

WWC Review of the Report “Large-scale Randomized Controlled Trial with 4th Graders Using Intelligent Tutoring of the Structure Strategy to Improve Nonfiction Reading Comprehension”¹

The findings from this review do not reflect the full body of research evidence on *Intelligent Tutoring of the Structure Strategy* for reading comprehension.

What is this study about?

The study examined the effects of a web-based tutoring program, *Intelligent Tutoring of the Structure Strategy (ITSS)*, on the reading comprehension of fourth-grade students in language arts classrooms. The analysis included 1,875 to 2,371 fourth-grade students from 100 to 117 classrooms in Pennsylvania elementary schools (sample sizes varied across outcome measures).

Schools volunteered to participate in the study. Within each school or group of similar schools, researchers randomly assigned 131 classrooms to either participate in *ITSS* or serve as the comparison group, which followed the regular school curriculum for language arts. Students in the *ITSS* classrooms used the system for one class period a week for 6–7 months as a partial substitute for their regular language arts curriculum (i.e., time spent using *ITSS* replaced regular instructional time).

This study assessed the effectiveness of *ITSS* immediately after the end of the intervention by comparing the reading comprehension of students in the *ITSS* classrooms with students in the comparison classrooms. Reading comprehension was measured with a standardized test (the Gray Silent Reading Test, or GSRT) and five researcher-designed measures on two types of text structures: comparison type texts and problem/solution type texts.²

What did the study find?

The study authors reported, and the WWC confirmed, that *ITSS* had a statistically significant

positive effect on the reading comprehension of fourth-grade students as measured by the researcher-designed tests. The study authors also reported, and the WWC confirmed, no statistically significant effects of the intervention on the GSRT.

WWC Rating

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

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

Features of *Intelligent Tutoring of the Structure Strategy (ITSS)*

ITSS is a one-on-one, web-based intelligent tutoring system which models a “structure strategy” technique, provides practice opportunities, and gives immediate feedback to students. Structure strategy is a method for explicitly using knowledge about the text structure to increase reading comprehension of nonfiction texts. Students are taught to (a) classify the text by identifying signaling words that clue arguments, (b) write a main idea using a pattern specific for that type of text, and (c) recall the information from the text using the signaling words and main idea to prompt their recollection in an organized manner. The system has a book-like interface and an animated intelligent tutor who guides the learner through the exercises with a human voice.

Appendix A: Study details

Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, 60(6), 987–1013.

Setting The study took place in fourth-grade classrooms in rural and suburban Pennsylvania elementary schools that volunteered to participate in the study.

Study sample Within each school, fourth-grade classrooms were randomly assigned to either the *ITSS* intervention group or the comparison group. If a school did not have enough classrooms, schools with similar characteristics were grouped together to form a “site” before random assignment. The initial sample included 131 classrooms with 3,152 students. The final research sample varied by outcome measures because two large schools were not able to complete all the measures. The final research samples varied from 100 classrooms with 1,875 students to 117 classrooms with 2,371 students across outcomes.

Intervention group Students in the intervention group used *ITSS* as a partial substitute for 30–45 minutes a week for the regular language arts curriculum for 6–7 months (i.e., time spent using *ITSS* replaced regular instructional time). *ITSS* is a one-on-one web-based intelligent tutoring system for learning structure strategy, a method for strategically using knowledge about text structure to increase reading comprehension. The text structure knowledge is designed to improve encoding and information retrieval from nonfiction texts.

The structure strategy has three steps: (a) identify signaling words to classify the text, focusing on top-level structure, and creating strategic memory representations, (b) write a thorough main idea using the main idea pattern for the particular text structure, and (c) write a full organized recall of the passage using signaling words and the main idea. The tutoring system models the steps, provides practice, assesses, and gives feedback.

Comparison group Students in the comparison group participated in their school’s standard language arts curriculum. Total daily and weekly amounts of language arts instruction were the same for the comparison group and the intervention group.

Outcomes and measurement Students were tested in reading comprehension before and after the intervention using a standardized assessment (Gray Silent Reading Test, or GSRT), and five researcher-designed measures. The researcher-designed measures tested students on two types of texts: comparison structure passages and texts with problem and solution structures. For the comparison texts, there were three subtests included in this WWC report: main idea quality (use of a comparison when writing a two-sentence main idea), total recall, and comparison competency (using comparison structures for the recall task). For the problem/solution passages, there were two subtests: total recall and problem/solution competency. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

After random assignment was complete, the researchers conducted professional development sessions for the intervention group teachers in each school. The researchers also reviewed the weekly computer usage logs and mailed biweekly reports to the *ITSS* teachers on student progress.

Reason for review

This study was identified for review by the WWC because it was supported by a grant to Pennsylvania State University (Principal Investigator: Kay Wijekumar; Award Number: R305A080133) from the National Center for Education Research (NCER) at the Institute of Education Sciences (IES).

Appendix B: Outcome measures for the reading comprehension domain

Reading comprehension	
<i>Comparison Text:</i> ^a <i>Comparison Competency Test</i>	In this researcher-designed measure, students were scored on competency of using the comparison structure to organize the recall responses. Scores ranged from one to eight. Inter-rater agreement was 89%.
<i>Comparison Text:</i> ^a <i>Main Idea Test</i>	In this researcher-designed measure, students were asked to write a two-sentence main idea while the text was available to consult. Scores ranged from one to six based on how well the comparison structure was used. Inter-rater agreement was 93%.
<i>Comparison Text:</i> ^a <i>Total Recall Test</i>	In this researcher-designed measure, students were asked to write as much as they could recall from a text while it was out of sight. Recalls were scored as the number of ideas remembered. Inter-rater agreement was 99%.
<i>Gray Silent Reading Test (GSRT)</i>	The GSRT measures silent reading comprehension through a series of progressively difficult passages. Each passage is followed by five questions designed to gauge comprehension of the passage. Form B was given as a pretest, and Form A was used as a posttest.
<i>Problem/Solution Text:</i> ^b <i>Problem Solution Competency Test</i>	In this researcher-designed measure, students were scored on competency of using the problem/solution structure to organize the recall responses. Scores ranged from one to eight. Inter-rater agreement was 89%.
<i>Problem/Solution Text:</i> ^b <i>Total Recall Test</i>	In this researcher-designed measure, students were asked to recall as much as they could while the text was out of sight. Recalls were scored as the number of ideas remembered. Inter-rater agreement was 98%.

Table Notes: In addition to the outcomes described above, the study also examined impacts on a Signaling Test. This assessment was determined to be overly aligned with the intervention and, therefore, is not included in this SSR.

^a *Comparison Text* is a researcher-created assessment based on two similar comparison structure texts with 128 words in each. Each domain was based on student tasks after these texts were read.

^b *Problem/Solution Text* is a researcher-created assessment based on two similar problem/solution texts of 98 words each. Each problem/solution text described a problem and a related solution. Each domain was based on student tasks after these texts were read.

Appendix C: Study findings for the reading 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	
Reading comprehension								
<i>Comparison Text: Comparison Competency Test</i>	ITSS vs. Comparison	100 classrooms/ 1,877 students	3.79 (0.80)	3.38 (0.70)	0.41	0.18	+7	0.01
<i>Comparison Text: Main Idea Test</i>	ITSS vs. Comparison	100 classrooms/ 1,875 students	3.22 (0.57)	2.44 (0.49)	0.78	0.49	+19	0.00
<i>Comparison Text: Total Recall Test</i>	ITSS vs. Comparison	100 classrooms/ 1,900 students	21.21 (5.52)	19.57 (4.69)	1.64	0.11	+4	0.02
<i>Gray Silent Reading Test (GSRT)</i>	ITSS vs. Comparison	117 classrooms/ 2,371 students	28.93 (4.37)	27.86 (3.89)	1.07	0.09	+4	0.08
<i>Problem/Solution Text: Problem Solution Competency Test</i>	ITSS vs. Comparison	100 classrooms/ 1,904 students	3.07 (0.62)	2.79 (0.62)	0.28	0.13	+5	0.01
<i>Problem/Solution Text: Total Recall Test</i>	ITSS vs. Comparison	100 classrooms/ 1,910 students	15.36 (2.85)	13.48 (3.18)	1.88	0.18	+7	0.00
Domain average for reading comprehension						0.20	+8	Statistically significant

Table Notes: Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor 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 WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC; the study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

Study Notes: The adjusted mean differences, effect sizes, and p-values were reported in the study and are based on a three-level hierarchical linear model (HLM), which accounts for the clustering of students into classrooms and classrooms into schools. The sample sizes for each comparison were obtained through an email request to the author. The WWC calculated the intervention group mean by adding the adjusted mean difference estimates reported in the study to the unadjusted comparison group means reported in Tables 1 and 2. The means and standard deviations reported in Tables 1 and 2 represent classroom level means and standard deviations. As a result, the WWC did not calculate effect sizes relative to the standard deviations reported above and, instead, relied on the study calculations of these measures for all outcomes except for the *GSRT*, since this effect size was reported in terms of standard deviation units on the pretest, not the posttest assessment. In order to calculate the WWC effect size for the *GSRT* outcome, the WWC used the unconditional within-classroom variance component of 136.99 shown in MO of Table 4. By taking the square root of this variance component, and calculating the mean difference relative to this standard deviation of the posttest, the WWC was able to estimate a revised ES of 0.09 posttest standard deviation units, which is reported here in favor of the study-presented value of 0.10 pretest standard deviation units. Corrections for multiple comparisons were needed but did not affect the significance of the reported results.

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 review protocol, version 2.0.

² There was one outcome included in the study that is not described in this WWC report. See the table notes in Appendix B for more information.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, July). *WWC review of the report: Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension*. 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 are 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.