pretest and posttest on the composite and factor scores are presented in Tables 6-8. Table 6 displays means and *SDs* for students in both 4th and 5th grades by study group. Tables 7-8 display means and *SDs* separately for students in grade 4 and grade 5, respectively. Additional tables with means and *SDs* can be found in Appendices A-F. Means and *SDs* for the composite and factor scores for all tutored students combined (i.e., Comp and Comp+Transfer) and control students can be found in Appendices A (combined grades) and B (separated by grade). Means and *SDs* for individual measures for both grades combined are in Appendices C (combined active treatments and control) and D (separate study groups). Means and *SDs* for individual measures for each grade separately are in Appendices E (combined active treatments and control) and F (separate study groups).

Word Reading and Near Transfer Knowledge Acquisition

The unconditional model for the Word Reading composite indicated the need for the following random effects: school, tutoring pair for Comp and Comp+Transfer, and separate error terms for the three groups. Interaction terms for treatment by grade level were not statistically

significant (Model 1), so those terms were removed from the final model (Model 2), although the Comp x grade interaction was marginally significant ($p = 0.053$). Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Results for this model are in Appendix G. Controlling for pretreatment score and grade, neither students in the Comp+Transfer treatment nor in the Comp treatment outperformed students in the control group, Est. = 0.13, $SE = 0.22$, $p = .58$, and Est. = 0.09, $SE = 0.19$, $p = .65$, respectively. Additional results were obtained for similar models comparing all active treatment students to those in the control group. These results indicated no statistically significant differences between students at grades 4 or 5, Est. = 0.34, $SE = 0.20$, $p = .10$, and Est. = -0.17, $SE = 0.28$, $p = .54$, respectively.

The unconditional model for the Near Transfer Knowledge Acquisition measure indicated the need for the following random effects: school, teacher, tutoring pair for Comp and Comp+Transfer, and separate error terms for the three groups. Interaction terms for treatment by grade level were not statistically significant (Model 1), so those terms were removed from the final model (Model 2), although the Comp x grade interaction was marginally significant ($p = 0.054$). Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Results for this model are in Table 9. Controlling for pretreatment Near Transfer Knowledge Acquisition score, pretreatment TOWRE-2 score, and grade, students in the Comp+Transfer and Comp treatments each outperformed students in the control group, Est. = 5.24, $SE = 0.37$, $p < .001$, and Est. = 5.58, $SE = 0.36$, $p < .001$, respectively. Additional models comparing all active treatment students to controls also yielded results in favor of the treatment students at both grades 4 and 5, Est. = 6.05, $SE = 0.50$, $p < .001$, and Est. = 4.86, $SE = 0.44$, $p < .001$, respectively.

Reading Comprehension

Near Transfer. The unconditional model for the Near Transfer Reading Comprehension composite indicated the need for the following random effects: school, teacher, tutoring pair for Comp and Comp+Transfer, and separate error terms for the three groups. Interaction terms for treatment by grade level were not statistically significant (Model 1), so those terms were removed from the final model (Model 2). Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Results for this model are in Table 10. Controlling for pretreatment Near Transfer score, pretreatment TOWRE-2 score, and grade, students in the Comp+Transfer and Comp treatments each outperformed students in the control group, Est. = 1.51, $SE = 0.23$, $p < .001$, and Est. = 1.31, $SE = 0.25$, $p < .001$, respectively. Additional models comparing all active treatment students to controls also yielded results in favor of the treatment students at both grades 4 and 5, Est. = 1.73, $SE = 0.31$, $p < .001$, and Est. = 1.25, $SE = 0.26$, $p < .001$, respectively.

Mid Transfer. The unconditional model for the Mid Transfer Reading Comprehension measure indicated the need for the following random effects: school, teacher, tutoring pair for Comp and Comp+Transfer, and separate error terms for the three groups. Interaction terms for treatment by grade level were not statistically significant (Model 1), so those terms were removed from the final model (Model 2), although the Transfer x grade interaction was marginally significant ($p = 0.06$). Level-1 residuals from the final model met assumptions of normality and homoscedasticity.

Results for this model are in Table 11. Controlling for pretreatment Mid Transfer score, pretreatment TOWRE-2 score, and grade, students in the Comp+Transfer treatment outperformed students in the control group, Est. = 1.18, $SE = 0.42$, $p < .001$, but students in the

Comp treatment did not outperform students in the control group, Est. = 0.45, *SE* = 0.44, *p* = .32. Additional models comparing all active treatment students to controls also yielded results in favor of the treatment students at grade 4, Est. = 1.57, *SE* = 0.56, *p* < .01, but not at grade 5, Est. = 0.27, *SE* = 0.54, *p* = .61.

Far Transfer. The unconditional model for the Far Transfer Reading Comprehension factor indicated that no random effects were necessary, nor were separate residuals by group. Additionally, interaction terms for treatment by grade level were not statistically significant, so the final model was a simple multiple regression (Model 2). Residuals from the final model met assumptions of normality and homoscedasticity after removing four multivariate outliers (standardized residual = 2.58 and -2.58 and below).

Results are in Table 12. Controlling for pretreatment Far Transfer score, pretreatment TOWRE-2 score, and grade, neither students in the Comp+Transfer treatment nor in the Comp treatment outperformed students in the control group, Est. = 0.07, *SE* = 0.13, *p* = .56, and Est. = 0.08, *SE* = 0.13, *p* = .51, respectively. Likewise, additional models comparing all active treatment students to controls indicated no statistically significant differences between students at either grades 4 or 5, Est. = 0.05, *SE* = 0.18, *p* = .78, and Est. = 0.10, *SE* = 0.14, *p* = .50, respectively.

Effect Sizes

Hedges' *g* effect sizes for the Word Reading composite, Far Transfer factor, Mid Transfer measure, Near Transfer composite, and Near Transfer Knowledge Acquisition measure are in Table 13 and Figure 2. Within each grade level, effect sizes are shown for each treatment separately and combined, compared to the control group. Across all permutations, effect sizes are large and statistically significant for the Near Transfer composite and Near Transfer Knowledge

Acquisition measure, ES = 0.67-1.30 and ES = 2.03-2.85, respectively. Additionally, effect sizes on the Mid Transfer measure are statistically significant for 4th graders in the Comp+Transfer treatment, ES = .71, and combined across active treatments, ES = .55. However, there is no statistically significant effect size on the Mid Transfer for 4th graders in the Comp treatment. There are no other statistically significant effect sizes for any other measure by treatment or grade, although there is a marginally significant effect for the Word Reading composite for 4th graders in the Comp treatment, ES = 0.28, $p = 0.051$.

Discussion

The purpose of this experimental study was to extend previous work (see Fuchs, Hendricks, et al., 2018) in developing a nonfiction comprehension tutoring program for at-risk students in grades 4-5. The current study involved 189 students in 97 classrooms across 13 schools in a Southeastern urban school district. The screening criteria were designed to identify students whose word reading was in the average range but whose reading comprehension was poor. Students were randomly assigned in equal numbers to one of two tutoring conditions, Comp and Comp+Transfer, and a control group. All tutored students received intensive, strategy-based comprehension instruction in pairs three times a week for 14 weeks, with an average session length of 43 minutes. Student pairs were allowed to complete lessons at a pace that seemed comfortable for them, but they were also guided and prompted by the RAs to ensure adequate content coverage.

Both versions of the reading comprehension program addressed strategies students could use before reading (e.g., defining vocabulary words), during reading (e.g., clarifying confusing concepts), and after reading (e.g., making main idea statements). To differentiate the two treatment groups, students in Comp practiced recalling from memory the main idea statements

they had created during the lesson, whereas students in Comp+Transfer practiced activities designed to facilitate transfer, such as using strategy checklists, identifying question types, and completing Reading Challenges. The efficacy of the two versions of the comprehension program was assessed with commercially-developed and researcher-created reading comprehension measures at varying levels of transfer.

Results suggest that the program holds promise for promoting many students' comprehension of nonfiction passages similar to those they read in tutoring. Across grades 4 and 5, tutored students significantly and dramatically outperformed controls on the Near Transfer Knowledge Acquisition test and the Near Transfer Reading Comprehension composite. Of particular note is the tutored students' performance on the two tests that comprised the composite (i.e., the Near-Transfer Reading Comprehension measure and Near-Transfer Main Idea generation items). We considered both tests "near transfer" because they were intentionally created to resemble passages and questions that students encountered during tutoring. As a result, the tests were attuned to the strategies and skills taught across both treatments; e.g., students who could successfully apply the Main Idea strategy would most likely correctly recognize and answer Main Idea questions. However, the tests featured novel passages, which is to say that they were unfamiliar to the tutored students. Thus, they were still required to transfer their learned strategies and skills to a somewhat different context.

Student performance on the Mid-Transfer Reading Comprehension test suggests an important difference between the two versions of the comprehension program. The 4th-grade students in the Comp+Transfer program significantly outperformed controls whereas Comp 4th graders did not. Neither tutored group at 5th grade demonstrated better-than-controls

performance. These results suggest the value of additional explicit instruction in transfer, but why only at grade 4?

Two explanations seem plausible. First, both versions of the program seemed more beneficial for 4th graders than 5th graders (see especially the Near Transfer results). Although we cannot adequately describe classroom instruction in nonfiction reading comprehension for control students (a study limitation), there was likely a qualitative difference in that instruction across the two grade levels. The school district in which the study took place educates 4th graders in elementary buildings and 5th graders in middle schools, and differences in curricula at the two grades (as outlined on the school district's website) most likely permitted the older control students more opportunities to read nonfiction texts compared to the younger control students. Second, an important differentiating component of the Comp+Transfer program was the Reading Challenges, which allowed Comp+Transfer students more opportunities on novel informational texts to independently practice their strategies, self-monitor their strategy use, and receive feedback from RAs. Again, these opportunities were arguably more valuable for the 4th graders than the 5th graders because 5th grade children were more likely to obtain classroom-based instruction and practice with informational texts.

For Far Transfer, results are more equivocal. No statistically significant differences were found for either tutored group at the two grade levels. Lack of significance notwithstanding, the pattern of effect size differences was reversed by grade, such that the Comp+Transfer group at 4th grade appeared to perform better than controls, whereas the Comp group at 5th grade seemed to performed better. The largest effect sizes on the Far Transfer factor are in line with those found previously for this program on similar measures (Fuchs, Hendricks, et al., 2018). They approach the 0.25 threshold that is considered "substantively important" for education research

(Lipsey et al., 2012; What Works Clearinghouse, 2017, p. 14). However, the inconsistent pattern of effects across the two grades and two study groups defies a straightforward explanation, especially in light of the lack of statistical significance.

Teaching and Measuring Transfer of Learning

So, although results from this study suggest our transfer instruction had benefit beyond the base reading comprehension program, we do not consider the benefit particularly strong or robust. Our goal was to intensify our base program *qualitatively* by modifying the operationalization of the program; we knew that intensifying the base program quantitatively (e.g., tutoring at a 1:1 ratio, delivering 60-minute sessions, etc.) would render it unfeasible in our school district (and, we assume, many others). We also suspected that teaching for transfer could strengthen a greater number of cognitive processes than addressing just one process (e.g., working memory training). And we attempted to ensure that the embedded transfer activities were not just test-taking practice to develop "testwiseness," but rather would promote self-sustaining strategy use.

To some, it may seem difficult, if not impossible, to teach for transfer while avoiding also teaching test-taking skills. In other words, they are one and the same. In the case of our program (and perhaps many reading comprehension programs), the best evidence for transfer of learning would be unremarkable and unobservable: The student would read a passage, understand its factual contents, and consider its inferential implications while internally following the taught strategies as needed. Instead, we must rely on structured demonstrations of reading comprehension via commercially-developed and experimenter-created tests. And as discussed previously, the use of non-commercial tests of near- and mid-transfer has been discouraged (AERA, APA, & NCME, 2014; What Works Clearinghouse, 2017). Thus, the close-to-moderate

(near-transfer to mid-transfer) alignment with program components would be dismissed as "teaching to the test."

We would argue, however, that results from this study demonstrate the usefulness of such tests in evaluating program efficacy. Considering the comprehension outcomes separately at each level of (near-, mid-, and far-) transfer, three different impressions of efficacy emerge. Taken together, however, the results tell a more complete story: Both versions of our program are likely to help students comprehend nonfiction passages similar to those they reading in tutoring; neither program appears to strengthen understanding of far-transfer passages; and embedding transfer instruction seems to promote 4th grade students' comprehension of mid-transfer, nonfiction passages dissimilar from those encountered during instruction.

Limitations

Several study limitations should be noted. As mentioned above, commercially-available, norm-referenced tests tend to be considered the "gold standard" in intervention research, but there are important reasons to avoid relying exclusively on such tests. Although we took steps to mitigate the influence of any particular test by combining multiple far-transfer measures into a factor score (and by administering additional tests of near and mid-transfer), there is evidence that a sizable number of items on those tests are not aligned with the evidence-based strategies that make up our interventions (Fuchs, Hendricks, et al., 2018). However, the low Cronbach's alpha values on our own Far Transfer Reading Comprehension measure also indicate that our test may not have been consistently measuring the same construct. The Far-Transfer test represented our first attempt at creating such a comprehension measure of our program, and we may have gone too far to make the passages and questions too disparate from each other (for one example, by constructing questions with five different response formats compared to just one for the Near-

Transfer test). Thus, the seemingly incompatible results found on the Far Transfer factor across grades and treatments should be interpreted with caution.

Another limitation noted previously was the lack of information collected on classroom instruction concerning reading for understanding in nonfiction texts. Generally, we found that most students in the two treatment groups received tutoring during their schools' designated intervention times. However, we did not collect information from teachers on what exactly the control students were doing during those times, nor did we conduct observations of any classrooms to determine the degree of overlap between our programs and the teachers' instruction. Such information would have positioned us to better isolate the unique dimensions of our programs and possibly their "active ingredients."

An additional limitation concerns how we grouped students for tutoring. First, a 2:1 student to tutor ratio is unrealistic in most school settings, and we have no evidence to suggest that our interventions could be equally effective for larger student groups. Second, although we attempted to pair students on the basis of their pretest TOWRE-2 SWE scores to form groups of relatively equal readers, we were not always successful in doing so. Anecdotal reports from RAs revealed several pairs of students who appeared mismatched, which occasionally created a difficult tutoring environment. When this did happen, typically the more fluent reader in the pair would become impatient and the less fluent student would become discouraged and frustrated, which would subsequently negatively affect their use of the taught comprehension strategies and their ability to work together. In the future, we would likely do better to try to match students on both word reading (via measures of word reading in connected text) and reading comprehension ability, whenever possible.

We also did not assess tutoring fidelity beyond overall adherence, as acknowledged previously, due to the loss of many of our checklists prior to item-by-item entry. While overall adherence is important to document, we were not able to assess fidelity of implementation of the Comp+Transfer-specific components across all times and grades. Having this information, for example, would have allowed us to better understand the apparent lack of added benefit of the transfer activities at grade 5.

Finally, in our desire to compare two versions of our intervention, we were faced with a choice. On one hand, we could allow students in each group to complete the activities as-is, with more students in the Comp group likely to finish their lessons more quickly than their Comp+Transfer counterparts. On the other, we could ensure students in both groups received relatively equal amounts of time in tutoring, at the risk of students in each group receiving unequal amounts of content coverage. We decided to equalize tutoring time. By allowing the student pairs to work at their own pace, the amount of content coverage by the Comp students tended to exceed that of the students in the Comp+Transfer group because the Comp students had fewer activities to complete per lesson. On average, Comp pairs completed a total of 29.3 lessons' worth of content across their sessions, but the average Comp+Transfer pair completed 26.2 lessons. With greater resources, we could have established a second Comp+Transfer group in which students would complete the same amount of *lessons* as the Comp students, which would have allowed us another means of determining the potential added benefit of the transfer activities.

Limitations notwithstanding, we have reason to believe that both versions of our intervention improved students' nonfiction reading comprehension, more so at 4th grade than at 5th. Furthermore, for the 4th grade students, there is evidence to suggest that our additional

transfer instruction and practice strengthened performance on measures of mid- and far transfer. In future work, we will attempt to explore the potential of transfer in a different way. We have preliminary data from the most recent school year (2018-19) in which we provided some students with an intervention similar to the Comp+Transfer described here, while other were given a "slim" version with a smaller number of strategies to learn and practice. Having fewer strategies to remember, our thinking goes, may help promote more effective strategy use (and transfer) for these students. Our goal remains, however, to create the best possible version of the program for students and teachers alike.

References

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364-373. https://doi.org/10.1598/RT.61.5.1

Alloway, T., Bibile, V., & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior*, *29*(3), 632–638. https://doi.org/10.1016/j.chb.2012.10.023

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Archer, A. L. & Hughes, C. A. (2010). *Explicit instruction: Effective and efficient teaching*. New York: Guilford Press.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128(4)*, 612-637. https://doi.org/10.1037/0033-2909.128.4.612

Barth, A. E., & Elleman, A. (2017). Evaluating the impact of a multistrategy inference intervention for middle-grade struggling readers. *Language, Speech & Hearing Services in Schools (Online), 48*(1), 31-41. https://doi.org/10.1044/2016_LSHSS-16-0041

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*(2), 137-164. https://doi.org/10.1080/02702710801963951

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind,*

experience, and school (expanded ed.). Washington, DC: National Academy Press.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76,* 683–696. https://doi.org/10.1348/000709905X67610

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension–age match design. *Reading and Writing, 12*, 31–40. https://doi.org/10.1023/A:1008058319399

Catts, H., & Kamhi, A. (2017). Prologue: Reading comprehension is not a single ability. *Language, Speech, and Hearing Services in Schools*, *48*, 73–76. https://doi.org/10.1044/2017_LSHSS-16-0033

Collins, A. A., Lindström, E. R., & Compton, D. L. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities*, *51*(2), 108–123. https://doi.org/10.1177/0022219417704636

Compton, D., Miller, A., Elleman, A., & Steacy, L. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? *Scientific Studies of Reading, 18*, 55-73. https://doi.org/10.1080/10888438.2013.836200

Cutting, L., & Scarborough, H. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*, 277–299. https://doi.org/10.1207/s1532799xssr1003_5

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed

for reading comprehension. *Journal of Educational Psychology*, *104*(3), 515–528. https://doi.org/10.1037/a0027182

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*(1), 174–206. https://doi.org/10.3102/00028312034001174

Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Zhang Tracy, W., . . . Peng, P. (2018), Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near-transfer. *Learning Disabilities Research & Practice, 33,* 11-23. https://doi.org/10.1111/ldrp.12162

Fuchs, D., Kearns, D. M., Fuchs, L. S., Elleman, A. M., Gilbert, J. K., Patton, S., . . . Compton, D. (2018). Using moderator analysis to identify the first-grade children who benefit most and least from a reading comprehension program: A step towards aptitude-by-treatment interaction. *Exceptional Children*, *85*(2), 229–247. https://doi.org/10.1177/0014402918802801

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*(3), 210–225. https://doi.org/10.1177/00222194070400030301

Gambrell, L. B., Koskinen, P. S., & Kapinus, B. A. (1991). Retelling and the reading comprehension of proficient and less-proficient readers. *Journal of Educational Research, 84*(6), 356-362. https://doi.org/10.1080/00220671.1991.9941816

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.

https://doi.org/10.1177/001440290507100202

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. https://doi.org/10.3102/00346543071002279

Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly, 48*, 135–154. https://doi.org/10.1002/rrq.45

Gilmour, A., Fuchs, D., & Wehby, J. (2018). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*. https://doi.org/10.1177/0022219409359341

Goldman, S. R., & Rakestraw, J. A. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Erlbaum.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Hagaman, J. L., & Reid, R. (2008). The effects of the paraphrasing strategy on the reading comprehension of middle school students at risk for failure in reading. *Remedial and Special Education, 29*(4), 222-234. https://doi.org/10.1177/0741932507311638

Hall, K. M., Sabey, B. L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology, 26,* 211–234. https://doi.org/10.1080/02702710590962550

Hebert, M., Bohaty, J. J., Nelson, J. R., & Roehling, J. V. (2018). Writing informational text using provided information and text structures: An intervention for upper elementary struggling writers. *Reading and Writing: An Interdisciplinary Journal, 31*(9), 2165-2190. https://doi.org/10.1007/s11145-018-9841-x

Institute of Education Sciences & National Science Foundation. (2013). Common guidelines for education research and development. Washington, DC: Author. Retrieved from https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

Keenan, J., Betjemann, R., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281–300. https://doi.org/10.1080/10888430802132279

Keenan, J., & Meenan, C. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*, 125–135. https://doi.org/10.1177/0022219412439326

Koskinen, P. S., Gambrell, L. B., Kapinus, B. A., & Heathington, B. S. (1988). Retelling: A strategy for enhancing students' reading comprehension. *The Reading Teacher, 41*(9), 892-896. https://www.jstor.org/stable/20199962

Lipsey, M., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., . . . Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. (NCSER 2013–3000). Washington, DC: U.S. Government Printing Office.

MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: Riverside Publishing Company.

Manset-Williamson, G., & Nelson, J. M. (2005). Balanced, strategic reading instruction for upper-elementary and middle school students with reading disabilities: A comparative

study of two approaches. *Learning Disability Quarterly, 28*(1), 59-74. https://doi.org/10.2307/4126973

McKeown, M. G., Crosson, A. C., Moore, D. W., & Beck, I. L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal, 55*(3), 572-616. https://doi.org/10.3102/0002831217744181

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education, 4*(1), 229–257.

McNeish, D., & Wentzel, K.R. (2017). Accommodating small sample sizes in three level models when the third level is incidental. *Multivariate Behavioral Research*, *52*, 200–215. https://doi.org/10.1080/00273171.2016.1262236

Meneghetti, C., Carretti, B., & De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Learning and Individual Differences*, *16*, 291–301. https://doi.org/10.1016/j.lindif.2006.11.001

Miller, B., McCardle, P., & Hernandez, R. (2010). Advances and remaining challenges in adult literacy research. *Journal of Learning Disabilities*, *43*, 101–107. https://doi.org/10.1177/0022219409359341

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Author. Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf

National Reading Panel & National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for*

reading instruction (NIH Publication No. 00–4769). Washington, D.C.: U.S. Government Printing Office. Retrieved from https://www.nichd.nih.gov/publications/product/247

Paris, S. G., & Stahl, S. A. (Eds.). (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Erlbaum.

Patton, S. A., & Fuchs, D. (2018). Experimenter-created tests of reading comprehension interventions in the elementary grades: A systematic literature review. Manuscript in preparation.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices: past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp 13-69). Mahwah NJ: Lawrence Erlbaum Associates.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., . . . Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144,* 48–76. https://doi.org/10.1037/bul0000124

Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist, 47*(3), 248-258. https://doi.org/10.1080/00461520.2012.693354

Polya, G. (2014). How to solve it: *A new aspect of mathematical method* (2nd. ed. reprint). Princeton, NJ: Princeton University Press.

Powell, S. R., & Fuchs, L. S. (2010). Contribution of equal-sign instruction beyond word-problem tutoring for third-grade students with mathematics difficulty. *Journal of Educational Psychology, 102*(2), 381-394. https://doi.org/10.1037/a0018447

Pressley, M. (2002). Comprehension strategies instruction: A turn-of-the-century status report. In C.C. Block & M. Pressley (Eds.), *Comprehension instruction: Research–based best*

practices (pp.11-27). New York: Guilford.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/explicit teaching. *Reading & Writing Quarterly, 25*, 125-138, https://doi.org/10.1080/10573560802683523

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Scammacca Lewis, N. K. (chair), Clemens, N., & Roberts, G. (2019). *Best practices for developing proximal measures of intervention effects*. Symposium conducted at the Pacific Coast Research Conference, Coronado, CA.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340*(7748), 698-702. https://doi.org/10.1136/bmj.c332

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/practiceguide/14

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2010). Effective reading programs for the elementary grades: A best-evidence synthesis. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education.

Snow, C. (2002). Reading for understanding: Toward an R&D program in reading comprehension. Washington DC: Rand Corporation. Retrieved from

https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf

Sterba, S. K. (2017). Partially nested designs in psychotherapy trials: A review of modeling developments. *Psychotherapy Research, 27*(4), 425-436. https://doi.org/10.1080/10503307.2015.1114688

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency* (2nd ed.). Austin, TX: Pro-Ed.

U.S. Department of Education. (2017). *NAEP reading report card*. Washington, D.C.: Institute of Education Sciences, National Center for Education Statistics. Retrieved from https://www.nationsreportcard.gov/reading_2017/nation/achievement?grade=4

Vadasy, P. F., & Sanders, E. A. (2008). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education, 29*(4), 235-249. https://doi.org/10.1177/0741932507312013

Wanzek, J., & Vaughn, S. (2009). Students demonstrating persistent low response to reading intervention: Three case studies. Learning Disabilities Research & Practice, 24, 151–163. https://doi.org/10.1111/j.1540-5826.2009.00289.x

Wechsler, D. (2009). *Wechsler individual achievement test* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence* (2nd ed.). San Antonio, TX: NCS Pearson.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Standards Handbook (Version 4.0)*. Retrieved from http://whatworks.ed.gov

Williams, J. P., Pollini, S., Nubla-Kung, A. M., Snyder, A. E., Garcia, A., Ordynans, J. G., &

Atkins, J. G. (2014). An intervention to improve comprehension of cause/effect through expository text structure instruction. *Journal of Educational Psychology, 106,* 1–17. https://doi.org/10.1037/a0033215

Woodcock, R. W. (2011). *Woodcock reading mastery tests* (3rd ed.) San Antonio, TX: Pearson.

Table 1

Descriptive Statistics for Teacher Demographics by Grade

	Grade 4 (n = 49)		Grade 5 (n = 48)	
Variable	n	%	n	%
Female	39	79.59	26	54.17
African American	8	16.33	7	14.58
Asian	0	--	0	--
Caucasian	35	71.43	21	43.75
Hispanic	0	--	1	2.08
Biracial	0	--	0	--
Other	0	--	0	--
No Data	6	12.24	19	39.58
Highest Educational Degree				
B.S./B.A.	17	34.69	8	16.67
B.S./B.A. +	1	2.04	2	4.17
M.Ed./M.S.	15	30.61	13	27.08
M.Ed./M.S. +	8	16.33	3	6.25
Ed.S.	2	4.08	1	2.08
Ed.D/Ph.D.	0	--	2	4.17
No data	6	12.24	19	39.58
Elementary Certification	43	87.76	23	47.92
ELL Certification	20	40.82	4	8.33
Reading Certification	5	10.20	11	22.92
Special Ed. Certification	1	2.04	5	10.42
	M	*SD*	*M*	*SD*
Years in current position	5.07	4.73	6.34	6.00
Years in teaching profession	13.40	7.68	13.69	10.03

Note. Percentages were calculated based on available data.

Table 2
Descriptive Statistics for Student Demographics by Grade

	Grade 4 (n = 87)		Grade 5 (n = 102)	
Variable	*N*	%	*n*	%
Female	44	50.57	50	49.02
African American	32	37.78	33	32.35
Caucasian	13	14.94	20	19.61
Hispanic	30	34.48	35	34.31
Other	10	11.49	4	3.92
Free/Reduced Price Lunch	46	52.87	51	50.00
Individualized Education Plan	1	1.15	3	2.94

Note. Percentages were calculated based on available data.

Table 3

Tutoring Fidelity Adherence by Treatment and Time, Grades 4-5

	Comp	Comp+Transfer
Time 1 (in-school)	94.3%	96.0%
Time 2 (in-school)	98.3%	94.4%
Time 3 (audio)	92.9%	92.9%

Note. Adherence was determined as a percentage of the correctly-implemented behaviors out of all observed behaviors, based on experimenter-created checklists.

Table 4

Tutoring Fidelity Adherence by Treatment Component, Times 2-3, Grade 4 Only

	Comp	Comp+Transfer	
	Base Program	Base Program	Transfer Activities
Time 2 (in-school)	100.0%	94.4%	92.3%
Time 3 (audio)	91.8%	89.3%	87.5%

Note. Adherence was determined as a percentage of the correctly-implemented behaviors out of all observed behaviors, based on experimenter-created checklists.

Table 5
Testing Inter-scorer Reliability by Measure and Time, Grades 4-5

	Pretest	Posttest
TOWRE-2 SWE	98.7%	98.7%
WRMT ORF	96.7%	97.2%
WIAT-3 RC	90.1%	92.4%
NT MI Generation	90.6%	90.9%
WASI-2 MR	95.2%	n/a
WASI-2 Vocabulary	85.5%	n/a

Note. Percentages were calculated as the number of items matched by two separate scorers over the total number of items attempted. WASI-2 Matrix Reasoning and Vocabulary subtests were administered at pretest only.

Table 6

Means (and SDs) for Pre- and Posttreatment Combined Measures by Treatment, Grades 4-5

	Comp			Comp+Transfer			Control		
Variable	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	60	-0.09	1.65	64	0.03	1.83	62	0.04	1.81
Po WR (composite)	59	-0.09	1.43	65	0.05	1.87	64	0.04	2.04
Pr NT RC (composite)	60	-0.04	1.31	65	-0.03	1.56	64	0.07	1.67
Po NT RC (composite)	60	0.35	1.44	65	0.57	1.42	64	-0.91	1.47
Pr FT RC (factor)	60	0.01	1.00	65	-0.03	0.98	62	0.02	1.03
Po FT RC (factor)	59	0.05	0.87	65	0.01	1.05	64	-0.05	1.07

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Table 7

Means (and SDs) for Pre- and Posttreatment Combined Measures by Treatment, Grade 4

	Comp			Comp+Transfer			Control	
Variable	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	29	-0.19	1.56	29	-0.09	1.69	29	0.05	1.80
Po WR (composite)	29	-0.03	1.50	29	-0.20	1.28	29	-0.31	1.77
Pr NT RC (composite)	29	-0.36	1.21	29	-0.13	1.64	29	-0.62	1.37
Po NT RC (composite)	29	0.20	1.34	29	0.41	1.46	29	-1.52	1.31
Pr FT RC (factor)	29	0.16	0.84	29	0.06	0.87	28	-0.09	0.89
Po FT RC (factor)	29	-0.12	0.70	29	0.14	1.18	29	-0.27	1.00

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Table 8

Means (and SDs) for Pre- and Posttreatment Combined Measures by Treatment, Grade 5

	Comp			Comp+Transfer			Control		
Variable	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	31	-0.01	1.75	35	0.13	1.96	33	0.03	1.85
Po WR (composite)	30	-0.15	1.39	36	0.25	2.24	35	0.33	2.22
Pr NT RC (composite)	31	0.27	1.36	36	0.04	1.51	35	0.63	1.71
Po NT RC (composite)	31	0.49	1.54	36	0.70	1.39	35	-0.39	1.41
Pr FT RC (factor)	31	-0.12	1.13	36	-0.10	1.07	34	0.11	1.14
Po FT RC (factor)	31	0.49	1.54	36	-0.10	0.93	35	0.12	1.10

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Table 9

Multilevel Model Results for Posttreatment Near Transfer Knowledge Acquisition (n=188)

	Model 1				Model 2			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Fixed Effect								
Intercept	7.20	1.85	3.90	0.00	7.90	1.83	4.31	0.00
Pretreatment Score	0.37	0.06	6.56	0.00	0.37	0.06	6.62	0.00
Pretreatment TOWRE-2 SS	-0.00	0.02	-0.04	0.97	-0.00	0.02	-0.08	0.94
Comp+Transfer vs Control	5.92	0.53	11.07	0.00	5.24	0.37	14.29	0.00
Comp vs Control	6.36	0.52	12.15	0.00	5.58	0.36	15.39	0.00
Grade 5	1.39	0.61	2.27	0.03	0.29	0.31	0.93	0.37
Comp+Transfer x Grade 5	-1.26	0.73	-1.73	0.09	--	--	--	--
Comp x Grade 5	-1.44	0.72	-2.01	0.05	--	--	--	--
Random Effect								
School	0.02	--			0.00	--		
Classroom	0.09	--			0.13	--		
Pair (Comp)	0.00	--			0.00	--		
Pair (Comp+Transfer)	0.00	--			0.00	--		
Residual (Comp+Transfer)	2.62	--			2.58	--		
Residual (Comp)	2.11	--			2.10	--		
Residual (Control)	5.22	--			5.44	--		

Table 10

Multilevel Model Results for Posttreatment Near Transfer Reading Comprehension Composite (n=188)

	Model 1				Model 2			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Fixed Effect								
Intercept	-1.39	1.25	-1.12	0.27	-1.19	1.24	-0.96	0.34
Pretreatment Score	0.39	0.06	6.12	0.00	0.40	0.06	6.38	0.00
Pretreatment TOWRE-2 SS	0.00	0.13	0.09	0.93	0.00	0.01	0.04	0.97
Comp+Transfer vs Control	1.73	0.34	5.15	0.00	1.51	0.23	6.72	0.00
Comp vs Control	1.64	0.37	4.47	0.00	1.31	0.25	5.16	0.00
Grade 5	0.67	0.32	2.07	0.05	0.41	0.26	1.61	0.13
Comp+Transfer x Grade 5	-0.41	0.46	-0.88	0.38	--	--	--	--
Comp x Grade 5	-0.63	0.51	-1.24	0.22	--	--	--	--
Random Effect								
School	0.08	0.10			0.08	0.10		
Classroom	0.04	0.10			0.04	0.09		
Pair (Comp)	0.29	0.44			0.31	0.44		
Pair (Comp+Transfer)	0.47	0.32			0.43	0.31		
Residual (Comp+Transfer)	1.09	0.29			1.10	0.29		
Residual (Comp)	1.97	0.53			1.98	0.53		
Residual (Control)	1.04	0.22			1.04	0.21		

Note. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation.

Table 11

Multilevel Model Results for Posttreatment Mid Transfer Reading Comprehension (n=188)

	Model 1				Model 2			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Fixed Effect								
Intercept	6.66	2.33	2.86	0.01	7.28	2.33	3.13	0.00
Pretreatment Score	0.57	0.06	9.12	0.00	0.57	0.06	9.20	0.00
Pretreatment TOWRE-2 SS	-0.03	0.02	-1.09	0.28	-0.03	0.02	-1.18	0.24
Comp+Transfer vs Control	2.02	0.61	3.32	0.00	1.18	0.42	2.84	0.01
Comp vs Control	0.92	0.65	1.43	0.16	0.45	0.44	1.01	0.32
Grade 5	1.00	0.66	1.52	0.14	0.11	0.44	0.24	0.81
Comp+Transfer x Grade 5	-1.61	0.82	-1.95	0.06	--	--	--	--
Comp x Grade 5	-0.85	0.89	-0.95	0.34	--	--	--	--
Random Effect								
School	0.12	0.28			0.07	0.28		
Classroom	0.19	0.42			0.35	0.46		
Pair (Comp)	0.13	1.18			0.01	0.60		
Pair (Comp+Transfer)	0.00	0.00			0.00	--		
Residual (Comp+Transfer)	4.45	0.89			4.44	0.91		
Residual (Comp)	5.35	1.47			5.28	1.31		
Residual (Control)	5.34	1.04			5.41	1.07		

Table 12

Multilevel Model Results for Posttreatment Far Transfer Reading Comprehension Factor (n=186)

	Model 1				Model 2			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Fixed Effect								
Intercept	0.57	0.69	0.82	0.41	0.55	0.69	0.80	0.42
Pretreatment Score	0.69	0.05	12.97	0.00	0.69	0.05	12.88	0.00
Pretreatment TOWRE-2 SS	-0.01	0.01	-1.02	0.31	-0.01	0.01	-1.02	0.31
Comp+Transfer vs Control	0.22	0.19	1.18	0.24	0.07	0.13	0.59	0.56
Comp vs Control	-0.11	0.18	-0.59	0.56	0.08	0.13	0.66	0.51
Grade 5	0.14	0.17	0.82	0.42	0.17	0.11	1.60	0.11
Comp+Transfer x Grade 5	-0.27	0.25	-1.05	0.30	--	--	--	--
Comp x Grade 5	0.37	0.25	0.82	0.41	--	--	--	--
Random Effect								
Pair (Comp)	0.04	0.94						
Pair (Comp+Transfer)	0.00							
Residual (Comp+Transfer)	0.57	0.10						
Residual (Comp)	0.40	0.11						
Residual (Control)	0.44	0.08						

Note. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Table 13
Effect Sizes by Treatment Group and Combined

	4th Grade			5th Grade		
	Comp+Transfer	Comp	Combined	Comp+Transfer	Comp	Combined
Word Reading	0.14	0.28+	0.22	0.02	-0.14	-0.09
Far RC	0.21	-0.08	0.05	-0.06	0.23	0.08
Mid RC	0.71**	0.33	0.55**	0.13	0.04	0.09
Near RC	1.30***	1.23***	1.25***	0.95***	0.67*	0.86***
NT Knowledge	2.61***	2.82***	2.85***	2.03***	2.14***	2.32***

+ $p = 0.051$; * $p < .05$; ** $p < .01$; *** $p < .001$

Note. **Bold text** indicates a factor or composite score. Effect sizes are Hedges' *g*. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation.

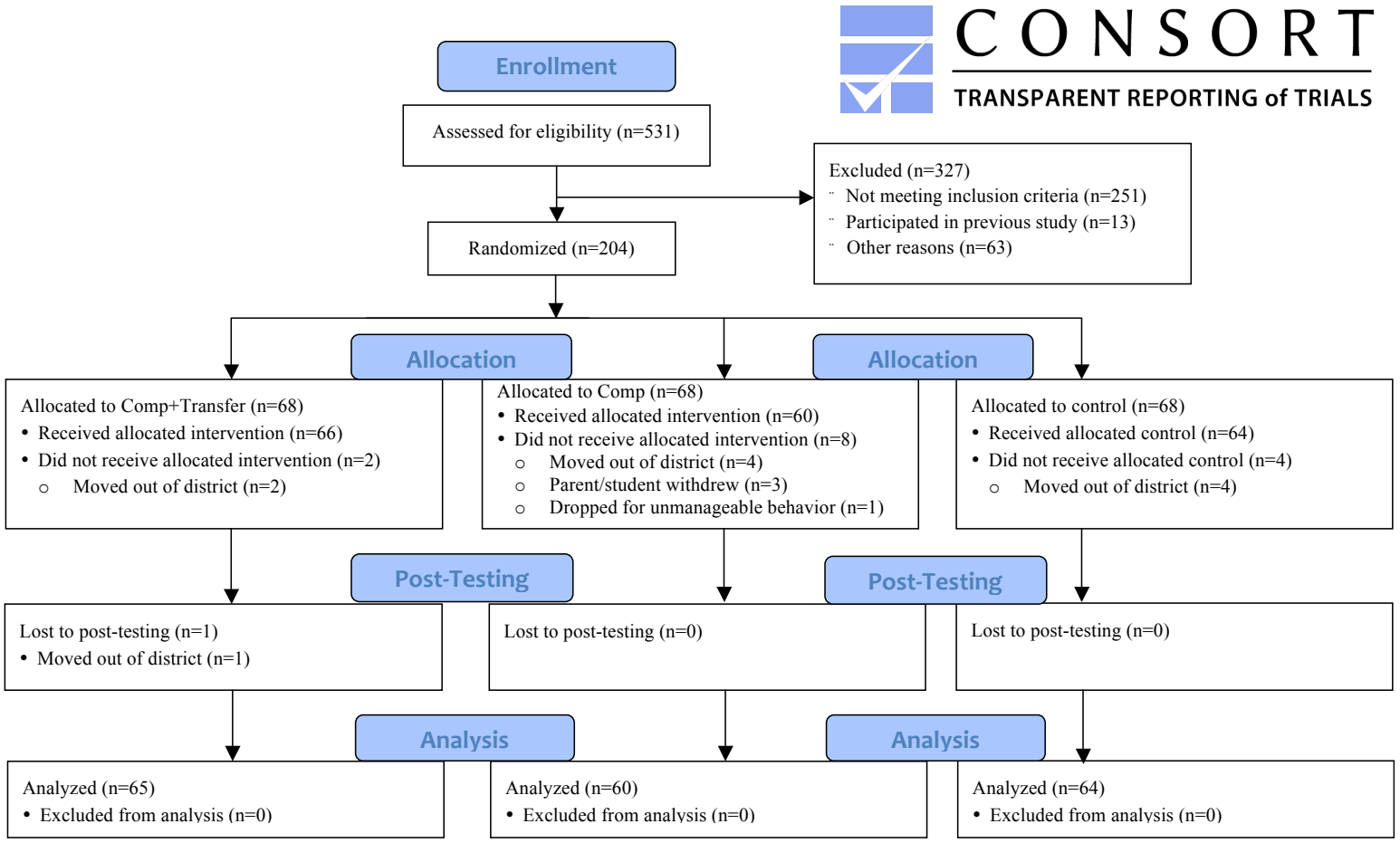
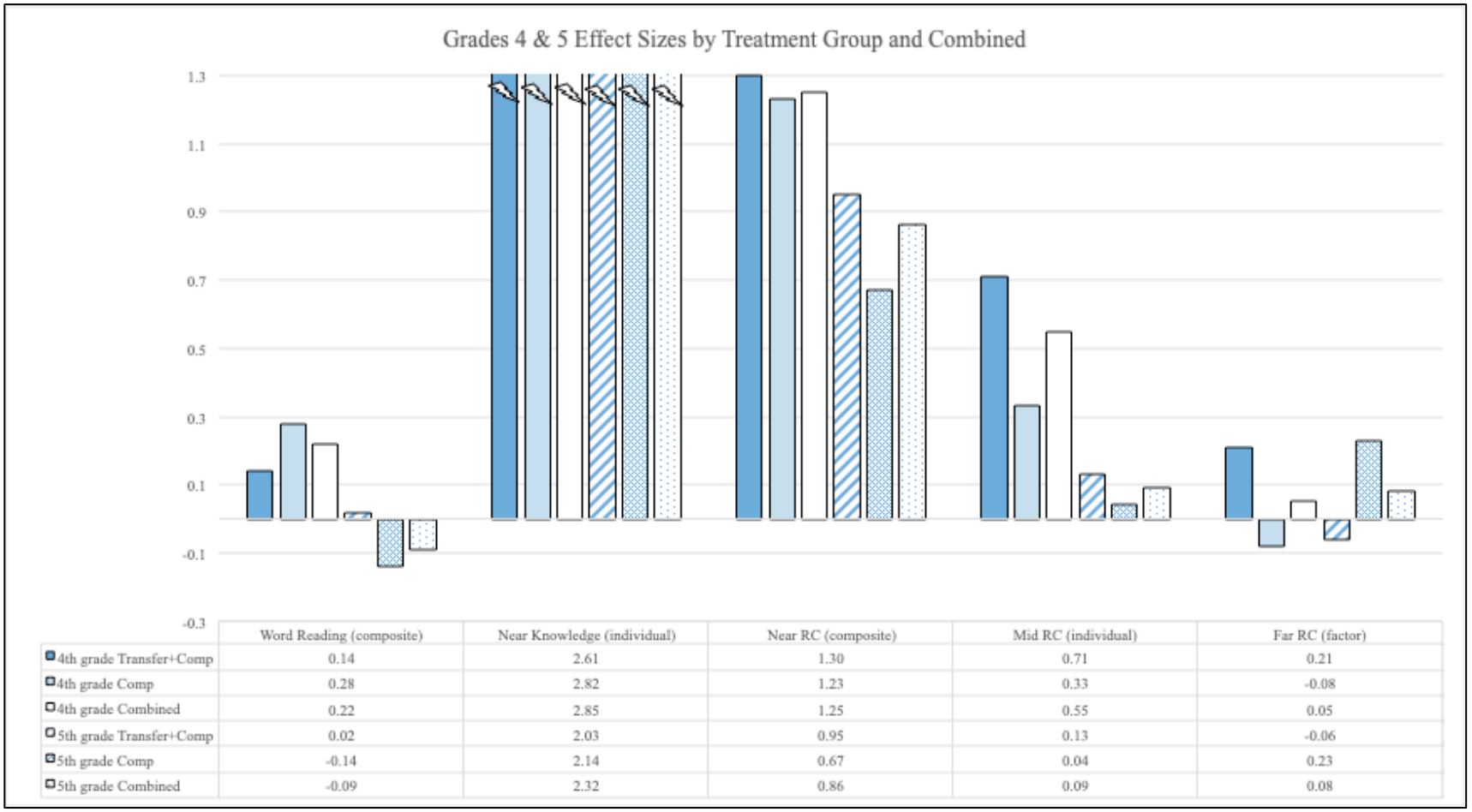


Figure 1. CONSORT flow diagram (Schulz et al., 2010) of student participants.



	Word Reading (composite)	Near Knowledge (individual)	Near RC (composite)	Mid RC (individual)	Far RC (factor)
4th grade Transfer+Comp	0.14	2.61	1.30	0.71	0.21
4th grade Comp	0.28	2.82	1.23	0.33	-0.08
4th grade Combined	0.22	2.85	1.25	0.55	0.05
5th grade Transfer+Comp	0.02	2.03	0.95	0.13	-0.06
5th grade Comp	-0.14	2.14	0.67	0.04	0.23
5th grade Combined	-0.09	2.32	0.86	0.09	0.08

Figure 2. Effect sizes (Hedges' *g*) by treatment group and combined.

Appendix A

Means (and SDs) for Pre- and Posttreatment Combined Measures, Combined Treatments, Grades 4-5

	Combined TRTs			Control		
Variable	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	124	-0.03	1.74	62	0.04	1.81
Po WR (composite)	124	-0.01	1.68	64	0.04	2.04
Pr NT RC (composite)	125	-0.03	1.44	64	0.07	1.67
Po NT RC (composite)	125	0.46	1.43	64	-0.91	1.47
Pr FT RC (factor)	125	-0.01	0.99	62	0.02	1.03
Po FT RC (factor)	124	0.03	0.97	64	-0.05	1.07

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Appendix B

Means (and SDs) for Pre- and Posttreatment Combined Measures, Combined Treatments by Grade

Table B1

Means (and SDs) for Pre- and Posttreatment Combined Measures, Combined Treatments, Grade 4

	Combined TRTs			Control		
Variable	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	58	-0.14	1.62	29	0.05	1.80
Po WR (composite)	58	-0.11	1.38	29	-0.31	1.77
Pr NT RC (composite)	58	-0.25	1.43	29	-0.62	1.37
Po NT RC (composite)	58	0.31	1.40	29	-1.52	1.31
Pr FT RC (factor)	58	0.11	0.85	28	-0.09	0.89
Po FT RC (factor)	58	0.01	0.97	29	-0.27	1.00

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Table B2

Means (and SDs) for Pre- and Posttreatment Combined Measures, Combined Treatments, Grade 5

	Combined TRTs			Control		
Variable	*n*	Mean	SD	*n*	Mean	SD
Pr WR (composite)	66	0.07	1.85	33	0.03	1.85
Po WR (composite)	66	0.07	1.90	35	0.33	2.22
Pr NT RC (composite)	67	0.15	1.43	35	0.63	1.71
Po NT RC (composite)	67	0.60	1.45	35	-0.39	1.41
Pr FT RC (factor)	67	-0.11	1.09	34	0.11	1.14
Po FT RC (factor)	66	0.04	0.97	35	0.12	1.10

Note. WR = Word Reading; RC = Reading Comprehension; NT = Near Transfer; FT = Far Transfer. The Word Reading composite includes TOWRE-2 and WRMT ORF. The Near Transfer RC composite includes Near Transfer RC and Near Transfer WM Main Idea Generation. The Far Transfer RC factor includes Gates, WIAT, and Far Transfer RC.

Appendix C

Means (and SDs) for Pre- and Posttreatment Individual Measures, Combined Treatments, Grades 4-5

	Combined TRTs			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	125	93.13	7.39	64	93.48	7.78
Po TOWRE-2 (ss)	125	96.76	9.12	64	96.58	10.14
Pr WRMT ORF (ss)	124	93.24	6.59	62	93.60	7.23
Po WRMT ORF (ss)	124	95.94	8.31	64	96.69	10.03
Pr Gates (nce)	125	35.65	10.83	64	36.34	8.49
Po Gates (nce)	125	42.38	11.87	64	43.45	11.76
Pr WIAT (ss)	125	93.66	6.83	62	93.39	7.46
Po WIAT (ss)	124	98.44	7.73	64	97.25	8.09
Pr Far RC (raw)	125	6.95	2.15	64	7.00	2.73
Po Far RC (raw)	125	8.35	2.33	64	7.98	2.54
Pr Mid RC (raw)	125	9.47	3.00	64	9.95	2.96
Po Mid RC (raw)	125	10.96	2.70	64	10.39	3.23
Pr Near RC (raw)	125	15.18	3.50	64	15.48	4.32
Po Near RC (raw)	125	18.98	3.34	64	17.32	3.59
Pr NT MI gen (raw)	125	1.91	1.96	64	1.95	1.85
Po NT MI gen (raw)	125	3.88	2.31	64	1.71	1.88
Pr Near Knowledge (raw)	125	9.72	2.51	64	10.05	2.63
Po Near Knowledge (raw)	125	16.93	1.77	64	11.59	2.76

Appendix D

Means (and SDs) for Pre- and Posttreatment Individual Measures by Treatment, Grades 4-5

	Comp			Comp+Transfer			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	60	92.57	7.06	65	93.64	7.71	64	93.48	7.78
Po TOWRE-2 (ss)	60	96.23	8.19	65	97.25	9.94	64	96.58	10.14
Pr WRMT ORF (ss)	60	93.33	6.64	64	93.16	6.59	62	93.60	7.23
Po WRMT ORF (ss)	59	95.71	6.79	65	96.15	9.53	64	96.69	10.03
Pr Gates (nce)	60	36.03	10.12	65	35.30	11.51	64	36.34	8.49
Po Gates (nce)	60	43.10	10.24	65	41.71	13.25	64	43.45	11.76
Pr WIAT (ss)	60	93.17	7.14	65	94.12	6.54	62	93.39	7.46
Po WIAT (ss)	59	97.92	6.94	65	98.91	8.40	64	97.25	8.09
Pr Far RC (raw)	60	7.12	2.15	65	6.8	2.16	64	7.00	2.73
Po Far RC (raw)	60	8.48	2.43	65	8.23	2.25	64	7.98	2.54
Pr Mid RC (raw)	60	9.8	2.67	65	9.17	3.27	64	9.95	2.96
Po Mid RC (raw)	60	10.85	2.92	65	11.06	2.51	64	10.39	3.23
Pr Near RC (raw)	60	15.53	3.09	65	14.86	3.82	64	15.48	4.32
Po Near RC (raw)	60	18.72	3.46	65	19.22	3.23	64	17.32	3.59
Pr NT MI gen (raw)	60	1.73	1.96	65	2.08	1.96	64	1.95	1.85
Po NT MI gen (raw)	60	3.78	2.35	65	3.97	2.29	64	1.71	1.88
Pr Near Knowledge (raw)	60	9.63	2.50	65	9.8	2.53	64	10.05	2.63
Po Near Knowledge (raw)	60	17.08	1.77	65	16.78	1.78	64	11.59	2.76

Appendix E

Means (and SDs) for Pre- and Posttreatment Individual Measures, Combined Treatments, by Grade

Table E1
Means (and SDs) for Pre- and Posttreatment Individual Measures, Combined Treatments, Grade 4

	Combined TRTs			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	58	93.09	7.16	29	93.90	7.41
Po TOWRE-2 (ss)	58	96.34	8.03	29	95.59	9.17
Pr WRMT ORF (ss)	58	92.55	5.90	29	93.14	7.02
Po WRMT ORF (ss)	58	95.53	7.02	29	94.48	8.08
Pr Gates (nce)	58	38.99	7.46	29	38.64	7.08
Po Gates (nce)	58	42.97	11.65	29	40.11	10.22
Pr WIAT (ss)	58	94.50	6.87	28	94.21	7.50
Po WIAT (ss)	58	99.33	7.75	29	98.38	8.76
Pr Far RC (raw)	58	6.52	2.00	29	5.72	2.05
Po Far RC (raw)	58	7.86	2.30	29	7.10	2.16
Pr Mid RC (raw)	58	8.97	3.32	29	9.00	2.67
Po Mid RC (raw)	58	10.78	2.72	29	9.31	3.01
Pr Near RC (raw)	58	14.34	3.05	29	13.83	3.70
Po Near RC (raw)	58	18.72	3.24	29	15.76	3.38
Pr NT MI gen (raw)	58	1.93	2.13	29	1.48	1.68
Po NT MI gen (raw)	58	3.67	2.37	29	1.31	1.54
Pr Near Knowledge (raw)	58	9.12	2.36	29	9.31	2.07
Po Near Knowledge (raw)	58	16.69	1.84	29	10.59	2.56

Table E2

Means (and SDs) for Pre- and Posttreatment Individual Measures, Combined Treatments, Grade 5

	Combined TRTs			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	67	93.16	7.65	35	93.14	8.16
Po TOWRE-2 (ss)	67	97.12	10.00	35	97.40	10.95
Pr WRMT ORF (ss)	66	93.85	7.13	33	94.00	7.50
Po WRMT ORF (ss)	66	96.30	9.33	35	98.51	11.19
Pr Gates (nce)	67	32.76	12.41	35	34.44	9.17
Po Gates (nce)	67	41.87	12.13	35	46.22	12.37
Pr WIAT (ss)	67	92.94	6.76	34	92.71	7.47
Po WIAT (ss)	66	97.65	7.68	35	96.31	7.49
Pr Far RC (raw)	67	7.33	2.22	35	8.06	2.79
Po Far RC (raw)	67	8.78	2.23	35	8.71	2.63
Pr Mid RC (raw)	67	9.91	2.64	35	10.74	2.99
Po Mid RC (raw)	67	11.12	2.70	35	11.29	3.18
Pr Near RC (raw)	67	15.91	3.71	35	16.86	4.35
Po Near RC (raw)	67	19.19	3.44	35	18.63	3.26
Pr NT MI gen (raw)	67	1.90	1.82	35	2.34	1.91
Po NT MI gen (raw)	67	4.06	2.26	35	2.06	2.09
Pr Near Knowledge (raw)	67	10.24	2.54	35	10.66	2.90
Po Near Knowledge (raw)	67	17.13	1.70	35	12.43	2.67

Appendix F

Means (and SDs) for Pre- and Posttreatment Individual Measures, by Treatment and Grade

Table F1
Means (and SDs) for Pre- and Posttreatment Individual Measures by Treatment, Grade 4

	Comp			Comp+Transfer			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	29	92.83	6.96	29	93.34	7.46	29	93.90	7.41
Po TOWRE-2 (ss)	29	96.17	8.36	29	96.51	7.84	29	95.59	9.17
Pr WRMT ORF (ss)	29	92.45	5.99	29	92.66	5.91	29	93.14	7.02
Po WRMT ORF (ss)	29	96.45	6.88	29	94.62	7.17	29	94.48	8.08
Pr Gates (nce)	29	38.88	6.64	29	39.10	8.32	29	38.64	7.08
Po Gates (nce)	29	41.43	8.80	29	44.51	6.47	29	40.11	10.22
Pr WIAT (ss)	29	94.45	7.36	29	94.55	6.47	28	94.21	7.50
Po WIAT (ss)	29	97.97	6.00	29	100.69	9.07	29	98.38	8.76
Pr Far RC (raw)	29	6.79	2.13	29	6.24	1.86	29	5.72	2.05
Po Far RC (raw)	29	7.83	2.24	29	7.90	2.40	29	7.10	2.16
Pr Mid RC (raw)	29	9.55	2.89	29	8.38	3.66	29	9.00	2.67
Po Mid RC (raw)	29	10.59	2.87	29	10.97	2.60	29	9.31	3.01
Pr Near RC (raw)	29	14.31	2.74	29	14.38	3.38	29	13.83	3.70
Po Near RC (raw)	29	18.59	2.65	29	18.86	2.22	29	15.76	3.38
Pr NT MI gen (raw)	29	1.72	2.05	29	2.14	2.22	29	1.48	1.68
Po NT MI gen (raw)	29	3.52	2.29	29	3.83	2.48	29	1.31	1.54
Pr Near Knowledge (raw)	29	8.52	1.88	29	9.72	2.66	29	9.31	2.07
Po Near Knowledge (raw)	29	16.69	1.85	29	16.69	1.85	29	10.59	2.56

Table F2

Means (and SDs) for Pre- and Posttreatment Individual Measures by Treatment, Grade 5

	Comp			Comp+Transfer			Control		
Measure	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Pr TOWRE-2 (ss)	31	92.32	7.25	36	93.89	8.01	35	93.14	8.16
Po TOWRE-2 (ss)	31	96.29	8.16	36	97.83	11.43	35	97.40	10.95
Pr WRMT ORF (ss)	31	94.16	7.19	35	93.57	7.16	33	94.00	7.50
Po WRMT ORF (ss)	30	95.00	6.74	36	97.39	11.02	35	98.51	11.19
Pr Gates (nce)	31	33.37	12.04	36	32.24	12.86	35	34.44	9.17
Po Gates (nce)	31	44.66	11.34	36	39.47	12.42	35	46.22	12.37
Pr WIAT (ss)	31	91.97	6.84	36	93.78	6.67	34	92.71	7.47
Po WIAT (ss)	30	97.87	7.85	36	97.47	7.64	35	96.31	7.49
Pr Far RC (raw)	31	7.42	2.16	36	7.25	2.30	35	8.06	2.79
Po Far RC (raw)	31	9.10	2.47	36	8.50	2.12	35	8.71	2.63
Pr Mid RC (raw)	31	10.03	2.47	36	9.81	2.81	35	10.74	2.99
Po Mid RC (raw)	31	11.10	2.98	36	11.14	2.47	35	11.29	3.18
Pr Near RC (raw)	31	16.68	2.99	36	15.25	4.16	35	16.86	4.35
Po Near RC (raw)	31	18.84	4.12	36	19.50	2.74	35	18.63	3.26
Pr NT MI gen (raw)	31	1.74	1.91	36	2.03	1.75	35	2.34	1.91
Po NT MI gen (raw)	31	4.03	2.42	36	4.08	2.16	35	2.06	2.09
Pr Near Knowledge (raw)	31	10.68	2.59	36	9.86	2.46	35	10.66	2.90
Po Near Knowledge (raw)	31	17.45	1.63	36	16.86	1.74	35	12.43	2.67

Appendix G

Multilevel Model Results for Posttreatment Word Reading Composite

	Model 1				Model 2			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Fixed Effect								
Intercept	-0.35	0.20	-1.70	0.10	-0.16	0.17	-0.94	0.36
Pretreatment Score	0.78	0.05	16.32	0.00	0.78	0.05	16.16	0.00
Comp+Transfer	0.22	0.33	0.66	0.51	0.13	0.22	0.56	0.58
Comp	0.48	0.27	1.75	0.08	0.09	0.19	0.46	0.65
Grade 5	0.46	0.28	1.63	0.11	0.10	0.18	0.58	0.58
Comp+Transfer x Grade 5	-0.18	0.45	-0.40	0.69	--	--	--	--
Comp x Grade 5	-0.75	0.38	-1.98	0.05	--	--	--	--
Random Effect								
School	0.00	0.04			0.00	0.04		
Pair (Comp)	0.05	0.18			0.06	0.19		
Pair (Comp+Transfer)	0.34	0.29			0.32	0.29		
Residual (Comp+Transfer)	1.24	0.31			1.24	0.31		
Residual (Comp)	0.81	0.23			0.83	0.23		
Residual (Control)	1.18	0.22			1.20	0.23		

Note. The Word Reading composite includes TOWRE-2 and WRMT ORF. $n = 185$.