

Appendix A: CONSORT Criteria (Law & Plunkett, 2006)

Headings	Subheadings	Garner & Bochna, 2004	Paris & Paris, 2007	Short & Ryan, 1984	Westerveld & Gillon, 2008
Title and abstract		The title does not provide information about study design, but the abstract indicates that a comparison group was used.	The title does not provide information about study design, but the abstract indicates that a comparison group was used.	The title does not provide information about study design, but the abstract indicates that a comparison group was used.	The title does not provide information about study design, but the abstract indicates that a comparison group was used.
Introduction		The paper's introduction provides the research base on story grammar development and instruction and a rationale for the current research.	The introduction provides a detailed discussion of the research in narrative instruction and story grammar. It also provides a clear rationale for the current study.	The introduction provides a detailed discussion on meta-cognitive strategies including information on the two types of interventions: attribution training and story grammar training.	The introduction provides a detailed discussion of the research in narrative structure comprehension and production. It also provides a clear rationale for the current study.
Methods	Participants	Selection criteria are not clearly presented in the paper. Treatment group n=37, comparison group n not clear. Authors reports n=16 girls but not total n. The setting of the intervention was not stated in the paper.	One hundred and twenty-three students in six first-grade classrooms in one elementary school. Demographic data is provided in Table 1. All students whose parents gave consent for the study participated.	Forty-two 4th-grade, less skilled readers were randomly assigned to one of three treatment groups. Fourteen 4th-grade skilled readers did not receive treatment and served as a comparison group using posttest measures.	Ten children identified as poor readers between the ages of 7:11 and 9:2 from the original longitudinal study who were available for intervention. Five of the children formed the treatment group. The remaining five children served as the poor reading control group and received delayed treatment. Ten typically developing children from the longitudinal study served as the control group.

Appendix A: CONSORT Criteria (Law & Phunkett, 2006)

Headings	Subheadings	Garner & Bochna, 2004	Paris & Paris, 2007	Short & Ryan, 1984	Westerveld & Gillon, 2008
	Interventions	A brief explanation of the intervention in both groups was presented. The intervention condition utilized direct explanation and guided practice to teach students: main character, setting, problem, attempted solution, and solution. The comparison group used their language arts basal series to read and listen to stories in comparable frequency to the intervention group. No references to story structure were used. The intervention program received 15-20 minutes of instruction, including 8-10 teacher-read stories per element, twice daily for 16 weeks. The authors did not specify the intensity of the comparison program.	The NSI (Narrative Strategy Instruction) condition consisted of five specific NSI units that were created for the study. Each unit was taught over two lessons in a one-week period. Each class in the NSI condition received 10 lessons, each 45-minutes long, for a total instructional time of 450 minutes. The same instructor (one of the researchers) provided all lessons. A detailed explanation of the instructional sessions is also presented by the authors. The comparison condition utilized the same amount of instructional time in the same manner using the poetry genre instead of the narratives used in the intervention condition.	The total training group received story grammar strategy training and attribution training (designed to increase awareness of effort in efficient reading). The strategy group received only story grammar strategy training. The control group of less skilled readers received only attribution training. A detailed explanation of the instructional sessions is presented by the authors. All treatment groups received three sessions occurring over one week, each 30-35 minutes long. The experimenter provided the intervention.	The treatment group received 12 one-hour group sessions over a six-week period. Each intervention session consisted of instruction of a particular story element. The students listen to part of a story, identify the story element using a story map, and retell the story with feedback provided by other group members.
	Objectives	The objectives are stated clearly as the following research questions: (a) What, if any, pre-existing story grammar knowledge exists? (b) What is the benefit of direct instruction over and above mere exposure to narrative stories, and on what tasks might new knowledge be brought to bear?	The objectives are stated explicitly in the form of five research questions: (a) Did NSI have positive effects on children's narrative comprehension skills? (b) Did NSI generalize to expository text? (c) Did the effects of the treatment generalize to listening comprehension? (d) Did the effects of the treatment generalize to oral production? (e) Did NSI benefit all students?	The objectives are stated clearly as the following research questions: (a) Do strategy-trained, less skilled readers differ from skilled readers in their ability to utilize story schemata to aid their comprehension of new information? (b) Do less skilled readers benefit from training with a self-instructional story grammar strategy? (c) Are the benefits reaped from strategy training enhanced by attribution training?	The objectives are stated clearly as the following research questions: a) Does oral narrative intervention enhance the narrative production and comprehension skills of children with a mixed reading ability? (b) Does oral narrative intervention improve these children's reading comprehension performance?

Appendix A: CONSORT Criteria (Law & Plunkett, 2006)

Headings	Subheadings	Garner & Bochna, 2004	Paris & Paris, 2007	Short & Ryan, 1984	Westerveld & Gillon, 2008
	Sample size	Treatment group n=37, comparison group n not clear. Authors reports n=16 girls but not total n.	Treatment group n=83 in four classrooms, comparison group n=40 in two classrooms.	Authors initially report that 14 skilled and 42 less skilled readers participated, but later reported 14 skilled and 39 less skilled readers.	Treatment group n=5, control group of poor readers n=5, control group of typically developing students n=10.
	Randomization	Students were randomly allocated to either oral (n=8) or silent reading (n=19) treatment groups or a comparison group (14) Discrepancy from sample size not described by authors.	Six classrooms were randomly assigned to either the intervention or comparison condition (four classrooms in the experimental NSI condition, two classrooms in the comparison condition).	The less skilled readers were randomly assigned to one of three treatment groups. However, the authors report that they ensured the groups were racially balanced. It is unclear how they did this.	Ten poor readers were allocated to two groups based on the geographical location of their school. The two groups were randomly assigned to the treatment or delayed treatment groups.
	Blinding	No information was provided on blinding in the paper.	No information was provided on blinding in the paper.	No information was provided on blinding in the paper.	No information was provided on blinding in the paper.
	Statistical methods	ANOVA conducted for Reading Comprehension and Listening Comprehension. Subgroup analysis for correct responders to story element measure.	The study's data were analyzed using a 2 (time) x 2 (condition) repeated measures analyses of covariance (ANCOVA & MANCOVA) to test for interaction effects. There was a mismatch of assignment and analysis.	The study used a multivariate analysis of variance with group as a factor to analyze the data.	The study used a non-equivalent pre-test/posttest control group design in which one group received delayed treatment and served as the control group. There was a mismatch of assignment and analysis.
Results	Participant flow	Discrepancy between sample size and analysis n values not described.	It appears that all participants were present during all stages of the study.		It appears that all participants were present during all stages of the study.
	Recruitment	Recruitment of the participants was not specified in the paper.	Recruitment of the participants was not specified in the paper other than that they were all first graders at a single elementary school whose parents gave consent for participation in the study.	Recruitment of the participants was not specified in the paper other than that they were 4th-grade boys from four public and six private schools selected based on their reading comprehension scores from the Stanford Diagnostic Reading Test. All treatment participants scored at or below the 50th percentile on this measure.	The participants were originally recruited for participation in a longitudinal study being conducted by the authors. The 10 poor readers were those who were available for intervention. The 10 typically developing readers also were recruited originally for participation in the longitudinal study and matched the 10 poor readers on age, gender, ethnicity, and grade level.

Appendix A: CONSORT Criteria (Law & Plunkett, 2006)

Headings	Subheadings	Garner & Bochna, 2004	Paris & Paris, 2007	Short & Ryan, 1984	Westerveld & Gillon, 2008
	Baseline data	Baseline demographic and performance data were not provided in the paper.	Baseline demographic data were provided in Table 1 for all students. Pre-test data were provided on all participants in all evaluation measures (Narrative Comprehension task, Expository Comprehension task, Narrative Production task, Expository Production task, Listening Comprehension tasks, Woodcock-Johnson III Picture Vocabulary test, Michigan Literacy Progress Profile Rhyming, Blending, and Phoneme Segmentation, Teacher Ratings, and Literacy Habits Interview).	Some baseline demographic data were provided (i.e., Stanford Diagnostic Reading Test, Comprehension Score were used for all children to determine whether they were considered “skilled” or “less skilled” readers. Additional pre-test measures included for only the less skilled readers included free recall and probed recall.	Baseline demographic and pre-test data were provided for all participants. Pre-test data were provided in the following evaluation measures: Neale Analysis of Reading Ability, Clinical Evaluation of Language Fundamentals, 3rd Edition, Test of Nonverbal Intelligence, 3rd Edition, Peabody Picture Vocabulary Test, 3rd Edition.
	Numbers analyzed	The study reported that 35 students were assigned to the intervention group and 31 students to the comparison group for Reading Comprehension and Listening Comprehension analysis at the posttest. Discrepancy between participants at the pre-test and posttest not described.	The study reported that 83 students in four classrooms were assigned to the intervention group and 40 students in two classrooms were assigned to the comparison condition. It appears that all students were present for pre-testing and posttesting.	Though there was initially a discrepancy in the number of participants in the less skilled group, it appears that ultimately 39 students in the treatment group and 14 students in the control group were present for pre-testing and posttesting. The authors did note that one child was excluded from the recall analyses due to equipment failure.	The study reported that five students were assigned to the treatment group, five students were assigned to the poor reader control group, and 10 children were assigned to the typically developing control group. It appears that all students were present for pre-testing and posttesting.
	Outcomes and estimation	Outcome measures included: Reading Comprehension (reading ability, story structure, free recall and prompted recall) and Listening Comprehension (free recall and prompted recall). The means, standard deviations, and p-values for composite scores are provided.	Outcome measures included: Narrative Comprehension task, Expository Comprehension task, Narrative Production task, Expository Production task, Listening Comprehension tasks, Memory Questionnaire, and Effort and Enjoyment Questionnaire. Some variables were excluded from analysis due to ceiling effects, inter-correlation, and leptokurtic distribution of participant performances.	Outcome measures included: free recall, probed recall, note-taking, Reading Concept Inventory, test of error detection and correction.	Outcome measures included: Neale Analysis of Reading Ability, Oral Narrative Production, Oral Narrative Comprehension.

Appendix A: CONSORT Criteria (Law & Plunkett, 2006)

Headings	Subheadings	Garner & Bochna, 2004	Paris & Paris, 2007	Short & Ryan, 1984	Westerveld & Gillon, 2008
	Ancillary analyses	t-test for story element in Tx group	Teacher Ratings and a Literacy Habits Interview were used at pre-testing. A Memory Questionnaire and an Effort and Enjoyment Questionnaire were used at posttesting.	The information subtest of the Wechsler Intelligence Scale for Children-Revised Edition was used to control for any discrepancy between groups in cognitive ability. Also, the comprehension subtest of the Stanford Diagnostic Reading Test was used as pre-testing only.	
	Adverse events	No details for adverse effects.	No details for adverse effects.	No details for adverse effects.	No details for adverse effects.
Discussion	Interpretation	The authors interpret the results in reference to their hypotheses and research questions. They also acknowledge additional areas of research that remain unanswered by their study.	The authors interpret the results in detail, addressing their findings' relevance to the research base they discussed in the introduction. They address and answer each study question in depth and acknowledge the limitations of the study.	The authors interpret the results in reference to their research questions. They also acknowledge the limitations of the study.	The authors interpret the results in reference to their research questions. They also acknowledge the limitations of the study.
	Generalizability	The paper implies generalization of the story grammar instruction to beginning readers when using the same instructional techniques.	The study addresses the question of whether the instructional skills could be generalized. The paper discusses the results and also how to generalize effects in other classroom settings.	The authors call one of the outcome measures "Strategy Generalization" but it appears to be more of a measure of maintenance as it occurs three days later. The authors discuss maintenance and the anecdotal evidence for generalization but also limitations of the interpretation.	The study addresses the question of whether the instructional skills could be generalized to reading comprehension. The authors discuss these results, including the lack of transfer of the skills, and offer some explanation.
	Overall evidence	The findings of the study are discussed relative to the research question that they answered. Suggestions for future research are made in relation to questions that the authors were not able to answer.	The findings of the study are discussed relative to the research question that they answered. Suggestions for future research are made in relation to other populations, ages, and other areas that are lacking a research base.	The findings of the study are discussed relative to the research question that they answered. Suggestions for future research are made in relation to questions that the authors were not able to answer.	The findings of the study are discussed relative to the research question that they answered. However, suggestions for future research are not made.