Learning Disabilities Research & Practice, 0(0), 1–15

DOI: 10.1111/ldrp.12276

An Experimental Study to Strengthen Students' Comprehension of Informational Texts: Is Teaching for Transfer Important?

Samuel Allen Patton

Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, United States

Douglas Fuchs

Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, United States and American Institutes for Research, Washington D. C., United States

Emma L. Hendricks

Martinez Unified School District, Martinez, CA, United States

Annie J. Pennell

Washington Office of the Superintendent of Public Instruction, Olympia, WA, United States

Meagan E. Walsh

Michigan MTSS Technical Assistance Center, Holland, MI, United States

Lynn S. Fuchs

Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, United States and American Institutes for Research, Washington D. C., United States

Wen Zhang Tracy

Foundations Early Intervention Services, Dickson, TN, United States

Loulee Yen Haga

Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN, United States

Reading nonfiction texts with understanding is important to school success, yet many students struggle to do so. This randomized controlled trial extends previous research by contrasting an earlier iteration of a comprehension tutoring program (Comp) against a variant with strategies for transferring learning (Comp+Transfer). Participants were 189 fourth and fifth graders with weak reading comprehension. To evaluate their efficacy, we used commercially developed far-transfer measures and experimenter-made near- and mid-transfer measures of reading comprehension. In contrast to controls, students in both programs significantly improved their understanding of near-transfer passages. Additionally, students in Comp+Transfer improved performance on mid-transfer passages. These findings suggest the value of teaching for transfer and the importance of measuring program efficacy with researcher-made tests alongside commercial tests.

Reading informational texts with understanding is necessary in all but the earliest grades. Yet, evidence indicates many children and youth do not adequately comprehend what they read. Only 35% of fourth-grade students scored at or above "Proficient" on the National Assessment of Educational Progress (NAEP); only 12% of fourth graders with disabilities did so (U.S. Department of Education, 2019). Moreover, these figures likely overestimate the proportion of students who comprehend nonfiction texts because the NAEP indexes comprehension of both nonfiction and fiction texts, and the latter tends to be easier to understand (Best et al., 2008; Hall et al., 2005; McNamara et al., 2011).

In recent years, there has been a nationwide focus on reading informational texts with understanding, as emphasized in the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). With this focus has come

404-402-8983, 615-343-4782

Requests for reprints should be sent to Sam Patton, 110 Magnolia Circle, Room 417A Nashville, TN 37212, sam.patton@vanderbilt.edu and Doug Fuchs, 110 Magnolia Circle, Room 417C Nashville, TN 37212, doug.fuchs@vanderbilt.edu

greater recognition of the consequences of failing to understand such texts (Meneghetti et al., 2006; Miller et al., 2010). The consequences may be especially dire for students with disabilities, many of whom are much poorer readers than their classmates (e.g., Gilmour et al., 2018). Hence, there is an obvious need for programs that help all students comprehend informational texts. Development of these programs is challenging, however, because comprehension depends on a complex interaction of student, text, and tasks, with many opportunities for failures to occur (Cain et al., 2000; Catts & Kamhi, 2017; Clemens & Fuchs, in press).

STRATEGY TRAINING AND SKILLS INSTRUCTION

Two well-known approaches to improve reading comprehension are strategy training and skills instruction. In strategy training, students are typically taught step-by-step cognitive routines, which have been shown to strengthen reading comprehension (e.g., Gajria et al., 2007; Gersten et al., 2001; Wanzek & Vaughn, 2009) as well as mathematics performance (e.g., Powell & Fuchs, 2010) and writing (e.g., Hebert et al., 2018). With respect to reading comprehension, strategies can be taught for retelling (e.g., Koskinen et al., 1988), paraphrasing (e.g., Hagaman & Reid, 2008), inference-making (e.g., Barth & Elleman, 2017), and related activities. For students with disabilities, teaching these strategies explicitly may be more beneficial than teaching them implicitly (Manset-Williamson & Nelson, 2005).

The more general intent of strategy training across reading, mathematics, and writing is to help students solve problems deliberately, systematically, and independently. Skills instruction, by contrast, targets more automatic processes such as reading texts with accuracy and fluency (Afflerbach et al., 2008). Notwithstanding this distinction, strategy training and skills instruction may make use of similar activities, which is to say they are not mutually exclusive of each other. Moreover, with practice, well-learned strategies may be used with as little conscious effort as mastered skills. Interventions combining strategy training and skills instruction have improved comprehension (Fuchs et al., 2018; Williams et al., 2014).

Strategy training and skills instruction protocols reflect decades of hard work by many reading researchers. That said, at least two caveats in connection with this work require mentioning. First, whereas both approaches benefit many students' reading comprehension, the phrase "many students" should not be understood as "all students" (Compton et al., 2014; Gilbert et al., 2013). Some children do not demonstrate improvement even after participating in a series of consecutive intensive explicit interventions (Wanzek & Vaughn, 2009). For students with serious and persistent comprehension deficits, something more or something else is needed. Second, the benefit of these interventions dissipates for many students, especially when they encounter problems in novel contexts. In other words, many children have difficulty generalizing or transferring acquired strategies and skills to passages and to questions that deviate from their instructional experiences (e.g., Fuchs et al., 2018; Williams et al., 2014).

A CASE FOR FOCUSING ON TRANSFER

The "something else" in the instruction many children may require is combining strategy training and skills instruction with activities to strengthen specific cognitive processes. Such processes may include (but not be limited to) executive functioning (EF; Eason et al., 2012), working memory (WM; Peng et al., 2018), and metacognition (Pressley, 2002). Consistent with this suggestion is that many who struggle to understand informational texts demonstrate weaknesses in one or more of these (or other) processes (Cain & Oakhill, 2006), in addition to inadequacies in vocabulary, decoding, fluency, background knowledge, and so on. With such children in mind, an obvious question is, "Which cognitive process or processes might be targeted by instruction?"

Alternatively, a case may be made for the importance of strengthening the *transfer of learning* (Bransford et al., 1999); that is, facilitating the generalization of taught strategies and skills to new contexts rather than attempting to train specific cognitive processes. Agreement on a definition, or a consensual description, of transfer has proved elusive, as evidenced in the large volume of research conducted on transfer during the past 100 years. But Barnett and Ceci's (2002) two-dimensional approach to understanding transfer is helpful. The first dimension, the *content* of transmitted knowledge (what is transferred), is measured on a continuum from near to far transfer across learned skills, performance changes, and memory demands. The second dimension, the *context* of transmitted knowledge (when and where knowledge is transferred), is measured on the same continuum across knowledge domains, physical, temporal, functional, and social context, and modalities. Targeting transfer, according to Barnett and Ceci, need not encompass all of these considerations.

With respect to reading comprehension, successful transfer of learning from the instructional setting to other settings is likely to require cognitive processes, which is to say that promoting transfer and strengthening cognitive processes are not mutually exclusive endeavors. As students are introduced to new strategies and skills-to-be-mastered, transfer instruction could serve a metacognitive, or self-regulatory, function (Eason et al., 2012). In other words, students would be responsible for transferring learning. Later, as they employ taught strategies with increasing automaticity, passage-level factors might dictate the focus of transfer instruction.

For example, when applying a summarization strategy to paragraphs of text in unfamiliar genres (e.g., persuasive essay, non-linear narrative), students may continue to rely on the deliberate, purposeful, self-regulatory function of explicit transfer instruction. But when they read passages more like those encountered previously and, therefore, requiring less transfer, the presumed increase in efficiency (automaticity) in the use of cognitive processes might compensate for weaknesses in EF, WM, or other cognitive process (e.g., Alloway et al., 2013).

So, providing explicit transfer instruction may be more efficient than training a relatively isolated cognitive process. That said, transfer subsumes applications to highly familiar contexts and to strongly dissimilar contexts. Students may be capable of transfer in some circumstances and may require considerable assistance in others (e.g., Rupley et al., 2009). Therefore, instruction should be sequenced thoughtfully and scaffolded appropriately to support students who require it (e.g., Archer & Hughes, 2010). For example, when introducing a new reading comprehension strategy, students may practice it more efficiently with texts with familiar text structures and vocabulary. Reducing cognitive demand by manipulating passage-level features so they are more familiar to instructional passages might lessen the need for transfer and increase the chances of successful strategy implementations. As students master a given strategy, they would be encouraged to apply it in increasingly dissimilar texts while getting instruction in transferring learning to those unfamiliar texts.

MEASURING READING COMPREHENSION ACROSS THE SPECTRUM OF TRANSFER

As noted, research indicates that not all students transfer strategies and skills to unfamiliar contexts even after intensive intervention. Commercially developed, norm-referenced tests are likely part of the reason for this. Whereas such tests are viewed by many as the "gold standard" for evaluating the efficacy of reading comprehension programs (see 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; Clemens & Fuchs, in press; What Works Clearinghouse [WWC], 2017), they often require students to use multiple strategies and skills that may not have been addressed in an instructional program. Similarly, they may require students to read passages in different genres, and with dissimilar text structures and greater complexity, than those used for instruction. Indeed, most commercial tests of reading comprehension are constructed so they do *not* align with instructional programs because the test developers' intent is to provide an unbiased estimate of comprehension defined very broadly (Clemens & Fuchs, in press; Schneider, 2020; Slavin et al., 2010).

The use of commercial tests has increased in recent years coincident with a push for more rigorous reading research (Scammacca et al., 2015). Notwithstanding such popularity, there is reason to question the reasonableness of using these tests exclusively. Because reading comprehension is a latent process, students must demonstrate their understanding through observable means (Pearson & Hamm, 2005), such as answering questions or oral summarization. Test developers' definition of reading comprehension can have an appreciable impact on test performance beyond a student's ability, knowledge, or skill level (Cutting & Scarborough, 2006; Keenan et al., 2008). This may be especially true for children and youth with reading comprehension deficits (Collins et al., 2018).

A more meaningful approach, perhaps, when exploring the efficacy of a reading comprehension program might be to include more proximal (near- and mid-transfer) measures of learning—not to supplant but to supplement the commercially developed far-transfer tests (cf. Gersten et al., 2005; Institute of Education Sciences & National Science Foundation, 2013). The use of near- and mid-transfer measures might mitigate the operationalization problems noted by including a variety of item types and response formats. Near- and mid-transfer measures might also provide students an opportunity to demonstrate not only *if* they can transfer learning to new contexts, but *how far* they can do so. Indexing performance on measures aligned in varying degrees to instruction can also promote theory-building by clarifying (via mediator analyses) the processes or mechanisms responsible for performance changes on the far-transfer commercial tests (Clemens & Fuchs, in press; Scammacca Lewis et al., 2019).

THE CURRENT STUDY

For several years, we have been developing an instructional program to improve intermediate-grade children's comprehension of informational texts (Fuchs et al., 2018). To build on—and hopefully strengthen—its effects, we recently developed two versions of the program. The first represents a slight modification of an earlier iteration, hereafter referred to as "Comp." A second version, "Comp+Transfer," combines Comp with instruction to promote transfer of taught strategies to new contexts.

We had three related questions when developing Comp+Transfer. First, could we teach transfer strategies in the same way we taught other reading comprehension strategies? Second, could we embed this instruction within the Comp program without sacrificing key components, and would the resulting program still provide sufficient practice time? Third, would Comp+Transfer prove beneficial above and beyond Comp? Previous work led us to expect that the Comp-only version of the program would be effective for many students, a prerequisite to explore the added value of transfer instruction.

We relied on far-transfer measures and near- and mid-transfer measures for two reasons. We expected some, but not many, students to improve their performance on the far-transfer measures (cf. Barnett & Ceci, 2002; Fuchs et al., 2018). And as explained, we believed the near- and mid-transfer measures would help us explore the degree of learning transfer. We regarded the near- and mid-transfer measures as representing variations on several of Barnett and Ceci's (2002) content and context dimensions: learned skill, memory demands, modality, and temporal context. The near-transfer test would be very similar to instruction in all but passage content, and students would perform well by applying strategies they had learned in tutoring. The mid-transfer test, by contrast, would require students to first recognize the applicability of their learned strategies to passages that differed from instructional passages. Then they would need to modify their use of the taught strategies to suit differing response types. Our research questions were as follows:

TABLE 1
Descriptive Statistics for Teacher Demographics by Grade Levels and Grade Levels Combined

	Grade 4 (n = 49)		*Grade 5 (n = 48)*		*Combined (n = 97)*	
Variable	N	%	N	%	N	%
Female	39	79.59	26	54.17	65	67.01
African American	8	16.33	7	14.58	15	15.46
Asian	0	–	0	–	0	–
Caucasian	35	71.43	21	43.75	56	57.73
Hispanic	0	–	1	2.08	1	1.03
Biracial	0	–	0	–	0	–
Other	0	–	0	–	0	–
No Data	6	12.24	19	39.58	25	25.77
Highest Educational Degree						
B.S./B.A.	17	34.69	8	16.67	25	25.77
B.S./B.A. +	1	2.04	2	4.17	3	3.09
M.Ed./M.S.	15	30.61	13	27.08	28	28.87
M.Ed./M.S. +	8	16.33	3	6.25	11	11.34
Ed.S.	2	4.08	1	2.08	3	3.09
Ed.D/Ph.D.	0	–	2	4.17	2	2.06
No Data	6	12.24	19	39.58	25	25.77
Elementary Certification	43	87.76	23	47.92	66	68.03
ELL Certification	20	40.82	4	8.33	24	24.74
Reading Certification	5	10.20	11	22.92	16	16.49
Special Ed. Certification	1	2.04	5	10.42	6	6.19
	M	*SD*	*M*	*SD*	*M*	*SD*
Years in Current Position	5.07	4.73	6.34	6.00	5.58	5.27
Years in Teaching Profession	13.40	7.68	13.69	10.03	13.51	8.64

Note. Percentages were calculated based on available data.

1. Do children in the Comp program demonstrate stronger understanding of informational texts than comparable controls on commercially developed and researcher-made tests?
2. Do Comp+Transfer children outperform controls on commercial and researcher-made tests of reading comprehension in informational texts?
3. Do children in the Comp+Transfer program significantly outperform Comp program children on commercial and researcher-made tests in informational texts?

METHOD

Participants

Teachers

There were 97 teachers of 4th and 5th grade classrooms in 20 schools in a large school district in the Southeastern United States. Most teachers were female and Caucasian and nearly half held graduate degrees. Approximately one quarter of them were certified to teach English Language Learners (ELLs), 16% were certified in reading, and 6% were certified as special educators (see Table 1). Teachers of the tutored students did not implement either version of the tutoring program. Instead, they distributed and collected study-relevant paperwork, worked with the research team to schedule testing and tutoring sessions, and completed surveys. Control teachers collected study-relevant paperwork, helped to schedule the testing of students, and completed surveys. All study teachers were compensated equally for their cooperation and help.

Students

Recruitment and Selection. Our intention was to recruit students whose word reading was generally "low-average" but whose reading comprehension was lower. This was because our program was designed to improve reading comprehension, not word reading or fluency. After telling teachers the purpose of our study and asking them to distribute parental/guardian consents to students with relatively poor reading comprehension, we got back affirmative written responses from 547 households.

In the first test session of a gated screening process, 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 et al., 2012) and Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence-2 (WASI-2; Wechsler, 2011). We tested 531 students instead of 547 because 16 were no longer available for testing. Following this test session, 112 students were excluded because they scored below the 14th percentile on the TOWRE-2 SWE, which was our minimum criterion. Fifteen more students were lost because of scheduling changes. The remaining 404 children were screened on the Reading Comprehension subtest of the Gates-MacGinitie Reading

TABLE 2
Descriptive Statistics for Student Demographics and Screening Performance by Grade Levels and Grade Levels Combined

	Grade 4 (n = 87)		*Grade 5 (n = 102)*		*Combined (n = 189)*	
Variable	N	%	N	%	N	%
Female	44	50.57	50	49.02	94	49.74
African American	32	37.78	33	32.35	65	34.39
Caucasian	13	14.94	20	19.61	33	17.46
Hispanic	30	34.48	35	34.31	65	34.39
Other	10	11.49	4	3.92	14	7.41
Free/Reduced-Price Lunch	46	52.87	51	50.00	97	51.32
Individualized Education Plan	1	1.15	3	2.94	4	2.12
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
TOWRE-2 SWE raw score	62.55	5.60	65.25	6.03	64.01	5.97
TOWRE-2 SWE SS	93.36	7.21	93.16	7.79	93.25	7.51
TOWRE-2 SWE %ile of Mean SS	32nd		32nd		32nd	
Gates-MacGinitie RC Raw Score	20.26	3.80	17.78	4.74	18.93	4.50
Gates-MacGinitie RC NCE	38.87	7.29	33.34	11.38	35.89	10.08
Gates-MacGinitie RC %ile of Mean NCE	30th		25th		25th	

Note. Percentages were calculated based on available data. SS = standard score. NCE = normal curve equivalent.

Tests-4 (Gates-MacGinitie; MacGinitie et al., 2000), which was administered in 45 min to groups of 2–4 students. An additional 133 students were excluded because they scored above our maximum Gates-MacGinitie criterion of the 50th percentile. Forty-seven more students were eliminated because of changes to school schedules or school transfers.

Thus, 224 students were individually administered the Vocabulary subtest of the WASI-2. An additional eligibility requirement was that students score at or above a T-score of 37 on the Vocabulary subtest or Matrix Reasoning subset of the WASI-2. Six students with T-scores below 37 on both, indicating a Full-Scale IQ score below the “low average” threshold, were removed from the sample. Fourteen more students were excluded due to parent requests, school transfers, or frequent absences. The final selection pool consisted of 204 students who were given all remaining tests (see Table 2 for student demographics and screening data).

Attrition. Of the 204 students who were randomly assigned to Comp, Comp+Transfer, and Control study groups, complete data were collected on 189, a loss of 15 students for an overall attrition rate of 7.4%. Attrition rates by study group were 11.8%, 4.4%, and 5.9% for Comp, Comp+Transfer, and Control, respectively. A CONSORT flow diagram (Consolidated Standards of Reporting Trials; Schulz et al., 2010) is in the Appendix. It provides specific numbers of student participants, by study group, from initial screening to data analysis.

Procedures

Random Assignment

Eligible students were first grouped by school. Then, within schools, they were grouped by what time in the day their teachers permitted tutoring to occur. Within each of these time blocks and grade levels (fourth or fifth), students were randomly assigned to the three study groups. Group comparability was examined for each grade on TOWRE-2 SWE (raw scores and standard scores), Gates-MacGinitie (raw scores and NCE scores), and WASI-2 Matrix Reasoning (T-scores). No significant between-group differences were found. Finally, students were paired within treatment groups. At some schools, there was only one pair. When four or more students in a school were assigned to the same treatment group, they were paired such that each pair shared similar (or as similar as possible) TOWRE-2 SWE scores, reflecting our preference for more homogeneous dyads to facilitate peer interaction.

Project Staff and Timeline

All tutoring and testing sessions were conducted by 28 graduate students, who were hired as research assistants (RAs), and two project coordinators. Two RAs were doctoral students in special education. The remaining were master’s students in education and counseling. The two project coordinators and two doctoral students trained and supervised the RAs, along with carrying out their own tutoring and testing responsibilities. Pretreatment testing occurred between late August and early October. Tutoring began in late October. Posttreatment testing was conducted from late February to late March.

Similarities Between Comp and Comp+Transfer Tutoring

Pacing

Tutoring occurred three times weekly for 14 weeks, with each session lasting 43 min on average. Thus, total tutoring time was about 30 hours for both treatments. During each session, student pairs completed as much of a lesson

Before Reading

1) Vocabulary
- Identify bold words and review in glossary

2) Text Features
- Review pictures, captions, headings, etc., for clues about important information in the text

3) Text Structure
- Hunt for Text Structure Words to determine one of four structures:
  - Descriptive
  - Sequence
  - Compare-Contrast
  - Problem-Solution

4) Background Knowledge
- Discuss what is already known about passage topic
- Add to knowledge using media from a curated library of videos

5) Big Idea Prediction
- Predict most important idea of the passage

During Reading

1) Clarify
- Students identify (or tutor prompts for) confusing concepts
- Clarify confusing concepts by doing one of the following:
  - Reread relevant text
  - Activate background knowledge
  - Review glossary
  - Ask for outside help

2) Connect
- Deepen understanding of text using one of three prompts:
  - I wonder …
  - I'd like to know more about ___ ...
  - ____ made me think of …
- Connect to previous texts, videos, and/or background knowledge

After Reading

1) Main Idea/Big Idea
- Two components:
  - Most Important Who/What
  - Most Important Thing About the Who/What
- Main Idea: summarize *paragraph* in one sentence:
- Big Idea: summarize *passage* in one sentence

2) In or Out: Answering Questions
- Find the Key Words in question
- Find the Key Words in passage
- Read around the Key Words in passage
- Prove with evidence in passage to answer "In" questions
- Brainstorm, then combine evidence with background knowledge to answer "Out" questions

FIGURE 1 Comprehension Strategies Taught in Comp and Comp+Transfer Treatments.

as possible within the allotted time. The RAs followed written guidelines to ensure that no lesson components lasted too long. RAs conducted lessons using scripts but, to foster natural interactions with the children, they did not read from them. With the exception of four training lessons that introduced the comprehension strategies, tutoring lessons were grouped into five themed units (Explorers, Animals, Olympics, Change Your World, and Ancient Egypt). The units ranged from 5–11 nonfiction texts of 4–6 paragraphs in length. RAs were instructed to spend no more than two sessions on a lesson. When pairs worked slowly, the RAs were directed to increase pacing by limiting the amount of student-guided error correction or by skipping portions of lessons. So, although all tutored students received roughly the same amount of tutoring, they completed from 29 to 34 lessons across the 42 tutoring sessions.

Materials

For each themed unit (Explorers, Animals, Olympics, etc.), students were given an attractive workbook that contained the texts, accompanying text features (e.g., pictures and maps), and a glossary of uncommon words used in the unit. Students also received separate lesson worksheets that included comprehension questions to be completed after they read the texts. The texts consisted of original and adapted stories created or modified by the research team to ensure an appropriately challenging reading level for the students, with easier texts in the early units and more difficult texts introduced 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 et al., 2011). Across texts, Flesch-Kincaid grade-level scores ranged from 3.7 to 6.4. Lexile scores ranged from 500 to 830.

Strategy Instruction Overview

Students were introduced to the various strategies, which were organized and presented to them as before-, during-, and after-reading activities. Whereas lessons were mostly scripted, the RAs were trained to conduct them flexibly as they encouraged the children to use the strategies when reading. When introducing a strategy, the RAs provided explicit direction for how to complete each step. As students worked through the lessons, the RAs scaffolded their support, providing guidance and corrections as needed. A summary of strategies may be found in Figure 1, and a more detailed description may be found in Fuchs et al. (2018).

Differences Between Comp and Comp+Transfer Tutoring

Comp Treatment

The strategies described above and in Figure 1 were used by students in both treatments. Students in the Comp treatment engaged in one after-reading activity that was unique: Main Idea Recall. Students in that condition 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 unique activities designed to facilitate generalization of strategies and skills presented 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 one as they used it during the lesson. This was to serve as a self-regulatory prompt when they implemented the strategies with greater independence. Second, the Comp+Transfer students were encouraged to silently use the Main Idea strategy, in contrast to the Comp students, who (in pairs) created main ideas aloud. Third, when answering In or Out questions after reading, the Comp+Transfer pairs were instructed to identify the question type as "factual," "inferential," or "main idea." The pairs marked each question with an F, I, or M on their lesson-specific worksheets. Finally, beginning with Lesson 18, they practiced strategy use independently by completing a Reading Challenge. The Reading Challenges were administered every third lesson and consisted of short passages of varying genres with accompanying comprehension questions. Students were given their checklists and they were encouraged to use them to understand the passage and answer the questions. Afterward, they were encouraged to discuss their strategy use.

Additionally, in each lesson, the RAs explained the importance of transferring learning to forge a stronger connection between the strategies used in the tutoring context and their applicability to other contexts. This effort was meant to cultivate a "learning culture of demand" (Perkins & Salomon, 2012), with the hope that students would recognize patterns across contexts and demonstrate their learned strategies appropriately.

Broader Description of Study Participants' Reading Instruction

Tutored and non-tutored students participated in their schools' core literacy programs. The district required use of *Journeys* (Houghton Mifflin Harcourt, 2017) as a core curriculum, although schools were permitted discretion in supplementing it. At Grades 4 and 5, the scope and sequence of *Journeys* designates 75 instructional texts for use across 30 lessons, excluding texts for performance tasks or extended learning. Nonfiction texts comprise 30 and 27 of the 75 texts at Grades 4 and 5, respectively.

Both study groups also participated in their schools' daily "intervention blocks" (also known as "Response to Intervention" periods). We delivered our tutoring program during intervention block time. We assume most control students also received reading instruction during intervention block time because we and their teachers knew them to be weak readers who needed additional instruction. We emphasize that this is an assumption, not a fact. If control students received reading instruction during intervention block time, then tutored and non-tutored children obtained an equal amount of reading instruction. If the non-tutored children were instructed in mathematics, say, rather than reading, then the tutored and non-tutored students may have had unequal amounts of reading instruction.

Measures

Commercially Developed Reading Comprehension Measures

Two standardized commercial 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 three texts and answer open-ended factual and inferential questions about them. Questions are read aloud by the tester, and students may view the texts as they answer the questions. Sample-based reliabilities were not obtained because, within grade levels, not all fourth and fifth graders were assessed on identical passages and items due to a "passage reversal rule" (i.e., students performing poorly at grade level were subsequently given easier passages). The WIAT manual provides internal reliabilities of .85 and .91 for students in Grades 4 and 5, respectively. On the Gates-MacGinitie, students are given 35 min to read 11 short passages and to produce written answers to multiple-choice questions. Sample-based Cronbach's alpha for the Gates-MacGinitie at pre- and posttreatment was .61 and .81, respectively. The test's manual provides internal reliabilities of .93 and .92 for students in Grades 4 and 5, respectively.

Experimenter-Made Reading Comprehension Measures

Three additional near-transfer tests of reading comprehension were created and administered by the research team. First, a Knowledge Acquisition test required students to answer 20 multiple-choice questions about vocabulary, facts, and ideas in the instructional passages used in tutoring. The tester read aloud questions and answer choices, proceeding one question at a time so students could mark an answer in their test booklets. Sample-based Cronbach's alpha for 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. Passages and questions were similar in presentation (e.g., layout and design) to those used during tutoring. No passage had been seen previously by the students, but their content was drawn from topics featured in the instructional texts (e.g., the Civil Rights Movement). Sample-based Cronbach's alpha at pre- and posttreatment was .69 and .73, respectively.

Third, Near-Transfer Main Idea required students to read two passages aloud, each consisting of four paragraphs. Students orally provided a main idea statement following each paragraph. Cronbach's alpha for the sample at pre- and posttreatment was .71 and .79, respectively.

The research team also created a mid-transfer test of reading comprehension. The Mid-Transfer Reading Comprehension test consisted of two nonfiction passages about topics *not* addressed in tutoring (clean energy and Thanksgiving). Students gave written responses to multiple-choice questions (some of which required multiple answers) and fill-in-the-blank questions. Additional questions required them to mark answers directly in the passage. Like the Near-Transfer Reading Comprehension test, presentation of the Mid-Transfer Reading Comprehension passages and questions was similar to that used in tutoring. Sample-based Cronbach's alpha at pre- and posttreatment was .65 and .66, respectively.

Word Reading

Word reading was assessed with the TOWRE-2 SWE (Sight Word Efficiency) and the 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 sec from a list of words that increase in difficulty. Oral Reading Fluency asks students to read 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 and alternate form and test-retest reliabilities of .84 and .80, respectively, for students in Grades 3–8.

IQ

We used the Matrix Reasoning and Vocabulary subtests from the WASI-2 to measure IQ. 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, and the test manual provides internal reliabilities of .93, .89, and .92 for children ages 9, 10, and 11, respectively. The Vocabulary subtest evaluates expressive vocabulary, verbal knowledge, and foundational information. For each item, students identify a picture or provide a definition of a word read by the tester. Sample-based Cronbach's alpha was .66; the test manual provides internal reliabilities of .88, .92, and .88 for children ages 9, 10, and 11, respectively.

Fidelity

Test Administration Training

Prior to pretreatment testing, RAs were trained to administer and score all assessments in accordance with explicit test-developer guidelines. Over four weeks, RAs received approximately nine hours of training and, for an additional three hours, were required to practice administering and scoring the tests with a partner. For each test, RAs had to demonstrate at least 90% adherence to the standard administration and scoring rules during a formal fidelity check before testing students. If they failed the fidelity check, they were tested again until 90% adherence or better was demonstrated. These fidelity evaluations were conducted by the project coordinators and doctoral students using explicit and comprehensive checklists. Before post-treatment testing, RAs received an additional three hours of training and were required to meet the same fidelity criterion set prior to pretreatment testing.

Tutor Training

Before tutoring, RAs were prepared to administer lessons in standard fashion. They received eight hours of training across two days and practiced lesson administration with a partner for an additional two hours. Each RA had to achieve a fidelity score of 90% or higher. During tutoring, three additional fidelity checks were conducted on each RA, two by in-school observations and one by audio recording. All tutoring checks were conducted by a project coordinator or a doctoral student and scored with checklists reflecting all program components, including lesson preparation, implementation, and behavior management.

Adherence was operationalized as a percentage, such that the sum of correctly implemented behaviors was the numerator, and correct, incorrect, and unobserved behaviors were in aggregate the denominator. For Comp, fidelity for the two in-school observations and audio assessment was 94.3%, 98.3%, and 92.9%, respectively; for Comp+Transfer, it was 96.0%, 94.4% and 92.9%. Averaged across the two treatment conditions, fidelity for the two in-school checks and audio check was 94.4%, 95.2%, and 92.9%, respectively.

In addition, fidelity data were calculated separately for the tutoring components of Comp (common to the Comp and Comp+Transfer treatments) and Transfer (unique to the Comp+Transfer treatment). Because a portion of the fidelity checklists was destroyed before they could be entered item-by-item into a database, percentages reflect data collected on only fourth-grade students during the second (in-school) and third (audio) fidelity checks. For Comp alone, fidelity of Comp component implementation was 100% (in-school) and 91.8% (audio). For Comp+Transfer, fidelity of Comp component implementation was 94.4% (in-school) and 89.3% (audio). On only the Transfer components of Comp+Transfer, fidelity was 92.3% (in-school) and 87.5% (audio).

Inter-rater agreement on 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 in-school fidelity checks. Again, due to the loss of a portion of the checklists, those used for calculating inter-rater agreement overrepresented fourth-grade students at the middle and end of the treatment period. Overall inter-rater agreement was 89.7%.

Data Analysis

Reliabilities for Scoring Tests and Entering Data

Student test performance was double-scored by two scorers working independently of each other, and the data were then double-entered into identically organized databases, again by two RAs working independently. Scoring discrepancies were resolved by a project coordinator or doctoral student who went back to original test protocols. Inter-scorer agreement was calculated by a third RA (i.e., not the original scorer or the double-scorer), who listened to the audio files and completed a blank test protocol. This second protocol was then compared item-by-item to the original protocol. Agreement was calculated as a percentage of matched scores (numerator) over all recorded scores (denominator). Agreement was calculated only on tests for which students provided oral responses (e.g., WIAT), not written responses (e.g., Gates-MacGinitie). Thus, excluding the written tests, inter-scorer agreement was conducted on 22.4% of all test sessions. Inter-scorer reliability exceeded 90% for all measures at pre- and posttreatment with the exception of WASI-2 Vocabulary at pretreatment (85.5%).

Analytic Plan

We created three composite scores: (a) Word Reading (TOWRE-2 SWE and WRMT ORF); (b) Near-Transfer Reading Comprehension (Near-Transfer Reading Comprehension and Main Idea tests); and (c) Far-Transfer Reading Comprehension (Gates-MacGinitie and WIAT-3 Reading Comprehension subtests). The Mid-Transfer Reading Comprehension and Knowledge Acquisition measures were analyzed separately.

We accounted for potential clustering effects by conducting multilevel models, one for each outcome of interest: Knowledge Acquisition test, Near-Transfer Reading Comprehension composite, Mid-Transfer Reading Comprehension test, Far-Transfer Reading Comprehension composite, and Word Reading composite. 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 in each final model for accurate estimation of standard error. Teachers and pairs were cross-classified at Level 2 for students in the two treatment groups. Control students were nested in teachers with 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 case 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 using restricted maximum likelihood estimation procedures and Kenward-Roger standard error adjustment (McNeish & Wentzel, 2017).

Once the variance components were established for each outcome in the unconditional models, we added fixed effects to the final models. Our interest focused on (a) the academic benefit of our two program variants (Comp and Comp+Transfer) beyond typical schooling effects and (b) the potential added benefit of Transfer instruction. Thus, the treatment groups were compared to the control group and to each other with dummy variables. First, Comp alone was compared to control with d_Comp (Comp = 1; Trans = 0; Control = 0), and Comp+Transfer alone was compared to control with d_Trans (Comp = 0; Trans = 1; Control = 0). Then, both treatments combined were compared to control with d_trt (Comp = 1; Trans = 1; Control = 0). Finally, the treatments were compared to each other with d_comptran (Comp = -0.5; Trans = 0.5). In addition, three variables were included to reduce error variance in the outcome: grade (Grade 4 = -0.5, Grade 5 = 0.5), pretreatment TOWRE-2 SWE SS (the word-reading screening variable), and the pretreatment score associated with the outcome variable. Interactions between grade and treatment were also checked for each outcome. 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 coefficients (WWC, 2017).

RESULTS

Table 3 displays pre- and post-treatment means and standard deviations for individual tests and study groups. Final model results for the Knowledge Acquisition test and the reading comprehension outcomes are shown in Table 4. Effect sizes (Hedges' g) for all outcomes and comparisons may be found in Table 5.

Word Reading and Near-Transfer Knowledge Acquisition

The unconditional model for the Word Reading composite indicated a need for the following random effects: school, tutoring pair, and separate error terms for the three study groups. Interaction terms for treatment by grade level were not statistically significant, so those terms were removed from the final model. Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Students in neither Comp nor Comp+Transfer outperformed controls, Est. = 0.09, $SE = 0.19$, $p = .65$, $g = 0.05$, and Est. = 0.13, $SE = 0.22$, $p = .58$, $g = 0.07$, respectively. Similar

TABLE 3
Means and SDs for Pre- and Posttreatment Individual Measures by Study Group

	Comp			*Comp+Transfer*			*Control*		
Measure	N	Mean	SD	N	Mean	SD	N	Mean	SD
Pr TOWRE-2 SWE (SS)	*60*	*92.57*	*7.06*	*65*	*93.65*	*7.71*	*64*	*93.48*	*7.78*
Po TOWRE-2 SWE (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 RC (NCE)	60	36.03	10.12	65	35.30	11.51	64	36.34	8.49
Po Gates RC (NCE)	60	43.10	10.24	65	41.71	13.25	64	43.45	11.76
Pr WIAT RC (SS)	60	93.17	7.14	65	94.12	6.54	62	93.39	7.46
Po WIAT RC (SS)	59	97.92	6.94	65	98.91	8.40	64	97.25	8.09
Pr Mid RC (raw)	60	9.80	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	3.83	64	15.48	4.32
Po Near RC (raw)	60	18.72	3.46	65	19.22	3.24	64	17.33	3.59
Pr Near MI (raw)	60	1.73	1.96	65	2.08	1.96	64	1.95	1.85
Po Near MI (raw)	60	3.78	2.35	65	3.97	2.29	64	1.72	1.88
Pr Knowledge Acq. (raw)	60	9.63	2.50	65	9.80	2.53	64	10.05	2.63
Po Knowledge Acq. (raw)	60	17.08	1.79	65	16.78	1.78	64	11.59	2.76

Note. Pr = pretest. Po = posttest. SS = standard score. NCE = normal-curve equivalent.

nonsignificant results were obtained on the Reading Composite when the two treatment groups combined were compared to controls, Est. = 0.10, *SE* = 0.18, $p = .56$, $g = 0.06$, and when they were compared to each other, Est. = 0.05, *SE* = 0.21, $p = .81$, $g = 0.03$ (favoring the Comp+Transfer group).

The unconditional model for Knowledge Acquisition indicated a need for random effects for school, teacher, tutoring pair, and error terms for each group. Interaction terms for treatment by grade level were not statistically significant and were removed from the final model. Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Students in the Comp and Comp+Transfer groups outperformed controls, Est. = 5.58, *SE* = 0.36, $p < .001$, $g = 2.40$ and Est. = 5.24, *SE* = 0.37, $p < .001$, $g = 2.25$, respectively. Also, students in the two tutored groups combined performed more strongly than controls, Est. = 5.41, *SE* = 0.33, $p < .001$, $g = 2.50$. There was no statistically significant difference between the Comp and Comp+Transfer groups, Est. = −0.36, *SE* = 0.29, $p = .22$, $g = 0.20$, favoring the Comp group.

Reading Comprehension

Near Transfer

The unconditional model for the Near-Transfer Reading Comprehension composite indicated a need for random effects for school, teacher, tutoring pair, and error terms for each group. Interaction terms for treatment by grade level were not statistically significant and removed, and Level-1 residuals met assumptions of normality and homoscedasticity. The Comp and Comp+Transfer groups outperformed controls, Est. = 1.31, *SE* = 0.25, $p < .001$, $g = 0.88$ and Est. = 1.51, *SE* = 0.23, $p < .001$, $g = 1.04$, respectively. Comparing all tutored students to controls also yielded results favoring the tutored students, Est. = 1.43, *SE* = 0.19, $p < .001$, $g = 0.98$. There was no statistically significant difference between the Comp and Comp+Transfer groups, Est. = 0.21, *SE* = 0.28, $p = .45$, $g = 0.15$, favoring the Comp+Transfer group.

Mid Transfer

The unconditional model for the Mid-Transfer Reading Comprehension test indicated a need for random effects for school, teacher, tutoring pair, and error terms for each group. Interaction terms for treatment by grade level were not statistically significant and were removed. Level-1 residuals from the final model met assumptions of normality and homoscedasticity. Students in Comp+Transfer performed significantly better than controls, Est. = 1.18, *SE* = 0.42, $p = .006$, $g = 0.40$, whereas students in Comp did not, Est. = 0.45, *SE* = 0.44, $p = .29$, $g = 0.15$. An additional model comparing all tutored students to controls yielded significant results in favor of the tutored students, Est. = 0.85, *SE* = 0.38, $p = .025$, $g = 0.29$. A direct comparison between the Comp and Comp+Transfer groups yielded no significant difference, Est. = 0.57, *SE* = 0.47, $p = .19$, $g = 0.21$, favoring the Comp+Transfer group.

Far Transfer

The unconditional model for the Far-Transfer Reading Comprehension composite showed a need for the following random effects: teacher, tutoring pair, and error terms for each group. Interaction terms for treatment by grade level were not statistically significant and were removed; residuals from the final model met assumptions of normality and homoscedasticity after removing four multivariate outliers

TABLE 4
Multilevel Model Results for Posttreatment Measures

Measure/Fixed Effect	*Estimate*	SE	T	p
Knowledge Acquisition ($N = 188$)				
Intercept	7.90	1.83	4.31	0.00
Pretreatment Score	0.37	0.06	6.62	0.00
Pretreatment TOWRE-2 SS	−0.00	0.02	−0.08	0.94
Comp+Transfer vs. Control	5.24	0.37	14.29	0.00
Comp vs. Control	5.58	0.36	15.39	0.00
Grade 5	0.29	0.31	0.93	0.37
Near-Reading Comp. ($N = 188$)				
Intercept	−1.19	1.24	−0.96	0.34
Pretreatment Score	0.40	0.06	6.38	0.00
Pretreatment TOWRE-2 SS	0.00	0.01	0.04	0.97
Comp+Transfer vs. Control	1.51	0.23	6.72	0.00
Comp vs. Control	1.31	0.25	5.16	0.00
Grade 5	0.41	0.26	1.61	0.13
Mid-Reading Comp. ($N = 188$)				
Intercept	7.28	2.33	3.13	0.00
Pretreatment Score	0.57	0.06	9.20	0.00
Pretreatment TOWRE-2 SS	−0.03	0.02	−1.18	0.24
Comp+Transfer vs. Control	1.18	0.42	2.84	0.01
Comp vs. Control	0.45	0.44	1.01	0.32
Grade 5	0.11	0.44	0.24	0.81
Far-Reading Comp. ($N = 186$)				
Intercept	0.41	0.53	0.78	0.44
Pretreatment Score	0.92	0.10	9.35	0.00
Pretreatment TOWRE-2 SS	−0.01	0.01	−0.97	0.33
Comp+Transfer vs. Control	0.03	0.10	0.29	0.77
Comp vs. Control	0.02	0.10	0.22	0.82
Grade 5	0.13	0.10	1.32	0.19

Note. The Near-Reading Comprehension composite includes the Near-Transfer Reading Comprehension and Main Idea tests. The Far-Reading Comprehension composite includes Gates-MacGinitie and WIAT Reading Comprehension subtests. TOWRE-2 is the Sight Word Efficiency subtest of the Test of Word Reading Efficiency, 2nd edition. Necessary random effects were included in each model but not presented due to space considerations.

TABLE 5
Effect Sizes for Study Group Contrasts

Outcome	*Comp+Transfer vs. Control*	*Comp vs. Control*	*Combined Treatments vs. Control*	*Comp+Transfer vs. Comp*
Word Reading	0.07	0.05	0.06	0.03
Knowledge Acq.	2.25***	2.40***	2.50***	−0.20
Near RC	1.04***	0.88***	0.98***	0.15
Mid RC	0.40**	0.15	0.29*	0.21
Far RC	0.04	0.03	0.03	0.00

* $p < .05$; ** $p < .01$; *** $p < .001$.
Note. **Bold text** indicates a composite score. Effect sizes are Hedges' g. The Word Reading composite includes TOWRE-2 SWE and WRMT ORF. The Near RC composite includes the Near-Transfer Reading Comprehension and Main Idea tests. The Far RC composite includes Gates-MacGinitie and WIAT Reading Comprehension subtests.

(standardized residuals = 2.58 and -2.58 and below). Neither students in Comp nor Comp+Transfer outperformed controls, Est. = 0.02, $SE = 0.10$, $p = .82$, $g = 0.03$, and Est. = 0.03, $SE = 0.10$, $p = .77$, $g = 0.04$, respectively. Comparison of all tutored students to controls also indicated no statistically significant difference, Est. = 0.02, $SE = 0.08$, $p = .80$, $g = 0.03$. Moreover, Comp and Comp+Transfer performed similarly, Est. = −0.00, $SE = 0.10$, $p = .99$, $g = 0.00$.

DISCUSSION

The purpose of this experimental study was to extend previous work (see Fuchs et al., 2018) on a tutoring program aimed at strengthening at-risk intermediate-grade children's understanding of informational texts. Both versions of the tutoring program evaluated in this study included strategies for students to use before, during, and after reading. Comp students practiced recalling from memory the main idea statements they had created during the lesson, whereas Comp+Transfer students engaged in activities designed to facilitate the transfer of strategies addressed during tutoring (e.g., using strategy checklists, identifying question types, and completing Reading Challenges). Otherwise, the two program variants were the same. The efficacy of the two programs was assessed using commercially developed and researcher-made comprehension measures, which collectively tested near-to-far levels of transfer of learning.

Results suggested that the two programs promoted many students' comprehension of nonfiction passages. Specifically, both groups of tutored students significantly and dramatically outperformed controls on the Knowledge Acquisition test and the Near-Transfer Reading Comprehension composite. The Knowledge Acquisition test narrowly focused on factual retention. The texts of the measures comprising the Near-Transfer composite were written to represent nonfiction texts—similar in content to those read during tutoring sessions but never seen by the students. These "familiar but different" texts were the basis for the kinds of comprehension questions students encounter in their classrooms and beyond. Their performance on the Near-Transfer composite generally suggested that they learned the strategies and skills taught during tutoring and applied them as they read previously unseen content.

For teachers of students who have difficulty comprehending nonfiction texts, this finding may be of most interest. Tutoring, we believe, strengthened students' reading comprehension because (a) we integrated multiple evidence-based strategies for them to use before, during, and after reading a passage; (b) instruction was tutor-driven, systematic, and explicit until students demonstrated mastery and independent use of the strategies; (c) students were engaged by means of peer-mediated learning, and attractive and novel instructional materials, including curated video clips to build background knowledge; and (d) because of a generally collaborative and supportive context for learning (Best et al., 2008; Hall et al., 2005; McNamara et al., 2011). Although the study was implemented with student pairs, the strategies and collaborative teamwork described could be applied to small groups of children as well (with the caveat that this would be beyond the scope of the original intervention).

On the Mid-Transfer Reading Comprehension test, students in the Comp+Transfer program significantly outperformed controls whereas the students in Comp did not. However, there was no statistically significant difference between the Comp and Comp+Transfer groups. We offer several reasons for this finding. First, we believe an important instructional component of Comp+Transfer was Reading Challenges. This activity gave students greater opportunity to independently practice and monitor strategy use and to get frequent feedback from the tutors. While we did not have the resources to conduct a component analysis of Comp+Transfer, post-intervention interviews with the tutors suggested the activity's usefulness.

Second, the lack of a significant difference between the two tutored groups may reflect the limited utility of the other transfer instructional components: the strategy checklists, the construction of Main Ideas, and the labeling of question types. Because we did not want to take much time from our base Comp program—because of its relative efficacy with similar students in prior studies (e.g., Fuchs et al., 2018)—we may inadvertently have given short shrift to the very activities that we added to make Comp+Transfer unique. (An exception is the Reading Challenges activity.)

Third, the measure we created to assess Mid-Transfer required considerably greater transfer than the tests making up our Near-Transfer composite, which more closely resembled the content of the tutoring sessions. To the extent that we helped Comp+Transfer students strengthen their capacity to transfer learning, the Mid- and Far-Transfer tests were the more likely measures to reflect that.

For the Far-Transfer Reading Comprehension composite, between-group differences were equivocal and nonsignificant, although effect sizes were in line with those found previously for this reading program on similar measures (Fuchs et al., 2018). The two tests in the Far-Transfer composite were not aligned with either program variant. For most students, we suspect, the content and context of these tests represented "a bridge too far." For example, more than half of the passages that students read in each of the Gates-MacGinitie and the WIAT-3 Reading Comprehension subtests were not nonfiction (e.g., folk tales, first-person narratives). The Mid-Transfer test was our attempt to "split the difference" between the two ends of the transfer spectrum. That is, we hoped it would give students a chance to apply their strategies and skills to nonfiction passages that superficially resembled those from tutoring but for which they had not received any in-program practice or background knowledge. Although this effort met with some success, as the Comp+Transfer students outperformed the control group, their performance was insufficiently strong to significantly distinguish them from the Comp students.

Teaching and Measuring Transfer of Learning

Results from this study suggest our transfer instruction benefited some study participants beyond the effects of the base Comp program. Whereas we do not wish to minimize or devalue this finding, it is important to acknowledge that the benefits were not particularly strong or robust. Our goal was to intensify our base program qualitatively by modifying its operationalization. We knew that by only intensifying it quantitatively, by means of 1:1 teaching ratios or 60-minute sessions, we would render it unfeasible in our school district. We also suspected that teaching for transfer could strengthen a greater number of cognitive processes than training one process (e.g., working memory), and we attempted to ensure that the embedded transfer activities were not just "teaching to the test," but that they would promote self-sustaining strategy use.

To some, it may seem difficult if not impossible to teach for transfer without also "teaching to the test." That is, the two may be seen as one and the same. In the case of our reading program, the best evidence for learning transfer would be unremarkable and unobservable: the student would read a passage, understand its factual content, and consider its inferential implications while internally following the taught strategies as needed. Instead, for purposes of program evaluation, we must make use of structured demonstration of reading comprehension via commercially-developed and experimenter-made tests. And, as discussed, the use of non-commercial tests of near and mid transfer has been discouraged by many (e. g., Slavin, 2019; WWC, 2017).

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

Study Limitations

The relatively low sample-based Cronbach's alphas of several of our measures, commercially developed and self-made, pose limitations on our study. Although we engaged the RAs in rigorous training, preparing them to accurately administer and score all the measures, the fact remains that if some of these measures had been more internally reliable, our findings may have been clearer and more consistent in terms of both comparing Comp and Comp+Transfer to controls and comparing the two to each other. A second limitation is that we obtained insufficient information on classroom reading instruction. If we had had the resources, we could have described and compared the reading experiences of the three study groups, which may have permitted us to (a) dismiss potential confounds and (b) isolate unique instructional dimensions of our tutoring programs and understand their "active ingredients."

As explained, we also had an incomplete set of fidelity data due to the loss of a portion of our fidelity checklists. Whereas we provided data on overall adherence to the two tutoring programs, we could not assess fidelity of the Comp+Transfer components at all data collection points and grade levels. More complete fidelity data may have permitted a better understanding of the non-significant differences between the two tutoring groups, particularly on the Near-Transfer Reading composite and Mid-Transfer Reading Comprehension test.

Finally, in comparing the two versions of our tutoring program to each other, we faced a hard choice: equalize instructional time or content coverage. On one hand, we could have permitted students in the Comp+Transfer group more time per session than those in the Comp group to ensure fuller participation in the additional (and presumably beneficial) transfer activities. This would have given them more instructional time, and greater instructional time would have become a competing explanation for their possible superior performance on reading outcomes.

On the other hand, we could have made certain that both groups obtained equal instructional time, risking the probability that they would get unequal content coverage. Given these two options, we decided to ensure equal tutoring time. Thus, the amount of content covered by the Comp students exceeded that of students in the Comp+Transfer group because Comp students had fewer activities to complete per lesson. Specifically, Comp pairs completed 29.3 lessons (on average) across their sessions, whereas the Comp+Transfer pairs completed 26.2 lessons. In principle, and with greater resources, we could have established a second Comp+Transfer group (representing a third active treatment group) for students who completed the same number of lessons as the Comp students. This would have allowed another means of determining the added value of transfer instruction.

Study limitations notwithstanding, we believe both versions of our tutoring program improved many students' nonfiction reading comprehension. Furthermore, we obtained evidence (albeit limited) to suggest that transfer instruction and practice strengthened performance on our measure of mid transfer. In future work, we will explore the potential of transfer instruction in a different way by providing similarly at-risk students with one of two interventions: one similar to the Comp+Transfer described here, and another version with fewer strategies to learn and practice. With fewer strategies to learn and remember, our thinking goes, the reduced program may promote more effective strategy use and transfer by the students. A secondary benefit of this briefer or abbreviated program may be an increase in ease of implementation for our RAs and, eventually, teachers and other interventionists. Our foremost goal remains to create an efficacious version of the program for at-risk readers that their teachers can implement with fidelity in authentic settings.

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*. AERA.

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

Clemens, N., & Fuchs, D. (in press). Commercially-developed tests of reading comprehension: Gold standard or fool's gold? *Reading Research Quarterly*.

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., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Zhang Tracy, W., Patton, S., Davis-Perkins, N., Kim, W., Elleman, A. M., & Peng, P. (2018). Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near-transfer. *Learning Disabilities Research & Practice*, *33*(1), 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., Peng, P., & 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

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

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

Houghton Mifflin Harcourt. (2017). Journeys *[Scope and sequence*]. Houghton Mifflin Harcourt. https://www.hmhco.com/documents/journeys-scope-sequence-grades-k-6

Institute of Education Sciences & National Science Foundation. (2013). *Common guidelines for education research and development*. 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

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

MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). 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

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*. http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf

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). Lawrence Erlbaum Associates.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & 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

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). 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

Schneider, M. (2020, May 5). *Making common measures more common*. Institute of Education Science*s*. https://ies.ed.gov/director/remarks/5-05-2020.asp

Slavin, R. (2019, October 24). *Developer- and researcher-made measures*. https://robertslavinsblog.wordpress.com/2019/10/24/developer-and-researcher-made-measures

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

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.). Pro-Ed.

U.S. Department of Education. (2019). *NAEP reading report card*. National Center for Education Statistics, Institute of Education Sciences. https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4

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.). Psychological Corporation.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). NCS Pearson.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Standards handbook (Version 4.0)*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

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.). Pearson.

APPENDIX A

CONSORT Flow Diagram for Student Enrollment

CONSORT
TRANSPARENT REPORTING of TRIALS

Enrollment

Assessed for eligibility ($N = 531$)

Excluded ($N = 327$)
- Not meeting inclusion criteria ($N = 251$)
- Participated in previous study ($N = 13$)
- Other reasons ($N = 63$)

Randomized ($N = 204$)

Allocation

Allocated to Comp+Transfer ($N = 68$)
- Received allocated intervention ($N = 66$)
- Did not receive allocated intervention ($N = 2$)
  - Moved out of district ($N = 2$)

Allocated to Comp ($N = 68$)
- Received allocated intervention ($N = 60$)
- Did not receive allocated intervention ($N = 8$)
  - Moved out of district ($N = 4$)
  - Parent/student withdrew ($N = 3$)
  - Dropped for unmanageable behavior ($N = 1$)

Allocation

Allocated to control ($N = 68$)
- Received allocated control ($N = 64$)
- Did not receive allocated control ($N = 4$)
  - Moved out of district ($N = 4$)

Posttesting

Lost to post-testing ($N = 1$)
- Moved out of district ($N = 1$)

Lost to post-testing ($N = 0$)

Posttesting

Lost to post-testing ($N = 0$)

Analysis

Analyzed ($N = 65$)
- Excluded from analysis ($N = 0$)

Analyzed ($N = 60$)
- Excluded from analysis ($N = 0$)

Analysis

Analyzed ($N = 64$)
- Excluded from analysis ($N = 0$)

About the Authors

Samuel Allen Patton is a Research Associate at Vanderbilt University.

Douglas Fuchs is an Institute Fellow at the American Institutes for Research and Research Professor at Vanderbilt University.

Emma L. Hendricks is an Instructional Support Teacher in the Martinez Unified School District.

Annie J. Pennell is the Program Supervisor for the English Language Arts Learning Assistance Program in the Washington Office of the Superintendent of Public Instruction.

Meagan E. Walsh is an Intensification Specialist at the at Michigan MTSS Technical Assistance Center.

Lynn S. Fuchs is an Institute Fellow at the American Institutes for Research and Research Professor at Vanderbilt University.

Wen Zhang Tracy is an Early Interventionist at Foundations Early Intervention Services.

Loulee Yen Haga was a Project Coordinator at Vanderbilt University.

This research was conducted at the National Center on Accelerating the Academic Achievement of Students with Severe and Persistent Learning Disabilities, which is funded by Grant No. R324D130003 from the National Center on Special Education Research, Institute for Educational Sciences.