



Using strategy instruction to help struggling high schoolers understand what they read



Institute of Education Sciences
U.S. Department of Education



Using strategy instruction to help struggling high schoolers understand what they read

October 2007

Prepared by

Helen Apthorp

Mid-continent Research for Education and Learning

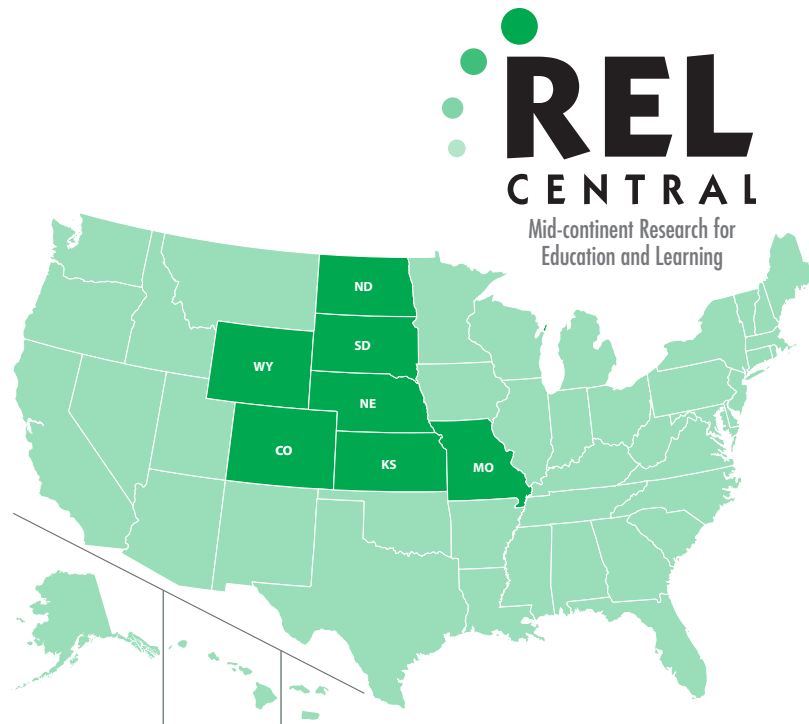
Tedra Clark

Mid-continent Research for Education and Learning



Institute of Education Sciences

U.S. Department of Education



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

October 2007

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-C0-0023 by Regional Educational Laboratory Central administered by Mid-continent Research for Education and Learning. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Apthorp, H., & Clark, T. (2007). *Using strategy instruction to help struggling high schoolers understand what they read* (Issues & Answers Report, REL 2007–No. 038). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from <http://ies.ed.gov/ncee/edlabs>

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

Summary

Using strategy instruction to help struggling high schoolers understand what they read

The evidence indicates that peer-assisted learning can have a substantively important positive effect on struggling high school students' reading comprehension. But reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention in the one study that met evidence standards.

This review sought to locate and summarize findings from rigorous, scientifically based studies of the effectiveness of strategy instruction—teaching students to use and articulate strategies that foster active, competent, self-regulated, and intentional learning—for helping struggling high school students improve their reading comprehension. The goal was to address information needs in the Central Region by identifying evidence-based practices intended to help high school teachers teach struggling readers.

Extensive searches were conducted for relevant studies, which were then screened against rigorous evidence standards. To identify the least biased estimates of the effect of strategy instruction, the evidence screens required that only exposure to strategy instruction distinguish the intervention and comparison

groups. By establishing this single difference, outcomes can with some confidence be attributed to strategy instruction. One study, on peer-assisted learning strategies, passed the relevance and rigorous evidence screens.

Fuchs, Fuchs, and Kazdan (1999) studied the effects of peer-assisted learning strategies on struggling high school readers' reading comprehension. Using a quasi-experimental design, the study passed each What Works Clearinghouse evidence screen, meeting evidence standards with reservations. The improvement index, a measure that helps in attributing practical significance to study findings, was 13 percentile points—meaning that the student at the midpoint in the control group distribution would have gained at least 13 percentile points in achievement if exposed to the intervention.

The study showed that peer-assisted learning improved performance in reading comprehension when the struggling readers in the sample engaged in three key collaborative activities: reading passages aloud with a partner modeling and coaching, formulating a general understanding of what they read by asking and answering questions about each paragraph (with their partners), and predicting and confirming or disconfirming predictions

of what would be learned next (again, with their partners).

The evidence indicates that peer-assisted learning can have a substantively important positive effect on struggling high school students' reading comprehension. But reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention.

In addition, the extent of evidence was very small: 102 students, primarily students with disabilities in 10 high schools in only the southeastern United States. Comprehension was measured using only researcher-developed assessments and only one type of text: folktales with grade 2.5 readability. With only these assessments, instructional settings, and students in the evidence base the finding of positive effects for peer-assisted learning has limited applicability. Further research is needed to strengthen the evidence base. A first

step would be to randomly assign students to the intervention.

Better and broader outcome measures are also needed to assess the effect of peer-assisted learning. Outcome measures in future studies could include passages with grade 7 or higher readability. A state high school reading assessment could also be included to evaluate whether benefits might accrue for high-stakes indicators of achievement. To accommodate struggling readers' difficulties and reduce barriers to reading, studying the effects of peer-assisted learning on listening comprehension would also be valuable.

The findings in this report are limited by the scope of the searches and the research available. Findings and conclusions about the effectiveness of strategy instruction for struggling high school readers may change as additional research is identified or completed.

October 2007

TABLE OF CONTENTS

Overview	1
Why this report	3
Results from one study of peer-assisted learning	5
Study design	5
School and student samples—10 high schools and 102 students	6
The intervention: peer-assisted learning strategies	6
Effects on reading comprehension	6
Better research on strategy instruction	7
Appendix A Methodology	8
Appendix B Students with disabilities in the Central Region by state and disability category	15
Appendix C Search terms and results for literature searches	16
Notes	18
References	19
Box 1 Methods and definitions of key terms	2
Tables	
A1 Summary of Fuchs, Fuchs, & Kazdan (1999)	14
B1 Students ages 12–17 with disabilities in the Central Region by state and disability category	15
C1 Bibliographic database search terms and results	16
C2 Publication list search results	17

The evidence indicates that peer-assisted learning can have a substantively important positive effect on struggling high school students' reading comprehension. But reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention in the one study that met evidence standards.

OVERVIEW

This review sought to locate and summarize findings from rigorous, scientifically based studies of the effectiveness of strategy instruction—teaching students to use and articulate strategies that foster active, competent, self-regulated, and intentional learning—for helping struggling high school students improve their reading comprehension. The goal was to address information needs in the Central Region by identifying evidence-based practices intended to help high school teachers teach struggling readers.

Extensive searches were conducted for relevant studies, which were then screened against rigorous evidence standards. To identify the least biased estimates of the effect of strategy instruction, the evidence screens required that only exposure to strategy instruction distinguish the intervention and comparison groups. By establishing this single difference, outcomes can with some confidence be attributed to strategy instruction. One study, on peer-assisted learning strategies, passed the relevance and rigorous evidence screens.

Fuchs, Fuchs, and Kazdan (1999) studied the effects of peer-assisted learning strategies on struggling high school readers' reading comprehension. Using a quasi-experimental design, the study passed each What Works Clearinghouse evidence screen, meeting evidence standards with reservations (see box 1 and appendix A for methods and definitions). The effect size was 0.34. Although not statistically significant, this effect size is substantively important and positive by What Works Clearinghouse standards, discussed later in the report. The improvement index for this effect size is 13 percentile points—meaning that the student at the midpoint in the control group distribution would have gained at least 13 percentile points in achievement if exposed to the intervention.

The study showed that peer-assisted learning improved performance in reading comprehension when the struggling readers in the sample engaged in three key collaborative activities: reading

BOX 1

Methods and definitions of key terms

The four stages of What Works Clearinghouse's study review process were used to locate and screen relevant studies. These stages include defining relevant studies for review, searching for and screening research reports for relevance, systematically applying evidence standards and categorizing studies accordingly, and, for studies that met evidence standards either with or without reservations, systematically describing participant and other study characteristics, intervention characteristics, outcomes, and findings.

Defining relevant studies for review

Topical criteria were used to locate studies on strategy instruction. Studies were included if their purpose was to examine the effectiveness of strategy instruction for improving the reading comprehension of struggling high school readers, including students with learning disabilities. Studies had to involve students in grades 9, 10, 11, or 12 and list reading comprehension as an outcome measure. Only studies published over 1975–2007, after the passage of the original Individuals with Disabilities Education Act, were included.

Research design criteria were used to locate studies that examined the effectiveness of strategy instruction. To draw causal conclusions about the effects of strategy instruction, researchers for this report searched for and included studies using either randomized controlled trials or quasi-experimental designs.¹ In randomized

controlled trials, participants are randomly assigned to different experimental groups. By placing participants in groups randomly (for example, assuring that each student has an equal chance of being placed in one group or another), the assignment reduces the risk of bias and establishes that the only systematic difference between the groups is the presence or absence of the intervention.

Quasi-experimental designs do not randomly assign participants to intervention and comparison groups, but the groups are matched or shown to be equivalent before the intervention. Establishing the pre-intervention equivalence in a quasi-experimental design gives some confidence, though not as much as with a randomized controlled trial, that the measured effects are due to the intervention.

Search strategy for relevant studies

Searches for relevant randomized controlled trials and quasi-experimental design studies were conducted using two types of sources: bibliographic databases and lists of publications, either in prior reviews or available on relevant web sites. Completed between March 2006 and February 2007 were 26 searches, 12 of bibliographic databases and 14 of lists of publications. Searches included some of the gray literature (Dissertation Abstracts International), but they did not include reports recommended by experts or in conference proceedings.

Evidence screen and adequacy of data review

The intent of this report was to use a standard, replicable set of criteria and

procedures for evaluating the technical and methodological quality of relevant studies. The report therefore adopted What Works Clearinghouse standards of evidence and rules for categorizing randomized controlled trials and quasi-experimental design studies into one of three levels of technical quality: meets evidence standards, meets evidence standards with reservations, and does not meet evidence screens.

The protocol for categorizing studies involved eight evidence screens: adequacy of the outcome measures, baseline equivalence, attrition, differential attrition, potential intervention contamination, potential teacher–intervention confound, potential school–intervention confound, and adequacy of reporting for computing effect sizes. Whenever there was not enough information in the study report to accurately assess against an evidence screen, authors were contacted for the necessary information. The studies were assessed according to the authors' responses.

Coding and analysis of key study characteristics

Key information was coded from the one study that met evidence standards, Fuchs, Fuchs, and Kazdan (1999). This information included the locale of the participating schools and district; students' grade level, diagnostic category, and average pretest intelligence quotient and reading achievement scores; and the intervention and comparison condition, intervention duration, and setting characteristics.

This information was used to compute an effect size and an improvement index.

Effect size. An effect size—a standardized mean difference—expresses in standard deviation units the increase or decrease in achievement of the intervention group compared with that of the control or comparison group using a “difference in differences” approach.

Improvement index. The improvement index—ranging from –50 to +50 percentile points, with positive values favoring the intervention group—represents the difference in percentile

rank between the average students in the intervention and comparison groups. An improvement index of 13 percentile points, for example, means that the average student in the control group would have gained at least 13 percentile points in achievement if exposed to the intervention.

What Works Clearinghouse decision rules for the importance of study effects. What Works Clearinghouse decision rules were used to describe the importance of the study effects:

statistically significant and positive, substantively important and positive, indeterminate, substantively important and negative, or statistically significant and negative.

Note

1. Regression discontinuity and single subject, multiple baseline designs were not included because a replicable set of criteria and procedures for rating the technical adequacy of these designs was not available from What Works Clearinghouse (2006a) at the time of the initiation of the report.

passages aloud with a partner modeling and coaching, formulating a general understanding of what they read by asking and answering questions about each paragraph (with their partners), and predicting and confirming or disconfirming predictions of what would be learned next (again, with their partners).

Although the evidence indicates that peer-assisted learning can have a substantively important positive effect on struggling high school readers’ reading comprehension, reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention.

WHY THIS REPORT

Educators in the Central Region states (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming) have requested scientifically valid research on practices and policies for teaching diverse groups of high school students (Mid-continent Research for Education and Learning, 2006). This study identifies evidence-based practices intended to help high school teachers teach struggling readers.

High school texts are usually more abstract, dense, and complex than elementary texts. But high school

teachers typically are not attuned to helping their students with the more demanding reading tasks, such as preparing for class discussion based on reading textbook passages and other written materials (Balfanz, McPartland, & Shaw, 2002; Langer, 1999). As stated in a recent report, it is critical that secondary school teachers “better understand and teach specific literacy strategies to help students read and extract meaning from the written material used to teach the course content” (National Association of Secondary School Principals, 2005, p. 1).

For this report, struggling readers are defined broadly, using a definition from the U.S. Department of Education’s Striving Readers Program, which characterizes struggling readers as students who read at least two years below grade level (Federal Register/Vol. 70, No. 189/Friday, September 30, 2005/Notices, p. 57257). In many schools struggling readers make up one-quarter to two-thirds of entering freshmen (Neild & Balfanz, 2006). Strategic engagement in reading is necessary to meet the complex learning demands of the high school curriculum (Scruggs & Mastropieri, 2004). For struggling readers the problem is often that they do not recognize when reading comprehension breaks down and so do not actively move beyond the problem with alternative strategies (Rose, 2000).

The focus of this report is struggling high school readers, including those with learning disabilities.

Although students with learning disabilities may appear to function competently and independently, the majority of them have severe reading difficulties (Wagner, Newman, Cameto, & Levine, 2006), and they make up the largest subset by far of students with disabilities (see appendix B). Struggling high school readers are also eligible for the U.S. Department of Education’s Striving Readers Program. A goal of this report and of the Striving Readers Program is to leverage and expand the research base of practices that strengthen the literacy skills of adolescent readers.

The literacy skill of interest here is reading comprehension: understanding what is read, learning new concepts, getting deeply involved in reading, critically evaluating text, and applying new knowledge to solve intellectual and practical problems, as reflected in high school state standards in the Central Region and as defined in key publications on reading comprehension (Anders, 2002; RAND Reading Study Group, 2002). According to the National Assessment of Educational Progress Reading Framework, reading comprehension comprises four aspects: “forming a general understanding, developing interpretation, making reader/text connections, and examining content and structure” (National Assessment Governing Board, 2005, p. 7).

Relevant interventions for struggling readers include strategy instruction (Mastropieri, Scruggs, Bakken, & Wheadon, 1996)—which includes reciprocal teaching (Weedman & Weedman, 2001) and peer-assisted learning (Topping, 2001)—skills training (Mastropieri et al., 1996; Swanson, 1999), text enhancements (Gersten, Fuchs, Williams, & Baker, 2001), and wide reading (Fisher & Ivey, 2006). Strategy instruction helps students become more active, organized, and thoughtful readers and learners (Gersten et al., 2001). Strategy instruction was selected for this report because of its alignment with high school reading standards, which expect that high school students can “use a range of automatic monitoring

and self-correction methods (for example, rereading, slowing down, sub-vocalizing, consulting resources, questioning)” (Mid-continent Research for Education and Learning, 2007).

This report uses the definition of strategy instruction in Trabasso and Bouchard (2002), two members of the comprehension subgroup of the National Reading Panel (2000). Strategies are “specific, learned procedures that foster active, competent, self-regulated, and intentional learning” (Trabasso & Bouchard, 2002, p. 176). Strategy instruction is teaching students to use and articulate these strategies (Trabasso & Bouchard, 2002). Strategy instruction shifts the role of the teacher away from asking comprehension questions to helping students acquire specific strategies for fostering and monitoring comprehension (Roshenshine & Meister, 1994). Palincsar and Brown (1988) are credited with prompting this shift with their development of reciprocal teaching, which involves teaching students to use explicit questioning, predicting, clarifying, and summarizing strategies to foster reading comprehension and giving students opportunities to learn the strategies in cooperative reading groups.

Previous meta-analyses suggested that the effects of strategy instruction on reading comprehension are large. For example, mean effect sizes are 0.88 or higher for training in self-questioning and other strategies (Mastropieri et al., 1996; Swanson, 1999) and reciprocal teaching (Rosenshine & Meister, 1994). Although previous reviews found large and positive effects of strategy instruction on reading comprehension, their searches were limited. They included studies not designed to assess causal inferences, and they mixed different types of strategy instruction, outcome measures, and student grade levels in their analyses. Rigorous application of research design criteria is needed to address the effectiveness of strategy instruction for helping struggling high schoolers understand what they read. This report applies What Works Clearinghouse evidence screens to a literature base defined after an extensive search.

Strategy instruction helps students become more active, organized, and thoughtful readers and learners

This review was intended to provide evidence to high school educators to guide decisions about selecting practices and programs for teaching reading. It was also intended to summarize for researchers the extent and strength of research on strategy instruction for struggling high school readers and suggest directions for future research. It found, however, only one study that met evidence standards—on peer-assisted learning.

RESULTS FROM ONE STUDY OF PEER-ASSISTED LEARNING

The keyword and publication list searches located 1,423 nonduplicative study reports. After a quick review of their abstracts, 215 were held for further relevance screening. Of these, 208 were excluded as not relevant, and seven reports were screened against evidence standards. Study reports were excluded as not relevant because they did not involve high school students (many of the studies involved middle school students), strategy instruction, reading comprehension as an outcome measure, or a valid design for making causal inferences (a randomized controlled trial or a quasi-experimental design).

Of the seven studies included as relevant and held for evidence screening, one passed the What Works Clearinghouse screens. Six failed. In one study, designed to examine the effects of strategy instruction in text structure use, the impact of the intervention could not be separated from that of the teacher who delivered it (Gallini, Spires, Terry, & Gleaton, 1993). In another, designed to examine the effects of reciprocal teaching, the impact of the intervention could not be separated from that of the particular school and school context where the intervention was used (Alfassi, 1998). Two studies, one designed to examine the effects of instruction in semantic-feature analysis (Bos, Anders, Filip, & Jaffe, 1989) and another to examine the effects of instruction in text structure and graphic organizers, did not have adequate reporting for computing effects (Boyle & Weishaar, 1997). The fifth study, designed to examine the effects of strategy

instruction in text structure and graphic organizers, did not include evidence demonstrating the outcome measure's reliability (the measure had too few items, and no reliability coefficients were reported or provided by the study author when requested; Boyle, 2000).

The sixth study, designed to examine the effects of using peer tutoring to teach comprehension strategies in world history, did not pass the adequacy of outcome measure evidence screen (Mastropieri, Scruggs, Spencer, & Fontana, 2003).¹

The one study that passed the evidence screens was Fuchs et al. (1999). An explanation of why this study passed each evidence screen is in appendix A.

Study design

To examine the effects of strategy instruction in peer-assisted learning on student reading comprehension, Fuchs et al. (1999) randomly assigned 18 high school remedial reading and special education teachers either to use peer-assisted learning strategies to teach reading or to continue “business as usual.” Reading comprehension was measured in pretests and post-tests. But at the end of the experiment, data from only a subset of students were used in the analysis of the effects of peer-assisted learning on reading. Teachers used a combination of the prior year's standardized test results and professional judgment to identify the subset of students whose data would be included in the analysis, identifying students whose reading levels were between grades 2 and 6 (Lynn S. Fuchs, personal communication, June 1, 2007). This selection of students for inclusion disrupted the randomization process, creating a potential bias. So, the study was assessed as a quasi-experimental design. It passed each of the remaining evidence screens and thus was categorized as meeting evidence standards with reservations.

Study reports were excluded as not relevant because they did not involve high school students, strategy instruction, reading comprehension as an outcome measure, or a valid design for making causal inferences

One study, Fuchs et al. (1999), passed the evidence screens, but the selection of students for inclusion disrupted the randomization process, creating a potential bias, so the study was assessed as a quasi-experimental design

To measure reading comprehension, Fuchs et al. (1999) used a researcher-developed test, the Comprehension Reading Assessment Battery. Students read traditional folktales aloud and then give short answers to 10 comprehension questions that require recall of the thematically important information. The test passages have a readability level of grade 2.5 (Fuchs et al.).²

School and student samples—10 high schools and 102 students

The study took place at 10 high schools in one metropolitan southeastern school district. It involved nine intervention classrooms and nine comparison classrooms. Across these two groups of classes, 102 high school students in remedial or special education classes participated. The average age of the students was just under 16. Students in both groups of classes were equally distributed across four diagnostic categories: remedial readers, learning disabilities, mild mental retardation, and other disabilities. The two groups were equivalent on the reading comprehension pretest.

The intervention: peer-assisted learning strategies

Half the classes were taught reading with peer-assisted learning strategies, a form of classwide tutoring in which pairs of higher and lower performing readers work together on three structured activities: partner reading, paragraph shrinking, and prediction relay. The three activities focus on developing oral reading, questioning, and the ability to make and confirm or disconfirm predictions to help students identify and summarize a passage's main ideas.

In partner reading the higher performing reader reads first, modeling a read-aloud of the text. Then the lower performing reader reads aloud. In paragraph shrinking the students ask and answer two questions to identify main ideas and summarize paragraphs: "(a) who or what is the paragraph

mainly about, and (b) what is the most important thing about the who or what?" (Fuchs et al., 1999, p. 312). In prediction relay the student pair works through four steps: predicting what will be learned in the next half page, reading the half page aloud, confirming or disconfirming the prediction, and summarizing the main idea of the half page.

Throughout the three activities, the members of each pair monitor each other's accuracy and offer corrections as needed. Pairs earn points for such accomplishments as accurate oral reading, paragraph shrinking, and confirming or disconfirming predictions. The points earned are turned in each month, allowing students to participate in a lottery. Earning more points increases a student's probability of winning.

High school teachers in special education and remedial reading implemented peer-assisted learning in their classrooms two to three times a week over 16 weeks, for a total of 23.3 hours. Teachers in control classrooms implemented the same amount of reading instruction, using their conventional programs and practices but without peer-assisted learning.

Effects on reading comprehension

Computed for this report was an effect size for peer-assisted learning of 0.34, which was not statistically significant (see box 1 for definitions).³

What Works Clearinghouse decision rules were used to describe the importance of the study effects. These decision rules are based on the level of statistical significance, the magnitude of the effect (equal to or greater than the minimum effect size of +0.25), and the direction of the effect (positive or negative). The minimum effect size of +0.25 represents the "smallest positive value at or above which the effect is deemed substantively important with relatively high confidence for the outcome" (What Works Clearinghouse, 2006b, p. 3). These rules characterize the study effect as statistically significant and positive, substantively important and positive, indeterminate, substantively important and negative, or statistically significant and

negative (What Works Clearinghouse, 2006b). Because the effect of peer-assisted learning was greater than 0.25, but was not statistically significant, the effect in Fuchs et al. (1999) was classified as substantively important and positive.

An improvement index was calculated to understand the practical implications of this effect size. The improvement index ranges from -50 to +50 percentile points, with positive values favoring the intervention group. It represents the difference in percentile rank between the average students in the intervention and comparison groups. The improvement index for an effect size of 0.34 is 13 percentile points, meaning that the average student in the control group would have gained at least 13 percentile points in achievement if exposed to the intervention (see table A1).

BETTER RESEARCH ON STRATEGY INSTRUCTION

The evidence indicates that peer-assisted learning can have a substantively important positive effect on struggling high school readers' reading comprehension. But reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention.

In addition, the extent of evidence was very small: 102 students, primarily students with disabilities in 10 high schools in only the southeastern United States. Comprehension was measured using only

researcher-developed assessments and with only one type of text: folktales with grade 2.5 readability. With only these assessments, instructional settings, and students in the evidence base, the finding of positive effects for peer-assisted learning has limited applicability. Further research is needed to strengthen the evidence base. A first step would be to randomly assign students to the intervention.

Reservations remain about attributing improved comprehension to peer-assisted learning because the students were not randomly assigned to the intervention, and the extent of evidence was very small

Better and broader outcome measures are also needed to assess the effect of peer-assisted learning. Different outcome measures in future studies could include passages with grade 7 or higher readability. A state high school reading assessment could also be included to evaluate whether benefits might accrue for high-stakes indicators of achievement. To accommodate struggling readers' difficulties and reduce barriers to reading, studying the effects of peer-assisted learning on listening comprehension would also be valuable.

The findings in this report are limited by the scope of the searches and the research available. Findings and conclusions about the effectiveness of strategy instruction for struggling high school readers may change as additional research is identified or completed.

APPENDIX A METHODOLOGY

The four stages of What Works Clearinghouse's study review process were used to locate and screen relevant studies. These stages include defining relevant studies for review, searching for and screening research reports for relevance, systematically applying evidence standards and categorizing studies accordingly, and, for studies that met evidence standards either with or without reservations, systematically describing participant and other study characteristics, intervention characteristics, outcomes, and findings (What Works Clearinghouse, 2006e).

Defining relevant studies for review

Topical and research design criteria were established for locating relevant studies. Studies had to meet both criteria to be relevant to this review.

Topical criteria were used to locate studies on strategy instruction. Studies were included if their purpose was to examine the effectiveness of strategy instruction for improving the reading comprehension of struggling high school readers, including students with learning disabilities. Studies had to involve students in grades 9, 10, 11, or 12 and list reading comprehension as an outcome measure. Only studies published over 1975–2007, after the passage of the original Individuals with Disabilities Education Act (Education for All Handicapped Children Act, PL 94–142), were included.

Studies were excluded if they met any of the following topical criteria:

- Not published between 1975 and 2007.
- Did not include students with learning disabilities, remedial readers, or “struggling readers” (students reading two or more years below grade level or one or more standard deviations below the mean).
- Did not include high school students in grades 9, 10, 11, or 12 and of ages 14 to 21 years—or

results were not disaggregated for these students if the sample included a broader range of grade levels or ages.

- Reading comprehension was not measured as an outcome.
- Intervention did not include strategy instruction for reading comprehension.
- A more recent version of the same work was available and obtained.

Research design criteria were used to locate studies that examined the effectiveness of strategy instruction. To draw causal conclusions about the effects of strategy instruction, we searched for and included studies using either randomized controlled trials or quasi-experimental designs.⁴ In randomized controlled trials, participants are randomly assigned to different experimental groups. By placing participants in groups randomly (for example, assuring that each student has an equal chance of being placed in one group or another), the assignment reduces the risk of bias and establishes that the only systematic difference between the groups is the presence or absence of the intervention. If the experiment has been well implemented, we can confidently infer that the measured effects are due to the intervention rather than unintended group differences (such as willingness to volunteer).

Quasi-experimental designs do not randomly assign participants to intervention and comparison groups, but the groups are matched or shown to be equivalent before the intervention. Establishing the pre-intervention equivalence in a quasi-experimental design gives some confidence, though not as much as with a randomized controlled trial, that the measured effects are due to the intervention (Flay et al., 2005; Shadish, Cook, & Campbell, 2002; U.S. Department of Education, 2003).

Studies were excluded from consideration if the research design was neither a randomized controlled trial nor a quasi-experimental design.

Search strategy for relevant studies

Searches for relevant randomized controlled trials and quasi-experimental design studies were conducted using two types of sources: bibliographic databases and lists of publications, either in prior reviews or available on relevant web sites. Completed between March 2006 and February 2007 were 26 searches, 12 of bibliographic databases and 14 of lists of publications. Although we searched some of the gray literature (Dissertation Abstracts International), we did not search reports recommended by experts or in conference proceedings.

Bibliographic database searches. Searches of ERIC, PsychINFO, C2 Spectr, EBSCOhost professional development collection, and Dissertation Abstracts International were conducted to identify studies published since 1975, the year after the passage of the original Individuals with Disabilities Education Act. After the passage of the act, we expected research and publications of research on the effectiveness of practices to include individualized education programs for students with learning and reading disabilities. We used different combinations of key words (for example, “reading,” “secondary students,” “high school,” “comprehension,” “learning disabilities,” “disab*,” “remedial reading,” and “struggling readers”) and the years 1975 to 2007. We narrowed the results of searches by subject area: “reading strategies” or “learning disabilities” and “reading comprehension.” Additional searches were conducted on specific, individual strategies (such as mental imagery, summarization, and reciprocal teaching). The key terms used for each of the 12 searches and the number of reports identified for each search are in appendix C, table C1.

Lists of publications. Potentially relevant studies were identified in the reference lists of three previous reviews, Kim, Vaughn, Wanzek, and Wei (2004), Mastropieri, Scruggs, Bakken, and Whedon (1996), and Swanson (1999). As recommended by the What Works Clearinghouse search protocol, we also searched recent tables of contents of

key journals. We searched the following tables of contents from 2000 to 2007 for potentially relevant studies: *Journal of Learning Disabilities*, *Learning Disabilities Research and Practice*, *Exceptional Children*, *The Journal of Special Education*, *Annals of Dyslexia*, *Reading Research Quarterly*, *Remedial and Special Education*, *Reading and Writing Quarterly*, and *Journal of Educational Psychology*. Last, to locate other research on the effectiveness of strategy instruction for improving adolescents’ reading comprehension, we reviewed studies listed on the University of Kansas Center for Research on Learning for possible inclusion. The number of reports identified in each of the 14 publication searches is presented in Appendix C, table C2. Results of each search were stored in Endnote and duplicates were identified and eliminated.

Evidence screen and adequacy of data review

We wanted this report to use a standard, replicable set of criteria and procedures for evaluating the technical and methodological quality of relevant studies. We adopted for this report the What Works Clearinghouse standards of evidence and rules for categorizing randomized controlled trials and quasi-experimental design studies into one of three levels of technical quality (What Works Clearinghouse, 2006a). These categories are listed below, representing the highest technical quality to the lowest:

- Meets evidence standards.
- Meets evidence standards with reservations.
- Does not meet evidence screens.

A protocol was developed based on the What Works Clearinghouse standards of evidence and rules for categorizing studies. The protocol involved eight decision trees (sequences of yes/no questions), one for each of eight evidence screens: adequacy of the outcome measures, baseline equivalence, attrition, differential attrition, potential intervention contamination, potential teacher–intervention confound, potential

school–intervention confound, and adequacy of reporting for computing effect sizes. Whenever there was not enough information in the study report to accurately assess against an evidence screen, authors were contacted for the necessary information. The studies were assessed according to the authors' responses.⁵

Decision rules for each of the eight evidence screens were stated explicitly and applied as follows:

1. *Adequacy of outcome measures.* All standardized measures of reading comprehension (for example, state level or nationally normed tests) were judged as having adequate reliability and validity, and so checking for reliability or face validity was unnecessary when a study's outcome measure was standardized. Curriculum-based and teacher- or researcher-developed measures of reading comprehension were assessed against technical adequacy criteria. A measure was determined valid when the original study authors presented significant correlations on tests of concurrent validity, predictive validity, factor analysis, and so on—or the measure, including items, was documented and described in a way that provided evidence of face validity.

Review team members judged face validity by determining whether the measure, including items, adequately resembled an acceptable measure of student ability to read a text and perform one or more of the four aspects of reading identified in the 2005 National Assessment of Educational Progress Reading Framework: form a general understanding, develop interpretation, make reader/text connections, and examine content and structure (National Assessment Governing Board, 2005).

In addition, assessments that did not require students to read new passages, instead requiring students to answer questions about passages read and studied during the instruction

or training, were judged to be assessments of memory and not to be valid assessments of reading comprehension. Adequate reliability evidence constituted either reports of internal consistency (coefficient of 0.60 or higher) or test-retest stability (coefficient of 0.40 or higher). If the measure required judgment, evidence of inter-rater reliability (coefficient of 0.50 or higher) or at least an average of 90 percent agreement among raters for the measures was required. Studies with no technically acceptable outcome measures were categorized as *does not meet evidence screens*, regardless of whether they were randomized controlled trials or quasi-experimental design studies.

2. *Baseline equivalence.* Although random assignment helps establish that the only systematic group difference is the presence or absence of the intervention, the goal was to ensure that this was in fact so for the randomized controlled trials. There are possible sources of contamination or disruption to the randomization process. If the study authors provided and acknowledged baseline data, we checked baseline equivalence.

For both randomized controlled trials and quasi-experimental designs, we checked baseline equivalence between groups on measures of reading comprehension pretests and on the age/grade level and diagnostic labels of the students. For reading comprehension, we used the following pretest condition as the criterion to confirm baseline equivalence of groups: a statistically nonsignificant difference between groups, defined as less than 0.5 standard deviations on a measure of reading comprehension. We used two pretest conditions as criteria to disconfirm baseline equivalence: patently nonequivalent groups, defined as a statistically significant difference equal to or greater than 0.5 standard deviations, and lesser nonequivalence, defined as a statistically significant difference between groups of less than 0.5 standard deviations.

If baseline equivalence was patently non-equivalent, neither a randomized controlled trial nor a quasi-experimental design met this evidence screen and the study was categorized as *does not meet evidence screens*. If there was lesser nonequivalence, it must have been statistically corrected in both randomized controlled trials and quasi-experimental designs; and if it was corrected statistically (for example, with ANCOVA or MANCOVA), the study was categorized at best as *meets evidence standards with reservations*. If there was lesser nonequivalence and it was not corrected statistically, it was categorized as *does not meet evidence screens*.

For randomized controlled trials where no baseline data were reported, we assessed the randomization process for potential disruption or contamination. If there was no indication of potential disruption or contamination, we assumed baseline equivalence and did not downgrade the study. If there was an indication of potential disruption or contamination to the randomization process (for example, teachers selected some but not all students to participate after classrooms were randomly assigned), the study was categorized as a quasi-experimental design and screened for baseline equivalence and against the remaining five evidence screens as a quasi-experimental design.

For demographic variables (age/grade and diagnostic label), we assessed whether groups were comparable using the following criteria: the groups must have been drawn from the same school, and the groups must have had the same distribution of students across age/grade levels and diagnostic categories.

3. **Attrition.** We examined overall sample attrition to protect against sample bias. For a randomized controlled trial to be categorized as *meets evidence standards* and for a quasi-experimental design to be categorized as *meets evidence standards with reservations*, attrition had to be less than 20 percent overall in both the intervention and comparison

groups (not severe). When a final sample lost more than 20 percent (severe attrition), we did not compute or use effect sizes for any outcomes derived from that sample. If there was some attrition that was not severe (for example, 15 percent), the authors had to demonstrate post-attrition baseline equivalence between the groups for a study not to be downgraded and for the outcomes derived from that sample to be included in computing effect sizes. For both randomized controlled trials and quasi-experimental designs, if attrition was not reported, we adopted the What Works Clearinghouse default of assuming that there was no attrition problem.

4. **Differential attrition.** We also assessed differential attrition to protect against sample bias. Attrition was said to be differential between the intervention and comparison groups if it was greater than 7 percent (Sanchez, Herman, Song, & Maynard, 2006). If less than 7 percent differential attrition was either reported in manuscript, indicated by results (degrees of freedom), or indicated through contact with the first author, the study was not downgraded. For randomized controlled trials where there was attrition greater than 7 percent, the study was categorized at best as *meets evidence standards with reservations*. For example, when the differential attrition was greater than 7 percent and post-attrition equivalence on the pretest was confirmed, the randomized controlled trial was categorized as *meets evidence standards with reservations*.

For quasi-experimental designs where there was greater than 7 percent attrition, post-attrition equivalence on the pretest must have been confirmed for the study to be included and categorized as *meets evidence standards with reservations* (if post-attrition equivalence on the pretest was not confirmed, a study using a quasi-experimental design was categorized as *does not meet evidence standards*). In studies reporting multiple outcomes, differential attrition was assessed for each outcome

and corresponding particular analytic sample. If a particular analytic sample had more than 7 percent differential attrition and post-attrition equivalence on the pretest was not confirmed, the outcomes derived from the sample were not included in computing effect sizes. If a particular analytic sample had differential attrition greater than or equal to 7 percent and post-attrition equivalence on the pretest was confirmed, the outcomes derived from this sample were included in computing effect sizes.

5. *Intervention contamination.* If there was evidence of any contaminant (for example, changed expectancy, local history, novelty, or disruption) after the beginning of the intervention that differentially affected outcomes, randomized controlled trials were categorized as *meets evidence standards with reservations* and quasi-experimental designs were categorized as *does not meet evidence screens*. For example, if intervention teachers were replaced with more experienced teachers partway through the study, the disruption would contaminate the intervention.
6. *Teacher–intervention confound.* The review team specified criteria for classifying studies based on the What Works Clearinghouse teacher–intervention confound decision rules (What Works Clearinghouse, 2006d). The first set of decision rules applied only to randomized controlled trials. When a randomized controlled trial involved one teacher per condition (teacher A teaching the intervention, $n = 1$, and teacher B teaching the control condition, $n = 1$) and students were randomly assigned to the teachers or conditions, the effects of the intervention cannot be separated from the particular teacher teaching it. A randomized controlled trial with $n = 1$ was categorized as *does not meet evidence screens* unless the teacher had a limited role in the intervention implementation (for example, an intervention that was “predominantly computer-based;” What Works Clearinghouse,

2006d, p. 2). In that case, the study was downgraded to *meets evidence standards with reservations*.

When a quasi-experimental design involved one teacher per intervention or comparison condition, the problem is similar: the effects of the intervention cannot be separated from the particular teacher teaching it or from the particular composition of the class. Therefore, quasi-experimental designs with one teacher per condition were categorized as *does not meet evidence screens*.

A second set of What Works Clearinghouse (2006d) teacher–intervention confound decision rules applied to randomized controlled trials or quasi-experimental designs. This set of decision rules applied when the study involved one teacher implementing both the intervention and comparison conditions ($N = 1$). A randomized controlled trial with an $N = 1$ was downgraded to *meets evidence standards with reservations*, and a quasi-experimental design was categorized as *does not meet evidence screens*, unless the study authors demonstrated that the teacher had equal motivation and ability to implement each condition (for example, authors reported equal training in each condition or demonstrated that teacher effects were negligible). If teacher effects were not negligible in a randomized controlled trial (for example, the teacher had prior experience using one intervention and not the other), the randomized controlled trial was categorized as *does not meet evidence screens*.

7. *School–intervention confound.* For studies involving one school (school A) as the site for the intervention and another school (school B) as the site for the control/comparison condition (to which students were not randomly assigned), the effects of the intervention cannot be separated from those of the students’ school environment. According to What Works Clearinghouse guidance, if the two schools differed appreciably in their location and student characteristics, it is unlikely that matching or statistical equating could assure that final groups

of children would yield a statistically unbiased estimate of the intervention effect. Therefore, in this case, a randomized controlled trial was downgraded to *meets evidence standards with reservations*, and a quasi-experimental design was categorized as *does not meet evidence screens*. But if the authors provided compelling evidence such as comprehensive demographic information (in the original article or through personal communication) that the schools did not differ appreciably and that the equating process was adequate, the study was not downgraded. This rule also applied to teachers and districts at the cluster level with one cluster per condition, when students were not randomly assigned.

8. *Adequacy of reporting for computing effect sizes.* To provide a standardized index of intervention effects, we required adequate reporting for computing effect sizes, including group mean post-test scores, standard deviations, and sample sizes. In some cases, other essential parameters were required for computing effect sizes, such as unadjusted group mean pretest and post-test scores and standard deviations. When the essential parameters were missing from a study report, we contacted authors to request them.

Coding and analysis of key study characteristics

We coded key information in Fuchs, Fuchs, and Kazdan (1999). We identified the locale of the participating schools and district. We coded students' grade level, diagnostic category, and average pre-test intelligence quotient and reading achievement scores. We briefly described the intervention and

comparison condition, intervention duration, and setting characteristics.

To provide a standardized index of the intervention's effectiveness, we computed an effect size, which expresses in standard deviation units the increase or decrease in achievement of the intervention group compared with that of the control or comparison group. Given the design and data reported by Fuchs et al. (1999), we used a difference-in-differences approach to computing the effect size, which is appropriate when the pre-test and post-test are the same test (What Works Clearinghouse, 2006e). A difference-in-differences approach compares the post-test–pretest difference of the intervention group with the post-test–pretest difference of the control group (What Works Clearinghouse, 2006e). In addition, because there was a mismatch between the unit of assignment (teachers) and analysis (students), we corrected for clustering when assessing the statistical significance of the difference. We did not need to correct for multiple comparisons or small sample sizes.

To indicate the practical importance of the effect size, we computed an improvement index for the reading comprehension outcome (table A1). This involved determining Cohen's U_3 , the proportion of the area under the normal curve for the z-score value corresponding to the effect size (Glass & Hopkins, 1996; p. 615). The improvement index was computed by subtracting 50 percent from Cohen's U_3 (What Works Clearinghouse, 2006e). It ranges from –50 to +50 percentile points, with positive values favoring the intervention group, and represents the difference in percentile rank between the average students in the intervention and comparison groups.

TABLE A1

Summary of Fuchs, Fuchs, & Kazdan (1999)

Study design	Quasi-experimental design				
Technical quality	Meets evidence standards with reservations				
Study sample	102 high school students (74 with learning disabilities, 22 remedial readers, 4 with mild mental retardation, and 2 with other disabilities) clustered in 18 remedial reading/special education classes				
Intervention	Peer-assisted learning strategies: partner reading, paragraph shrinking (summarization), and prediction relay (formulation/disconfirmation of predictions) embedded in peer tutoring				
Outcome measure	Comprehension Reading Assessment Battery (CRAB)				
Mean outcome (standard deviation)			Effect calculations		
Intervention group (n = 52, clustered in 9 classes)	Comparison group (n = 50, clustered in 9 classes)	Mean difference	Effect size	Statistical significance ^a	Improvement index
Pretest: 5.88 (2.56) Post-test: 7.22 (2.23)	Pretest: 6.10 (2.49) Post-test: 6.64 (2.44)	Pretest: -0.22 Post-test: 0.58	0.34 ^b	Not significant (after correcting for clustering ^c)	+13

Note: Pretest is the pre-intervention test; post-test is the immediate post-intervention test.

a. Statistical significance was assessed at the 0.05 alpha level.

b. The effect size was computed using a difference-in-differences approach using formula 7 (What Works Clearinghouse, 2006e, p. 4). When incorporating the pretest–post-test relationship (that is, the test–retest reliability correlation of 0.92) using formula 8 (What Works Clearinghouse, 2006e, p. 5) to gain a more precise estimate of the intervention’s effect, the effect was 0.34. The following difference-in-differences formula was used:
Hedges $g = (X1 - X1\text{-pre}) - r(X2 - X2\text{-pre}) / \sqrt{[S21(n1 - 1) + S22(n2 - 1)] / [n1 + n2 - 2]}$.

c. Because students were nested in classrooms and the analysis was conducted at the student level without appropriately addressing the nested structure of the data (say, with hierarchical linear modeling), adjustments were made for the nested structure of the data. A t-statistic was applied for this effect size that ignored clustering to check the statistical significance of the difference (What Works Clearinghouse, 2006e).

APPENDIX B

STUDENTS WITH DISABILITIES IN THE CENTRAL REGION BY STATE AND DISABILITY CATEGORY

TABLE B1

Students ages 12–17 with disabilities in the Central Region by state and disability category

	Colorado	Kansas	Missouri	Nebraska	North Dakota	South Dakota	Wyoming	Total/ average
Total number of students with disabilities	34,512	26,399	63,312	18,171	6,026	6,207	5,346	159,973
<i>Students with disabilities by category (percent of all students with disabilities)</i>								
Specific learning disabilities	52.2	55.3	63.0	51.2	53.9	63	59.0	56.8
Speech/language impairments	8.9	3.7	6.2	10.3	11.7	2.5	7.4	7.2
Mental retardation	5.0	10.4	10.2	16.4	10.1	10.4	6.0	9.8
Emotional disturbance	16.8	9.3	8.3	7.8	12.5	7.8	12.2	10.7
Multiple disabilities	4.0	4.7	0.7	0.9	0.0	4.5	0.5	2.2
Hearing impairments	1.8	1.0	0.9	1.6	1.1	0.9	1.3	1.2
Orthopedic impairments	9.8	0.5	0.4	0.9	0.8	0.6	0.9	2.0
Other health impairments	0.0	13.5	8.4	9.2	8.2	8.0	10.8	8.3
Visual impairments	0.4	0.3	0.4	0.5	0.4	0.3	0.6	0.4
Autism	0.7	1.0	1.2	0.9	1.0	1.7	0.7	1.0
Deaf-blindness	0.1	0.0	0.0	0.0	0.00	0.02	0.00	0.0
Traumatic brain injury	0.5	0.4	0.3	0.5	0.3	0.4	0.84	0.5

Source: U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, 2006.

APPENDIX C

SEARCH TERMS AND RESULTS FOR LITERATURE SEARCHES

TABLE C1

Bibliographic database search terms and results

Search engine	Database	Search terms	Report yield
First Search	ERIC	"secondary school+" AND achieve* AND disab* in key term field	154
First Search	ERIC	"high school+" AND achieve* AND disab* in key term field	(559)
		Limited by subject heading:	(414)
		<ul style="list-style-type: none"> • academic achievement • disabilities • high schools • high school students 	
		Search in Endnote for "read" anywhere in any field	98
First Search	ERIC	"high school+" AND achieve* AND "remedial read*" in key term field	107
First Search	ERIC	"secondary school+" AND achieve* AND "remedial read*" in key term field	31
American Psychological Association membership web site fielded search	PsychInfo	"secondary school+" AND achieve* AND disab* in key term field	177
American Psychological Association membership web site fielded search	PsychInfo	"high school+" AND achieve* AND disab* in key term field	(412)
		Search in Endnote for "read" anywhere in any field	138
American Psychological Association membership web site fielded search	PsychInfo	"secondary school+" AND achieve* AND remedial read* in key term field	12
American Psychological Association membership web site fielded search	PsychInfo	"high school+" AND achieve* AND remedial read* in key term field	143
Reference Web Poster	C2 Spectr	Secondary school* in Keywords (15) AND achievement in all indexed fields	1
Advanced Search for References (http://geb9101.gse.upenn.edu/RIS/RISWEB.ISA)	All databases	Reading comprehension (43) AND secondary education in all indexed fields	1
		High school* in keywords (62) AND achieve* (2) in all indexed fields	2
		Reading comprehension (43) AND high school* in all indexed fields	1
EBSCOhost	Professional development collection	"secondary school+" AND achieve* AND disab*	10
		"high school+" AND achieve* AND disab* in key term field	51
		"secondary school+" AND achieve* AND remedial read* in key term field	0
		"high school+" AND achieve* AND remedial read* in key term field	1

TABLE C1 (CONTINUED)

Bibliographic database search terms and results

Search engine	Database	Search terms	Report yield
UMI ProQuest	Dissertation abstracts	"secondary school" AND achieve* AND disab* in citation and abstract fields	42
		"high school" AND achieve* AND disab* in citation and abstract field	230
		"Secondary school" AND achieve* AND "remedial reading"	3
		"high school" AND achieve* AND "remedial reading"	23
First Search	ERIC	"graphic organizer," "mnemonics," "mental imagery," "reciprocal teaching," "summarization," "questioning," "monitoring," "story structure," "text elements," "text structure," "prior knowledge," AND "high school," AND "reading comprehension"	114
Database subtotal			1,339

Note: Numbers in parentheses are initial search yields prior to the use of search terms to identify relevant reports held in Endnote libraries.

TABLE C2

Publication list search results

List type	Source	Number
Reference lists	Gersten, Fuchs, Williams, & Baker (2001)	9
	Kim, Vaughn, Wanzek, & Wei (2004)	8
	Mastropieri, Scruggs, Bakken, & Whedon (1996)	9
	Swanson (1999)	8
	Reference list subtotal:	34
Tables of contents (2000–07)	Journal of Learning Disabilities	3
	Learning Disabilities Research and Practice	2
	Exceptional Children	0
	The Journal of Special Education	0
	Annals of Dyslexia	0
	Reading Research Quarterly	1
	Remedial and Special Education	9
	Journal of Educational Psychology	12
	Reading and Writing Quarterly	6
Tables of contents subtotal:	33	
Web sites	University of Kansas Center for Research on Learning (retrieved February 2, 2007, from http://www.ku-crl.org/library/pub_author.shtml)	17
Publication lists subtotal		84
Total for all search methods		1,423

NOTES

This review would not have been possible without the expertise and assistance of Ms. Terry Young, Mid-continent Research for Education and Learning (McREL) Resource Center Program Coordinator, who located and obtained the study reports screened and included in this review. The authors also gratefully acknowledge the contributions of Ms. Mya Martin-Glenn, McREL Researcher, and Ms. Carolyn Woempner, McREL Senior Consultant, who assisted with instrumentation, conducted relevance and evidence screens, coded studies, and prepared results tables and draft summaries of the findings.

The authors gratefully acknowledge the feedback and recommendations of anonymous Institute of Education Sciences reviewers—and consultation with Dr. Yael Kidron, Senior Research Analyst, American Institutes of Research (AIR), and Dr. Rebecca Herman, also of AIR, for their guidance on applying What Works Clearinghouse decision rules and procedures. The authors are grateful for statistical consultations with Dr. Bruce Randel and the helpful review by Dr. Andrea Beesley, McREL Senior Researcher, and Dr. Bob St. Pierre, Director, STP Associates. They also thank Kirsten Miller, McREL Senior Consultant of Publications, and Dr. Zoe Barley, McREL Senior Research Fellow, for editorial assistance.

1. Two reading comprehension measures were assessed in Mastropieri et al. (2003). In one measure comprehension was confounded with strategy explanation. In the other students were not required to read and understand

passages that were not already studied during instruction and training.

2. Test-retest reliability (0.92) and validity evidence (correlation with Stanford Achievement Test in reading comprehension was 0.82) for the number of comprehension questions answered correctly on the Comprehension Reading Assessment Battery supported the technical adequacy of this measure.
3. To compute an effect size, the post-test–pretest difference of the peer-assisted learning group and that of the control group were compared to chance differences, referred to as a “difference-in-differences” approach (What Works Clearinghouse, 2006e). The approach is appropriate when the pretest and post-test are the same.
4. We did not include regression discontinuity and single subject, multiple baseline designs because a replicable set of criteria and procedures for rating the technical adequacy of these designs was not available from the What Works Clearinghouse (2006a) at the time of the initiation of this study.
5. Some single studies were handled as two studies; for example, we treated a multiple-arm study as two separate studies. So if the study authors randomly assigned students to one of three conditions, including one control and two different strategy instruction interventions, we treated each comparison between the control and one of the interventions as a single study when assessing against the evidence screens.

REFERENCES

- Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. 3, pp. 285–310). Mahwah, NJ: Erlbaum.
- Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, 35(2), 309–332.
- Anders, P. L. (2002). Toward an understanding of the development of reading comprehension instruction across the grade levels. In C. M. Roller (Ed.), *Comprehensive reading instruction across the grade levels* (pp. 111–133). Newark, DE: International Reading Association.
- Balfanz, R., McPartland, J. M., & Shaw, A. (2002). *Re-conceptualizing extra help for high school students in a high standards era*. Baltimore, MD: Center for Social Organization of Schools, Johns Hopkins University.
- Bos, C. S., Anders, P. L., Filip, D., & Jaffe, L. E. (1989). The effects of an interactive instructional strategy for enhancing reading comprehension and content area learning for students with learning disabilities. *Journal of Learning Disabilities*, 22(6), 384–390.
- Boyle, J. R. (2000). The effects of a Venn diagram strategy on the literal, inferential, and relational comprehension of students with mild disabilities. *Learning Disabilities—Multidisciplinary*, 10(1), 5–14.
- Boyle, J. R., & Weishaar, M. (1997). The effects of expert-generated versus student-generated cognitive organizers on the reading comprehension of students with learning disabilities. *Learning Disabilities Research & Practice*, 12(4), 228–235.
- Cook, L., & Friend, M. (1995). Co-teaching: Guidelines for creating effective practices. *Focus on Exceptional Children*, 28(3), 1–16.
- Fisher, D., & Ivey, G. (2006, November). Evaluating the interventions for struggling adolescent readers. *Journal of Adolescent and Adult Literacy*, 50(3), 180–189.
- Flay, B. R., Biglan, A., Boruch, R. F., Gonzalez Castro, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175.
- Fuchs, L., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education*, 20(5), 309–318.
- Gallini, J. K., Spires, H. A., Terry, S., & Gleaton, J. (1993). The influence of macro and micro-level cognitive strategies training on text learning. *Journal of Research and Development in Education*, 26(3), 164–178.
- 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.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology*. Needham Heights, MA: Allyn & Bacon.
- Kendall, J. S., Snyder, C., Schintgen, M., Wahlquist, A., & Marzano, R. J. (1999). *A distillation of subject-matter content for the subject areas of language arts, mathematics, and science*. Aurora, CO: Mid-continent Research for Education and Learning.
- Kim, A., Vaughn, S., Wanzek, J., & Wei, S. (2004). Graphic organizers and their effects on the reading comprehension of students with LD: A synthesis of research. *Journal of Learning Disabilities*, 37(2), 105–118.
- Langer, J. A. (1999). *Beating the odds: Teaching middle and high school students to read and write well*. Retrieved February 16, 2001, from <http://cela.albany.edu/eie2/main.html>
- Mastropieri, M. A., Scruggs, T. E., Blakken, J. P., & Whedon, C. (1996). Reading comprehension: A synthesis of

- research in learning disabilities. *Advances in Learning and Behavioral Disabilities*, 10B, 201–227.
- Mastropieri, M. A., Scruggs, T. E., Spencer, V., & Fontana, J. (2003). Promoting success in high school world history: Peer tutoring versus guided notes. *Learning Disabilities Research and Practice*, 18(1), 52–65.
- Mid-continent Research for Education and Learning. (2006, January). *Technical proposal for tasks 1-5: Best and final offer* (Central Region Educational Laboratory Solicitation #ED-05-R-0006). Aurora, CO: Author.
- Mid-continent Research for Education and Learning. (2007). *Content knowledge*. Retrieved May 31, 2007, from <http://www.mcrel.org/compendium/reference.asp?item=benchmark&BenchmarkID=1249&subjectID=7>
- National Assessment Governing Board. (2005). *Reading framework for the 2005 National Assessment of Educational Progress*. Retrieved May 1, 2007, from http://www.nagb.org/pubs/r_framework_05/ch1.html
- National Association of Secondary School Principals. (2005). *Creating a culture of literacy: A guide for middle and high school principals*. Reston, VA: NASSP.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups*. Jessup, MD: National Institute for Literacy.
- Neild, R. C., & Balfanz, R. (2006). An extreme degree of difficulty: The educational demographics of urban neighborhood high schools. *Journal of Education for Students Placed at Risk*, 11(2), 123–141.
- Palincsar, A. S., & Brown, A. L. (1988, January/February). Teaching and practicing thinking skills to promote comprehension in the context of group problem solving. *Remedial and Special Education*, 9(1), 53–59.
- RAND Reading Study Group. (2002). *Reading for understanding: toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Rose, A. (2000, November/December). Literacy strategies at the secondary level. *Leadership*, 30(2), 12–16.
- Rosenshine, B., & Meister, C. (1994, winter). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530.
- Sanchez, S., Herman, R., Song, W., & Maynard, R. (2006). What Works Clearinghouse. Presentation to Regional Educational Laboratory Directors' Meeting, Washington, DC.
- Scruggs, T. E., & Mastropieri, M. A. (2004). Recent research in secondary content areas for students with learning and behavioral disabilities. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Research in secondary schools: Advances in learning and behavioral disabilities* (pp. 243–263). Oxford: Elsevier Ltd.
- Shadish, W. R., Thomas D. Cook, & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Swanson, H. L. (1999, November/December). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities*, 32(6), 504–532.
- Swanson, H. L. (2001). Research on interventions for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities*, 32(6), 504–532.
- Topping, K. (2001). *Peer assisted learning*. Newton, MA: Brookline Books.
- Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In C. C. Block and M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176–200). New York: The Guilford Press.
- U.S. Department of Education (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Washington, DC.

- U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2006). *26th Annual (2004) Report to Congress on the Implementation of the Individuals with Disabilities Education Act* (Vol. 2). Washington, DC: Author.
- Wagner, M., Newman, L., Cameto, R., & and Levine, P. (2006). *The academic achievement and functional performance of youth with disabilities: A report of findings from the National Longitudinal Transition Study-2 (NLTS2)*. Menlo Park, CA: SRI International. Retrieved from www.nlts2.org/reports/2006_07/nlts2_report_2006_07_complete.pdf
- Weedman, D. L., & Weedman, M. C. (2001). When questions are the answer. *Principal Leadership*, 2(2), 42–46.
- What Works Clearinghouse. (2006a, September). *Evidence standards for reviewing studies*. Retrieved from <http://whatworks.ed.gov/reviewprocess/standards.html>
- What Works Clearinghouse. (2006b). *Intervention rating scheme*. Retrieved May 7, 2006, from http://whatworks.ed.gov/reviewprocess/rating_scheme.pdf
- What Works Clearinghouse. (2006c). *Improvement index*. Retrieved May 7, 2006, from <http://whatworks.ed.gov/reviewprocess/workpapers.html>
- What Works Clearinghouse. (2006d, September 13). *Teacher-intervention confound*. 2006. Retrieved from <http://whatworks.ed.gov/reviewprocess/teacher-confound.pdf>
- What Works Clearinghouse. (2006e, September). *Technical details of WWC-conducted computations*. Retrieved January 2, 2007, from http://www.whatworks.ed.gov/reviewprocess/conducted_computations.pdf