Final Report for Period:10/2001 - 09/2007 Submitted on:12/11/2007
Principal Investigator: Kintsch, Walter
Award ID: 0115419
Organization: University of Colorado Boulder
Title: Scalable and Sustainable Technologies for Reading Instruction and Assessment

Project Participants

Senior Personnel

Name: Kintsch, Walter
Worked for more than 160 Hours: Yes
Contribution to Project: Principal Investigator, Overall management of the Colorado Literacy Tutor (CLT) IERI program.
Name: Caccamise, Donna
Worked for more than 160 Hours: Yes
Contribution to Project: Co-PI. Program management of all (CLT)component elements. Designed and implemented scaling research, Co-managed Summary Street development team. Co-managed CLT Evaluation effort.
Name: Snyder, Lynn
Worked for more than 160 Hours: Yes
Contribution to Project: Co-PI. Project Manager for Assessment effort (*FtL*) and Project Manager for CLT Evaluation.
Name: Olson, Richard
Worked for more than 160 Hours: Yes
Contribution to Project: Co-PI. Reading expert “consultant” for any design issues for CLT
Name: Cole, Ronald
Worked for more than 160 Hours: Yes
Contribution to Project: Co-PI. Project Manager for Foundations to Literacy (*FtL*) component of the Colorado Literacy Tutor
Name: Kintsch, Eileen
Worked for more than 160 Hours: Yes
Contribution to Project:
Project manager for *Summary Street* development issues.
Name: Franzke, Marita
Worked for more than 160 Hours: Yes
Contribution to Project:
Headed projects such as copyright issues for text used in *Summary Street*, and other topics important to scalability. Helped with *Summary Street* development and assessment. 50% time
Name: Pellom, Brian
Worked for more than 160 Hours: Yes
Contribution to Project: Project manager for speech recognition software development

Name: Wise, Barbara
Worked for more than 160 Hours: Yes
Contribution to Project: Head of pedagogical development in foundational reading tutors
Name: Isidoros Doxos
Worked for more than 160 Hours: Yes
Contribution to Project: Project Manager for Supplement to investigate students' misconceptions in physics using LSA (see attached final report)
Name: Oliver, William
Worked for more than 160 Hours: Yes
Contribution to Project: Chief statistician
Name: Weston, Tim
Worked for more than 160 Hours: yes
Contribution to Project: Statistician

Post-doc
Name: Wade-Stein, Dave
Worked for more than 160 Hours: Yes
Contribution to Project: Manager for foundational tutor development
Name: Jones, Mike
Worked for more than 160 Hours: Yes
Contribution to Project: Assisted with LSA strategies, which underlie the *Summary Street* functionality.
Name: Hacioglu, Kadri
Worked for more than 160 Hours: Yes
Contribution to Project: part of the speech recognition software development

Graduate Student
Name: Davis, Gwendolyn
Worked for more than 160 Hours: Yes
Contribution to Project: Helps develop and implement various aspects of the tutor infrastructure. Worked with the *FtL* development group and later the *Summary Street* development group.
Name: Eckhoff, Angela
Worked for more than 160 Hours: Yes
Contribution to Project: participated in *Summary Street* field deployment and data collection. Angela also ran metacognitive study. Funded as a GRA
Name: Mangalath, Praful
Worked for more than 160 Hours: Yes
Contribution to Project: participating in final statistical analyses
Name: Corvey, Will
Worked for more than 160 Hours: Yes
Contribution to Project: Data cleaning/cataloging for *FtL* and *Summary Street*

Technician, Programmer

Name: Johnson, Nina
Worked for more than 160 Hours: Yes
Contribution to Project: worked full-time with our summary street development, implementation and assessment group. She took the lead in developing the summary street library.
Name: Tobias, Megan
Worked for more than 160 Hours: Yes
Contribution to Project: worked full-time with our summary street development, implementation and assessment group. She was our first field supervisor, coordinating the in-house and school-designated liaisons.
Name: Martin, Dian
Worked for more than 160 Hours: Yes
Contribution to Project: LSA programmer—needed for refinements to *Summary Street*.
Name: Ma, Jiyong
Worked for more than 160 Hours: Yes
Contribution to Project: Programmer for foundational tutors and interactive books
Name: Ngampatipatpong, N.
Worked for more than 160 Hours: Yes
Contribution to Project: Programmer for foundational tutors and interactive books
Name: Allen, Cory
Worked for more than 160 Hours: Yes
Contribution to Project: Field supervisor for scale up of both the *FtL* and *Summary Street*; data collection and organization
Name: Dehart, Maryanne
Worked for more than 160 Hours: Yes
Contribution to Project: participated in field deployment and data collection. Has become a graduate student in 2006-2007 and is assisting in data organization
Name: Gordon Golding
Worked for more than 160 Hours: Yes
Contribution to Project: IT professional responsible for field implementation of machines/software for *FtL*. Also serves as a statistician to *FtL* evaluation efforts.
Names: Guild, Nancy; Griem, Christen; Haight, Gretel; Johnson, Sarah; Kihn, Lisa; McClure, Erin; Rothenberg, Rebecca; Rumsey, Rebecca; Schooler, Vicki; Tepe, Katharine
Worked for more than 160 Hours: variable
Contribution to Project: *FtL* reading assessors—all students K-4 were tested prior to participation to determine their level of reading decoding, fluency, and phonemic awareness skills. This testing occurred pre-, mid-year, and post year for participating students.

Other Participants
Research Experience for Undergraduates

Werner, Jessica; Whitehurst, David; Fisher, Laura; Loveless, Stephanie. These undergraduates assisted in the development of texts and also helped to process pre/post summarization data from *Summary Street*. Ziccardi, Adam; Stanwood, Phillip; Mehring, Russell; Laurino, Lisandro; Hovey, Christopher; Hofmockel, Evan; Hofmockel, Amy; Brojde, Chandra; participated in various aspects of *FtL* content development.

Organizational Partners: The school districts of Strasburg and Ignacio, due to their distant locations, participated in our Liaison program, where we hired and trained 1-2 teachers in each district to serve as research liaisons for our implementation. These teachers oversaw (not in their own classrooms) research requirements such as participant permission form submissions, reading assessment and assignment for the *FtL* program, teaching teachers as necessary how to use both *FtL* and *Summary Street*, and answering teachers' questions. These liaisons interfaced between the teachers and researchers, providing the researchers timely feedback so that the researchers could address any problems as they arose. At Strasburg, these liaisons also had an active teaching load, while the Ignacio liaison was dedicated to this project.

Other Collaborators or Contacts
Center for Teaching, Learning, & Technology:
Julie O'Brian
Mike Lowe (originally with Colorado Dept. of Education)

Pearson Knowledge Analysis Technologies (Subcontract):
Scott Dooley
Tom Landauer, Ph. D.
Darrell Laham, Ph. D.
Lynn Streeter, Ph. D.

Boulder Valley School District (Subcontract):
Administration: Jean Riordan, Jana McMillan, Rosemary Bogart

Ftl's Interactive Books and Reading Tutors were used in kindergarten,1st, 2nd, and a few 3rd grade classes in four Boulder schools, Pioneer, Sanchez, Columbine, and Creekside Elementary. In these schools participation in the Free Lunch Program was 52%, and 38% of the students are non-native English speakers. *FtL* was also placed in 5 of the 7 participating Denver Archdiocese Schools, the Strasburg, and Ignacio School Districts. Over the course of three years, the majority of students in the *FtL* group whose data were included in the final analyses used the software an average of 6 hours.

Schools where *Summary Street* has been implemented and evaluated in Colorado:
9 districts (Archdiocese of Denver, Bennett, Adams County, Englewood, Boulder Valley, Ignacio, Denver, Steamboat Springs, and Strasburg); 22 schools; 82

teachers; a total of 4,274 students worked with Summary Street in 2003-04—2005-06 while a total of 2,409 students were part of the FtL effort in 2003-04—2005-06.

Other Schools:
Henderson, North Carolina: 1 middle school (Fall 2003)
Maryland: working with John Guthrie's CORI project; 4 - 5 teachers used *Summary Street* with their 4th grade classes (Summer and Fall 2003)

Researchers from other Countries:
Researchers from other countries visited our lab for up to a year to explore the LSA engine and *Summary Street.* A summary of their efforts is described below.

Activities and Findings

The goal of the present project was scale up the Colorado Literacy Tutor and to evaluate its performance. The tutor has two components, ***Foundations to Literacy***, designed to teach reading in grades K-3, and ***Summary Street***, designed to teach strategies for summary writing in grades 5-10. In the proposal we envisaged a closer integration between these two components by incorporating *Summary Street* in the reading tutors for the early grades. It became apparent, however, that this was not feasible: *Summary Street* is not suitable for grades earlier than 5 or 6 because it relies on written texts. We experimented with oral summaries, but it became quickly obvious that oral summarization required a quite different support system than that provided by *Summary Street* provides. Thus, the two components of the project were developed quite independently. It should be noted here that *Summary Street* is not suitable for students in grades 11 or higher either: typically, these students know already what *Summary Street* teaches. This does not mean that the use of LSA is impossible in support programs for younger or older students – merely that such programs would have to be quite different than *Summary Street*. We focused here on scaling up and evaluating *Summary Street*, instead.

Research involving *Summary Street*:

Summary Street is a software program that provides guided summarization practice for students writing summaries. It was originally developed by David Wade-Stein as his dissertation project (Steinhart, 2001) in cooperation with W. Kintsch (advisor), Tom Landauer, and E. Kintsch (E. Kintsch et al., 2000; E. Kintsch et al., 2007).

Summary Street is a computer tutor that offers a supportive context for students to practice summarizing. Students are guided through successive cycles of revising with feedback on the content of their writing. Students send their written drafts via the Internet for evaluation by Latent Semantic Analysis, which compares the similarity in meaning between the input summary and the source text from which it was derived. A graphic interface displays the feedback in an easy-to-grasp form that a student can use to revise his/her summary until it reaches the criterion for content coverage and appropriate length. Figure 1 shows an example of the feedback page: horizontal bars correspond to the section headings of the text to be summarized. The vertical bar on the right provides a length guideline. For each topic section, LSA computes a cosine as a measure of similarity between the information in the summary about that topic and the source text.

Students see improvement in content coverage in terms of how closely each horizontal bar approaches the vertical black threshold line. Color provides an additional cue: Initially red, each bar turns yellow, then green when each topic is adequately described. The length indicator uses color in a similar manner: read an d yellow indicate that that the summary is much -to-somewhat too short or too long, green that the length is in the appropriate range. Students may request additional checks to help correct spelling errors and to deal with problem sentences: When requested, sentences that are overly redundant with other sentences in the summary or that are not very relevant to the topic are marked by highlighting. Plagiarized sentences are similarly flagged. This prevents students from simply cutting and pasting sentences into their summary. It forces them to re-state ideas in their own words. Feedback from *Summary Street* is directed at the content of the writing and does not directly evaluate organization, writing style or mechanics (e.g., grammar and punctuation). *Summary Street* is designed to help students make sure the content of their summary is adequate before handling it to their teacher for final evaluation.

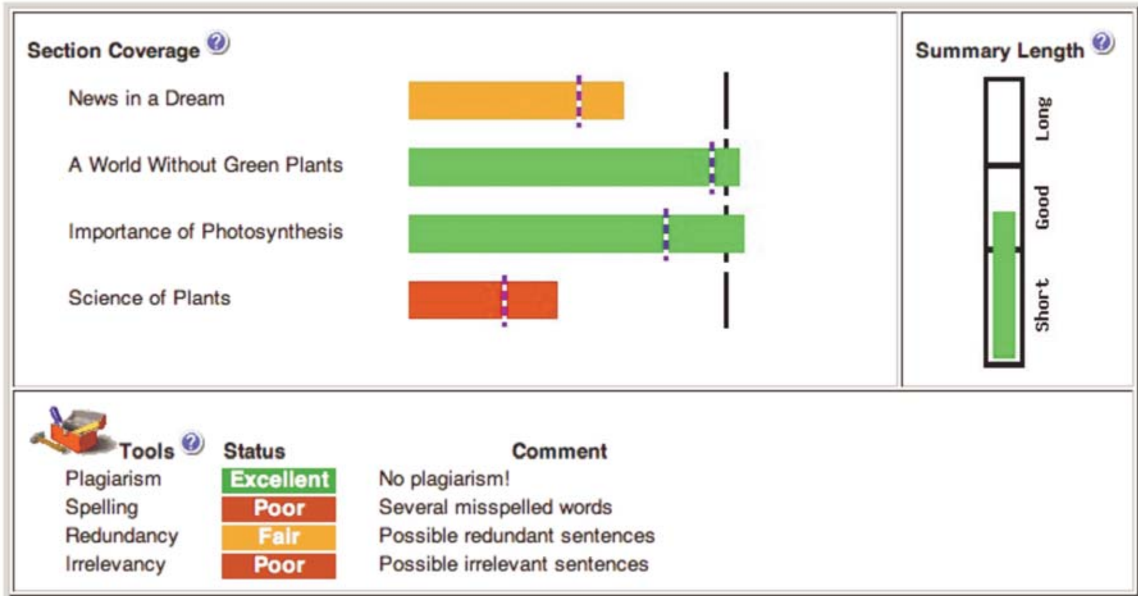

Figure 1. Screenshot of *Summary Street* feedback page.

An overview of our activities throughout the funding period is shown in Table 1. The first two years of the project were devoted to pilot testing *Summary Street* in schools. Only a few schools in the Boulder Valley School District (BVSD) were involved. In Year 3, we cautiously expanded the use of *Summary Street* to selected schools in Colorado, after familiarizing teachers and administrators with the tool in a series of workshops. Years 4 and 5 were the scale-up years: *Summary Street* was used widely in classrooms all over Colorado, sometimes with support from our staff, often without. Some teachers used *Summary Street* wisely (i.e., they integrated *Summary Street* into their instruction) and with success, others did not. Our presence in Colorado schools ended formally in the summer of 2006. Fortunately, Pearson Knowledge Technologies (PKT) at that time began to market their version of *Summary Street* commercially.

The basic software for *Summary Street* (based on scoring summaries via Latent Semantic Analysis, LSA; see Landauer et al., 2007) already existed at the start of the project. However, it had to be rewritten to make it school-proof. This was done in cooperation with the subcontractor, Knowledge Analysis Technologies (later PKT). Also, a number of subsidiary systems had to be developed for the efficient use of *Summary Street* in schools:

- An editing tool was created for teachers as well as researchers.
- A recording tool made it possible for teachers to keep track easily of what was happening in their classroom and which students needed special attention.

- A library of texts (eventually containing 120 items) was constructed based on topics that teachers had suggested and that fit into their regular curriculum.
- In order to use a text with SS, thresholds had to be set for each section of the text; an automatic tool for estimating these thresholds was developed.
- We began to integrate a vocabulary trainer with *Summary Street* (this work is being continued commercially by PKT);
- The lack of adequate comprehension tests for evaluating a program like *Summary Street* was a serious concern. Therefore we experimented with tests that differentiate between superficial levels of comprehension sufficient to reproduce a text, and deeper understanding that involves the construction of an adequate situation model to support inferential processes. Automatic methods to score such tests, based on LSA, also were developed. This work is described in detail in Mulligan, Rawson, Mangalath, & W. Kintsch (submitted);
- Absolutely crucial for the success of *Summary Street* was the website http://colit.org that allowed researchers, teachers, and administrators to experiment with SS. A corresponding website http://www.WriteToLearn.net is now maintained by PKT.

Table 1. Overview of *Summary Street* activities throughout the funding period 2001-2007.

	Activities	Development	Findings
2001/2	5 classrooms in BVSD (6th & 8th grade)	Editing tool; Reporting tool; Website	Usability testing
2002/3	6 classrooms in BVSD	7 workshops for teachers; Text library	Usability testing Experiment 1: Wade-Stein & E. Kintsch (2004)
2003/4	24 classrooms in BVSD, Denver parochial inner city schools, a Strasburg rural school district, and two suburban districts	Automatic threshold calculations for SS	Experiment 2: Franzke et al. (2005)
2004/5	7 school districts: 85 teachers 77 classrooms and 2,294 students	New PKT interface; Vocabulary trainer; Comprehension test (Mulligan et al., submitted)	Evaluation Study – Year 1: 1891 students use *Summary Street* for one year; pre-post-tests (Caccamise et al., 2007; E. Kintsch et al., 2007)
2005/6	7 school districts 120 classrooms (K-5, middle, high-schools) 3,000 students (diverse ethnicity, SES; urban, suburban, and rural)		Evaluation Study – Year 2:
2006/7	Data analysis for Evaluation Study	Commercial use of *Summary Street* by PKT	Caccamise et al. (2007) Caccamise et al. (in preparation)

Throughout the project period PKT (formerly KAT) has been involved as a subcontractor. Software development was done cooperatively between our group and PKT, with Scott Dooley as the main contact person in PKT. When *Summary Street* was being used in schools all student responses went via the internet to a server at PKT that provided instant feedback and recorded and saved all information. Currently, PKT is continuing the use of *Summary Street* which has been integrated into Pearson Prentice Hall's "Literacy Skill Builder Series" and "*WriteToLearn*" program. Workbooks from Prentice Hall and web-based literacy tools from PKT offer middle school students powerful tools for building reading and writing skills. Thus, the insights and experience gained with the present project will continue to be useful – *Summary Street* survives, although the name has changed.

Findings:

Six years of working on this project by our research group has been a source of rich and rewarding experiences for us. To turn personal experience into public knowledge, we report the results of two small-scale experimental studies and one large-scale evaluation study.

Study 1:

Two 6th grade classes, 60 students; 10 texts; each student wrote 2 summaries, one with and one without *Summary Street*, with order and text counterbalanced. The summaries written with the help of *Summary Street* received significantly higher grades than those written without. Positive effects were also found for summaries written later without the support of *Summary Street*. The biggest effects (one grade point) were found for the most difficult texts, and for average students. For the best students and easiest texts *Summary Street* was not needed; for the poorest students, the texts assigned to the class were too hard even with the support of *Summary Street.* (Wade-Stein & E. Kintsch, 2004).

Study 2:

Franzke, E. Kintsch, Caccamise, Johnson, & Dooley (2005) observed students writing summaries in twice weekly sessions for four weeks in four classrooms: half the students in each classroom used Summary Street and the other half did not. Their results replicate the findings of Wade-Stein and Kintsch, (2004), showing improvement in summaries written while supported by the content feedback that *Summary Street* provides. It also extends the findings of the earlier study in several ways.

In blind scoring summaries written with *Summary Street* were judged to be superior on several measures of writing quality: their overall quality, more complete content coverage, better organization, lack of low-level details, and stylistic quality. Note that *Summary Street* does not support organization or style directly; nevertheless, improving the content of the summaries also improved these characteristics of writing quality.

How does training with *Summary Street* transfer to comprehension in general? Overall scores on a comprehension test based on the standardized test required by the State of Colorado – (Colorado Student Assessment Program - CSAP) –) did not change from pre- to post-test for either the control or the *Summary Street* group, nor did we find group differences for test items unrelated to the training (inference, vocabulary, fact finding and other). However, we did observe a 25% gain among *Summary Street* users for test items requiring a summary response. Thus, summary practice with feedback that directs students to attend to relevant content can transfer to reading outside of the summary writing context. Simply practicing summary writing without such feedback is not sufficient. Our observations in the classroom suggest that students work harder when using *Summary Street*: they spend more time on their summaries, tend to be more engaged in the task and take pride in their efforts, all of which seemed to influence in a positive manner the way they subsequently read the expository texts they encountered on the CSAP post-test.

<u>Evaluation Study</u>:

A full account of this study will be given in Caccamise Snyder, Allen, E. Kintsch, W. Kintsch, Oliver & Thye (in preparation).

2,851 students in 184 classrooms, 21 schools, and 9 school districts in Colorado participated in this study. These students were drawn from a larger population of students (4,166 students) who used *Summary Street* as part of their normal classroom curriculum. However, those who did not give their consent were NOT included in the study. Students in grades 5 - 9 participated in the study; most participants were in grade 7. The actual number of students entering into a given analysis depended on the pattern of missing values. At the beginning of the year as well as the end of the year, students were graded on a summary they wrote without access to *Summary Street*. The pre-and post texts were two texts matched in length, complexity, lexile level and were chosen to be in the normal reading range for the subjects' grade level. Students also were tested with a short comprehension test (*TORC*) both at the beginning and end of the year. We also had available the results of standardized reading comprehension tests administered by individual schools. Scores on the CSAP test were also available for each student, which allowed us to remove the effects of general performance levels from our analyses.

During the year, students in the experimental condition used *Summary Street* and students in the control condition received traditional summarization instruction, as the teacher would normally have done in her classroom. How *Summary Street* was used – with which texts and how often – depended on the teacher and was different for each classroom.

In order to carry out the scaling study, certain special populations, such as students in a remote rural school, had to be treated somewhat differently than the majority of students in the study because of limitations on how they could use *Summary Street*. These data from special populations are invaluable, and are analyzed separately. However, to assess the overall effect of *Summary Street*, we only included the data for students who followed the standard method of *Summary Street* usage. Also, the data for students who did not complete both the pre and post-tests were not included in the analysis. After applying these constraints, there were 1,577 students who contributed data to the analysis.

Figure 2 shows the frequency distribution for *Summary Street* use in the experimental condition. The number of texts that a student studied usually corresponded to the number of times that *Summary Street* was used by that student. The average number of times the tool was used ranged from 0 to 12. Thus, quite a few students used *Summary Street* only minimally or not at all, for a variety of reasons (teacher provided few opportunities, they were absent, etc.). However, the majority of the students did get a reasonable amount of practice with *Summary Street*.

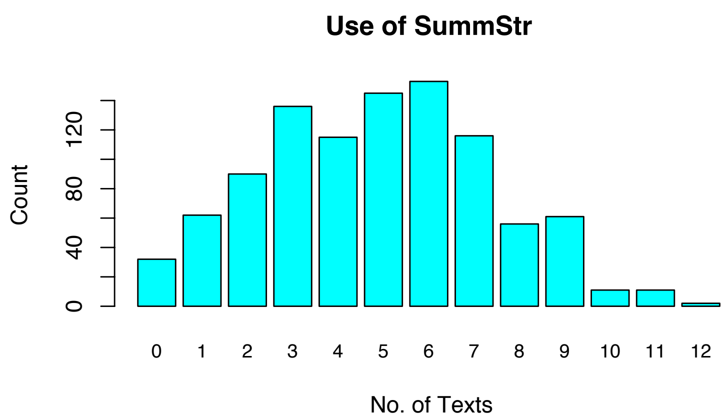

Figure 2. Distribution for the number of texts studied by students over the course of a school year. Data for academic years 2004/2005 and 2005/2006 are combined.

On the post-test, the experimental group outperformed the control group in summary writing for both years of the study. The dependent variable is the proportion of sections passed as measured by *Summary Street* (passratio). This same measure was used for giving students feedback during training. The scores were adjusted for text difficulty and differences in scale by standardizing the pre- and post-test scores on the basis of the pre-test scores. The effect sizes of condition on the standardized passratio as measured by Cohen's d were 0.31, $t(829) = 3.94$, $p < .001$, and 0.28, $t(711) = 3.77$, $p < 0.001$, for the two years, respectively. Figure 3 shows the overall difference between conditions (effect size = 0.31, $t(1542) = 5.81$, $p < 0.001$.

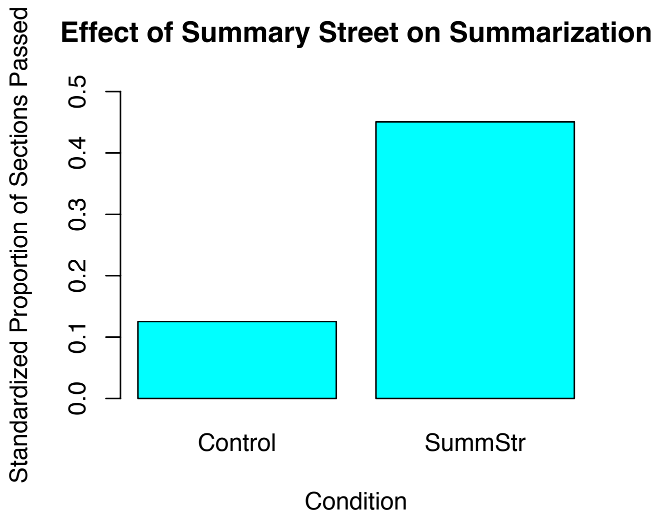

Figure 3. Standardized proportion of sections passed by students as a function of the Control condition and the Summary Street condition (SummStr). Data for academic years 2004/2005 and 2005/2006 are combined.

The effect was statistically significant for each academic year in separate HLM analyses that included the covariate adjusted standardized passratio, $F(1,61) = 7.28$, $p = 0.009$; $F(1,50) = 6.65$, $p = 0.013$; and $F(1,112) = 15.35$, $p < 0.001$, for academic years 04/05, 05/06, and both years combined, respectively. The effect of *Summary Street* on summary writing as measured by the standardized passratio is not large. This is not surprising because the data for the experimental group came from an extremely heterogeneous sample of students, including even students who never used *Summary Street*. A better picture of the effectiveness of *Summary Street* is given by Figure 4, where improvement in summarization is plotted as a function of the number of texts that a student studied with *Summary Street* during the year. Since the data for the two different years were much alike, the figure shows combined results for both years. Obviously, those students who actually used *Summary Street* improved their summarization skills quite a bit during the year, $t(171) = 3.23$, $p = 0.002$ and $t(77) = 3.07$, $p = 0.003$ for the two years, respectively. The error term was again based on classes, and there were the additional constraints that each class must include at least 4 participating students and that the number of texts studied by students could not exceed 9, so as to prevent highly influential observations from biasing the results. In addition, each student's pre-test standardized passratio score was included as a covariate in the analysis. Furthermore, the overall effect remains when the students' scores on the writing section of the CSAP are partialled out. Hence, it is unlikely that the "dose" effect of

Summary Street is merely a selection effect in that good students got better and used *Summary Street* more often.

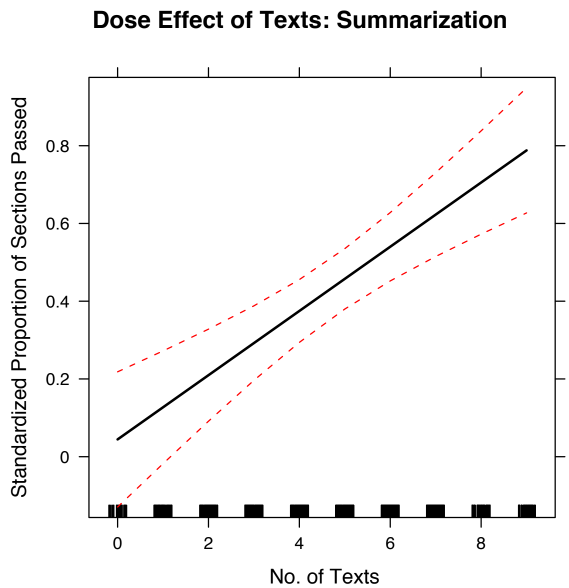

Figure 4. Standardized proportion of sections passed on the adjusted summarization post-test as a function of the number of texts studied with SS during the school year. The pre-test scores for the test were included in the analysis to control for student ability.

A very similar picture emerges when improvement in summarization is plotted as a function of other measures of *Summary Street* use, such as the total time that *SumStr* was used or the number of summarization attempts.

As mentioned, students were given a comprehension test (two subsections of the TORC) both at the beginning and end of the year. This test was administered only for the academic year 2005/2006. The improvement in comprehension scores was not statistically significant between experimental and control groups, but when comprehension was evaluated as a function of the amount of *Summary Street* use, a statistically significant effect was obtained, $t(79) = 3.70$, $p < 0.001$ (Figure 4). Improvement in scores on an objective comprehension test was a function of the number of times *Summary Street* was used. This is probably not a selection effect, since the effect remains significant statistically when CSAP scores are controlled statistically.

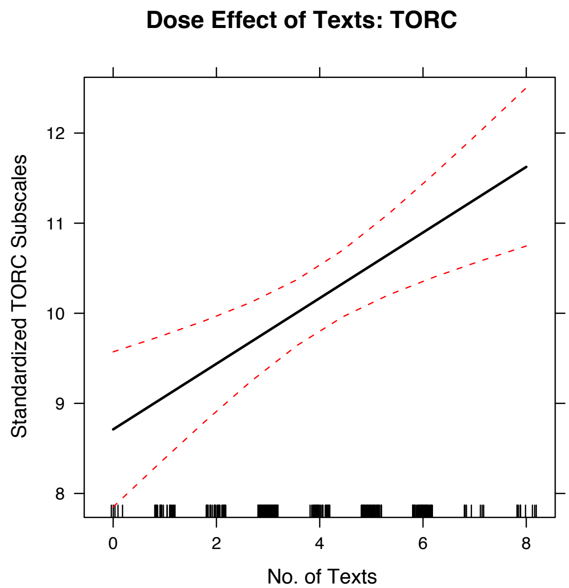

Figure 5. Performance on two subtests of the TORC (standardized scores) as a function of the number of texts studied with SS during the school year. The pre-test scores for the same test were included in the analysis to control for student ability.

The findings from the three studies reported here, therefore, are quite positive:

(1) All three studies show that *Summary Street* can effectively teach summarization to middle school students. Essentially, it appears that students need to use *Summary Street* 4 to 6 times to profit from it. Once they grasp the essentials of summarization – coverage of all topics, focus on gist, avoiding redundancy and irrelevancy for the sake of conciseness – they have learned what this system can teach them and are able to use this skill in the absence of the system.

(2) Moreover, the improvement in summarization skills as a consequence of practice with *Summary Street* has measurable positive effects on comprehension in general. Summarization is a very important comprehension skill, and teaching students to summarize helps their comprehension in general. There are comprehension strategies other than summarization, and there is more to comprehension teaching than summarization, but being able to achieve this crucial subskill of comprehension is a real achievement for *Summary Street*.

(3) The results reported here are supported by rigorous statistical analyses. In Experiment 2, students within classes were randomly assigned to control and experimental groups. In the Evaluation Study, the results were reliable when classrooms were used as the units of analysis, and remained reliable when general performance (the students' CSAP scores) was controlled for.

Projects outside Colorado:

Although the focus of this project was the scale-up of *Summary Street* in Colorado schools, the tool has been used elsewhere in a number of places, notably the project directed by Louis Gomez at Northwestern University. Work on three sites outside Colorado was directly coordinated with the present project:

A group of educators used *Summary Street* in rural schools in North Carolina; we ran a workshop for the teachers involved.

A group of German educators and psychologists at the University of Würzburg headed by Joachim Hoffmann and Wolfgang Schneider obtained grant support from the Deutsche Forschungsgemeinschaft to develop a German version of *Summary Street*. They sent two of their researchers, Wolfgang Lenhard and Herbert Baier to Boulder for extended visits. The group successfully implemented and tested a German version of LSA and designed a *Summary Street* prototype. At this point their software is undergoing preliminary evaluation in German schools (Lenhard, Baier, Hoffmann, Schneider, & Lenhard, A. 2007).

In France, Lemaire, Mandin, Dessus, and Denhière (2005) have proposed a model of how students write summaries. The model is based on LSA and focuses on assigning importance to sentences and the use of macrorules for summarization. This model provides the foundation for a learning environment for learning how to summarize. Drs. Denhière and Lemaire have both visited our lab in Boulder.

Minglei Chen, with support from the Taiwan National Science Foundation, spent a year in Boulder during which time she successfully established a Chinese version of LSA. Given the differences in the writing system this is a nontrivial achievement: to achieve meaningful results, the system could not be character based, but required the marking of word boundaries in the Chinese texts. She is now developing a Chinese version of *Summary Street*, which is to be used in Taiwan schools.

Thus, the IERI Summary Street project has had a broad impact in schools in the State of Colorado and well beyond its borders.

Summary of Major Findings from *Foundations to Literacy*, 2001-2007

The *Foundations to Literacy component* of the IERI project focused on the development and assessment of programs to help beginning and early readers improve their reading, to assist school districts meet the Federal mandate to help as many children as possible become readers by third grade (NCLB, 2001). Earlier research by Olson and Wise had established that theoretically grounded programs were highly effective with older poor readers in *Reading with ROSS* programs, which paired a) reading for meaning in context with speech and comprehension support and b) speech supported phonological exercises for development of word level skills (Wise & Olson, 1995; Wise, Ring, & Olson, 2000). In these studies, remedial readers in grades 2-5 made significant gains in reading, after reading with the computer programs between 22 and 40 hours. The children read in pull-out settings of 4-5 children, where each child had his own computer and a research assistant was present to read with each child weekly and to do some related small group instruction. Lower level and younger readers benefited especially from training with phonological exercises, which particularly helped their untimed word reading (Wise, Ring, & Olson, 2000). Readers at higher levels experienced extra benefits from more time reading accurately in context within interesting books, and this practice especially benefited their time-limited or fluent word reading. Other theoretically grounded programs have also shown effectiveness in controlled studies (see reviews in Blok, et al, 2002; Tijms, Hoeks, Hoogeboom, & Smolenaars, 2003; Olson & Wise, 2006.)

A recent study by the Institute for Education Science lends a cautionary note. It found that pedagogically sound software which had shown effectiveness in small studies proved only as helpful as the classroom teacher, in a large-scale study with little outside support for teachers (IES, 2007). It is important to note that even if reading software eventually proved only as helpful as classroom instruction, it could still provide an important service. Such software that is as effective as classroom use can definitely extend teachers' resources, allowing them to work with more children in 1:1 or small group situations. We also believe that software is proving effective among lower readers, when children have sufficient hours (Macaruso & Hook, 2007; Wise & Van Vuuren, 2007).

The *ROSS* programs were highly effective with the support of a research assistant, but they were not scalable without that support. When the project lost funding, some schools kept the programs. Without the research assistant present for support, the programs did not get used. The combination of findings of effectiveness with problems in scalability inspired the development of the *Foundations to Literacy* program. We hoped that developing an engaging and realistic animated Virtual Tutor would help children to get the assistance they needed with little support from a research assistant or teacher. We planned that the Tutor would play the role of coach, greeting, encouraging, and giving hints to the child, to provide some of the support that the trained research assistants had in the *ROSS* studies.

The first 2 years of the project were devoted to development of the tutors and books, with content and study plan development led by Barbara Wise, and pilot testing of the program. *Foundations to Literacy (FtL)* built on the best of the *ROSS* functionalities, improved comprehension support, and added independence and engagement through the use of an animated "coach." *FtL* was developed over the last 6 years, in federally funded studies at the Center for Spoken Language Research, at the University of Colorado, with grants to Ron Cole, Lynn Snyder, Barbara Wise and other colleagues, supported by the efforts of many members of the Center. Others on the IERI team helped with the development of comprehension support, and Lynn Snyder in coordination with

Donna Caccamise designed the evaluation studies, and coordinated the assessment and evaluation program with the assistance of Tim Weston from the Alliance for Technology, Learning and Society (ATLAS) to determine the benefits of the program. Drs. Snyder, Caccamise and Weston were independent evaluators; they were not part of the *FtL* program's development team.

Foundations to Literacy (FtL) works on phonological and word level skills and automaticity in foundational exercises. It also works on vocabulary, fluency, and comprehension in interactive books. For a full description of the program, see the chapter by Wise, Cole, Van Vuuren, Schwartz, Snyder, Ngampatipatpong, Tuantranont, and Pellom, (2007). In accordance with the ROSS findings above (Wise et al, 2000), *FtL's* study plan varied the balance of instruction according to reading ability, beginning with a balance towards foundational work (with books read aloud by the Tutor) and shifting with children's progress to ever increasing time in book reading with speech support for any difficult word and with comprehension support throughout (Van Vuuren & Wise, 2007). *FtL* has a speech recognition capability available (Van Vuuren, 2003; Hagen et al., 2004). In testing in the laboratory, it could note when children misread a word in a book and could provide the correct word to ensure accurate reading. However, this function did not become stable enough to deploy in schools over any significant time.

In *FtL*, the Virtual Tutor engaged children in exercises and in interactive books by giving them focused hints to self-correct their errors, rather than passively supplying answers. Students did exercises that adapted in difficulty in response to children's responses. In books, the Tutor engaged them with background questions and inference questions, and provided various kinds of support—pronouncing needed words, sentences, or reading the whole story. Most children liked her, calling her a good teacher (Cole, Wise, & Van Vuuren, 2007). *FtL* included frequent ongoing performance measures, built into an internal study plan that monitored and to a limited extend, adjusted a child's instruction according to the speed and accuracy of responses in foundational exercises and to changes in text-reading levels.

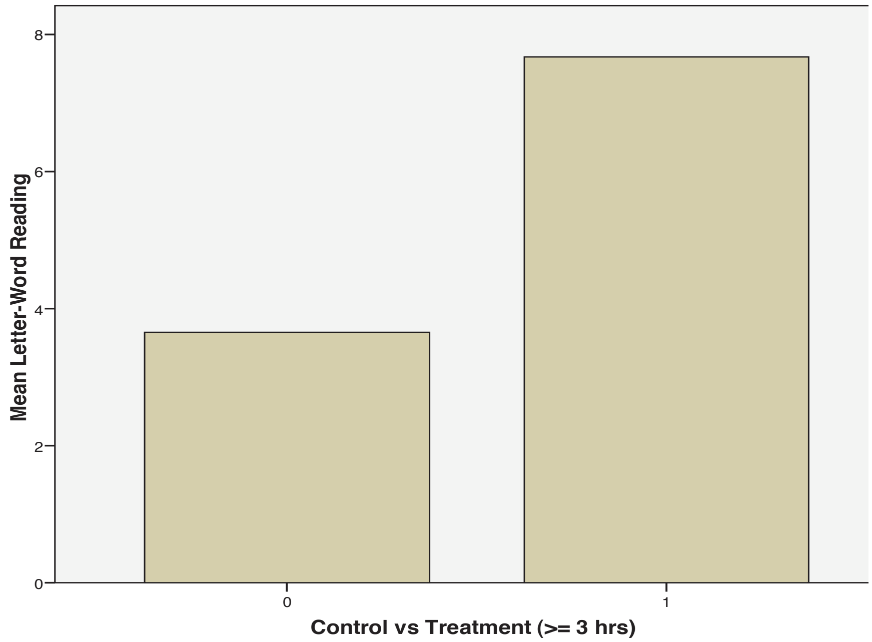

Figure 1. Woodcock Letter Word Reading Standard Score Gains by Group.

(FtL: N = 697, Classroom Control: N = 632)

FtL was deployed and assessed in schools across Colorado during 3 years: 2003-04, 2004-05, and 2005-2006. Kindergarten to third-grade children in 53 classrooms in 14 schools were assigned to read independently with *FtL* without the daily support of a research assistant that the ROSS programs had used. During the third year of the project, students were sampled by classroom into either the experimental or control groups because the participating school district would not permit random assignment. When the Archdiocese of Denver joined the project in January of that year, they permitted us to use random assignment to groups on a trial basis in one school. However, this effort was compromised by a compensatory equalization of treatment threat to validity, so we used assignment by school with the Denver Archdiocese Schools in the remaining years of the project. This was due to the fact that these schools were small with only one class per grade, making it not possible to assign students to group by classroom. In the fourth project year, students from all but 6 small schools were sampled into groups by class. In the final year, all participating schools, pressed by concerns generated by the NCLB mandates, indicated requested that their control groups needed to become part of the *FtL* group. Consequently, in the final year, we sampled by school and matched schools by demographic characteristics. In sum, the majority of the students in the study were sampled by classrooms into the *FtL* or control groups.

In most years, the students were pre-tested with a small battery of six subtests from standardized measures from late September through the beginning of November. The programs were deployed somewhere between November and

January, and children were post-tested from mid-April through late May. Implementation varied widely in classrooms, with students in some classrooms working few to zero hours with the programs.

We report analyses on the three combined years, after dropping outliers and the 414 students from the analyses who had less than 3 hours per school year on the programs. The 697 remaining children, who read at least 3 hours with *FtL*, were compared to 632 regular classroom control students in a quasi-experimental design (numbers vary in the figures below depending on missing data.).

The *FtL*-trained children in the analyses averaged about 6 hours with the programs (range 3-12 hrs). Unfortunately, the control children did start significantly higher than the *FtL*-trained children on letter-word reading and passage comprehension (F =53.5, p < .0001 for Letter Word, 19.5 p < .0001 for comprehension; *Woodcock Johnson III*, Woodcock, McGrew & Mather, 2001). This distribution led us to analyze the data with a repeated measures ANOVA in lieu of an ANCOVA. The findings indicated that the *FtL* students improved significantly more in these skills than the controls in (letter-word reading: F(1, 1342)= 46.5, p < .0001; and passage comprehension, F (1, 932) = 14.12, p < .0001). Standard Scores are the normalized measure of choice in educational research, reflecting a student's relative position in an age group, around a mean of 100. Trained children gained about 4 Standard Score points more than controls in word reading (Fig 1), and about 6 Standard Score points more in Passage comprehension (Fig 2). "Effect size" is a measure that suggests how much of an outcome difference can be attributed to treatment. Effect sizes were small to moderate (d=.24 for Passage Comprehension to .36 for Letter-Word Identification), but were encouraging given the low number of hours of training.

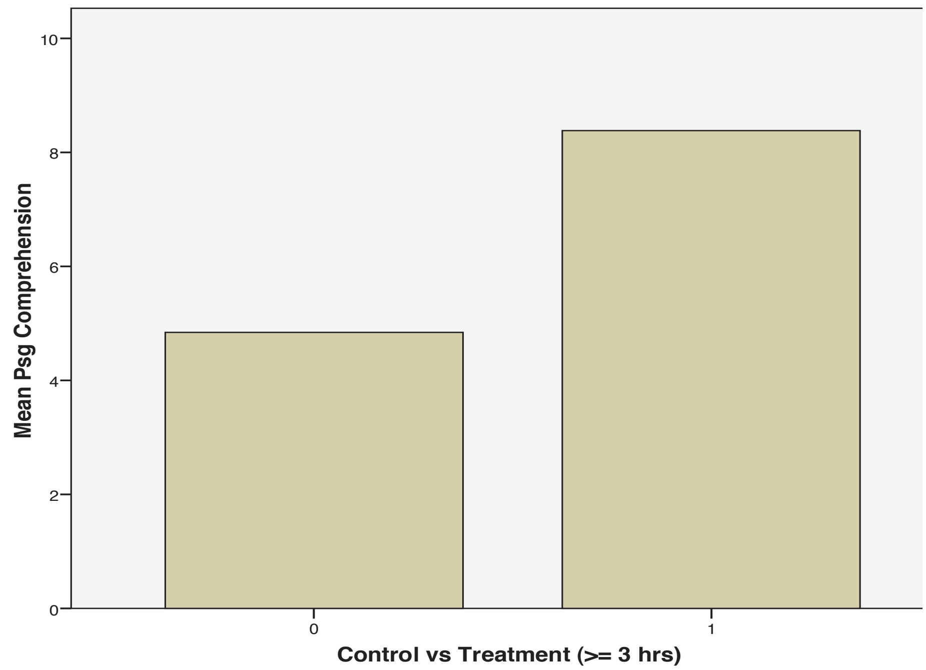

Figure 2. Passage Comprehension Standard Score Gains by Group.

(FtL: N = 510, Classroom Control: N = 420, some children didn't take this test)

The groups also differed on time-limited phonemic decoding, as measured by the *TOWRE* phonemic decoding test; however, this difference favored the control group (F(1, 934) = 5.0, p = .025*). The -.14 effect size was small, and we are not sure how to interpret this weak, contrary finding.

To be cautious, we conducted "propensity score analyses" as further support that the apparent difference was due to treatment and not due to the above-mentioned significant sample differences on pretests. These analyses work on the probabilities that a student with a given score, by decile, is in the treatment or control group, essentially comparing effects of treatment within decile groups in the two conditions. In a repeated measures ANOVA using these propensity scores as a factor, the main effect of training condition remained significant for Letter Word Identification (F (1, 1324) = 27.7, p< .0001). The treatment group outperformed the control group on 8 of the 10 decile blocks for letter word identification. The advantage for treatment over control was especially strong in the lower 5 decile groups, similar to other studies which have found computer training advantages for lower readers more than among higher readers (Macaruso & Hook, 2007.) For the *Woodcock Passage Comprehension* test, the main effect for condition also remained significant (F(1, 913) = 4.7, p = .03*). The experimental group did slightly better than the control group in seven out of ten blocks. The puzzling *TOWRE* Phonemic decoding advantage for the control group also remained significant in this analysis (F(1, 908) = 5.08, p = .015*).

Analyses also supported that effects differed by grades. A significant interaction suggested that *FtL* training gains for Kindergarten and first grade children were greater than for second grade children ($F = 5.3$, $p < .005$), with moderate effect sizes of .54 for kindergartners and .42 for first graders. Kindergartners showed Standard Score (SS) gains of 6 relative to controls, and first graders gained about 4 SS points relative to controls. Second graders did not differ by condition. This is not surprising because the *FtL's* foundational exercises mainly covered kindergarten and first grade skills. This finding is further support that treatment differences could be attributed to the program. The passage comprehension test also showed a significant interaction, with larger gains from *FtL* training for first graders over second graders ($F = 6.3$, $p < .0001$). First graders gained about 4 SS points more than control students did (Fig 4), with an effect size of .36 for the first graders. The Passage Comprehension test is not designed for nor given to Kindergartners, so they were not included in this analysis. Recent studies have shown that this test mainly reflects accurate and efficient word reading, so the results of these two measures confirm each other (Keenan & Olson, 2006). The consistency of theses measures with each other and with the strengths of the program suggests real benefits on post-tests from the training, despite the low training hours.

We conducted other exploratory analyses with a subset of the data utilizing hierarchical linear modeling (HLM). HLM is particularly suited to modeling intra- and inter-individual change using longitudinal data. In the current context, students can be nested within classrooms, which are nested within schools. In this study, positive changes due to time in school and to *FtL* treatment time would be indicated by fixed effects representing gains in test scores associated with time in school and number of hours of tutoring with the *FtL* computer tutor. These effects, as well as variance components representing individual differences in gains, are modeled via HLM. We attempted the HLM analyses with the second year cohort, where we hoped enough children had completed mid-year testing to allow these analyses. Based on the earlier results and due to low N for data for some measures, we focused on Passage Comprehension and Letter-Word Identification. We tested a number of models for each measure. Because of the known significant differences on pretests by schools, we created simple HLM models to see if it found the same effect. The HLM analyses did reveal this same effect. For example, using Passage Comprehension Standard Scores (SS), the intercepts for the control and experimental groups were 101.2 and 97.9, respectively: a difference of 3.3 SS ($p < .001$).

We also tested more complex models including variables associated with week of the school year as well as hours of *FtL* treatment. However, these more complex models encountered an identification problem, requiring one of the residual variance components to be held constant. Basically, the data set did not have sufficient degrees of freedom to allow for the analyses. This data set, with its low treatment hours was not robust enough to continue with more complex models. Iterative estimation and restriction of the various components of variance both showed great sensitivity. This precluded the use of such techniques to address the identification problem in the analyses. For these reasons, we could not analyze treatment differences with this data set with HLM. Current research efforts are aiming at ensuring that students receive more treatment hours. This should provide robust enough data to allow estimation of such complex models in future studies. Some trends in these HLM explorations were suggestive, but not significant in this data set. Nevertheless, the explorations and the knowledge of what a robust data set entails, have prepared us well for future analyses in our continuing studies with related computer-assisted treatments.

Despite the consistency of the findings from both the repeated measures and propensity analyses, we report our findings with caution. Difficulties with sampling discussed earlier leave the possibility open that the gains were related to school differences encountered during the final project year when schools were the unit of assignment (due to school district restrictions described earlier) rather than to treatment differences. While the study originally planned to assign treatment within classroom, we encountered restrictions imposed by the districts. Therefore, for part of the study, the unit of assignment to treatment was really at the school level for part of the study, and this confound cannot be removed from the study. On the other hand, the propensity score analyses do strengthen the suggestion that the condition differences were due to treatment, since there is no other reason to think that schools would differ for lower readers more than for higher readers. The findings that treatment gain differences occurred mainly with the lower half of readers, and that treatment differences occurred at grades where the program was strongest, both support the suggestion that significant differences in gains did relate to *FtL* treatment.

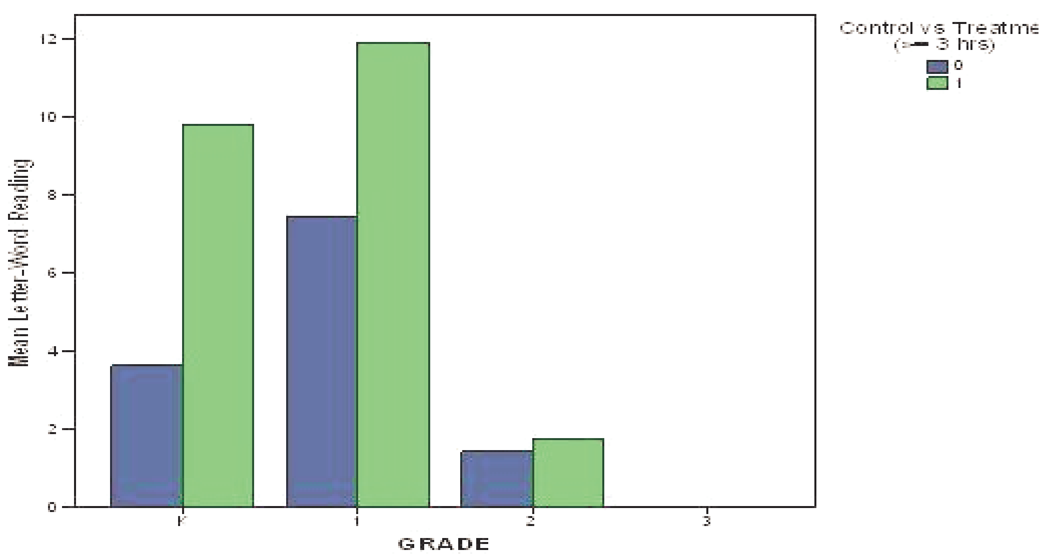

Figure 3. Letter Word ID Standard Score Gains by Grade.

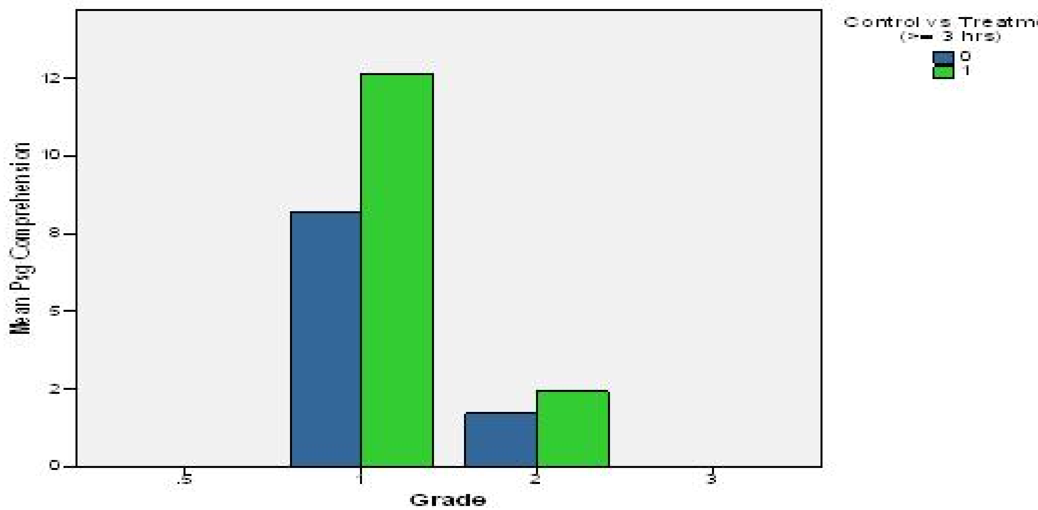

Figure 4. Passage Comprehension Standard Score Gains by Grade.

<u>Discussion and Future Directions for *FtL:*</u> The *Foundation to Literacy* studies suggest that adding an animated agent to theoretically grounded computer reading programs does improve engagement and independence over similar programs without such a helper. The studies also suggest that the program as developed was effective with kindergartners and first graders, especially with lower readers and in the areas best covered by the exercises. The major difficulty in the study was obtaining sufficient treatment hours, averaging only 6 hours, even after omitting children with fewer than 3 hrs from the analyses. *ROSS* studies mentioned earlier obtained strong and lasting gains with treatment hours of 22 to 40 hours, with the support of a research assistant. This difficulty in lack of hours limited the effectiveness of the programs and the possibilities for more complex analyses of effects for children of different profiles. Yet the technologies developed during the IERI period (example Van Vuuren, 2007), and the lessons learned, have led to very promising ongoing research in independent assessment, in response to computer-assisted instruction with stronger and more stable and better supported programs, and with plans for web delivery.

In our current studies, we build on the best of the functionalities from *ROSS* and from lessons learned in *Foundations to Literacy*. We are including an animated agent and some speech recognition in an independent assessment project called ICARE, for Independent Comprehensive Adaptive Reading Evaluation. ICARE is now in its 4th year of development, with a new project that extends its age range and looks at early predictors of later reading success (funded by IES, Wise, Sessions, Tomczyk, Snyder, Sager, Golding, Tuantranont, Ngampatipatpong, Van Vuuren, & Weston, 2007). ICARE could greatly reduce teacher time in testing in schools, and can also provide an initial instructional profile for struggling readers—useful in classrooms or in assigning computerized instructional plans.

We have also taken what we have learned from *FtL* to inform major improvements in a new intervention program called *Reading with RITA* (Reading with Intelligent Tutoring Assistance). *RITA* is being deployed in a preliminary study in 2007-08, as part of a study of Response to Computer-assisted Intervention, funded by NIH (Wise, Van Vuuren, & Byrne, 2007). *RITA* will be improved and stabilized this year, and then deployed in the full study with two cohorts of low readers in 2008-09 and 2009-10. The program uses ICARE testing to help with program placement and progress monitoring. *RITA* uses a talking coach who gives hints and support, as *FtL's* coach did. But she does more teaching within a vastly extended study plan that includes exercises from KG to 7^{th} grade, so children can continue learning at their own pace throughout the intervention time. It has a new study plan that is simple, yet integrated, so that it is easier for children and teachers to use and so children always have a variety of exercises available.

We certainly learned with *FtL* that a newly developed program was not suited to take to scale. The exercises worked well independently, as all progress depended on performance. But children could often skip through the books, without reading much of them in the ways we wanted. Therefore, in *RITA*, a research assistant attends at least weekly to the school to monitor children's performance, especially in books. *RITA* compares monitored and independent sessions and gives feedback to the children, so that both *RITA* and the assistant help the children learn to use the programs effectively even when independent. This kind of support was very helpful for helping children learn to use the programs optimally in *ROSS*. We expect better hours and stronger and more lasting gains with this support. We are also having the assistants record what kind of helping statements they give, and when, so we can continue improving the help and hints *RITA* gives, so that eventually the programs will indeed work with minimal assistance. We are currently designing 3 new ways to make children's progress in the books more conditional on performance, which should also strengthen effectiveness. A proof of concept grant we have just received from the Schwab Foundation will allow us to take portions of *RITA* to the web, using a simplified agent. e plan a final set of developments to improve *RITA's* ability to train comprehension, composition, and communication. The complete set of integrated programs that we are developing will provide a powerful research system for studying "what works best" for different children with different support. It may also one day help schools stretch teachers' resources to meet the needs of children while also improving their own teaching. The resources and findings of this IERI have grounded and empowered this whole effort.

- -

Training and Development:
Experience with reading comprehension and foundational reading skills research and assessment for graduate students Davis, Eckhoff, and Mangalath.

Outreach Activities:
A series of training workshops was held to familiarize teachers with Summary Street and its authoring and learning management tools. Discussion focused on ways to integrate the technology into regular instruction, and teachers provided valuable feedback to researchers on how the technology could be improved. Workshop participants thus become local experts for others at their schools who will be using it.
Spring and Summer workshops included:

- 5th-12th grade teachers from Boulder Valley and Englewood School Districts - 3-day workshop;
- Broomfield HS (BVSD) -1/2 day workshop;
- Strasburg School District - 1/2 day workshop.
- Englewood MS-HS (Englewood School District) -1/2 day workshop;
- Henderson, NC; full day workshop;
- Denver Public School District

Journal Publications
Caccamise, D., Snyder, L., Johnson, N., Allen, C., DeHart, M., Kintsch, E., Kintsch, W., & Oliver, R. (in preparation). Summary Street: Scale-up and evaluation.
Caccamise, D., Snyder, L. (2005) "Constructive Models of Reading Comprehension - The Situation Model and Related Constructs." *Topics in Language Disordes.*Vol. **25**, 1.
Cole, R., Wise, B., Van Vuuren., S. (Jan 2007). How Marni teaches children to read. *Educational Technology*, 24:1, 14-18.
Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary Street®: Computer Support for Comprehension and Writing. *Journal of Educational Computing Research, 33*, 53-80.
Kintsch, E., Steinhart, D., Stahl, G., Matthews, C., Lamb, R., & Group, L. R. (2000). Developing summarization skills through the use of LSA-backed feedback. *Interactive Learning Environments, 8*, 87-109.
Ma, J., Cole, R., Pellom, B., Ward, W., and Wise, B. (2004). Accurate Automatic Visible Speech Synthesis of Arbitrary 3D Models Based on Concatenation of Di-Viseme Motion Capture Data. Journal of Computer Animation and Virtual Worlds, Vol. 15(5), pp. 485-500.
Snyder, L., Caccamise, D., Wise, B. (2005). The assessment of reading comprehension: A cautionary tale. *Topics in Language Disordes.*Vol. **25**, 1.
Van Vuuren, S. (2007). Technologies that power pedagogical agents - vision and future. *Educational Technology*, 24:1, 4-10.
Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22*, 333-362.
Wise, B. & Van Vuuren, S. (2007) Choosing software gems to improve children's reading. *Perspectives*, 33 (3), 34-42.
Wise, B.; Cole, R.; van Vuuren, S.; Schwartz, S.; Snyder, L.; Ngampatipatpong, N.; Tuantranont, J.; & Pellom, B. (2007). Learning to Read with a Virtual

Tutor. In Kinzer, C. & Verhoeven, L. (Eds). *Interactive Literacy Education.* Mahwah, NJ: Erlbaum.

Books or Other One-time Publications

Caccamise,D., Snyder, L., & Kintsch, E., (2008) Constructivist Theory and the Situation Model: Relevance to Future Assessment of Reading Comprehension. Block, C. C., & Parris, S. R. (Eds.) *Comprehension instruction,* 2nd edition. New York, NY: Guilford

Caccamise, D., Franzke, M., Eckhoff, A., Kintsch, E., & Kintsch, W. (2007). Guided practice in technology-based summary writing. In D. S. McNamara (Ed.), *Reading comprehension strategies* (pp. 375-396). New York: Erlbaum.

Hagen, A., Pellom, B., Van Vuuren, S., and Cole, R. (2004). Advances in Children's Speech Recognition within an Interactive Literacy Tutor, in Proceedings of the Human Language Technology Conference / National Chapter of the Association of Computational Linguistics, Boston, May 2004.

Kintsch, E., Caccamise, D., Franzke, M., Johnson, N., & Dooley, S. (2007). Summary Street: Computer-guided summary writing. In T. K. Landauer, D. McNamara, S. Dennis & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 263-278). Mahwah, NJ: Erlbaum.

Kintsch, W. & Kintsch, E. (2005) Comprehension. In S. G. Paris & S. A. Stahl (Eds.) *Current issues on reading comprehension and assessment*. Mahwah, NJ: Erlbaum. Pp. 71-92.

Olson, R., & Wise B. (2006). Computer-based Remediation for Reading and Related Phonological Disabilities. In M. McKenna, L. Labbo, R. Kieffer, & D. Reinking (Eds.), *Handbook of Literacy and Technology, Vol. 2*. Mahwah, NJ: Lawrence Erlbaum.

Pellom, B. (2001) "SONIC: The University of Colorado Continuous Speech Recognizer", Technical Report TR-CSLR-2001-01, CSLR, University of Colorado, March.

Snyder, L. & Caccamise, D. (Submitted). Comprehension processes: Building meaning and making sense of things. In M. A. Nippold & C. M. Scott (Eds.) Expository Discourse in Children, Adolescents and Adults: Development and Disorders. Mahwah,NJ: Erlbaum/Taylor & Francis.

Van Vuuren, S. (2003). Interface and Method for Performing Tracking and Providing Feedback during Loud or Silent Reading of Electronic Texts. Invention disclosure filed with University of Colorado Tech Transfer Office.

Wise, B. (2004). Facts, Fictions, and Factions in the Reading Wars. In M. Joshi (Ed*), Dyslexia: Myths, Misconceptions, and Some Practical Applications.* Mahwah, NJ: Lawrence Erlbaum.

Wise, B., Rogan, L., & Sessions, L. (In Press). Training teachers in evidence-based intervention: The story of *Linguistic Remedies.* In Rosenfield, S. & Berninger, V. (Eds) *Translating Science-Supported Instruction into*

Evidence-Based Practices: Understanding and Applying the Implementation Process. Oxford University Press.

Presentations:

D. Caccamise, M. Franzke, E. Kintsch & N. Johnson, "Effect of using summary street to improve standardized test scores in comprehension", (2003), presentation/proceedings. Society for the Scientific Study of Reading, Boulder CO

Caccamise, D., Davis, N.,Franzke, M., Johnson, N., Kintsch, E. *Using LSA in the Classroom: Implementing Summary Street in Grades 6-12.* Presented at Annual Conference of the Society for the Scientific Study of Reading, June 2003.

Caccamise, D., Kintsch, W., Cole, R. *The Colorado Literacy Tutor: Making Colorado #1 in Student Achievement.* Invited address to the Colorado Legislature, Feb. 2003.

Caccamise, D., Snyder, L., & Kintsch, W. Improving High Stakes Testing Reading and Writing Outcomes through Summarization. Annual Meeting of the Society for the Scientific Study of Reading, Vancouver, Canada, 2006.

Caccamise, D., Snyder, L. Kintsch, E., Kintsch, W. *Summary Street:*Computer Support to enhance comprehension. International Dyslexia Association, Denver, CO, 2005

Caccamise, Franzke, Kintsch, E. *Summary Street*: LSA-Based Software for Comprehension and Writing. LSA Workshop, Boulder Colorado. 2004

Caccamise, D. , Kintsch, W., Cole, R., Snyder, L. The Colorado Literacy Project. International Reading Association, San Antonio, TX, 2005

Cole, R. van Vuuren, S., Ngampatipatpong, N., Tuantranont, J., Pellom, B., Hacioglu, K., Ma, J., Schwartz, S., Wade-Stein, D., Struemph, T., Yand, J., & Wise, B. (Sept 2003). Using human language technologies to teach reading through interactive books. Poster presented at annual meeting of IEEE.

Davis, N., Snyder, L., Wise, B., Olson, R., Wiley, E. & Finan, D. (Nov, 2005). Evaluating Phonological Sensitivity in Preschoolers: Study of methodology. Poster presented at Annual Meeting of International Dyslexia Association, Denver, CO.

Kintsch, E., D. Caccamise, W. Kintsch, L. Snyder. Workshop on Latent Semantic Analysis and Applications , University of Paris VIII, France, September

W. Kintsch. Presentations on Summary Street:
Bibliography:

- German Psychological Society, Berlin, 2002
- Society for Text and Discourse, Chicago, 2002
- IPMU, Annecy, France; 2002
- EARLI-Writing, Stafford, UK; 2002

- RCA, Ann Arbor MI, 2002
- Winter Text Conference, Jackson, WY; 2003
- LRDC, Pittsburgh PA, 2003
- Queen's University, Kingston, ONT, Canada, 2004
- Society for Computers in Psychology, Minneapolis; 2004
- International Reading Association, San Antonio, TX, 2005.
- Society for Mathematical Psychology, 2005
- University of Würzburg, Germany, 2005
- Symbols, Embodiment & Meaning Workshop, Tenerife, 2005
- Society for text & Discourse, Minneapolis, 2006
- Distinguished Visiting Lecture, Univ. Manitoba, 2007
- LRDC, Pittsburgh; 2007
- Tufts University, 2007
- Cognitive Science Society, Nashville, TN, 2007
- Sorbonne University, Paris, 2007
- University of Paris VIII, 2007

Schwartz, S., Cole, R., Van Vuuren, S., Wise, B., Ngampatipatpong, N., Tuantranont, J., Struemph, T., Snyder, L. & Golding, G. (Oct 2004). *Foundations to Literacy*: Improvements and Adaptations for Students with Cognitive Disabilities. Poster at Coleman Inst. Conference "Enhancing Quality of Life for People with Cognitive Disabilities through Technology." Denver, CO.

Snyder, L. (2004). Constructive comprehension processes in reading and listening. Calvin College, Dept. of Psychology, Grand Rapids, MI, November 2004.

Snyder, L. & Caccamise, D. (2005). Better reading comprehension: Taking students beyond story grammars. American Speech-Language-Hearing Association Convention, November 2005, SanDiego, CA.

Snyder, L. & Caccamise, D. (2005). Evidence based reading research: Lessons learned from www. WhatWorks.ed.gov. American Speech-Language-Hearing Association Convention, November 2005, SanDiego, CA.

Snyder, L. (2005). Promoting better reading with interactive computer based tools. Joint meeting of the Centre for Childhood Communicative Impairment and the Dept. of Experimental Psychology, University of Oxford, April 2005, Oxford, England, UK.

Wise, B. (Aug 2001). Hi-tech talking tutors: Promises and pitfalls. Nijmegen, Netherlands.

Wise, B. (Sept 2002). Effective Teaching for Children with Reading Disabilities. Invited address to the Japanese Association of Learning Disabilities, Tokyo.

Wise, B. (Apr 2003). Beyond Competence: Three paths to fluent independent reading. Invited keynote at annual meeting of Academic Language Therapy Association, Dallas, TX.

Tanaka, Y., Snyder, L., & Wise, B. (Nov 2003). Paper presented at the annual meeting of ASHA, Madison, WI.

Wise, B. (Nov 2003). “Linguistic Remedies.” Invited address as part of symposium on “Temporal Coordination of Orthographic, Phonological, and Morphologic Word Forms: Developmental and Instructional Perspectives on the Real Whole Language” at meeting of International Dyslexia Assoc., San Diego.

Wise, B. (Nov 2003). Facts, Fictions, and Factions in Remedial Reading. Invited address as part of a symposium on “Dyslexia: Myths, Misconceptions, and Some Practical Applications,” at annual meeting of International Dyslexia Assoc., San Diego.

Olson, R. & Wise, B. (Nov 2003). Intervention with Computer-Supported Reading with and without Training in Phonological Awareness and Decoding for 2nd to 5th Grade Poor Readers. Paper presented at Meeting of the Dutch Scientific Organization: Amsterdam.

Wise, B., Schwartz, S., Van Vuuren, S., Snyder, L., Struemph, T., & Cole, R. (Nov 2003). Launching the Colorado Literacy Tutor. Poster at the Meeting of the Dutch Scientific Organization: Amsterdam.

Wise, B.; Snyder, L.; Schwartz, S; Van Vuuren, S; & Cole, R. (June 29, 2004). Interactive Books and Tutors that run “by themselves(?)” in K-2 classrooms. Paper presented at Annual meeting of the Society for the Scientific Study of Reading, Amsterdam.

Wise, B., (Nov 2004). Past and present research with talking computers in Colorado. Paper presented as part of a symposium on “Computers & Reading Disabilities Research & Research-based Applications at the annual meeting of the International Dyslexia Association, Philadelphia.

Wise, B.; Olson, R.; & Cole, R, (July 2005). Computer-Assisted Instruction Informs the Debate and Helps Kids and Teachers. Part of an invited panel on Research to Practice: Advances in Reading and Literacy, at the International Dyslexia Association's Special Focus Conference.

Wise, B. (Aug 2005). Computer-Assisted Instruction Informs the Remediation of Reading Difficulties. Invited presentation at the Summer Institute on Neurodevelopmental Disorders, UC Davis MIND Institute and California State University, Sacramento.

Wise, B. (Nov, 2005). Colorado Reading Technology: Cutting-Edge Technology Weds Well-grounded Research. Introduction to Symposium of ICS and CLEAR researchers presented at Annual Meeting of International Dyslexia Association, Denver, CO.

Wise, B., Snyder, L,. Cole, R., Pellom, B., Van Vuuren, S., Sessions, Tomczyk, T. , Tuantranont, J. Ngampatipatpong, N., Hagen, A. (Nov 2005). Developing an Independent and Adaptive Reading Evaluation, as part of symposium on Colorado Reading Technology: Cutting edge technology weds well-grounded research, at Annual Meeting of International Dyslexia Association, Denver, CO.

Wise, B. (June, 2006). The ‘Scientific Reading’ Revolution, A Crucial time for Teachers and researchers. Invited presentation at the Washington State Dyslexia Summit, Seattle.

Wise, B. & Van Vuuren, S. (July, 2006). Opportunities and Challenges in Computer-reading Research. Invited talk at annual meeting of

International Association of Researchers in Learning Disabilities. In panel on Opportunities & Challenges of On-line Research in Learning Disabilities. Boulder CO.

Wise, B. (July, 2006). The 'Scientific Reading' Revolution, A Crucial time for Lexia. Invited keynote at the annual meeting of Lexia representatives, Providence, RI.

Wise, B. (Aug, 2006). Beyond Competence, to Application, Automaticity, and Transfer. Invited presentation at the Summer Institute on Neurodevelopmental Disorders, UC Davis MIND Institute and California State University, Sacramento.

Wise, B. (Aug, 2006). The 'Scientific Reading' Revolution: Opportunity and Challenge in Education. Invited presentation at the Summer Institute on Neurodevelopmental Disorders, UC Davis MIND Institute and California State University, Sacramento.

Wise, B. (January, 2007) Response to computer-assisted intervention for children with reading difficulties: Year 1. Paper presented as part of a panel on NIH Learning Disabilities Centers, at the Pacific Coast Research Conference, San Diego, CA.

Wise, B. (Oct, 2007). Recognizing, Teaching, and Empowering Children with Reading Disabilities. Keynote address at annual meeting of Young Presidents' Organization, San Francisco, CA.

Web/Internet Site

URL(s):

www.colit.org , lsa.colorado.edu

Description:

portal to using the various components of the Colorado Literacy Tutor, finding out details of the project, and accessing demo/s

Other Specific Products

Product Type:

Library Development: Summary Street Library. Interactive Book library

Product Description:

Library Development:

A library of texts is available for use with Summary Street. A total of 52 texts are organized by topic and grade level. Texts were chosen to represent a broad range of topics that would fit into the Grade 6-12 curriculum: History, Science, Social Science, and Current Issues. Text length ranged from about 500 words to 2000 words, long enough to pose a challenge, but short enough to be doable within a single (45 min) or double (one and a half hour) session. Longer texts had multiple subtopic sections; shorter ones only a single one.

New stories for K-5 level students were created to be put on the interactive

books.
Sharing Information:
via the Colorado Literacy Tutor, specifically in the use of Summary Street and the Interactive Books

Product Type:
Software (or netware)
Product Description:
Summary Street - a fully developed, classroom tested tool to help students write summaries
(Summary Street focus on comprehension instruction in grades 5-12).

Product Type:
Reading Tutors--Interactive Book - *Reading with RITA*
Product Description: (Interactive Books and Reading Tutors focus on reading instructions in grades 1-4 to build decoding skills). This is in a prototype stage and is being used in schools to also build in the assessment function (computerized), which was done by human assessors for this IERI grant. Assessing the students' reading skills is important for proper placement on the tutors and books. This computerized assessment work is being funded by the Dept. of Education (IES grants to Wise).

Sharing Information:
via the colit.org website, which is the portal to using the Colorado literacy tutor components.
Product Type:
Audio or video products
Product Description:
Caccamise Interview with "Colorado Matters" on public radio about the Colorado literacy tutor.
Sharing Information:
This interview is available via url: colit.org or on the Colorado PBS site.

Contributions
Contributions within Discipline:
Have refined LSA underpinnings of *Summary Street* to make it a scalable product. This effort is subject to continuous improvement. Developed reading assessment strategies with on-line presentation that our data shows do predict at-risk readers when they are in kindergarten. These measures are simple to administer and take a short time. Such assessments at this early age have allowed BVSD to assign at-risk children to special reading interventions much earlier than is typical.

We have made continued progress towards the development of the Colorado Literacy Tutor, a software system designed for literacy instruction and currently consisting of *Summary Street* in grades 5 to 12, and the Interactive Book, and various Reading Tutors for reading instructions in grades k-4. During the spring semester of 2003, Reading Tutors have been introduced in 6 classrooms. *Summary Street* has been used in several classrooms and resource rooms.

We have not only introduced the Colorado Literacy Tutor into schools, but we have also evaluated it. Formative evaluation has been performed for the Interactive Book and Reading Tutors. *Summary Street* has been evaluated in a randomized groups design, finding positive transfer effects to relevant items on a standardized test.

Contributions to Other Disciplines:
The Colorado Literacy Tutor has been introduced into several schools. Some of its components have been used widely and repeatedly; experienced teachers have used Summary Street on their own in their classrooms, without outside help. Other parts of the system have been as yet less well tested. Essentially, we are on schedule with this part of our project.

Summary Street is viewed by the teachers that have used it in their classroom as a breakthrough technology for teaching writing and comprehension strategies.

Contributions to Human Resource Development: none, but has the potential to improve literacy for general population of students.

Contributions to Resources for Research and Education:
Summary Street has been shared and similar programs are being developed by researchers in other nations, including Taiwan, Germany, and France as a direct result of this IERI work. In addition, our commercial partner, Pearson (K-tech) has commercialized the product as a stand alone product packaged with their essay grader, and is introducing the software as part of a science text book widely used in the U.S.

Barbara Wise, the content reading expert for *FtL*, has created a new program based on her scientifically supported Ross Reading program (also the basis for *FtL*) called *Reading with RITA*. Stemming from her work on *FtL*, this product has attracted 2 large grants for the Dept. of Ed. (IES) and one from NIH. This current works seeks to automatize the reading assessment process required to place the student in the right level of tutoring on this foundational reading program. (In this IERI project no such automated assessment existed, so for purposes of this project, a cadre of retired teachers were hired to systematically test students in grades K-4 to achieve proper placement in the *FtL* tutors as well has to provide benchmark ability levels and measures of improvement based on nationally standardized reading decoding tests.)

While *Summary Street* is now a commercially available product, the underlying LSA "engine" is being further refined to create educational reading tutors in support of science topics by Kintsch and Caccamise. This new work is funded by the McDonnell Foundation.

Contributions Beyond Science and Engineering:
In order to provide *Summary Street* technology on a broad scale, our subcontractor, Pearson Knowledge Technologies, *Summary Street* is now a part of a commercially available software package called Write to Learn.

Special Requirements
Special reporting requirements: None
Change in Objectives or Scope: None
Unobligated funds: None
Animal, Human Subjects, Biohazards: None
Categories for which nothing is reported:

Contributions: To Any Human Resource Development: These products provide literacy instruction to general population students in grades K-9.

****

Final Remarks on Supplementally Funded Project:

Computer-Assisted Misconception Discovery in the Sciences

The purpose of this supplement to grant REC-0115419, "Scalable and Sustainable Technologies for Reading Instruction and Assessment" was to demonstrate the technical ability of Latent Semantic Analysis (LSA) to classify student essays into predetermined categories that demonstrate particular misconceptions held by students in science. In work performed we have:
1) developed discipline specific corpora for physics/astronomy and for biology. The physics/astronomy corpus consists of textbooks under the Open Content license agreement (http://opencontent.org/opl.shtml) and contains 1465 documents. The biology corpus consists of a high school biology textbook and contains 3317 documents.
2) collected short (40-200 words) student essays for several questions in Physics, Astronomy, and Biology.
3) classified the student answers with two independent experts, and compared the correlation between the classifications of the experts and that of LSA using both a general English corpus (TASA) as well as that same corpus augmented by the corresponding discipline corpus.
4) explored metrics for the prevalence of misconceptions
5) Disseminated our results in conferences and journal publications

Collaborators
Physics and Astronomy:

Courtney Willis, University of Northern Colorado
Robert Walsh, University of Northern Colorado
Mark Moldwin, UCLA
Dick McCray, University of Colorado
Biology:
Michael Klymkowsky, University of Colorado
Richard Cyr, Pennsylvania State University
Cognitive Science:
William Oliver, University of Colorado
Simon Dennis, Ohio State University

Presentations/Publications

The dimensionality of Language, Doxas, I., S. Dennis, and W. Oliver, Proceedings of the National Academy of Science, submitted.

The dimensionality of Language, Doxas, I., S. Dennis, and W. Oliver, Cognitive Psychology 2007, Memphis, TN.

Doxas, I., M. Klymkowsky, K. Garvin-Doxas, and C. Willis, Developing a Space Physics Concept Inventory, American Geophysical Union, New Orleans, LA, 2006 (Invited).

Garvin-Doxas, K., M. Klymkowsky, and I. Doxas, Mass Producing Concept Inventories, American Association of Physics Teachers, Salt Lake City, UT, 2005 (Invited).

Klymkowsky, M., and K. Garvin-Doxas, The Biology Concept Inventory; Developing Tools to Identify Misconceptions, National Association of Science Teachers, Dallas, TX, 2004.

Garvin-Doxas, K., I. Doxas, and S. Dennis, Computer Assisted Misconception Identification and Evaluation in the Sciences, American Evaluation Association, Atlanta, GA, 2004.