What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

August 2012

WWC Review of the Report "Impact of the Thinking Reader® Software Program on Grade 6 Reading Vocabulary, Comprehension, Strategies, and Motivation"[1,2]

The findings from this review do not reflect the full body of research evidence on *Thinking Reader*®.

What is this study about?

The study of *Thinking Reader*® is a multisite cluster randomized controlled trial. Ninety-two reading/English language arts teachers from 32 elementary and middle schools were randomly assigned within their schools to either the *Thinking Reader*® condition or the comparison condition. The analysis sample consisted of 90 classes and 2,147 grade 6 students, with 1,156 students in the *Thinking Reader*® condition and 991 students in the comparison condition.[3]

Teachers in the *Thinking Reader*® condition supplemented their regular English language arts or reading instruction with one to three preselected *Thinking Reader*® novels that students were asked to read within the *Thinking Reader*® software program. Students in comparison group classrooms participated in the schools' regular curriculum.

The study assessed the effectiveness of *Thinking Reader*® by comparing the reading comprehension of students in the *Thinking Reader*® and comparison conditions at the end of the school year.[4]

What did the study find?

The study found no statistically significant differences on the comprehension outcomes of students in the *Thinking Reader*® classes, compared with students in the comparison classes.

WWC Rating

The research described in this report meets WWC evidence standards without reservations

Strengths: This study is a well-implemented randomized controlled trial.

Features of *Thinking Reader*®

Thinking Reader® is a software program that aims to motivate middle school students to read and to make self-directed use of seven target comprehension strategies: a) summarizing, b) clarifying, c) visualizing, d) reflecting, e) questioning, f) predicting, and g) feeling. Students listen to a novel while following highlighted text on a computer screen and then respond to questions about the story. The program applies reciprocal teaching methods through the use of animated coaches and peers to enhance comprehension strategies.

The *Thinking Reader*® instructional routine consists of three phases. In the first phase, teachers introduce students to the program through activities such as modeling a strategy. During the second phase, the teachers observe and review students' progress while students read a novel on the computer. For the third phase, teachers and students interact offline: they discuss the book, and then students complete an activity to demonstrate understanding. The program has five levels of interactive instructional support and allows students to progress to lower levels of support where they can independently select comprehension strategies.

Appendix A: Study details

Drummond, K., Chinen, M., Duncan, T. G., Miller, H. R., Fryer, L., Zmach, C., & Culp, K. (2011). *Impact of the Thinking Reader® software program on grade 6 reading vocabulary, comprehension, strategies, and motivation* (NCEE 2010-4035). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Setting

The study was conducted in 32 elementary and middle schools in 16 school districts in Connecticut, Massachusetts, and Rhode Island. More than one-third of the students at each school were eligible for free or reduced-price lunch.

Study sample

Researchers recruited schools that had at least two reading/English language arts teachers in grade 6. Within the 32 schools that agreed to participate, 92 teachers agreed to take part in the study and were randomly assigned within their school to either the *Thinking Reader®* condition or the comparison condition. Eligible students were defined as those enrolled in the classrooms of participating teachers at the time of the pretest who were not identified as having low English language levels or special Individualized Education Program (IEP) requirements that would have precluded them from testing. The original study sample consisted of 2,407 students: 1,286 in the *Thinking Reader®* condition and 1,121 in the comparison condition. The analysis sample consisted of 90 teachers and 2,147 students who remained at the end of the school year: 1,156 in the *Thinking Reader®* condition and 991 in the comparison condition.[5] The authors also reported subgroup impacts for students based on initial reading achievement levels, broken into tertiles (or achievement levels) of approximately 700 students per group.

Intervention group

Teachers in the intervention condition supplemented their regular English language arts or reading instruction with one to three preselected *Thinking Reader®* software-based novels, which were intended to be implemented during the school year over a period of 24 to 54 days (between 1,320 and 2,970 minutes of software-based instruction time). The actual implementation of the program included just over 1,000 minutes of software-based instruction over approximately 25 days. Most of the teachers initiated the first and second books, and just over half of the teachers initiated the third book. Student completion rates for books 1, 2, and 3 were 74%, 53%, and 9%, respectively.[4] Classroom observations showed that teachers did not follow the recommended three-phase instructional routine in 80% of observed lessons.

Comparison group

Students in comparison group classrooms participated in their schools' standard curriculum, which included English language arts curriculum activities, such as reading short stories, newspaper and magazine articles, and non-*Thinking Reader®* novels. Personnel at some participating schools expressed the desire that all students read the same novels. Thus, hard copy versions of the *Thinking Reader®* novels were provided to schools so that students in comparison group classrooms had access to the novels. Students in the comparison group classrooms, however, did not have access to the *Thinking Reader®* software.

Outcomes and measurement

Both prior to the introduction of the intervention and after its completion, students took the Vocabulary and Comprehension multiple-choice subtests from the standardized Gates-MacGinitie Reading Test (GMRT; MacGinitie et al., 1999).[6] For a more detailed description of these outcome measures, see Appendix B.

Two additional self-reported outcomes were examined in this study, but are not included in this report because they assess outcomes outside of the purview of the Adolescent Literacy protocol: students' use of comprehension strategies and students' motivation to read.

Support for implementation

Teachers assigned to the *Thinking Reader*® condition attended two group-session workshops during the year (lasting six hours each) and participated in three individual follow-up coaching sessions (lasting approximately eight hours combined). They also had opportunities to communicate with *Thinking Reader*® coaches throughout the school year.

Reason for review

This study was identified for review by the WWC because it is an Institute of Education Sciences (IES)-funded study conducted by 2006–11 Regional Education Laboratory Northeast and Islands at Education Development Center (EDC).

Appendix B: Outcome measures for each domain

Comprehension	
Gates-MacGinitie Reading Test (GMRT)—Vocabulary subtest	The Vocabulary subtest of the GMRT measures reading vocabulary by asking students to choose one word or phrase that means most nearly the same as a presented word. The test contains 45 questions.
GMRT—Comprehension subtest	The Comprehension subtest of the GMRT measures the ability of students to read and understand different types of prose. The test requires students to read passages of various lengths and subjects and answer a total of 48 questions based on these passages.

Appendix C: Study findings for the comprehension domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Comprehension								
GMRT—Vocabulary subtest	Grade 6	90 teachers/ 2,147 students	515.75 (34.86)	516.99 (34.86)	–1.24	–0.04	–1	0.35
GMRT—Comprehension subtest	Grade 6	90 teachers/ 2,140 students	507.42 (33.70)	506.52 (33.70)	0.90	0.03	+1	0.61
Domain average for comprehension						**0.00**	**0**	**Not statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The statistical significance of the study's domain average was determined by the WWC; a domain is characterized as not statistically significant when univariate statistical tests are reported for each outcome measure and each of the effects within the domain are not statistically significant. GMRT = Gates-MacGinitie Reading Test.

Study Notes: Hedge's *g* effect sizes were computed using a three-level model adjusted for multiple covariates, in which students were nested within teachers, who were nested within schools. The regression-adjusted means, pooled standard deviations, effect size, and *p*-values presented here were reported by the authors in the original study. Effect sizes were based on the pooled posttest standard deviations. A multiple comparison adjustment was made in the original study to account for the two comparisons. No additional corrections for clustering or multiple comparisons were needed.

Appendix D: Subgroup findings for the comprehension domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Comprehension								
GMRT—Vocabulary subtest	Grade 6 (achievement tertile 1)	685	461.38 (nr)	458.80 (nr)	2.58	0.10	+4	0.18
GMRT—Comprehension subtest	Grade 6 (achievement tertile 1)	740	459.29 (nr)	457.14 (nr)	2.15	0.07	+3	0.33
GMRT—Vocabulary subtest	Grade 6 (achievement tertile 2)	746	483.99 (nr)	484.92 (nr)	–0.93	–0.04	–2	0.60
GMRT—Comprehension subtest	Grade 6 (achievement tertile 2)	659	474.83 (nr)	478.44 (nr)	–3.61	–0.13	–5	0.10
GMRT—Vocabulary subtest	Grade 6 (achievement tertile 3)	716	516.05 (nr)	516.66 (nr)	–0.61	–0.02	–1	0.74
GMRT—Comprehension subtest	Grade 6 (achievement tertile 3)	741	505.91 (nr)	506.45 (nr)	–0.54	–0.02	–1	0.81

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. GMRT = Gates-MacGinitie Reading Test. nr = not reported.

Study Notes: The estimated impacts were computed using a three-level model adjusted for multiple covariates, in which students were nested within teachers, who were nested within schools. Hedge's *g* effect sizes were computed by the WWC based upon author reported sample sizes and *t*-statistics. The regression-adjusted means and *p*-values presented here were reported by the authors in the original study. The sample sizes presented here were provided by the authors and represent the number of students who had baseline and covariate data and were included in the analysis of each outcome. A multiple comparison adjustment was made in the original study to account for the two comparisons in each achievement tertile. No additional corrections for clustering or multiple comparisons were needed.

Endnotes

[1] Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the Adolescent Literacy topic area protocol, version 2.0. The WWC rating applies only to the results that were eligible under this topic area and met WWC standards without reservations or met WWC standards with reservations, and not necessarily to all results presented in the study.

[2] The Regional Educational Laboratory Northeast and Islands' (REL-NEI) technical working group provided insight and advice on the evaluation design of this study. The Regional Educational Labs were provided technical assistance by Mathematica Policy Research, which also operates the WWC. For this reason, this study was reviewed by staff from subcontractor organizations.

[3] These numbers reflect the overall analysis sample for the Gates-MacGinitie Reading Test (GMRT) Vocabulary subtest. The GMRT Comprehension subtest included 2,140 students (1,154 in the *Thinking Reader®* group and 986 in the comparison group).

[4] Two additional self-reported outcomes were examined in this study, but are not included in this report because they assess outcomes outside of the purview of the Adolescent Literacy protocol: students' use of comprehension strategies and students' motivation to read.

[5] In addition to less-than-intended program implementation, a small number of teachers (fewer than four) who were assigned to the *Thinking Reader®* condition chose not to implement the program, but agreed to allow the researchers to collect outcome data. A small number of teachers (fewer than four) also left the study schools after random assignment, due to layoffs or budget cuts. The schools distributed these teachers' students among other teachers in the school, and the study continued to track these students, maintaining the students' original group assignments. The analysis included all students with available data who were in classrooms that were randomly assigned to the intervention or comparison condition, regardless of program participation.

[6] MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (1999). *Gates-MacGinitie Reading Tests (4th ed.)*. Itasca, IL: Riverside Publishing.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, August). *WWC review of the report: Impact of the Thinking Reader® software program on grade 6 reading vocabulary, comprehension, strategies, and motivation.* Retrieved from http://whatworks.ed.gov.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Single-case design (SCD) A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.