Macmillan
McGraw-Hill

2 PENN PLAZA
NEW YORK, NY 10121-2298
800-442-9685
WWW.MHSCHOOL.COM

RD06 W 5284
1M 07/06

MACMILLAN/McGRAW-HILL *TREASURES* RESEARCH:

An Exploratory Study of the Effectiveness of Story Retelling as a Strategy to Enhance Reading Comprehension

A study conducted for

Macmillan/McGraw-Hill

by

Westat, Inc.

September 20, 2005

Macmillan McGraw-Hill

<NONE>BACKGROUND

What was the Purpose of the Study?

Macmillan/McGraw-Hill (MMH) has developed a new reading program called *Treasures.* As part of the development, MMH included strategies designed to improve acquisition of student reading comprehension skills. One of the strategies is the provision of retelling cards for grades 1 and 2. These retelling cards will be used to promote the understanding of story sequences and important points in the material the students have read. The purpose of this exploratory study is to examine the efficacy of the retelling component.

Who Was Involved?

The students in the sample were first graders from eight elementary schools, six from Atlanta Public Schools in Georgia and two from the St. Paul School District in Minnesota. Each school had a treatment and a comparison class, and 10 students from each class were chosen to participate in an assessment. Since students' participation in the study was based on their submission of parental consent forms, the sample was somewhat opportunistic. Because of transfers and other reasons, the final number of students receiving both pretests and posttests was 142.

All of the schools except one in Atlanta were located in the inner city. The majority of schools had high percentages of African American students. Data from the prestudy survey indicate that treatment class teachers have 6 to 15 years of experience in teaching. Four teachers had master's degrees; the other four had bachelor's degrees. In addition, six teachers reported using retelling in their previous teaching. All of the teachers have used small group instruction in their regular reading instruction before this study.

What is Retelling?

Retelling is a strategy to enhance comprehension, both the understanding of story sequences and the important points in materials read. In the present study, retelling cards with pictures showing the content of a story were used to support student discussion of what had been read. Six to eight cards were used in each package. On the front side there was a picture copied from a story in an instruction unit. On the back of the card, prompts were provided for teachers to use in guiding and modeling the retelling process.

For each reading unit teachers were expected to exercise three activities. First, they would use the prompts on the retelling cards to *model retelling* on the selection or to guide students' retelling of the selection. This activity usually occurred at the beginning of the unit. Next, they would reread the unit based on students' ideas and ask them if they would like to add any additional information. This is called *guided retelling* and was usually done with small groups of students after modeled retelling. The final activity is *discussing the retelling,* where the teacher asked students a series of more in-depth questions about the reading unit.

For this study, the reading selections used were taken from MMH's reading programs with 1998 and 2003 copyrights, respectively. The prompts on the cards were developed by the MMH editorial staff. Each picture was presented on a 5" x 7" card.

In the sections that follow, we summarize what we learned from this exploratory effort regarding the implementation and impacts of using retelling cards.

IMPLEMENTATION

A critical component of a study of program efficacy is the examination of the extent to which a program or strategy is implemented with fidelity to the program as designed. To understand how teachers were actually using the retelling cards, Westat conducted observations and interviews designed to address the following:

- Do teachers understand how to integrate the retelling component into their existing curricular material?
- Are teachers able to model retelling and effectively transfer retelling skills to their students?

In addition, we solicited teachers' assessments of the utility of the retelling component to determine the following:

- Do teachers feel that adding the retelling component serves to promote increased understanding on the part of their students? In what ways?

The interviews and classroom observations were conducted by Westat staff in April 2005, four to six weeks after the teachers began program implementation. Below is a brief overview of the interview and observation protocols used:

- The interview questions were organized into five main parts. First, teachers were asked eight general questions about their retelling card instruction (e.g., the length of a typical retelling card lesson). Teachers were then asked how they integrated the retelling cards into their schools' existing curricular material (5 questions), how they used different methods of retelling card instruction (modeled retelling, guided retelling, and retelling discussions) and the effectiveness of those methods in helping their students retell (11 questions), and their opinions about the retelling cards impact on their students' comprehension (5 questions). Finally, teachers were asked about the usefulness of the training on retelling cards they received at the end of February (Atlanta) or in early March (St. Paul).
- The observation protocol consisted of two main parts: a detailed description of classroom activities in the retelling lesson, including how students were grouped for instruction, and a summary assessment of the extent to which teachers implemented various elements of retelling cards and reading instruction during the observation.

On the day of the observed retelling lesson, three classes were on their first day of retelling card instruction for a unit, three were on their second day, and one was on its fifth day. The eighth class was described as "other" because the teacher did not use retelling cards the day of the observed lesson; rather, she used modeled and guided retelling on a story that did not have a set of retelling cards, then asked several students to retell the story or book they were reading independently that day. This teacher explained that she was trying something new, but that she had used the retelling cards earlier in the week.

What we heard and saw in the teachers' classrooms is described below.

Findings from the Interviews

How useful was the initial training provided?
While most teachers who attended the training said that they had sufficient instruction and materials to implement retelling cards, at least two teachers were unsure if they were implementing the retelling cards correctly. Teachers also had different understandings of how often they were supposed to implement retelling card instruction and the method they were supposed to use. Two of the teachers thought they were only supposed to model retelling, whereas the other teachers were using a combination of modeled and guided retelling, and to a lesser extent, retelling discussions. Of the seven teachers who were observed teaching with the retelling cards, all implemented the cards using the approaches suggested by the Macmillan/McGraw Hill trainers.[1]

Teachers suggested ways the training could be improved, such as seeing a video of what a typical lesson using retelling cards might look like, having more hands-on exercises, and having trainers come into classrooms and demonstrate a lesson.

[1] *The teacher who expressed the most doubt about her abilities did not teach with the retelling cards on the day of the observation.*

What retelling strategies were the teachers using?
Teachers generally used modeled retelling in whole group instruction first, with most using guided retelling in small group instruction the next day or later in the week. Teachers used retelling discussions less frequently. It seems that they had different understandings of the strategy: some thought it meant discussing the purpose and importance of retelling, while most thought it meant discussing the story more broadly than is done in a guided retelling.

How frequently and for how long were the retelling cards used during a week?
Four of the eight teachers said that they typically implement retelling card instruction two to three times a week, while two teachers said they use the cards once a week and two said they use the cards five times a week. About half of the teachers said they typically spend 15 to 20 minutes on a retelling card lesson, with two teachers reporting that they typically spend more than 20 minutes. One teacher explained that it depends on the method of retelling that she is using in a given lesson: while her typical modeled retelling card lesson is 15 minutes long, a guided retelling lesson can last up to 35 minutes.

How were the retelling cards integrated into existing curricular material?
All of the teachers integrated the retelling cards into their existing curricular materials. Many of them said it was easy to do so because they were already using retelling in their regular teaching. Implementing the retelling cards, however, allowed them to set a block of time aside more formally. In addition, they supplemented the guided retelling questions with comprehension questions of their own.

How frequently did teachers model retelling?
Modeled retelling most frequently occurred at the beginning of a story, but the number of times that a teacher modeled retelling with the cards varied. For example, two teachers used modeled retelling one session per book/story, while two other teachers modeled retelling every day, five times a week. One teacher modeled retelling the first two or three weeks after the pretesting, but said she's "fallen away from that" due to other things on her schedule such as testing. Most teachers noted that it took a few tries for them to engage students in the activity. One teacher's experience may be typical of this process:

- In the beginning students had difficulty retelling the story, rather than focusing on the cards. It "takes a lot of prompting to get them to retell the story" because some are focused on the pictures. At times, "it was like pulling teeth," trying to get students to "think better about the idea around the card." Things have gotten better with time. Even "some who are hesitant [to participate] have gotten a lot better." One student never would have participated before, but the cards have given her a way to be involved and feel more confident participating.

How was modeled retelling carried out?
Typically, teachers tried to use what was on the back of the cards to model retelling in the beginning. As time progressed, some of them used more of their own words, developed their own questions, and provided prompts at different times during a story. Instead of a straight retelling, teachers would often diverge to provide background information, for example, the meaning of a word, or to ask students questions about the story or vocabulary. One teacher began retelling instruction with visual aids and think-alouds before modeling the retelling.

How did teachers use guided reading?
With respect to guided retelling, teachers seemed to use a similar approach, which involved a combination of using the questions provided on the retelling cards and their own questions. Some teachers noted that they try to ensure that each student is provided an opportunity to retell some aspect of the story. In the interviews, most teachers said that they follow up modeled retelling with guided retelling in small groups, although during the observed lessons, guided retelling was used in whole group instruction. This suggests that teachers grouped their students in multiple ways for guided retelling.

How was retelling discussion used?
Teachers took different approaches in retelling discussions, and only one teacher said she did not use any retelling discussion with her students. One teacher said she talks to her students about genre (e.g., fiction), the main characters, the setting, the plot, and whether the scenario could have happened in real life. Sometimes she asks the students to make comparisons with other stories as well. Another teacher said she talks to her students about different elements of the story, and they are expected to be able to tell her the names of the author and the illustrator. Two more teachers said that they emphasize the sequence of a story and important events that happened in it, and they discuss the author's purpose with such aids as story maps. Teachers who did use retelling discussion said that the strategy enlightens the students and helps them know the story, recall information, and think about the story in sequence.

What was the best time to teach retelling?
In general, most teachers considered the time immediately following the first reading of a story to be the best time to teach retelling. As one teacher stated, "Once you model it, kids should be able to do it because you've shown them the procedure." Another teacher said that by the third day of a story and modeled retelling, she considers her students to be familiar enough with the story to try some of their own retelling. However, some teachers reported that while some students need modeled retelling only at the beginning, others need it all the time, and they adjust their instruction accordingly.

How did students react to the retelling cards?
Teachers reported that students reacted very positively to retelling cards. As mentioned earlier, it took a while for teachers to engage students in the activity, but after students were familiar with the routine, they better understood and enjoyed using the cards.

How did the retelling card instruction compare with other types of reading comprehension instruction?
Teachers responded that retelling card instruction is similar to instruction that calls for students to sequence events in a story. But while some teachers said they have used retelling strategies before, they also noted that their usage was less systematic or structured. They also said that the retelling cards provided in this study asked students for more detail and gave them more opportunities to speak and better understand the story.

Specifically, teachers named several strengths of the retelling cards:

- There is no text on the pictures, so students must talk about the story in their own words.
- They help students sequence events or ideas in a story.
- The questions that accompany the pictures are good at helping students think about the important story details and retell the story.
- Use of the cards helps improve students' reading comprehension.
- They provide an opportunity for students to express themselves and do so on a regular basis.
- Use of visual aids, such as the cards, is very important in student learning.

The most common weakness of the cards mentioned by teachers was their size; they were described as being too small for whole group instruction, a feature already revised in the program. In addition, one teacher mentioned that children could be "destructive" and damage the cards. She also mentioned some difficulties in handling and coordinating the cards.

To what extent did the retelling component work differently with students of different ability levels? Did the effects differ?
All of the teachers reported that the retelling cards helped students at all levels of reading ability. Teachers said that the retelling cards helped higher-skilled students provide more details from the story. The cards taught lower-skilled students how to get more information from the story, understand the story better, get more involved with the thinking process, and catch up to the progress of other students. One teacher also noted that for students who struggle with reading, the pictures help them make sense of the story they have read.

With respect to English Language Learners (ELL), one teacher who had a high proportion of ELL students in her class said the cards helped them stay on the same story as her other students. As an example, she mentioned a story that her class had just read called *On the Go.* She said the vocabulary is too difficult for her ELL kids, but the retelling cards gave them "a way to understand the story… I have kids participating that normally wouldn't." She mentioned one student in particular and said, "Give her a picture and she can tell you exactly what's going on." She described the "visual triggers" of the cards as very helpful.

Recommendations

Teachers also provided many thoughtful and practical recommendations to improve the retelling cards, including the following:

- Design some type of stand for the retelling cards so that teachers can display them in a line at eye level, and provide students with easy access to them.
- Develop a larger set of cards in addition to the smaller set.
- Provide higher-level questions such as compare/contrast, open-ended questions, and inferential questions.
- Number the cards.
- Make audiotapes to accompany the cards to which students could role play.
- Provide more-detailed prompts.
- Provide pre-story instruction, such as new vocabulary.
- Make cards as kid friendly as possible. For example, organize them in a retelling minibook or binder from which the cards can be removed.
- Have a more interactive training on retelling card instruction. For example, have trainers model retelling with students or develop a videotape of such modeled retelling.

Observations

The observations allowed us to get a firsthand look at how teachers were actually implementing the retelling component. Teachers were asked to schedule the observation on a day when they would be doing a retelling card lesson, either in whole group or small group, using modeled retelling, guided retelling, retelling discussions, or some combination of the three. The two observers were project staff who had designed the study, developed the data collection instruments, and attended the retelling card training in both districts. As a result, they were familiar with retelling card instruction and how teachers should be implementing it.

During the lesson, the observers filled out a detailed description of the lesson, which included how students were grouped for instruction. After the lesson, they completed an observation summary, which addressed student engagement, the amount of time spent in various instructional activities, and different topics covered by the retelling cards (e.g., main idea and setting) using the lesson description as supporting evidence.

What was the class like and for how long did retelling instruction take place?

The average number of students per class was 15, ranging from a low of 8 students[2] to a high of 18. Four of the seven teachers were on their fourth week of using the retelling cards at the time of the observation, with interruptions such as spring break and testing weeks preventing the use of the cards every week. One teacher was on her third week of using the retelling cards, while another had completed all six sets. In seven of the eight observed classrooms, retelling card instruction ranged from a low of 12 minutes, which was strictly modeled retelling, to a high of 50 minutes, which included choral reading, teacher-directed comprehension checks, and student retelling.

What types of groupings were observed?

Ninety-one percent of the observed retelling card instruction was held as whole group, with the remaining 9 percent held as individual seatwork, in which the teacher asked students to write a sentence for each card recounting the story. The observed instructional time included both modeled and guided retelling. This approach differs somewhat from the training teachers received. Based on what was presented in training, we expected that teachers would use modeled retelling in whole group instruction and guided retelling in small group instruction. Some teachers indicated in their interviews that they *do* use guided retelling in small groups, but not all the time. For example, one teacher explained that she did not use guided retelling with her students because she thought teachers were instructed to use modeled retelling only for the first six weeks of retelling card implementation.

[2] *The whole class retelling instruction involved 14 students initially, but 6 of those students left to attend an ELL class. The remaining 8 students then worked individually on a writing exercise involving the retelling cards.*

What instructional practices were being used?
We observed both teachers and students using retelling. While 29 percent of the observed retelling card instruction was spent with teachers modeling retelling, nearly half of the instruction (47 percent) was spent with students doing some level of retelling themselves. The remaining time was primarily spent on reading the story that would be used for the retelling (23 percent). Included in the latter group were time spent on the teacher reading the story (9 percent), the students reading the story silently (4 percent), the students reading in turns (1 percent), and the students reading as a group (9 percent). Only 3 percent of the observed time was spent on activities other than instruction.

How well were key features of the program being implemented?
The observation scale included a summary form used to rate the adequacy or quality of features identified as being key to program implementation.[3] After each observation, evaluators recorded their ratings on specific topics that teachers may or may not have covered during the retelling instruction. Westat evaluators held several discussions both before and after the observations in an effort to ensure the appropriate and consistent interpretation and application of items in the observation summary. When necessary, changes in coding were made based on these discussions and the detailed descriptive information recorded in the protocol. It is important to note that teachers were not asked or expected to cover several of the items in the summary such as author's purpose and cause and effect relationships. That said, teachers often did cover these topics as a way to improve their students' comprehension, and it is interesting to note that at least one teacher recommended that the retelling cards include questions that address such topics as compare and contrast and "why" questions.

Table 1 presents findings from the observations on prompted and unprompted student responses, student and teacher activities, and teachers' overall implementation of retelling card instruction. The section on prompted student responses, which addresses how well students covered topics in response to teacher-directed questions, indicates that the items teachers covered the most were those emphasized by the retelling cards. These include main idea, average rating 3.43; sequence of ideas or events, 3.29; and important details from the story, 3.14. These were also items that students were most likely to cover on their own, without prompting from their teachers (each with an average rating of 2.00). Given that teachers were only observed once, it is not possible to say whether the observed retelling card lesson was a typical one for them; however, the results seem to indicate that students are better able to address (with and without prompting) those topics emphasized by their teachers.

Six of the seven teachers used guided retelling in their own words during the observed lessons, and it is not surprising that students were able to cover topics much more effectively (or at all) when prompted by their teachers. In both modeled and guided retelling, teachers generally used a combination of the prompts as they were written on the backs of the retelling cards and their own words. In the interviews, some teachers explained that the language used on the retelling cards was too formal for their classrooms, and rephrasing the modeled retelling and the prompts using language more informal and familiar to their kids was more effective.

Most of the teachers strategically used modeled retelling, guided retelling, and retelling discussion with a several-day arc of instruction in mind. Five of the seven observed lessons showed teachers integrating the retelling cards with reading instruction. For example, some teachers had students retell as they went through the story, others after the story was read in entirety, and all teachers posed additional comprehension-oriented questions apart from those on the retelling cards during reading and retelling card instruction. In general, teachers used modeled retelling during the first retelling card lesson of the week then moved toward guided retelling and retelling discussions in the next lesson(s). As a result, a rating of "not at all" may reflect where the lesson falls along this arc of retelling card instruction, or that a teacher does not cover a particular topic or do a particular activity. For example, in three classes, only the teachers' use of one retelling method (e.g., modeled or guided) was observed, even though two of those teachers use multiple retelling card strategies during the week. More hypothetically, if a teacher was on her second or third retelling card lesson, and the story had been read several times leading into the lesson, it may no longer have been necessary to cover certain topics such introducing or reviewing story-related vocabulary words.

[3] *The summary was developed by Westat staff based on discussions with MMH personnel. It is thus "quality" as perceived through Westat's understanding of the program and its key features.*

Table 1: **Item Ratings from the Retelling Card Observation Summary**

Item	Average Rating (1-4)	Frequency (Number of Classes) Not At All (1)	Slight Extent (2)	Moderate Extent (3)	Great Extent (4)
Overall rating on implementation					
Successful implementation of retelling instruction[1]	3.29	0	0	5	2
Unprompted student responses					
Main idea	2.00	2	3	2	0
Sequence of ideas or events	2.00	2	3	2	0
Important details (e.g., setting)	2.00	2	3	2	0
Conclusions and inferences	1.57	4	2	1	0
Author's purpose	1.43	5	1	1	0
Connections to students' lives or other stories	1.57	3	4	0	0
Comparisons	1.29	5	2	0	0
Cause and effect relationships	1.14	6	1	0	0
Prediction of outcomes and consequences	1.43	5	1	1	0
Prompted student responses					
Main idea	3.43	0	1	2	4
Sequence of ideas or events	3.29	1	0	2	4
Important details (e.g., setting)	3.14	1	1	1	4
Conclusions and inferences	2.86	1	1	3	2
Author's purpose	2.00	4	0	2	1
Connections to students' lives or other stories	2.57	1	2	3	1
Comparisons	2.00	1	5	1	0
Cause and effect relationships	2.00	2	4	0	1
Prediction of outcomes and consequences	1.71	3	3	1	0
Teacher activities					
Modeled retelling	2.14	3	1	2	1
Guided retelling as written on the retelling cards	2.71	2	0	3	2
Guided retelling in their own words	2.86	1	1	3	2
Involved all students equitably	3.57	0	0	3	4
Introduced new vocabulary	1.43	5	1	1	0
Reviewed previously introduced vocabulary	1.71	4	1	2	0
Asked questions calling for student predictions about the story	2.57	2	1	2	2
Made connections to students' prior experiences	2.57	1	2	3	1
Asked students to do a written retelling exercise	1.43	6	1	0	0
Engage students in a discussion of retelling	2.00	4	2	1	0
Assist students with words in the story they did not know	1.43	5	1	1	0
Student activities					
Were engaged in the story	3.57	0	0	3	4
Covered the main idea	2.00	2	3	2	0
Covered the essential story elements	2.86	0	2	4	1
Asked questions about the story	1.43	4	3	0	0
Made connections to the story with little or no prompting	1.57	3	4	0	0
Made connections to the story with prompting	2.29	2	1	4	0
Used manipulatives (e.g., graphic organizers)	1.43	6	1	0	0
Compared ideas and themes	2.00	2	3	2	0
Retold card elements more than story elements	2.00	2	3	2	0
Needed assistance with words in the story they did not know	1.57	4	2	1	0
Raised cause and effect relationships	1.00	7	0	0	0
Made predictions about the story	1.57	4	2	1	0

[1] *For this item, a rating of 4 represents teachers who really took ownership of the retelling cards and effectively adapted them for their students; a rating of 3 represents teachers who implemented the cards well, but in a more basic, standard way. None of the seven teachers who used the retelling cards during an observed lesson were in need of significant improvement in terms of their implementation.*

NOTE: This table shows results from the seven classrooms in which the retelling cards were used. Ratings indicate the extent to which various items were covered.

With respect to the eighth observed lesson, which did not include the use of retelling cards, the teacher did two general retelling lessons. During the first lesson, she read a story to her students, asking them numerous comprehension questions throughout. After the story was finished, the teacher had students retell the story, calling on different students to fill in different story elements. Students were extremely engaged during this lesson, with most eager to participate in the retelling. After this lesson, students were directed to work independently or in small groups on a reading-related activity. Reading time was concluded with an additional lesson, in which several students were asked to retell the stories that had just been read. Students struggled more in this lesson, and the teacher explained that she was trying something new with them. She explained that she had planned to use the retelling cards during the observed lesson, but her class was starting a new story and she was concerned they might have difficulty with it because it was longer than the usual stories.

Unlike the guided retelling she used during the observed lessons, this teacher noted in her interview that she had only used modeled retelling with the cards. It was clear during the first retelling lesson that her students were familiar with retelling, and she noted during her interview that retelling is one of the standards students must meet at her school, so the cards were easy to incorporate into what she was already doing. However, she also expressed considerable doubt about whether she was implementing the cards correctly, was somewhat unfamiliar with the guided prompts on the back of the cards, and stated that other instructional priorities (e.g., testing) had pushed retelling card instruction on a back burner.

Were students engaged during the retelling component?

Students' engagement level during the observed retelling card lessons corresponds with the positive reaction of students to the cards described by teachers in their interviews. All observed lessons were judged to have shown a medium to high level of student engagement. In three classes, students were generally participating in the lesson, following the content, and completing their assignments (medium level of engagement). In the remaining four classes, students were actively participating in the lesson, seeking out knowledge, and making connections with prior lessons (high level of engagement).

CONCLUSIONS

Our interviews and observations indicate that teachers are enthusiastic about the retelling component and feel that it is beneficial to their students. Use of the retelling cards clearly engaged the students and, according to their teachers, was particularly useful with lower-ability and ELL students.

The seven teachers who were observed teaching a retelling card lesson did it well. They were comfortable with the retelling routine, equitably involved their students in the lesson, which kept them engaged, and covered the content of the cards, in some cases, adding their own comprehension questions as well. Nonetheless, teachers did have some different understandings about how the cards were to be used, with at least two teachers saying they had only used modeled retelling because they thought that was what they were instructed to do at the training they attended. It should be recognized that these teachers had only a limited amount of training before using the cards and did not have materials to refresh and support their understanding as the retelling component was actually implemented.

Some teachers adapted the use of the cards in creative ways, e.g., using graphic organizers, having students sequence cards in addition to retelling, etc., while others followed the modeled and guided retelling more closely. The teachers offered a number of suggestions about how an already useful strategy could be made even more effective.

IMPACTS

Overview of the Assessment

The study provided a preliminary look at the impacts of the retelling card component on students' early literacy skills. While we had reservations about the possible impact of the component, given the short time that teachers were able to use the cards, we examined changes in student performance as part of this exploratory effort.

An exploratory assessment approach was used to examine the impacts of retelling cards on students' literacy skills. Two of the measures are commonly used to examine early comprehension skills. The third measure was a study-developed assessment of ability in retelling per se. The instruments were

- Section A: Woodcock-Johnson III Test of Achievement, Story Recall, (WJ-Sec. A),
- Section B: Woodcock-Johnson III Test of Achievement, Passage Comprehension, (WJ-Sec. B), and
- Section C: Retelling Card Assessment (RC).

Section A of the Woodcock-Johnson III consists of two stories, in which students listened to a two to three sentence tape-recorded story and retold it back to the assessor. Each story is divided into several eleme-nts, with one element worth one score point. Section B of the Woodcock-Johnson III involved six passage comprehension items in which students were asked to read a word (or words) and point to the picture that depicts the word(s). A student earned one point when he or she made the correct connection of words with the picture. These same eight items from the Woodcock-Johnson III were administered to students for both the pre- and posttests.

For the third section of the assessment, students were asked to first read a book and then answer a series of questions based on the retelling cards presented to them. The retelling cards feature the same pictures as those in the story book. A fiction book, *Buying a New House,* was used for the pretest, and a nonfiction book, *Harvest Mouse,* was the selection for the posttest. For each retelling card, two types of questions were asked: a main question and two to four subsequent prompts. Main questions were generally phrased as, "What is happening in this part of the story?" Prompts asked for more specific information about the content of the specific stories, for example, "What creatures hunt harvest mice?" MMH subject experts selected the story and developed the prompts. Westat developed the procedures and scoring guides based on discussions with MMH and other experts in the use of retelling cards.

How were the data analyzed?

Results showed that the Woodcock tests were too easy for the students. With the ceiling effect that occurred at pretest, there was little opportunity to show growth over time. Our discussion of prepost differences therefore focuses on the retelling assessment.

Our first step was to look at the outcomes of the retelling assessment for all students to examine general performance and the properties of the test.

- Separate item calibrations were run for the pre- and the posttests. We wanted to see if they measured students' reading abilities consistently and could be used to rank item difficulty. Items that were common in both tests were also used to link the two tests by adjusting their item difficulties.
- Having successfully linked the two assessments, we then looked at student performance on the main questions and the questions with guided prompts. We examined estimates of student performance in terms of standard error of measurement (SEM). The smaller the error, the better the estimate of student ability and the more information we can obtain about a student at a certain ability level.

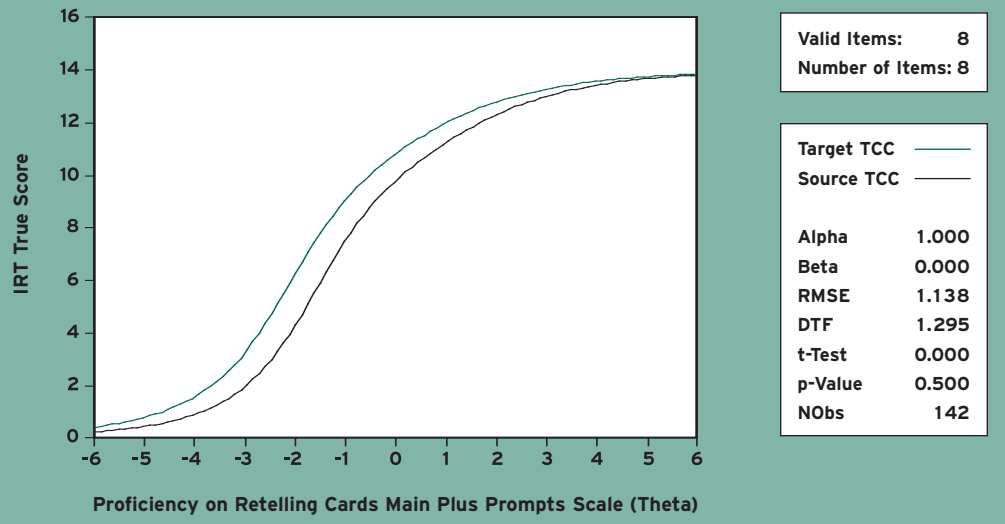
Once these analyses were completed and overall change from pretest to posttest was examined, we calculated the means and standard deviations (SD) for the two groups at the two testing times and calculated their effect sizes for individual schools and for the aggregated groups. To compare the growth rate from the pretest to the posttest between the treatment and control groups, we used Hedge's g_u, an unbiased estimate of effect size for small sample sizes to measure the change differences.

How were the pretest and the posttest data compared?

In order to compare the change from the pretest to the posttest for all students, we examined whether the pre- and posttests yield consistent and reliable ability estimates in this particular population. We addressed this issue by determining whether or not item responses in the sample are sufficiently consistent to effectively permit the ranking of items and of students. Our separate item calibration used the item response theory 2-parameter model (IRT-2PL). We found that both tests do scale well in this sample. Item response functions can be estimated, even with this relatively small sample. Our calculations yielded average standard errors of measurement as small as 0.25 of a population standard deviation across all levels of ability and reliability estimates as high as $r_{xx} = 0.90$, which are generally considered appropriate for clinical purposes.

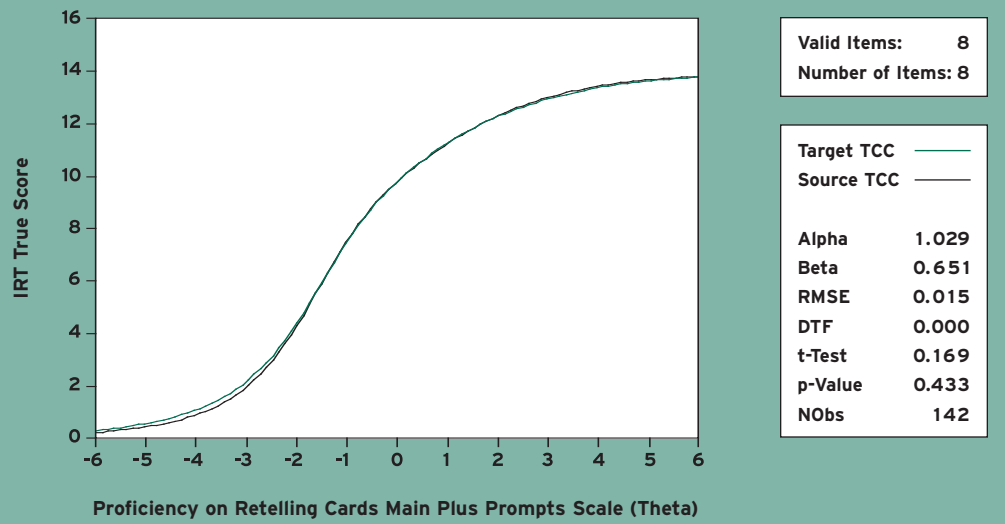
Having satisfactorily scaled both the pretest and posttest, we then placed ability estimates on a consistent scale metric to compare student reading abilities. We used the eight Woodcock-Johnson items in Sections A and B that appeared on both tests as the anchor items to link the pre- and the posttests. This step enabled us to calibrate both tests on a common scale metric for the purpose of comparison. The item response functions for these anchor items are used to carry out IRT true-score linking. This procedure allows the linear transformation of origin (intercept) and scale (slope) that the test characteristic curves on the two tests align. True-score linking in this case was quite precise, leaving weighted mean discrepancy of only root mean square error = 0.15 raw score points across the entire ability range. Coefficients from this linear transformation were used to transform posttest ability estimates so that they line up with the pretest scale metric. With this transformation, ability estimates from the two assessments are directly comparable (Figures 1 and 2).

Figure 1: **Pretest and Posttest Characteristic Curves before Linking**



NOTE: TCC = test characteristic curve. It is the average of item response functions for the n items in a test. The TCC expresses the monotonic relationship between students and their performance in the test.

Figure 2: **Pretest and Posttest Characteristic Curves after Linking**



NOTE: TCC = test characteristic curve. It is the average of item response functions for the n items in a test. The TCC expresses the monotonic relationship between students and their performance in the test.

Figures 3 and 4 show the standard error of estimates (SEM) for students' proficiency in reading comprehension for the pretest based on the responses to the **main questions only** and to the **main questions with prompts**. As noted above, the proficiency scores presented here were standardized. We found that SEMs were related to the ability level of the student. We obtained the smallest SEM for the most able students; the SEMs for average or less able students were much higher. Since Table 2 shows that this latter group makes up the majority of the students (about 90 percent), and the greater the SEM the poorer the estimation of student reading ability, the use of the responses from the main questions only is problematic.

Fortunately, when we examined the data using the prompts, the situation improved. When we included student responses to prompts in the analysis, we found that the SEMs decreased considerably from .65 to less than .30 for the average and less able students. This means that we could obtain much more precise estimates of these students' levels of retelling proficiency.

Figure 3: **Standard Error of Measurement for Retelling Cards Section of the Pretest, Main Questions Only**

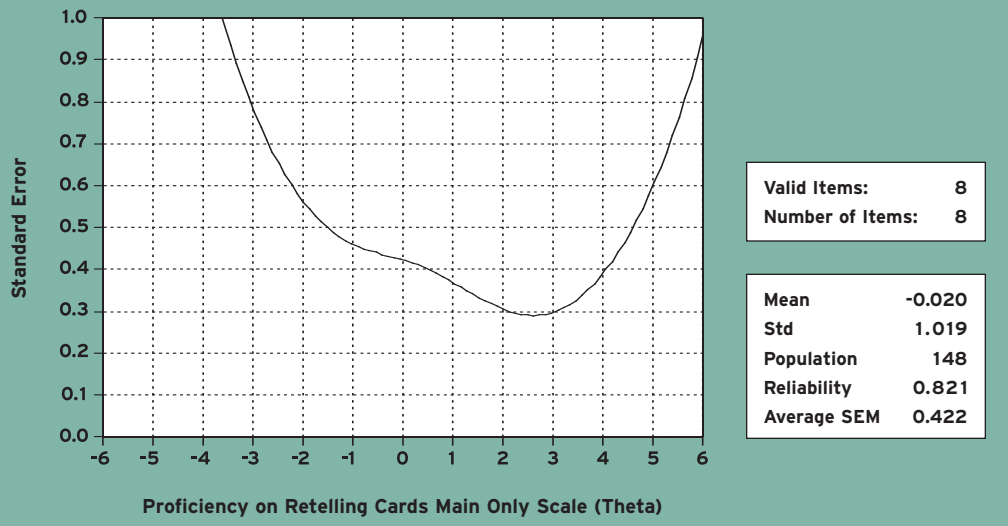
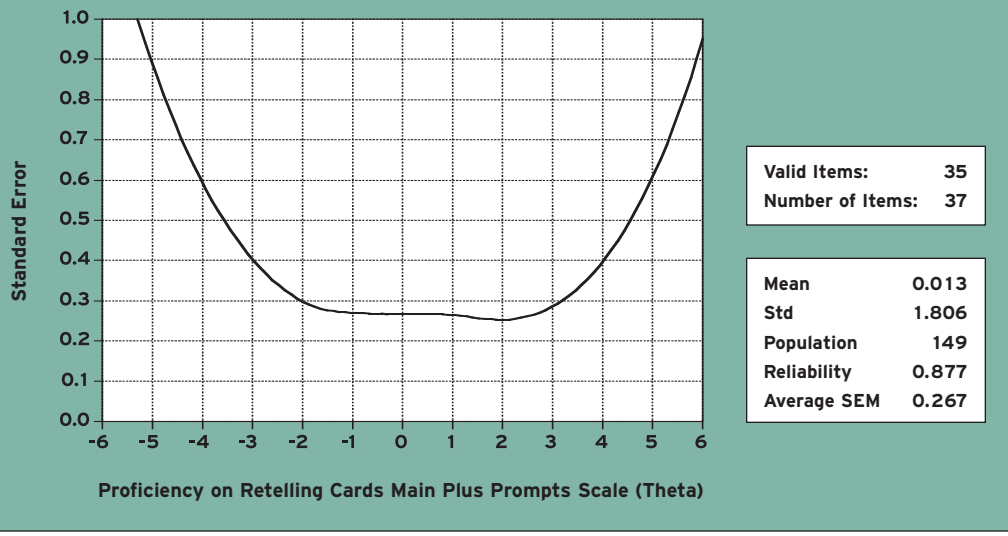


Figure 4: **Standard Error of Measurement for Retelling Cards Section of the Pretest, Main Questions and Prompts**



The above finding is also true for the analysis of the posttest results. Figures 5 and 6 again show that if we only take the student responses to the main questions of retelling cards, the SEMs for more able students in the posttest were much smaller than that for average and less able students.

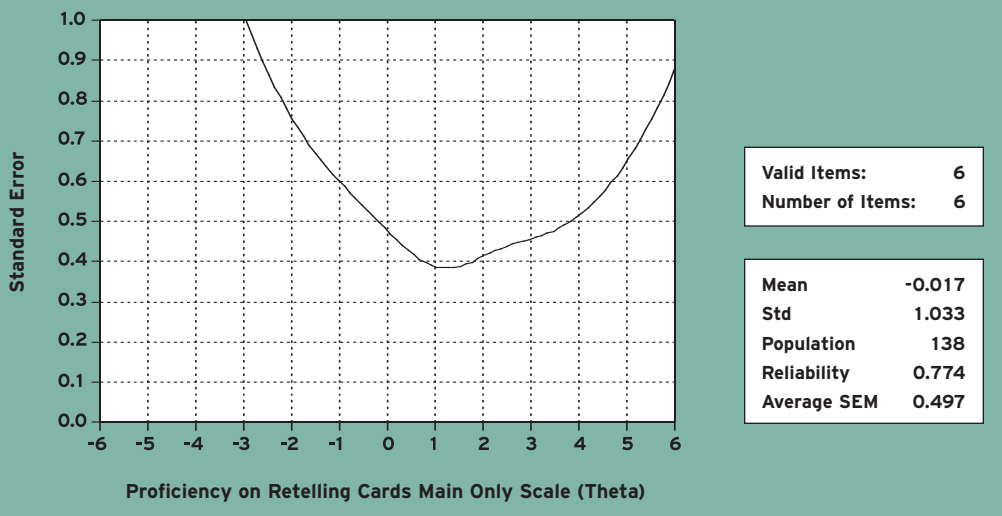
If we considered student responses to the main questions with accompanying prompts (Figure 6), we decrease the SEM to less than .30 and expand the range of more precise estimates of ability scores from -2 to +2.

The reliability estimates of the assessments also increased when the analyses included student responses to the prompts. The reliability for the pretest was .877 with a standard error of .262 when students were prompted compared to a reliability of .821 with a standard error of .408 without prompts. The reliability for the posttest was .892 with a standard error of .256 with prompts, compared to a reliability of .770 with an average standard error of .484 without prompts.

Table 2: **Range of Standardized Ability Scores for the Pre- and Posttests**

Pretest		Posttest	
Range	**Number of Students**	**Range**	**Number of Students**
-2.000 to -1.000	9	-1.300 to -1.000	1
-0.999 to 0.000	65	-0.999 to 0.000	30
0.001 to 1.000	65	0.001 to 1.000	67
1.001 to 2.000	10	1.001 to 2.000	40
		2.001 to 2.500	4

Figure 5: **Standard Error of Measurement for Retelling Cards Section of the Posttest, Main Questions Only**



What was change in performance from pre- to posttest for the students overall?
Having determined that the data from the assessment using prompts provides the best estimate of change over time, we then calculated pre-post differences for the entire sample. Figure 7 shows the estimated growth for all students from the pretest to the posttest in reading comprehension proficiency. The first curve represents the student proficiency distribution for the pretest, and the second curve represents the distribution for the posttest. Both are expressed in standardized scores. Since raw scores are not reliable, we had to calibrate the students' item responses to produce derived scores that would not fluctuate too freely from sample to sample or from one set of items to another. We ran the item calibrations with the 2-parameter model of item response theory (IRT-2PL). The derived scores generated from these calibrations were standardized with the range from -∞ to +∞. Practically, however, the estimates of student abilities run from -5 to +5. The figure shows that on average, a student's proficiency in retelling increased .637 of a standard unit from the pretest (mean=.013, SD=.806) to the posttest (mean=.65, SD=.877). This difference between the pretest and the posttest is statistically significant.

Figure 6: **Standard Error of Measurement for Retelling Cards Section of the Posttest, Main Questions, and Prompts**

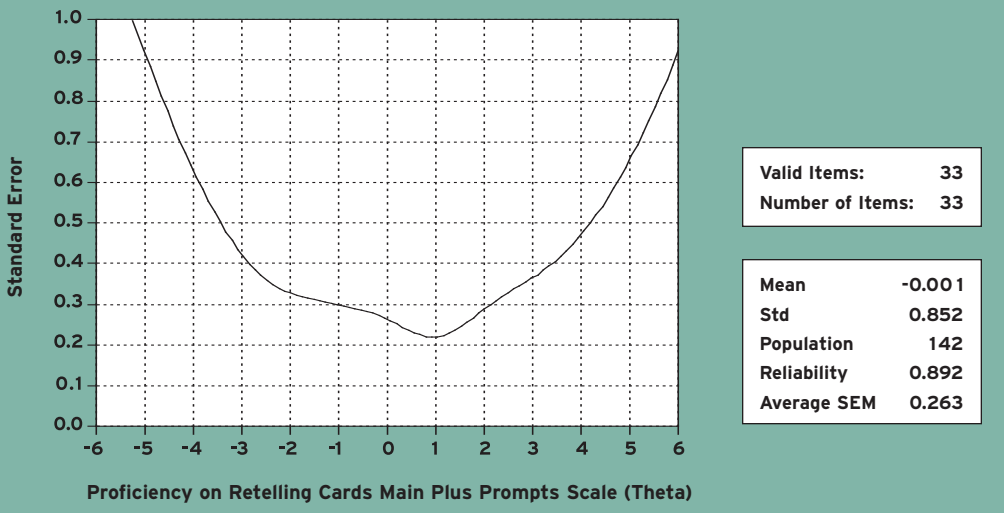
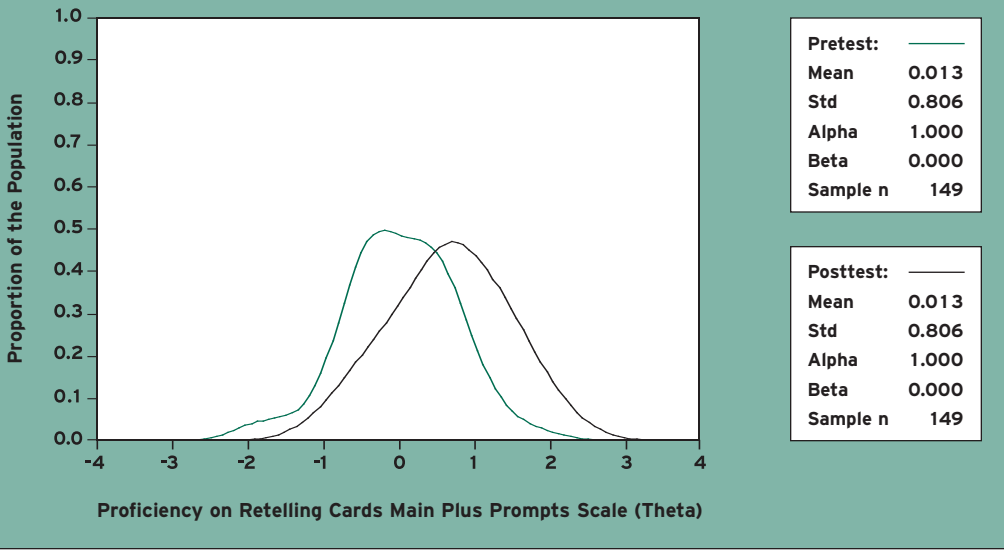


Figure 7: **Pre- and Posttest Distributions of Student Proficiency in Reading Comprehension**



Does the change in performance from pre- to posttest differ for the treatment and comparison students? Table 3 displays the pretest and the posttest standard mean scores for the treatment students and comparison students in individual schools and as aggregated groups. The results show that overall, students in each group improved in retelling proficiency from the pretest to the posttest, with the treatment group having slightly higher mean scores at both times.

Figures 8 through 10 show graphical representations of the comparisons between treatment and control students. The differences between the two groups of students on a school-by-school basis are shown first for the pretest (Figure 8) then for the posttest (Figure 9). As shown, the treatment and control classes were very similar at both pre- and posttest in some schools; in other schools, there were large deviations. By and large, the intervention did not change their positions.

Table 3: **Standardized Means, Standard Deviations, and Confidence Intervals between Treatment and Control Groups from Pretest to Posttest**

	Pretest									
	Control Pretest					Treatment Pretest				
School ID	N	Pretest Mean	Pretest SD	Lower CI	Upper CI	N	Pretest Mean	Pretest SD	Lower CI	Upper CI
12	10	0.224	0.469	-0.713	1.161	10	0.153	0.433	-0.714	1.019
13	6	0.254	1.019	-1.785	2.292	6	-0.148	1.035	-2.217	1.921
14	8	-0.078	0.723	-1.523	1.367	10	0.203	0.358	-0.513	0.920
15	10	0.252	0.484	-0.715	1.220	10	0.077	0.706	-1.335	1.488
16	8	-0.438	0.450	-1.337	0.461	8	-0.256	0.480	-1.216	0.703
17	10	-0.344	0.853	-2.050	1.363	10	-0.329	0.189	-0.708	0.050
18	9	-0.679	0.965	-2.608	1.250	8	0.236	0.667	-1.097	1.569
19	9	0.646	0.584	-0.521	1.814	10	-0.041	0.835	-1.712	1.629
Total	70	-0.020	0.693	-1.407	1.366	72	-0.013	0.588	-1.189	1.162
	Posttest									
12	10	0.853	0.368	0.118	1.589	10	1.260	0.505	0.250	2.270
13	6	0.745	0.638	-0.530	2.021	6	0.391	1.440	-2.490	3.272
14	8	0.497	0.812	-1.127	2.122	10	1.106	0.626	-0.145	2.358
15	10	0.829	0.693	-0.557	2.215	10	0.551	0.491	-0.432	1.533
16	8	0.024	0.374	-0.724	0.772	8	0.125	0.293	-0.460	0.710
17	10	0.224	0.785	-1.346	1.795	10	0.211	0.501	-0.791	1.213
18	9	0.147	0.722	-1.296	1.590	8	0.573	0.855	-1.136	2.283
19	9	1.374	0.471	0.431	2.316	10	0.783	0.954	-1.126	2.691
Total	70	0.587	0.608	-0.629	1.802	72	0.625	0.708	-0.791	2.041

CI = confidence interval; SD= standard deviation.

Figure 8: **Comparison of Pretest Means among Treatment and Comparison Classes**

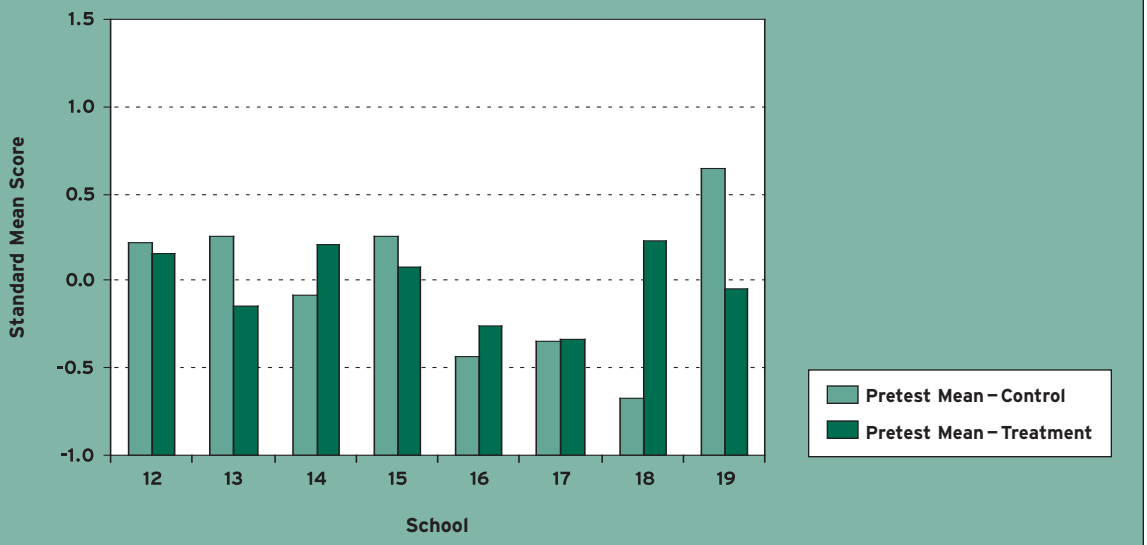


Figure 9: **Comparison of Posttest Means between Treatment and Comparison Classes**

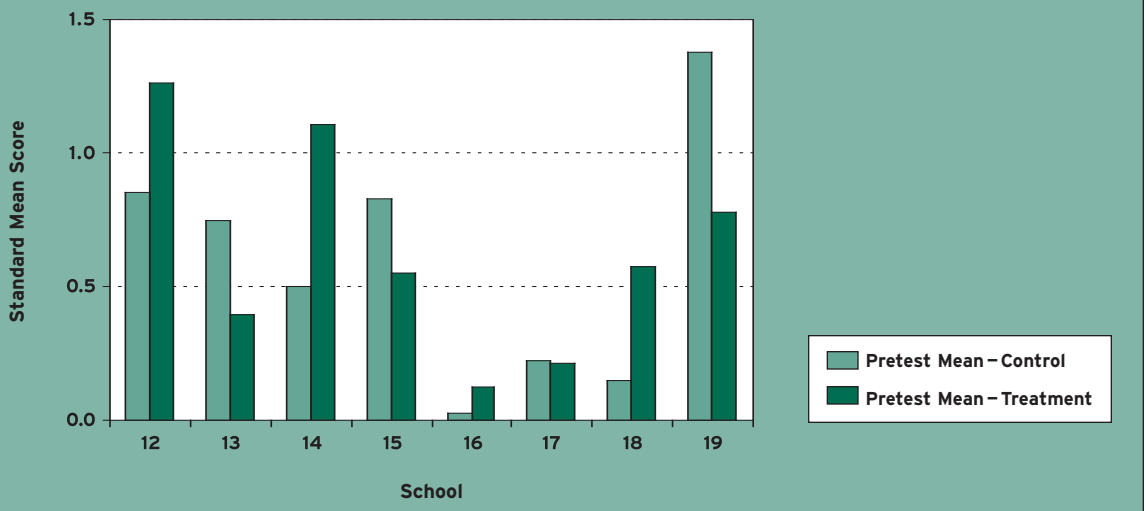


Figure 10: **Change of Mean Scores between the Treatment and Control Groups from the Pretest to the Posttest**

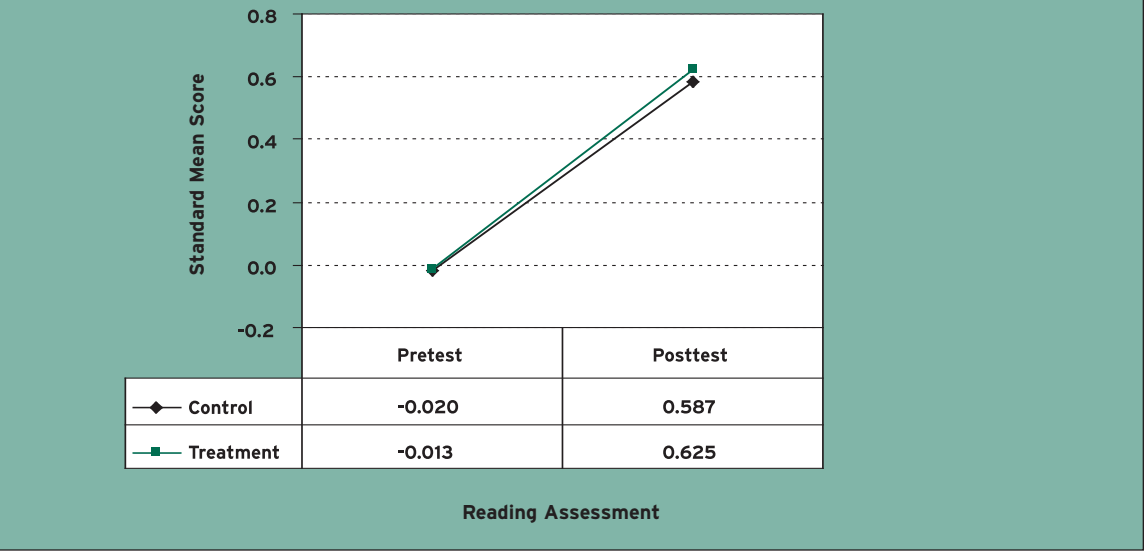


Figure 10 shows the pre- and posttest difference between the treatment and control students for all schools. Average performance for the two groups is very similar on the pretest; the difference between the two groups, while small, is larger at posttest.

To estimate how large the growth from the pretest to the posttest was for individual schools and for all schools overall between the treatment and the control groups, we calculated the effect sizes of the mean score differences. Specifically, we used Hedge's g_u, because this statistic is an unbiased estimate of effect size for small sample sizes.

The values in Table 4 show the differences between treatment and control mean growth in retelling proficiency from the pre- to the posttest, adjusted for test difficulty. The total effect size estimate was weighted for different numbers of participating students in each class. The lower and upper limits of the confidence intervals refer to the 95 percent confidence intervals for the unbiased estimates of effect sizes.

The analyses at the individual school level show a very mixed picture. There are negative effect sizes in four schools (schools 18, 16, 15, 17), indicating that growth for those four treatment classes was less than growth in the control classes in their schools. There are positive effective sizes in the remaining four schools (schools 13, 19, 14, 12), where the treatment classes outperformed the control classes in their schools in terms of growth between the pretest and posttest. The confidence intervals of the effect size estimates show that no individual school has a statistically significant effect size.[4]

When we examined the effect size for all schools combined, we find that that the treatment group as a whole outperforms the comparison group and that the effect size is significant. Practically speaking, however, the difference between the two groups is relatively small. The size of the impact found implies that 50 more students out of 1,000 would show a positive change in their reading comprehension if their teachers implemented retelling cards instruction over this short period of time.

Table 4: **Effect Sizes of Mean Differences and their Confidence Intervals**

School ID	Effect Size Estimate of Mean Score Change	Lower Limit of Confidence Interval	Upper Limit of Confidence Interval
18	-.657	-1.635	0.3204
16	-.200	-1.182	0.7826
15	-.147	-1.025	0.7308
17	-.063	-.9392	0.8142
13	0.054	-1.078	1.1858
19	0.138	-.7635	1.0398
14	0.384	-.5538	1.3224
12	0.809	-.1025	1.7209
Total	0.050	0.0398	0.0611

[4] *The confidence intervals show whether an effect size between a treatment class and a control class within a school is significant or not from the pre- to the post-test. A confidence interval that does not contain zero indicates that the effect size is significant.*

What have we learned from the examination of changes in treatment and comparison group scores? The results overall provide modest encouragement that the retelling cards are a beneficial instructional technique. While the overall differences are small, it must be remembered that our sample size is relatively small, and, perhaps even more important, the length of time that the retelling cards were used was very limited.

The individual school-level data require more exploration. While differences were not significant, the direction of change, with half favoring the comparison students and half favoring the treatment students, raises some questions. We do not have enough information about the individual schools to sort out whether there are some systematic reasons for the patterns observed or whether they are simply random variation. Clearly, in future studies it would be important to gather more data about the context of use of the retelling cards and what else is occurring in reading instruction in both the treatment and control classes.

CONCLUSIONS

This exploratory study has provided some important information about retelling cards, their use, and the assessment of student performance. Discussions with teachers and observations of instruction show that teachers find the retelling component useful and feel that it is helping students improve their comprehension skills. When retelling occurs, students are highly engaged and involved in learning.

Assessment of changes in students' ability to do retelling suggests that use of the cards can increase student skills. While students of all ability levels appear to profit from exposure to retelling cards, teachers feel that the less able students receive the greatest benefits. Important benefits for ELL students have also been reported. Our measures also show that over the short period of time the cards were used, students did not appear to gain the ability to do retelling without the support of prompts to guide their responses.[5]

The findings are encouraging, and the editors of *Treasures* should be encouraged by the results of the exploratory work. Nonetheless, more should be done to understand how contextual factors at the individual school level affect the efficacy of the retelling component and the extent to which the benefits of the strategy for students from different populations. We recommend assessment of the efficacy of retelling cards over a longer period of time with more detailed examination of the context of implementation and its utility for different student groups.

[5] *We are somewhat cautious in drawing this conclusion because we cannot rule out the possibility that the measures we used—being measures newly created for this study—might have been flawed. Indeed, we did note areas where improvements should be made.*

What Are Some Important Next Steps?

The new reading program, *Treasures,* has now been completed and is ready for implementation in the classroom. Based on what we learned from the current study, we propose to expand our scope of inquiry to address the implementation of several key features of the program, including the retelling component. Over the next three years, we will try to answer the following questions:

- Is the program design consistent with published research?
- Can teachers implement the program with fidelity and ease?
- Are there areas in which teachers need more help? Are there components that are left out?
- Do the lessons lead to the learning for which they are designed?
- When lessons are effectively implemented, what are the results?
- How does the program work with students from different populations—regular students, struggling readers, and students with special needs, such as English language learners?
- What is the efficacy of the program?

We plan to use a two-stage strategy to complete these tasks:

- Stage 1 (2005–06): Conduct design and learning validation studies. These studies will examine the implementation of key program elements through targeted examinations of varying duration in grades K–6.
- Stage 2 (2006–08): Conduct a scientific efficacy study that uses random assignment of treatment and control schools to document the outcomes of the *Treasures* reading program.

The efficacy study will target grades 1, 3, and 5. The key program elements to be studied are: as follows:

- Test preparation, including question-and-answer (QAR) techniques
- Strategy and skill instruction, the deliberate cognitive tactics that children use to analyze words and text.
- Classroom management of small groups using data to drive grouping decisions,
- Coverage and teacher support for addressing the five essential components of reading: Phonemic Awareness, Phonics, Vocabulary, Comprehension, and Fluency,
- Time for Kids, a specially written version of TIME magazine that focuses on expository text and incorporating science and social studies content,
- Paired selections, the inclusion of an additional genre to parallel the content of the main reading selection (e.g., expository text to fiction) and,
- Retelling Strategies, using the published retelling cards that accompany the program materials.