

Appendix

Appendix A1.1 Study characteristics: Brady, 1990

Characteristic	Description
Study citation	Brady, P. L. (1990). Improving the reading comprehension of middle school students through <i>reciprocal teaching</i> and semantic mapping strategies (Doctoral dissertation, University of Oregon, 1990). <i>Dissertation Abstracts International</i> , 52(03A), 230–860.
Participants	Eighteen Native Alaskan students in grades 5–8 were ranked by their baseline reading scores and then placed into groups of three starting with the three lowest-scoring students and ending with the three highest-scoring students. Students within each group of three were then randomly assigned to one of three study groups: (1) the <i>reciprocal teaching</i> group, (2) a group that received combined instruction in <i>reciprocal teaching</i> with instruction in semantic mapping strategies (SMART), and (3) a business-as-usual control group. ¹ This review focused on comparisons of the six students who received <i>reciprocal teaching</i> and the six students in the comparison group, and examined outcomes after the programs had been implemented for 25 days. Additional findings reflecting item-level scores on daily comprehension tests and students' outcomes 4.5 months after the start of the intervention can be found in Appendix A4.1 and Appendix A4.2.
Setting	The study was conducted in two multi-grade classrooms (one classroom contained seventh- and eighth-grade students, while the other contained fifth- and sixth-grade students) in a school in rural Alaska.
Intervention	The intervention group learned the four <i>reciprocal teaching</i> comprehension strategies of questioning, summarizing, clarifying, and predicting, as developed by Palincsar and Brown (1984). ² The study reported students' outcomes after 25 days of program implementation.
Comparison	The control group was not taught using <i>reciprocal teaching</i> or semantic mapping strategies. Students in the control group attended their regular basal reading classes during the study.
Primary outcomes and measurement	For both the pretest and the posttest, students took the Gates-MacGinitie Reading Test reading comprehension and vocabulary subtests. Students also took 35 daily comprehension tests during the course of the study. Scores from the first six daily tests administered were combined to yield an average pretest score. Scores from the six daily tests taken at the end of the 25-day program implementation period were combined to yield an average posttest score. Scores from the five daily tests administered 4.5 months after the start of the intervention were combined to yield an average follow-up score. Students also took two social studies comprehension tests and two science comprehension tests; these tests were administered at pretest (during the first week of instruction), at midpoint (during the third week of instruction), at posttest (during the last week of instruction), and at follow-up (4.5 months after the start of the intervention). The two scores for each time period were averaged to yield a single science comprehension score and a single social studies comprehension score for each time period. The reading passages for the social studies and science comprehension tests were drawn from sections of the texts not yet covered in class in order to avoid confounding the results with prior instruction. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	Teacher training was conducted by the study author.

1. This review did not focus on the SMART group as compared to the business-as-usual control group, because students in the SMART group received the *reciprocal teaching* portion of the intervention combined with semantic mapping, therefore, the measures of effectiveness could not be attributed solely to *reciprocal teaching*.
2. Palincsar, A. S., & Brown, A. (1984). *Reciprocal teaching* of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction*, 1(2), 117–175.

Appendix A1.2 Study characteristics: Dao, 1993

Characteristic	Description
Study citation	Dao, M. N. T. H. (1993). An investigation into the application of the <i>reciprocal teaching</i> procedure to enhance reading comprehension with educationally at-risk Vietnamese-American pupils (Doctoral dissertation, University of California–Berkeley, 1993). <i>Dissertation Abstracts International</i> , 55(06A), 105–1470.
Participants	Fifty-six fourth-, fifth-, and sixth-grade students from Vietnamese-American families with low socioeconomic status were randomly assigned to experimental and control groups. Participating students were fluent in English. The experimental group was divided into four reading groups, each consisting of six to eight students. The control group was divided into three groups, each consisting of six to eight students. After attrition, the analysis sample included 50 students: 29 students in the experimental group, and 21 students in the control group.
Setting	The study included two public schools from a school district in an urban area of northern California.
Intervention	Four experimental subgroups received <i>reciprocal teaching</i> instruction using the school's current reading materials. Each subgroup worked with the teacher to read and discuss a passage. When <i>reciprocal teaching</i> was first being implemented, the teacher assigned a passage to be read and played the role of the teacher or dialogue leader. The teacher and students read silently, and the teacher modeled comprehension-monitoring strategies by asking a question on the main idea, summarizing the content, discussing, clarifying any difficulties, and making a prediction about future content. Over time, students were encouraged to assume the role of dialogue leader/teacher with the teacher's support. The intervention was carried out over 20 consecutive days of instruction.
Comparison	The control group received the regular reading curriculum. Instructional activities in the three control subgroups included asking students to read aloud, asking vocabulary questions, asking questions based on the questions appearing at the end of the text, and assigning a paper-and-pencil task that required students to answer questions, draw a picture, or write a short paragraph.
Primary outcomes and measurement	For both the pretest and the posttest, students took the paragraph subtest of the Nelson Reading Comprehension Test. For the pretest, students also took the reading comprehension subtest of the California Test of Basic Skills. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	The teacher was credentialed and trained in <i>reciprocal teaching</i> . Details on teacher training were not provided.

Appendix A1.3 Study characteristics: Leiker, 1995

Characteristic	Description
Study citation	Leiker, L. (1995). <i>An investigation of the effects of reciprocal teaching on fifth graders' comprehension and comprehension monitoring</i> . Unpublished master's thesis, University of Kansas, Lawrence.
Participants	Random assignment was used to form the treatment and control groups. ¹ Thirty-nine fifth-grade students from two classrooms participated in the study. Forty-eight percent of the students were female. The author provided the instruction for both groups. The analysis sample consisted of 20 students in the <i>reciprocal teaching</i> group and 19 students in the business-as-usual control group.
Setting	The study was conducted in an urban elementary school in the midwestern United States.
Intervention	The classroom teacher introduced four <i>reciprocal teaching</i> strategies (clarification, questioning, summarization, and prediction) and instructed students to practice these strategies over a five-day period. For the next 20 days, students were split into groups of three to four and covered four chapters of their social studies textbook. Teaching materials were selected from <i>United States and Its Neighbors</i> in the Macmillan/McGraw-Hill Social Studies Series. ² Both the experimental and control groups studied the same social studies content on the American Revolution; however, the experimental group read two more chapters than the control group did. The experiment took place over a six-week period (25 school days).
Comparison	Students in the control group covered two chapters about the American Revolution in their social studies text book. Activities that occurred during this time included students reading and then discussing the material in small groups of three to four; outlining a lesson together as a group with the teacher modeling the procedure; using a cooperative learning strategy called the "jigsaw" (in which each student in a group of four seeks to become an expert on an issue, and the group members then come together to teach one another what they have mastered); and silent reading followed by answering comprehension questions.
Primary outcomes and measurement	During the pretest and posttest periods, students were asked to read ten short social studies passages and answer ten comprehension questions about each passage. The passages included in these assessments had not been seen or discussed by students before the tests were administered. For a detailed description of these researcher-designed measures, see Appendix A2.
Staff/teacher training	The study author, who taught both groups of students, studied the <i>reciprocal teaching</i> method in graduate school. No further information on training was provided in this study.

1. The WWC was unable to obtain information from the authors on whether students or classrooms were randomly assigned.
2. Banks; Beyer; Contreras; et al. (1993). *Social studies anthology with teaching strategies (The world around us)*. Macmillan/McGraw-Hill.

Appendix A1.4 Study characteristics: Lysynchuk, Pressley, & Vye, 1990

Characteristic	Description
Study citation	Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). <i>Reciprocal teaching</i> improves standardized reading-comprehension performance in poor comprehenders. <i>Elementary School Journal</i> , 90(5), 469–484.
Participants	Thirty-six fourth-grade students enrolled in six schools and 36 seventh-grade students enrolled in two schools participated in the study. Study participants were nominated by their teachers as students with adequate decoding skills (able to decode at least 80% of the words at their grade level as measured by their performance on the Diagnostic Reading Scales) but poor comprehension skills (scoring below the 50th percentile on the comprehension subtest of the Metropolitan Achievement Test for fourth grade, or the Gates-MacGinitie Reading Test for seventh grade). All participants were English-speaking Canadians. Students were ranked according to their pretest comprehension scores and then placed into pairs starting with the two lowest-scoring students and ending with the two highest-scoring students. Students within each pair were then randomly assigned to either the control group or the <i>reciprocal teaching</i> group. At the beginning of the study, participants in the treatment and control groups had comparable reading comprehension scores. The grade 4 analysis sample included 18 students who received <i>reciprocal teaching</i> and 18 control group students. The grade 7 analysis sample also consisted of 18 students who received <i>reciprocal teaching</i> and 18 control group students.
Setting	The study was conducted in eight schools in Canada.
Intervention	The intervention incorporated all the features of <i>reciprocal teaching</i> as described by Palincsar and Brown (1984), ¹ but the instruction lasted only 13 days instead of the recommended 15–20 days.
Comparison	Students in the control group were exposed to the same materials as students in the <i>reciprocal teaching</i> group but were not exposed to the strategy instruction that is part of the <i>reciprocal teaching</i> approach.
Primary outcomes and measurement	For both the pretest and posttest, students in grade 4 took the comprehension subtest of the Metropolitan Achievement test and the vocabulary subtest of the Canadian Test of Basic Skills. Students in grade 7 took the comprehension and vocabulary subtests of the Gates-MacGinitie Reading Test. Additional researcher-designed assessments, called the “Daily Retelling” and “Daily Questions” assessments, measured student comprehension after the first half and second half of each day’s instruction. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	No information on teacher training was provided in this study.

1. Palincsar, A. S., & Brown, A. (1984). *Reciprocal teaching* of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction*, 1(2), 117–175.

Appendix A1.5 Study characteristics: Martin, 1989

Characteristic	Description
Study citation	Martin, B. J. (1989). The effect of metacognitive strategy instruction on the problem-solving skills of disadvantaged/handicapped vocational students (Doctoral dissertation, Clemson University, 1989). <i>Dissertation Abstracts International</i> , 50(06A), 103–1627.
Participants	Ten districts were randomly selected to participate in the study. In nine of the ten districts, state vocational centers participated in the study; in the remaining district, a high school participated in the study. Twenty groups of students, two from each institution, were chosen by teachers to participate in the study. Teachers at each of the institutions randomly assigned their two intact groups of students to either the <i>reciprocal teaching</i> group or to the control group. Participating students (who were between the ages of 13 and 21) were handicapped and/or disadvantaged or enrolled in a vocational course or Job Training Partnership Act summer program. The analysis sample consisted of 59 students in the <i>reciprocal teaching</i> group and 59 students in the control group.
Setting	The study was conducted in nine different vocational centers and one high school in South Carolina.
Intervention	The intervention group received the <i>reciprocal teaching</i> intervention during 15 half-hour class sessions. In each class, a teacher and a group of students took turns leading a dialogue concerning the reading material. The dialogue included discussion, argument, and four comprehension activities: summarizing, questioning, clarifying, and predicting. The teacher initially acted as a model, then chose students to take over the teacher role. Both the experimental and control groups used the same reading materials, which were on approximately a sixth- or seventh-grade reading level and pertained to the development of job-seeking and job-keeping skills.
Comparison	Students in the control group were instructed using business-as-usual instructional methods for the same amount of time (and using the same reading materials) as the treatment group. The instructional methods used in the control group were left to the discretion of each individual teacher. For example, group discussion may or may not have occurred.
Primary outcomes and measurement	For the posttest, students took the comprehension subtest of the Peabody Individual Achievement Test. The posttest was given within a week after the end of the implementation of the <i>reciprocal teaching</i> program. For a more detailed description of this outcome measure, see Appendix A2. The similarities and picture-arrangement subtests of the Wechsler Intelligence Scale for Children and the Wechsler Adult Intelligence Scale Test were also used for the posttest, but these measures were not included in this report, as they were outside the scope of the Adolescent Literacy review protocol.
Staff/teacher training	Treatment group teachers were shown a video on the <i>reciprocal teaching</i> method and provided with a complete set of 15 lesson plans.

Appendix A1.6 Study characteristics: Westera & Moore, 1995

Characteristic	Description
Study citation	Westera, J., & Moore, D. W. (1995). <i>Reciprocal teaching</i> of reading comprehension in a New Zealand high school. <i>Psychology in the Schools</i> , 32(3), 225–232.
Participants	The authors selected 46 students to participate in the study. Thirty-five of these students came from five classes taught by four teachers who volunteered to implement <i>reciprocal teaching</i> as part of the study. The remaining 11 students (from the two remaining classrooms) comprised the comparison group. Treatment group teachers implemented the program to different degrees. Teachers that provided 12–16 sessions of <i>reciprocal teaching</i> were classified as part of the extended-duration program group (20 students in three classes), and teachers that provided 6–8 sessions of <i>reciprocal teaching</i> were classified as part of the short-duration program group (15 students in two classes). The study participants were students with adequate decoding skills but poor comprehension skills (on average, more than two age-equivalent years behind). About half of the participants were Maori or Pacific Islanders. This review focused on comparisons of the 15 students in the <i>reciprocal teaching</i> short-duration group and the 10 students in the comparison group. ¹
Setting	The study took place in one New Zealand high school.
Intervention	The students in the short-duration intervention group received six to eight <i>reciprocal teaching</i> intervention sessions over a five-week period (as opposed to the 15–20 sessions recommended by Palincsar). The teachers conducted the 30-minute <i>reciprocal teaching</i> sessions while an in-class assistant supervised the rest of the class. In classes with more than one <i>reciprocal teaching</i> group, the in-class assistant also taught a <i>reciprocal teaching</i> group. According to the authors, high interest and culturally relevant books and expository and narrative articles at the 11–13-year-old age-equivalent reading level were used in the <i>reciprocal teaching</i> sessions.
Comparison	The comparison group was exposed to business-as-usual instructional methods.
Primary outcomes and measurement	For both the pretest and posttest, students took the comprehension subtest of the Progressive Achievement Test (PAT). The pretest was form A, and the posttest was form B. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	The four participating teachers and two support staff received approximately three hours of <i>reciprocal teaching</i> training, which included an introduction to assessing students' reading levels and their understanding of class texts, as well as a discussion of scaffolding and ideal classroom interactions between a more-expert person and a less-expert person in a learning situation. The teachers were introduced to the rationale behind <i>reciprocal teaching</i> . They also observed, rehearsed, and were given feedback on the four comprehension-fostering strategies and the instructional process that together constitute <i>reciprocal teaching</i> .

1. Results from the extended-duration program were also reported, but they were not included in this report because the extended program group and the comparison group were not equivalent at baseline.

Appendix A2 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Reading comprehension construct</i>	
California Test of Basic Skills, reading comprehension subtest	According to the study author, the reading comprehension subtest of the California Test of Basic Skills assesses students' ability to comprehend text. No additional information on the assessment was provided (as cited in Dao, 2009).
Daily Questions	This researcher-developed measure was used in an alternating pattern with the Daily Retelling assessment (see below) to assess student comprehension. For both tests, students read a 200-word grade-appropriate passage. The Daily Questions test asked students to answer 10 short-answer questions about the passage. Five questions required students to recall something stated in the text, and the other five required students to report something implied in the text (as cited in Lysynchuk, Pressley, & Vye, 1990).
Daily Retelling	This researcher-developed measure was used in an alternating pattern with the Daily Questions assessment (see above) to assess student comprehension. For both tests, students read a 200-word grade-appropriate passage. The Daily Retelling test assessed whether students could retell the contents of passages from memory (as cited in Lysynchuk, Pressley, & Vye, 1990).
Gates-MacGinitie Reading Test, comprehension subtest	The Gates-MacGinitie Reading Test (GMRT) is used to assess a student's decoding, vocabulary, and passage-comprehension skills. The comprehension subtest of the GMRT measures each student's ability to answer comprehension questions based on brief passages, each typically a single paragraph. The content of the comprehension subtest was drawn from actual published sources. The passages were of four types: (1) narrative-descriptive, (2) social science, (3) science, and (4) art. Questions included both literal and inferential items (as cited in Brady, 1990).
Metropolitan Achievement Test, reading comprehension subtest	The Metropolitan Achievement Test is designed to test foundation skills, critical thinking processes, and strategies in reading and mathematics. The reading comprehension subtest measures reading achievement using informational selections, fiction, and functional texts and focuses on initial understanding, interpretation, and reflective thinking. The test includes several short passages with multiple-choice questions. The questions assess students' understanding of the passage and their ability to integrate the information it contains. Some items also explore students' understanding of inferences and implications presented by the text (as cited in Lysynchuk, Pressley, & Vye, 1990).
Nelson Reading Test, Revised Edition, paragraph subtest	The Nelson Reading Test, Revised Edition, is designed to assess reading comprehension in grades 3–9. The paragraph subtest consists of 75 items (25 paragraphs, each followed by three questions). Of the three questions, one pertains to students' understanding of the significance of the text, one to their knowledge of detailed information presented in the text, and one to their ability to predict probable outcomes from the situation depicted in the paragraph. The working time for the paragraph test is 20 minutes. The test spans multiple grade levels with one form, thereby eliminating a potential confound due to different tests at different grade levels (as cited in Dao, 2009).
Peabody Individual Achievement Test–Revised Normative Update, reading comprehension subtest	The Peabody Individual Achievement Test (PIAT) is an individually administered assessment that measures academic achievement in reading, mathematics, and spelling for students in kindergarten through twelfth grade. The reading comprehension subtest, which consists of 66 two-page items, provides a measure of problem-solving ability as it relates to text. The first page of each item contains a sentence that the student reads once silently. The second page, which is exposed to the student after the passage has been read, presents four alternate illustrations. The student is asked to select the one that best represents the meaning of the sentence just read (as cited in Martin, 1989).
Progressive Achievement Test, reading comprehension subtest	The reading comprehension subtest of the Progressive Achievement Test is designed to help classroom teachers determine students' reading comprehension achievement levels. This subtest uses a multiple-choice format to assess students' ability to construct meaning from text. Each test is organized around seven or eight extended texts, including narratives, poems, and transactional texts such as reports, explanations, and procedural and persuasive texts. Questions for each text passage are of three types: (1) retrieval questions, (2) local inference questions, and (3) global inference questions (as cited in Westera & Moore, 1995).
Science comprehension test	The content for this researcher-developed comprehension test was drawn from the district's science textbooks. The selected passages were 400–500 words each. Students studied the passages at their own pace, turned them in, and then received the assessment, which consisted of 10 open-ended questions intended to assess students' skills in identifying main ideas and important factual information in the passage (as cited in Brady, 1990).

(continued)

Appendix A2 Outcome measures for the comprehension domain *(continued)*

Outcome measure	Description
Daily comprehension tests	These daily comprehension tests were designed by Palincsar (1982) ¹ for original research on <i>reciprocal teaching</i> methods. The passages included in the tests were drawn from standard textbooks with a seventh-grade readability level. Students read passages of 400–475 words and answered 10 open-ended questions of three types: (1) text-explicit, (2) text-implicit, and (3) script-implicit. Four of the 10 questions were text-explicit questions based on information contained within a single sentence of the passage. Four of the questions were text-implicit, requiring students to combine information from two or more statements. Two of the questions were script-implicit, requiring students to combine new information from the text with their prior knowledge. Each question required students to write answers, sometimes in complete sentences or in phrases or words. No time limit was imposed. Four totals were computed for each daily test: (1) number of text-explicit items answered correctly, (2) number of text-implicit items answered correctly, (3) number of script-implicit items answered correctly, and (4) total number of items answered correctly (as cited in Brady, 1990).
Social studies comprehension test	The content for this researcher-developed comprehension test was drawn from the district’s social studies textbooks. The selected passages were 400–500 words each. Students studied the passages at their own pace, turned them in, and then received the assessment, which consisted of 10 open-ended questions intended to assess students’ skills in identifying main ideas and important factual information in the passage (as cited in Brady, 1990).
Teacher-generated assessment	This assessment was developed by the study author using passages from the students’ social studies textbook. Students were asked to read 10 passages between 500 and 600 words each and answer 10 questions on each passage. For each passage, five questions could be answered using one sentence of the text, and five required students to draw on several sentences in order to provide the correct answer. Of the 10 passages, five did not require any prior knowledge of the topic in order to answer the questions. Correctly answering the remaining questions required students to draw on a 5-minute explanation of the historical context surrounding the passage, as provided by teachers to both treatment and comparison group students (as cited in Leiker, 1995).
Outcome measure	Description
Vocabulary development construct	
Canadian Test of Basic Skills, vocabulary subtest	The Canadian Test of Basic Skills (CTBS), an adapted version of the Iowa Test of Basic Skills, is intended to assess basic curriculum skills taught in Canadian provinces. It assesses language (vocabulary, reading, spelling, capitalization, punctuation, and usage), mathematics (concepts, problems, and computation), and work-study skills (visual materials and references). The CTBS vocabulary subtest is a multiple-choice test designed to take 15 minutes. The student is given a phrase with one word in boldface print. The student indicates which one of the four answers has most nearly the same meaning as the word in boldface type (as cited in Lysynchuk, Pressley, & Vye, 1990).
Gates-MacGinitie Reading Test, vocabulary subtest	The Gates-MacGinitie Reading Test (GMRT) is used to assess a student’s decoding, vocabulary, and passage-comprehension skills. The vocabulary subtest of the GMRT measures a student’s reading vocabulary. The vocabulary test contains 45 questions which measure word knowledge by asking students to choose one word or phrase that means most nearly the same as a presented word. The vocabulary test takes 20 minutes to administer (as cited in Brady, 1990, and Lysynchuk, Pressley, & Vye, 1990).

1. Palincsar, A. S. (1982). *Improving the reading comprehension of junior high students through the reciprocal teaching of comprehension-monitoring strategies*. Unpublished doctoral dissertation, University of Illinois.

Appendix A3.1 Summary of study findings included in the rating for the comprehension domain

Meets WWC evidence standards	Comprehension domain
Brady (1990)	(+)
Dao (1993)	+
Leiker (1995)	Ind
Lysynchuk, Pressley, & Vye (1990)	Ind
Martin (1989)	(+)

Meets WWC evidence standards with reservations	Comprehension domain
Westera & Moore (1995)	(-)

Rating of effectiveness	mixed effects
-------------------------	---------------

+ = study finding was positive and statistically significant

(+) = study finding was positive and substantively important, but not statistically significant

(-) = study finding was negative and substantively important, but not statistically significant

ind = study finding was indeterminate; that is, neither substantively important nor statistically significant

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (Reciprocal teaching – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Reciprocal teaching group	Comparison group				
Brady, 1990^{7,8}								
Gates-MacGinitie comprehension tests	Grades 5–8	12 students	502.89 (56.28)	485.33 (61.89)	17.56	0.27	ns	+11
Gates-MacGinitie vocabulary tests	Grades 5–8	12 students	509.84 (27.35)	494.83 (46.12)	15.01	0.37	ns	+14
Science comprehension tests	Grades 5–8	12 students	3.09 (2.68)	1.75 (1.63)	1.34	0.56	ns	+21
Social studies comprehension tests	Grades 5–8	12 students	5.33 (2.25)	2.50 (1.38)	2.83	1.40	ns	+42
Daily comprehension tests	Grades 5–8	12 students	7.09 (1.71)	5.89 (1.74)	1.20	0.64	ns	+24
Average for comprehension (Brady, 1990)⁹						0.65	ns	+24
Dao, 1993^{7,10}								
Nelson Comprehension Test	Grades 4–6	50 students	35.07 (nr)	27.76 (nr)	7.31	na	Statistically significant	na
Average for comprehension (Dao, 1993)						na	Statistically significant	na
Leiker, 1995^{7,11}								
Researcher-designed social science comprehension questions	Grade 5	39 students	0.50 (0.22)	0.47 (0.22)	0.03	0.14	ns	+5
Average for comprehension (Leiker, 1995)						0.14	ns	+5

(continued)

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹ (continued)

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) ²		Mean difference ³ (Reciprocal teaching – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Reciprocal teaching group	Comparison group				
Lysynchuk, Pressley, & Vye, 1990^{7,12}								
Metropolitan Achievement Test, comprehension subtest	Grade 4	36 students	37.34 (21.34)	28.33 (12.87)	9.01	0.50	ns	+19
Canadian Test of Basic Skills, vocabulary subtest	Grade 4	36 students	40.62 (19.36)	31.44 (25.28)	9.18	0.54	ns	+21
Daily Retelling	Grade 4	36 students	25.07 (21.22)	22.90 (14.06)	2.17	0.12	ns	+5
Daily Questions	Grade 4	36 students	60.14 (12.76)	49.08 (20.92)	11.06	0.85	ns	+30
Gates-MacGinitie Reading Test, comprehension subtest	Grade 7	36 students	32.56 (24.56)	24.89 (16.92)	7.67	0.36	ns	+14
Gates-MacGinitie Reading Test, vocabulary subtest	Grade 7	36 students	19.44 (21.82)	30.89 (27.41)	-11.45	-0.62	ns	-23
Daily Retelling	Grade 7	36 students	8.17 (6.39)	8.86 (4.51)	-0.69	-0.12	ns	-5
Daily Questions	Grade 7	36 students	45.03 (18.78)	41.53 (14.96)	3.50	0.27	ns	+11
Average for comprehension (Lysynchuk, Pressley, & Vye, 1990)⁹						0.24	ns	+9
Martin, 1989⁷								
Peabody Individual Achievement Test, comprehension subtest	13–21 years old	20 groups (2 per school/vocational center)/ 118 students	47.10 (8.61)	44.66 (10.09)	2.44	0.26	ns	+10
Average for comprehension (Martin, 1989)						0.26	ns	+10
Westera & Moore, 1995^{7,13}								
Progressive Achievement Test, comprehension subtest	Grade 8	25 students	10.52 (6.67)	13.41 (4.00)	-2.89	-0.48	ns	-19
Average for comprehension (Westera & Moore, 1995)						-0.48	ns	-19
Domain average for comprehension across all studies⁹						0.16	na	+6

(continued)

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹ (continued)

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. Item-level and follow-up findings on daily comprehension tests from Brady (1990) are not included in these ratings but are reported in Appendix A4.1 and Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the cases of Dao (1993), Leiker (1995), and Westera and Moore (1995), no corrections for clustering or multiple comparisons were needed. In the cases of Brady (1990) and Lysynchuk, Pressley, and Vye (1990), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original studies. In the case of Martin (1989), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The intervention group values reported for Brady (1990) are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
10. The WWC could not calculate effect sizes for Dao (1993) in a way that was comparable to the other studies in this intervention report. This was because the WWC uses unadjusted standard deviations in the denominator of the effect size formula, whereas the author reported change scores' standard deviations.
11. The intervention and control group means reported for Leiker (1995) are ANCOVA-adjusted posttest scores. As the group standard deviations are not reported in the study, the pooled standard deviation (0.22) is derived from individual standard deviations of five posttest measures (Leiker, 1995:71).
12. The intervention group values reported for Lysynchuk, Pressley, and Vye (1990) are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
13. The intervention and control group means reported for Westera and Moore (1995) are ANCOVA-adjusted posttest scores, with pretest scores being treated as a covariate.

Appendix A4.1 Summary of item-level findings for the comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (Reciprocal teaching – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Reciprocal teaching group	Comparison group				
Brady, 1990^{7,8}								
Script-implicit items on daily comprehension tests	Grades 5–8	12 students	1.48 (0.22)	1.19 (0.26)	0.29	1.11	ns	+37
Text-explicit items on daily comprehension tests	Grades 5–8	12 students	2.93 (1.00)	2.61 (0.90)	0.32	0.31	ns	+12
Text-implicit items on daily comprehension tests	Grades 5–8	12 students	2.68 (0.76)	2.08 (0.80)	0.60	0.71	ns	+26

ns = not statistically significant

1. This appendix presents item-level findings for measures that fall in the comprehension domain. Total scale scores for daily comprehension tests were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Brady (1990), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
8. The intervention group values reported for Brady (1990) are the comparison group means plus the difference in mean gains between the intervention and comparison groups.

Appendix A4.2 Summary of follow-up findings for the comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (Reciprocal teaching – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Reciprocal teaching group	Comparison group				
Brady, 1990^{7,8}								
Script-implicit items on daily comprehension tests	Grades 5–8	12 students	1.21 (0.24)	0.90 (0.41)	0.31	0.85	ns	+30
Text-explicit items on daily comprehension tests	Grades 5–8	12 students	3.02 (0.78)	2.28 (0.69)	0.74	0.93	ns	+32
Text-implicit items on daily comprehension tests	Grades 5–8	12 students	3.00 (0.62)	2.42 (0.53)	0.58	0.93	ns	+32
Daily comprehension tests	Grades 5–8	12 students	7.23 (1.59)	5.60 (1.39)	1.63	1.01	ns	+34

ns = not statistically significant

1. This appendix presents follow-up findings for measures that fall in the comprehension domain. Follow-up data were collected over a two-week period three months after the end of the intervention. Posttest total scale scores on daily comprehension tests were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Brady (1990), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
8. The intervention group values reported for Brady (1990) are the comparison group means plus the difference in mean gains between the intervention and comparison groups.

Appendix A5 Reciprocal teaching rating for the comprehension domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of comprehension, the WWC rated *reciprocal teaching* as having mixed effects for adolescent learners. The remaining ratings (no discernible effects, potentially negative effects, and negative effects) were not considered, as *reciprocal teaching* was assigned the highest applicable rating.

Rating received

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Met. One of the six studies showed a statistically significant positive effect, two studies showed substantively important positive effects, and one study showed a substantively important negative effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. Three of the six studies showed statistically significant or substantively important positive effects, and two studies showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. One of the six studies showed statistically significant positive effects.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Not met. One study showed substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. Three of the six studies showed statistically significant or substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Not met. One study showed substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	na	na	na	na
Fluency	na	na	na	na
Comprehension	6	23	316	Medium to large
General literacy achievement	na	na	na	na

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see WWC Procedures and Standards Handbook, Appendix G.