Project Report:
MindStar Books

March 1, 2014

Ron Cole, PhD, Boulder Language Technologies

Wayne Ward, PhD, Boulder Language Technologies

Doris Luft de Baker, PhD, Southern Methodist University

Eric Hamilton, PhD, Pepperdine University

Daniel Bolanos, PhD, Boulder Language Technologies

Samantha Messier, PhD, Boulder Valley School District

Eric Borts, Boulder Language Technologies

William Devine, Boulder Language Technologies

GROMINDS: Improving Science Learning and Reading Proficiency

During the fifth year of the project the PI and Co-PI of the MyST project worked with Dr. Julio Lopez-Ferraro, the DRK-12 Program Director, to receive a supplement to the DRK-12 MyST grant to participate in the Science Across Virtual Institutions (SAVI) project entitled *Innovations in Education and Learning* (http://innovationsforlearning.net). The SAVI consisted of 8 teams in the U.S. and 8 teams in Finland who worked together on 8 different projects. The common theme across all projects is **engagement.** The different projects all seek to gain new knowledge about how student engagement in learning tasks can be measured and increased to improve learning. Our project, called GROMINDS, partners researchers at Boulder Language Technologies (Ron Cole, Eric Borts), Southern Methodist University (SMU; Doris Baker) and Pepperdine University (Eric Hamilton) with a team of researchers at AGORA center at the University of Jyväskylä in Jyväskylä Finland (Heikki Lyytinen, Ulla Richardson, Jarkko Hautala, *and Aleksi* Keurulainen). A major focus of this collaboration was development of **MindStars Books**, developed at BLT, and **Graphogame**, developed at the AGORA center, for use in both U.S. and Finnish elementary schools, . Participation by the U.S. team was supported by a one year supplement (September 2012-August 2013) to NSF grant 0733323: *Collaborative Research: Improving Science Learning in Inquiry-based Systems.*

Specific Aims of the GROMINDS Project

GROMINDS was planned as a two year SAVI project with the following objectives:

1. Design and test an initial prototype of the **MindStars System Architecture** (MSA) to provide 24/7 access to the resulting MindStars Science Books. The MSA will be configured as a set of Web services that support both Graphogame services (via servers in Finland) and MindStars Science Books. The goal is to provide web-based learning tools that are platform-independent, so they can be used on desktops, laptops, notebooks and mobile devices. In addition, we envision MSA as a free set of tools and technologies that researchers worldwide can use to replicate and extend our intelligent tutoring systems. We will provide the systems, documentation and tutorials so that others can develop and integrate BLT's speech recognition, natural language processing, dialog modeling and character animation technologies into tutorial dialogs in MindStars Books. These tools and technologies can also be used to develop speech corpora for training speech recognizers in different languages, so that Marni can interact with children in any language for which a recognizer has been developed.

2. Develop and investigate how English and Spanish versions of Graphogame (https://Graphogame.com/) can be used to help English learners and other struggling readers learn to read words in Spanish and English texts accurately and effortlessly. Graphogame has been shown to provide a powerful and flexible tool for helping students acquire *word reading automaticity,* a foundational reading skill that is necessary for reading texts fluently and with good comprehension. When students achieve word reading automaticity, they are able to devote their cognitive resources to making sense of the text; they change from students who are learning to read to students who are reading to learn (LaBerge & Samuels, 1974; Perfetti, 1985) Graphogame has been demonstrated to be a highly effective tool that can be integrated into classroom instruction to significantly improve children's ability to acquire the sound-letter

correspondences needed to recognize words accurately and automatically, with long term benefits demonstrated for students in different countries and languages. Our research will be the first to extend Graphogame to U.S. students, and investigate hypothesized benefits of combining Spanish and English versions of Graphogame to optimize acquisition of *word reading automaticity*, which predicts children's future reading fluency and comprehension.

MindStars Books

The Vision: MindStar Books represents an imaginative new generation of intelligent tutoring systems in science and in reading. Our vision builds on prior generations of intelligent tutoring systems, including significant foundational work carried out by the project team under NSF and IES support. We seek great strides in the quest to immerse students more effectively in multimedia learning activities in which they are challenged, motivated and empowered to acquire the knowledge and skills to learn science.

MindStar Books are thus designed to scaffold effective science learning with the following four aims: 1) They will enable students, especially including English language learners, to acquire the prerequisite vocabulary and concepts to listen to and understand science texts that are read aloud to them by a virtual tutor while they view illustrations that help them visualize the science being explained. 2) They will assess students' understanding of the science through spoken presentation of deep reasoning questions, challenging answer choices representing common misconceptions, and immediate formative feedback on their answer choices. and 4) MindStar Books will engage students in activities that lead to accurate, fluent and expressive reading of grade-level texts; skills that correlate highly with reading comprehension and future reading success (Baker et al., 2008; Fuchs, Fuchs, Hosp, & Jenkins, 2001; LaBerge & Samuels, 1974; Perfetti, 1985; Reynolds, 2000; Samuels, 1997; Stanovich, 2000). These are important and exciting aims based on prior research and development, and they are within grasp.

Scientific Foundations: MindStars Books are based on theory and evidence indicating that a student's ability to read and understand a text– their reading comprehension ability– consists of two component skills: *listening comprehension* and *word reading automaticity*. Listening comprehension is an individual's ability to listen to a text and answer spoken questions about it. Reading fluency is the ability to recognize words accurately and effortlessly. Research shows that students' reading comprehension abilities can be accurately predicted by independent measures of their listening comprehension skills and their ability to recognize words accurately and rapidly(Gough, Hoover, & Patterson, 1996; Gough & Tunmer, 1986; Hoover & Gough, 1990). MindStar Books are designed to help students develop these two essential skills.

MindStars Books Development Efforts

MindStars Books Toolkit: Research conducted during the GROMINDS project resulted in development of the MindStar Books Toolkit, an authoring environment for building, testing and publishing the MSBs. To date, 8 complete books have been developed and tested in schools. An additional eight books are under development. These books are aligned to Next Generation Science Standards, (NGSS), Colorado standards, Texas standards, and FOSS learning objectives for elementary school life science. Our near-term goal is to develop a complete sequence of books using the MSB toolkit for life science that are leveled to grades K-5. All but the kindergarten books will incorporate oral reading fluency training following listening comprehension training.

Design and Organization of MindStars Books

The MindStars Books Toolkit: The MindStars Books Toolkit was developed to provide an easy to use authoring environment for developing the listening comprehension activities in MS Books, and publishing the book in a library. The tool enables an author to (a) type in each sentence Marni will say, (b) record the sentences in English and record the Spanish translation of each sentence, (c) select a picture that will be presented with each narrated sentence (portions of pictures are highlighted using Photoshop), (d) include optional sound files into the narration, (e) design one or more multiple choice questions, with optional illustrations, that are presented after the page has been narrated, and (f) record the questions and answer choices in both English and Spanish. Once the listening comprehension activities have been developed, the oral reading fluency training activities, which follow listening comprehension, are generated automatically, using the text that is narrated by Marni during listening comprehension training. In May 2013, BLT hosted a workshop in which 7 research staff from BLT and SMU and a primary school teacher learned to use the authoring tools to create new books.

Listening Comprehension: In MS Books, Marni narrates each page of a science text while the student views illustrations that help them visualize the science. The narration is self-paced in alignment with research that indicates that self-paced presentations improve learning (Barker, 2003; Cole, Halpern, et al., 2007; Cole et al., 2003) . Students can stop and resume the narration after Marni speaks each sentence, and have Marni repeat the sentence in English or say a Spanish translation of the sentence. After listening to one or more pages of text, Marni presents students with multiple choice questions (MCQs) to assess their understanding of the vocabulary and concepts. These are deep reasoning questions with challenging answer choices that represent common misconceptions. Students can listen to the question and answer choices either in English or in Spanish as often as they like. After selecting an answer, the student receives immediate feedback about the answer they selected. Marni provides positive feedback to a correct answer. If the student selects an incorrect answer choice, Marni scaffolds learning by providing a hint; e.g., that spider has 8 legs, so it can't be an insect. After two tries, the correct answer is presented to the student, along with an explanation as to why the answer is correct. We note that during listening comprehension activities, words are not presented on the page, as the goal is to have students listen carefully while viewing illustrations; research indicates that printed words can distract the student's attention from the illustrations and reduce learning (Cole, Wise, & Vuuren, 2007) .

Oral Reading Fluency (ORF): ORF practice and training occurs immediately after the listening comprehension activities are completed; that is, after all pages of the science text have been narrated to the student and MC questions have completed. The goal of the ORF training is to help students learn to read grade level science texts accurately and fluently; oral reading fluency has been demonstrated to be a strong predictor of reading comprehension and later reading proficiency (LaBerge & Samuels, 1974; Perfetti, 1985; Ward et al., 2011; Ward et al., 2013). Fluency training occurs through repeated reading of each page of the science text. The student is presented with the first page of the text, with each sentence displayed on the page. The student can choose to practice reading the text, with support from Marni, before reading it independently. During practice, the student can listen to Marni read an entire sentence, or pronounce individual words in a sentence. The student can record themselves reading these sentences or words and play back their recordings to compare their reading with Marni's. During

playback of their recordings, each word is highlighted on the page as it spoken by the student. English learners can listen to Marni read a translation of the sentence in Spanish. When the student has finished practicing, they click an icon to read the page independently. Immediately after reading the page, the student receives feedback on the number of words they read correctly (out of the total number of words on the page), and their reading rate (relative to Marni's natural reading rate). The MindStar book highlights words that the speech recognizer scored as misread or skipped, so the student can practice reading these words and sentences. Repeating readings of the page, with practice before each reading and feedback on the student's reading performance immediately after independent reading, continues until the student achieves a criterion level of oral reading performance (90% word reading accuracy, reading speed within 10% of Marni's) or after three independent readings. Repeated reading of texts with feedback and practice following each reading has been shown to be a powerful tool to improve reading fluency, which correlates highly with reading comprehension (Baker et al., 2008; Fuchs et al., 2001; LaBerge & Samuels, 1974; Perfetti, 1985; Reynolds, 2000; Samuels, 1997; Stanovich, 2000).

MindStars Books Pilot Study

An initial set of eight books that supported listening comprehension activities in English, with Spanish translations of Marni's speech, which could be invoked by users, was tested in Kindergarten and first grade classrooms during May, 2013. The books were based on life science themes (e.g., Insects, Life Cycle of the Monarch Butterfly, What do Animals Need to Live?). Students were able to start and stop the (self-paced) multimedia presentation at any point, and repeat the entire presentation if they desired. They could also have Marni repeat questions and answer choices as often as they liked before choosing an answer. When a correct answer was chosen, Marni provided positive feedback to the student, and often expanded on the correct answer. When students chose an incorrect answer, Marni provided a hint (e.g., that spider has 8 legs, so it isn't an insect), and asked the student to choose again. If the second choice was incorrect, Marni explained why the choice was incorrect, and provided the student with the correct answer. English learners who spoke Spanish as their first language also listened to Marni in English during these activities, but had the option of clicking on an icon to hear Marni produce a Spanish translation.

The results of the pilot study provided initial evidence that children were highly engaged in using the books and that they were effective in helping students learn science vocabulary and concepts. Eight kindergarten and first grade students interacted with Marni in 6 to 8 different books. All together, the students were presented with 301 multiple choice questions. Across all students, 273 (90.7%) were answered correctly on their first choice. An additional 26 (8.3%) were answered correctly on students' second choice. While these results are encouraging, additional research is needed to determine if students will retain the knowledge acquired in the books, or will be able to transfer it to new contexts.

One English learner who used the books consistently invoked Spanish translations of prompts by Marni, and these occurred most often during their review of answer choices to multiple choice questions. The student who invoked Spanish translation most often—over 90% of the time during MCQs—was proficient in English. This student shared with us that he enjoyed listening to Marni explain science, ask questions and read answer choices in both English and Spanish because it helped him understand the science.

Summary of MindStars Books R&D

Research conducted during the SAVI project resulted in development of the MindStar Books Toolkit, an authoring environment for building, testing and publishing the MSBs. To data 8 complete books have been developed and published. An additional eight books are under development. These books are aligned to Next Generation Science Standards, (NGSS), Colorado standards, Texas standards, and FOSS learning objectives for elementary school life science. Our goal is to develop a sequence of books that are leveled to grades K-5. All but the kindergarten books will incorporate oral reading fluency training following listening comprehension training.

In addition, collaboration with researchers at the University of Jyvaskyla *resulted in Finnish versions of the eight MSBS developed thus far.* In these books, Marni speaks Finnish, and can produce English translations of Finnish prompts. The development of the Finnish speech recognition system, by the Finnish research team, was an important outcome of the collaboration. The Finnish research team is planning to evaluate the Finnish MSBs in third grade classrooms in Jyvaskyla schools in January, 2014. These books will include an initial version of the oral reading fluency component.

MindStars Books: Hopes for the Future

Development and evaluation of MindStars Books was planned as a two year project with our colleagues at the *University of Jyväskylä. In the context of the NSF SAVI project,* the *Finnish* Funding *Agency* for Technology and Innovation (TEKES), awarded two year research grants to each of the eight Finnish teams. All of the U.S. teams except BLT are funded for the second year of the SAVI project. BLT is actively exploring options for continuing this project. BLT submitted an ambitious proposal to the NSF Cyberlearning program to develop and assess MS Books. The proposal received strong reviews and was judged as competitive, but it was declined. In December 2013 BLT submitted a proposal to the NSF SBIR program to investigate the feasibility of commercializing the MS Books.
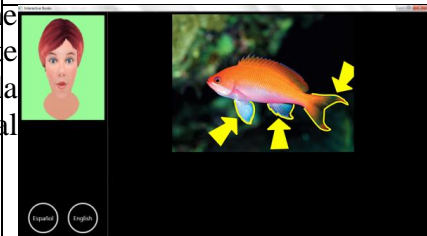
Graphogame

Graphogame was designed to help children acquire foundational reading skills that are taught in first, second and third grades. Many students in U.S. schools fail to acquire the foundational reading skills to read texts fluently and comprehend them. Students who do not learn to read words accurately and automatically have a difficult time reading with comprehension, and are cannot learn from texts as well as their more reading-
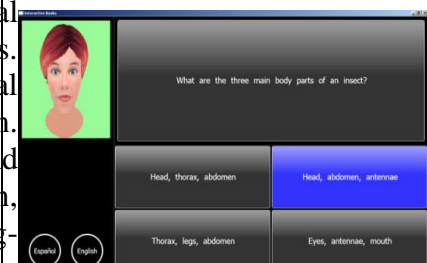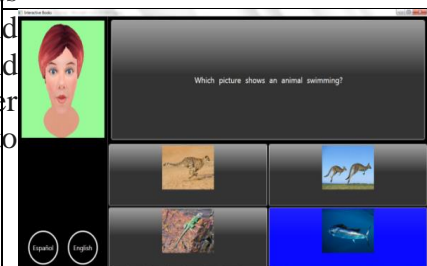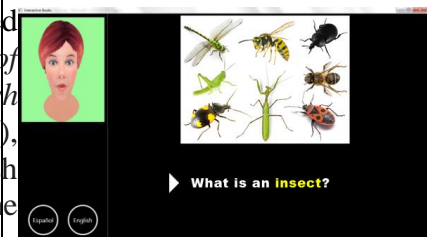
Figure 17 – MindStar Books UI
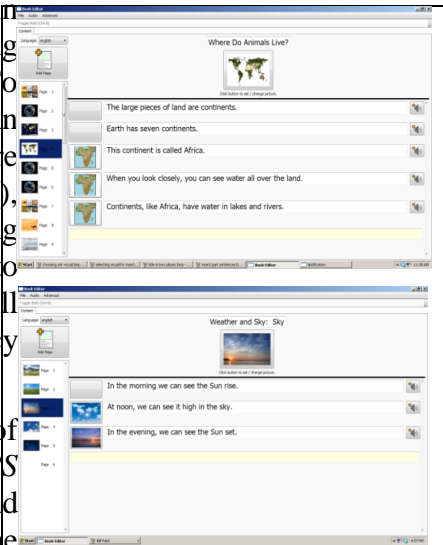
MS Authoring Tool


Marni Narrations

MCQ With Images and Text

6

proficient peers. One of the most fundamental skills in reading is the ability to read words accurately and effortlessly—i.e., acquiring word level automaticity. Graphogame (GG) was designed to enable children to learn sound-letter correspondences in different languages so they can learn to read words accurately and automatically

Graphogame is an online educational software for training and assessing students' reading skills. It was developed by the University of Jyväskylä (Finland) and the Niilo Mäki Foundation (Jyväskylä, Finland). Graphogame is currently available for research purposes outside of Finland. In Finland, the game is called Ekapeli, and it is delivered online free of charge to all learners (see http://info.Graphogame.com/).

GG consists of a number of different activities that build on each other to teach these skills. For example, in one exercise, GG presents young learners with two or more letters (or syllables or words) that fall from the top to the bottom of the screen (on a tablet computer in our research) while an auditory stimulus corresponding to the stimuli is presented to their ears. Their task is to choose (by pointing or shooting) the letter(s) corresponding to the spoken stimulus before the letters reach the bottom of the screen and disappear. Children's responses are recorded and the system adapts to each individual student's performance to optimize learning by adjusting the speed of the falling letters or the number of repetitions of the stimuli before introducing new choices.

Figure 18

GraphoGame Screenshots
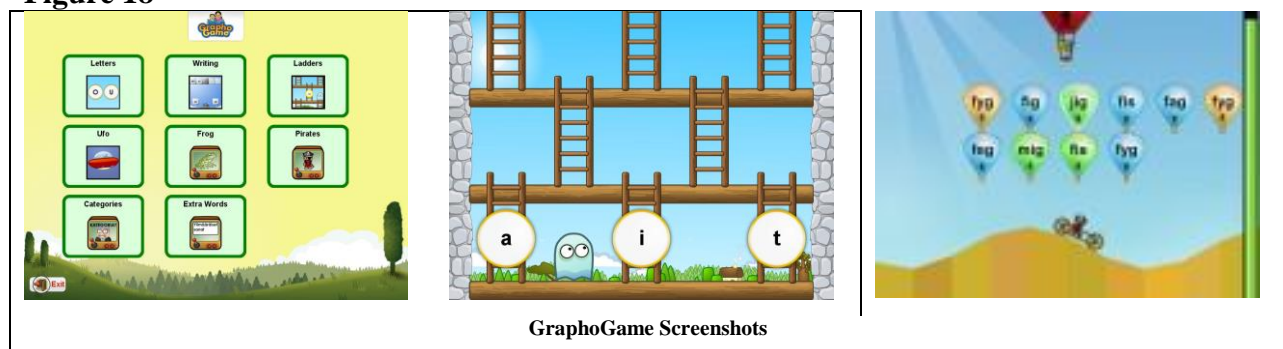
Graphogame has demonstrated success in helping students achieve word reading automaticity, an essential component of reading with comprehension. The game is very intuitive and all children appear to grasp it immediately. GG has been tested successfully by more than 200,000 children in Finland, Zambia, Great Britain, and Chile (Brem et al., 2010; Kujala, Lovio, Halttunen, Lyytinen, & Näätänen, 2012; Kyle, Kujala, Richardson, Lyytinen, & Goswami, 2013; Ojanen, Kujala, Richardson, & Lyytinen, 2013).

The aim of our research using Graphogame-English and Graphogame-Spanish was to optimize learning of word-level decoding skills by English-only students and English learners, so they can recognize English words in texts accurately and automatically. As we have noted, developing word recognition automaticity is a critical skill that children must acquire to read texts fluently and comprehend them.

Research Questions

The aim of the Graphogame study is to provide initial evidence of the usability, feasibility and promise of Graphogame to increase the decoding skills of first grade students in English and in Spanish. Specifically, we aimed at answering the following research questions.

1. Did students who participated in the Graphogame-English condition perform better on letter sound, word reading, and oral reading fluency (ORF) at posttest in English compared to students who did not play Graphogame taking pretest scores into account?
2. Did students who participated in the Graphogame-Spanish condition perform better on letter sound, word reading, and ORF at posttest in Spanish and English compared to students who did not play Graphogame taking pretest scores into account?
3. Did students find Graphogame engaging?
4. Did teachers find Graphogame easy to incorporate into their regular classroom routines?

We also included secondary questions that would help us identify potential moderator and mediator variables that could have affected results such as (a) individual differences in decoding skills at pretest, (b) classroom random assignment, and (c) amount of time playing the game.

GG Summative Evaluation

Research Design

We randomly assigned classrooms within Dallas Texas schools to playing Graphogame 10 minutes/day for 16 weeks versus business as usual instruction. The amount of time per day was determined based on previous studies on the efficacy of Graphogame (Kyle et al. 2013), and on the amount of time it was feasible for teachers to use the game during the English Language Arts block.

Prior to the beginning of the intervention, we met with teachers to explain the project and obtain their consent to participate. A total of 14 first grade classrooms in four schools agreed to participate. We randomly assigned these classrooms within schools to either a treatment condition or a control condition. Given that in one school there was only one English-only classroom and one bilingual classroom, both classrooms were in the treatment group after the random assignment.

Students in the treatment classrooms received headphones and general instructions on how to log into the game. Teachers used a timer to control the amount of time students used the game and to provide all students an opportunity to play the game. Students in the control classroom did not receive any opportunity to play the game during the 16-weeks of the study. However, after that time, teachers in the control condition were also eligible to use the game in their classrooms, if they wanted to.

All students in each of the 14 classrooms were eligible to participate in the study. Five English-only classrooms and three Spanish-English classrooms were in the treatment group and four English-only classrooms and two Spanish-English classrooms were in the control group. Spanish-English classrooms provided reading instruction in Spanish and English every other day to students who spoke Spanish at home as determined by a parent survey.

Participants

Schools: Four schools in a large metropolitan area in Texas agreed to participate. Three of the four schools were located in a high poverty neighborhood. The number of students in each

school ranged from 261 to 620. The fourth school served about 1,300 students in Grades K-12. Table 1 presents information about ethnicity, percentage of students on free and reduced lunch and percentage of students who received special education services at each school. Table 1 in Appendix A2 presents student demographics by school.

Students: 268 students participated in the Graphogame study. Fifty-four percent were male, 66% were economically disadvantaged, and 28% had limited English proficiency. Eighty-three students (31%) of the 268 were Spanish-native speakers and therefore were receiving Spanish and English reading instruction in a one-way dual language model. These students played Graphogame in Spanish.

Reading Instruction: In all schools students received approximately 60 minutes of daily reading instruction following the Journeys core reading program (Harcourt, 2010). The program has also a Spanish version, Senderos (HM, 2010), that was used in the Spanish language arts classroom. Teachers taught the reading programs following the specified scope and sequence, and the instructional guidelines suggested by the publisher. Students who were struggling with English also received supplemental instruction from their classroom teachers in an after-school care environment based on teacher recommendations. Supplemental instruction often included specific decoding and vocabulary lessons.

Professional Development: We trained teachers for half a day on the implementation of Graphogame. Teachers logged on to the game website, played the game, learned the different levels that students would progress through, and received detailed information about the best time to play the game. Teachers asked clarifying questions and received feedback and assistance from the research staff. Teachers and students also received technical support and Graphogame assistance by a research assistant in the beginning of the intervention. Research staff was also on call for questions from teachers during the 16-week study.

American English and American Spanish Graphogame Study Versions

Two different versions of the same Graphogame software platform were used in this study, the Graphogame US Rime English and the Graphogame US Spanish. In both versions the functionality and visuals of the central tasks are the same (i.e., the players are trained to connect spoken sounds to their corresponding written formats in either single letters or a longer sequence of letters). The task of the player is to first listen to a spoken sound, and then select the correct corresponding written target from several alternatives by using a computer mouse. After an incorrect selection, the game immediately shows the correct correspondence for the presented sound. The game never gives any negative feedback on the performance but instead provides the correct answer and/or provides positive feedback on the correct selections. Each game level is relatively short, lasting on average from one minute to three minutes.

Both game versions adapt to the performance of the player. In other words, the game constantly keeps logs and according to the performance in each particular trial, the program provides learning material in the following trials or levels that allow the player to achieve an average of 80% correct items. In addition, similar game levels are presented in several different types of graphic settings allowing students to practice the same task hundreds of times. The design of the game keeps students interested in playing the game and increases their engagement. Below we describe some of the differences between the English and the Spanish US version of Graphogame.

Graphogame English: The original version of Graphogame was developed with the support of British reading researchers (Kyle et al., 2013) and then modified by replacing the sounds of the letters in British English with the sounds of a general American accent recorded by a native American English speaker. The reward system in the Graphogame US English Rime game allows the player to select virtual stickers that can then be placed in a sticker book. The learning content in both of the game versions moves from small units to larger units. The learning content is organized into streams with several levels that explicitly instruct learners on orthographic rime units (Kyle et al., 2013). First, the game introduces the specific grapheme-phoneme connections (GPCs) that form specific rhyming word families. Second, players get to play with rime units thus, providing opportunities for players to reinforce their GPCs skills facilitating the recognition of psycholinguistically relevant reading and spelling units in English. The last step in a stream is to play with words that contain the rime units of the previous levels.). For example, in stream 1, a small set of seven single phonemes and graphemes is introduced (C, S, A, T, P, I, N). Students are then told to put these sounds together to make rime units. Once this is accomplished, students are told to put another sound in front of the rime units presented. The order in which the rime units are introduced is based on the phonological neighborhood density of the rime units according to a constructed database (deCara & Goswami, 2002)..

Graphogame Spanish: The original Spanish US game was developed for an intervention study in Chile. Sounds of letters that seemed unclear or confusing were rerecorded or deleted for the US Spanish version. Learning content in the US Spanish version is organized into dimensions that are listed by the order of difficulty. The game adaptation algorithm works in two phases (i.e., on the accuracy scores in the earlier play sessions, and in the last ten trials of the current run within each content type). The game levels are placed on a map where the player character moves. The performance in each level is rewarded with points that can be used for gaining more power, speed, style or protection for the player character. In addition, the player character gains game tokens that can be used to obtain various objects such as head wear, clothes, and toys.

Measures

Students receiving English instruction only were assessed with three measures. One measure was part of the GG game, the Graphogame US English assessment, and the other two were part of the Dynamic Indicators of Beginning Early Literacy Skills (DIBELS) (Good & Kaminski, 2002) system. Bilingual students receiving Spanish and English reading instruction were assessed in both, English and Spanish. In addition to the three English assessments, bilingual students were also assessed with one Graphogame Spanish assessment, and two measures of the Indicadores Dinámicos del Exito en la Lectura system (IDEL) (Baker, Good, Mross, et al., 2006). Appendix A provides detailed information about the English and Spanish measures.

Student Survey: The purpose of the student survey was to determine student self-reporting of their engagement and ease of use of the game. The student survey was adapted from a previous survey and consisted of 8 questions such as, *How much did you enjoy Graphogame?* To reduce the possibility of students guessing the answer to the questions because of having reading difficulties, a research assistant read the questions and asked students to color in icon that matched their thoughts. For example, a smiley face represented a lot, a straight smiley face represented a little, and a sad face represented none. For those questions that asked for other answers (e.g., When would you play Graphogame?) a picture of a house and a school were given. Appendix A has a sample of the teleform used to collect responses in English and in Spanish). We administered the survey to 144 students. Students could respond to the survey

either in Spanish or in English. In addition, 25% of the 144 students were randomly selected for an interview which allowed us to generalize outcomes for the entire sample in our study. An example of an open-ended question was: *How would you change the game to make it better?*

Teacher Survey: The purpose of the teacher survey was to learn more about the feasibility of Graphogame to be implemented in authentic settings as well as to better understand how teachers perceived student level of engagement in the game. The teacher survey was adapted from a previous survey (Ward et al., 2011). Teachers were assured anonymity in their responses both verbally and in written form. The questionnaire contained 29 questions, some included only rating items, others were open-ended. Seven questions were specifically related to teacher perception of student engagement (e.g., Graphogame impacted students. Potential ratings included *negative, no impact, or positive*). As a follow-up, teachers were asked an open-ended question such as *Please describe in detail the impact of the game for participating students.* Seventeen questions were related to the feasibility of implementing the game in an authentic classroom setting. Four additional questions were taken from a national survey on educational technology and were related to teacher perception of the use of this technology in their classrooms. Appendix XX includes the complete survey with teacher responses.

Fidelity of Implementation

We measured fidelity of implementation based on the amount of time students were engaged in playing the game excluding the amount of time they spent selecting stickers or tokens. Although we asked teachers to spend approximately 10 minutes per day for 16 weeks playing Graphogame, game time varied substantially among treatment classrooms (i.e., from 1.75 hours to 11 hours in the English-only classrooms, and from 4.2 hours to 14 hours in the Spanish-English classrooms). Amount of time playing Graphogame was weakly but significantly correlated with students' scores on the Graphogame Pseudoword ($r = .26$, p < .01) reading and Letter Sound subtests ($r = .35$, p < .01).

Results

Tables 2 and 3 in Appendix A2 present the means, standard deviations, and sample sizes for the reading outcomes of interest by language group and condition (i.e., treatment or comparison). We decided to report results by language group because we wanted to examine whether Graphogame had a differential effect in English on English-only students versus bilingual students given that the bilingual group received less English reading instruction than the English-only group (Baker, Park, & Baker, 2013). Based on the means and standard deviations in English and in Spanish across conditions, we can see that students in both conditions appeared to have made substantial growth in letter-sound recognition, pseudoword reading, and oral reading fluency in English and in Spanish. At posttest, English-only students in the treatment group had higher scores on DIBELS NWF and on ORF. On the other hand, bilingual students in the comparison group had higher scores on all DIBELS and IDEL measures. These differential scores were more apparent when examining the general outcome measures (i.e., DIBELS and IDEL) compared to the Graphogame measures where the difference between pretest and posttest scores was only 1-2 points.

Main Effects

To answer our first two primary research questions, we conducted Analyses of Covariance (ANCOVAs) with pretest scores as a covariate and condition (e.g., treatment or comparison) as a between-subjects factor. We analyzed the data for each outcome separately: DIBELS NWF, DIBELS ORF, IDEL FPS, IDEL FLO, Graphogame Letter Sound Evaluation (English and Spanish), Graphogame Phonology (Spanish), Graphogame Pseudoword Reading (English and Spanish), and Graphogame Word Recognition (English).

Results of the ANCOVAs for the English-only group: ANCOVA analyses indicate that there were no statistically significant differences in the posttest performance of students receiving English-only instruction in the treatment group compared to the comparison group after controlling for pretest differences.

Results of the ANCOVAs for the bilingual group: Similarly, for students receiving bilingual instruction, results indicated no statistically significant differences in the posttest performance of students in the treatment versus the comparison condition after taking pretest scores into account with one exception. The effect of condition was statistically significant for the English Graphogame Letter Sound Evaluation task, $F(1, 74) = 4.36$, $p = 0.04$, with higher adjusted means for students in the treatment group ($M_{Adj} = 17.38$) compared to the comparison group ($M_{Adj} = 15.83$). ANCOVA results also indicated that the Graphogame intervention explained only 5.6% of the variance in the Graphogame English Letter Sound scores for students receiving bilingual instruction.

Moderation Effects

We also tested a number of moderators to determine whether the Graphogame intervention was differentially effective depending on student decoding skills at pretest and amount of time playing the game. Tables 4 and 5 in Appendix A 2 present the results of the interactions between our outcome measures and our moderating variables. Below we describe the nature of these interactions.

Moderating Effects of Risk-Status at Pretest on Outcomes for English-Only Students: We grouped students receiving English-only instruction by risk status within conditions using the published benchmark goals for the DIBELS 6th edition measures. These goals indicate that students who earn a score between 0 and 29 on NWF are considered to be at-risk for later reading difficulties, students who earn a score between 30 and 49 on NWF are considered to be at some-risk for later reading difficulties, and students who earn a score of 50 and above are considered to be at low-risk for later reading difficulties. We present the results for all of the ANCOVAs examining the effect of condition and NWF risk status for all English reading measures in Table 4. For students receiving English only instruction, we found a significant interaction effect of risk status by condition ($F(2, 165) = 3.47$, $p = .03$, $\eta^2 = .04$) favoring students at low risk on NWF at pretest in the treatment group. In other words, level of risk on NWF at pretest moderated the effects of the Graphogame intervention for students receiving English only instruction at posttest. Students in the treatment group who were at low risk on NWF at pretest scored 17 points higher on NWF at posttest ($M_{Adj} = 68.60$) compared to students at low risk in the comparison group ($M_{Adj} = 51.06$).

Moderating Effects of Risk-Status at Pretest on Outcomes for Bilingual Students: Risk status for the bilingual group was determined using students' IDEL FPS score at pretest while for the

English measures risk status was determined using students' DIBELS NWF score at pretest. For students receiving bilingual instruction, results indicated that risk status on NWF at pretest moderated the effects of the Graphogame Spanish intervention for ORF scores at posttest after controlling for ORF pretest scores, $F(2,76) = 3.36$, $p = .04$, $\eta^2 = .08$. This moderating effect, however, appeared to favor students in the comparison group, meaning that the interaction between risk status and condition benefitted students in the comparison group. Scores for students at-risk in the comparison group earned scores approximately 7 points higher than students in the treatment group ($M_{Adj}= 44.70$ and 37.23, respectively) while some-risk students in the comparison group had scores approximately 10 points higher than their peers in the treatment group ($M_{Adj} = 45.26$ and $M_{Adj}=35.80$ respectively). Scores for students at low risk were similar in both, comparison and treatment groups ($M_{Adj} = 45.62$ and $M_{Adj} = 44.99$ respectively).

Results also indicated that, for students receiving bilingual instruction, risk status on FPS at pretest moderated the effects of the Graphogame Spanish intervention for students' scores on the Graphogame Spanish Phonology ($F(2, 45) = 3.19$, $p = .05$, $\eta^2 = .12$) and Spanish Pseudoword Reading ($F(2, 66) = 5.75$, $p = .005$, $\eta^2 = .148$) tasks at posttest after controlling for pretest scores. Results of the ANCOVA for the Graphogame Spanish Phonology task indicate that students categorized as being at low-risk on FPS at pretest had the highest Phonology scores at posttest ($M_{Adj} = 13.90$) compared to students at some risk ($M_{Adj} = 12.90$) and students at-risk ($M_{Adj} = 10.54$). Displaying a similar trend, results of the ANCOVA for the Graphogame Spanish Pseudoword Reading task indicated that students categorized as being at low-risk on FPS at pretest had the highest Spanish Pseudoword Reading scores at posttest ($M_{Adj} = 26.49$) compared to students at some risk ($M_{Adj} = 24.31$) and students at-risk ($M_{Adj} = 16.39$).

Moderating Effects of Spanish decoding skills on English outcomes: As noted earlier, we also conducted ANCOVAs for students receiving bilingual instruction to determine whether their Spanish decoding skills at pretest moderated the effects of the Graphogame intervention on their English literacy performance at posttest. Results from both ANCOVAs – one with NWF Correct Letter Sounds (CLS) as the outcome and the second with ORF as the outcome – revealed that this was indeed the case. For NWF CLS the effect of FPS risk status was significant, $F(2, 75) = 7.58$, $p = .001$, $\eta^2 = .168$) and examination of the adjusted means reveals that students at low-risk on FPS at pretest had the highest NWF CLS scores at posttest ($M_{Adj} = 80.34$), followed by students at some risk ($M_{Adj} = 61.73$). Students at-risk on FPS at pretest had the lowest NWF CLS scores at posttest ($M_{Adj} = 42.29$). In addition, 16.8% of the variance in students' English decoding scores at posttest was explained by their level of risk on FPS at pretest. For ORF the effect of FPS risk status was also significant, $F(2, 76) = 6.051$, $p = .004$, $\eta^2 = .137$, and examination of the adjusted means reveals that students at low risk on FPS at pretest had the highest ORF scores at posttest ($M_{Adj} = 46.09$). Interestingly, students categorized as being at-risk on FPS at pretest had higher ORF scores at posttest than did their peers at some risk ($M_{Adj} = 38.80$ and $M_{Adj} = 36.46$, respectively). Students' level of risk on FPS at pretest accounted for 13.7% of the variance observed in their English oral reading fluency scores at posttest.

Moderating Effects of Playing Time: The amount of time students in the treatment condition played Graphogame varied widely, as evidenced by the descriptors for amount of playing time, in minutes, by language of intervention and school. Given that the amount of time varied so widely from one school to the other and across the two languages, we were interested in whether the amount of time students spent playing Graphogame was related to their posttest performance. Correlations between English playing time and performance on English reading measures were

weak ranging from r = -.04 to .17, and none were statistically significant. Similar relations were observed for students receiving bilingual instruction who played Graphogame in Spanish, with correlations ranging from r - .17 to .36, with only the correlation between playing time and Spanish Letter Sound Evaluation at posttest being significant (r = .36, p < .01). Given that amount of playing time was not correlated with student performance at posttest, we did not examine further the moderating effect of time on the effect of the Graphogame intervention in either English or Spanish.

Student Engagement with Graphogame

 The common theme of the SAVI project is understanding and mediating the relationship between engagement in learning tasks and learning outcomes. To this end, we descriptively analyzed students' responses to the questions in the student survey. Of the students who played Graphogame in English, 93 students (86.9%) indicated they thought Graphogame helped them with reading words a lot, and 88 students (82.2%) indicated they enjoyed playing Graphogame a lot. When asked if they had a choice to play Graphogame at school, at home, or not at all, 70 students (65.4%) who played Graphogame in English indicated they would play it at school, 24 students (22.4%) indicated they would play at home, and three students (2.8%) indicated they would not play it at all. We also asked students to provide information about the amount of time they played Graphogame; 26 students (24.3%) who played in English felt they played Graphogame too much, 69 students (64.5%) felt they played just the right amount of Graphogame, and one student (0.9%) indicated that he/she felt they did not play Graphogame enough. Finally, when asked how they felt about reading after playing Graphogame, 89 students (83.2%) indicated they were more excited about reading, five students (4.7%) indicated they felt the same about reading, and only two students (1.9%) indicated they didn't like reading as much.

Feedback was similarly positive for students who received bilingual instruction and played Graphogame in Spanish. For example, 90% of students (n = 45) indicated they thought Graphogame helped them a lot with reading words and 92% of students (n = 46) indicated they enjoyed playing Graphogame a lot. Seventy-six percent of students (n = 38) indicated that, if they had their choice, they would play Graphogame at school while nine students (18%) indicated they would play Graphogame at home; no students indicated they would rather not play Graphogame at all. Students' ratings of the amount of time they played Graphogame was also positive overall, with 16 students (32%) indicating they felt they played Graphogame too much, 30 students (60%) indicating they felt they played just the right amount of Graphogame, and only two students (4%) indicating they felt they did not get to play Graphogame enough. Finally, 48 students (96%) indicated they felt more excited about reading after playing Graphogame.

Teachers' Perceptions of GG and Student Engagement: Overall**,** teachers impressions of GG were highly positive. Results of the teacher survey indicated that teachers thought students enjoyed playing the game and also showed more excitement about reading after being exposed to the game. One of the teachers said "My students were so excited to login and play Graphogame. Over time, I even saw that my Low-level students were excited to participate and identify letter names and sounds" (see Appendix B). 60% of teachers strongly agreed that Graphogame helped their struggling readers. 100% of teachers believed GG served as a useful tool for different forms of instruction in their language arts class. Finally, 100% of teachers expressed their willingness to continue using Graphogame in their classrooms.

Conclusions

The objectives of the Graphogame study were to investigate a) whether GG was an engaging learning game for students in the United States attending low-income schools, b) whether teachers believed it could be integrated into classroom instruction with benefits to their students, and c) whether using GG improved learning.

Our findings indicate that:

1. Students were able to use the game without difficulty and were highly engaged by it.
2. Teachers reported that they would like to use GG in their classrooms and believed that students would benefit from using it
3. Students who used GG did not increase reading skills we measured relative to students who did not use it. Subsequent analyses indicated that GG appeared to provide more benefit *for children at low-risk for reading difficulties in English and in Spanish* than children who are at-risk for reading difficulties. That is, GG did not help children at risk improve their letter-sound correspondence more than the regular reading instruction provided by the school.

Future Work: The SAVI project was envisioned and planned as a two year project. Year 1 was to focus on developing and investigating feasibility and promise of each project. In year 2, we plan *to integrate MindStars Books and Graphogame into a single reading program.* I think we should also add that the MS Books are really the follow-on of the Graphogame study. The programs, when used together, *are likely to provide a comprehensive and effective reading treatment,* covering word recognition automaticity (GG), fluent reading and comprehension (MSBs).

References:

Baker, Good, Mross, McQuilkin, Watson, Chaparro, & Sanford. (2006). Fluidez en la lectura oral idel *In indicadores dinámicos del exito en la lectura*. Eugene, OR: Dynamic Measurement Group.

Baker, Park, & Baker. (2013). Effect of initial status and growth in pseudoword reading on spanish reading comprehension at the end of first grade. *Psicothema*.

Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui, & Beck. (2008). Reading fluency as a predictor of reading proficiency in low-performing high poverty schools. *School Psychology Review, 37*, 18-37.

Barker, L. (2003). Computer-assisted vocabulary acquisition: The cslu vocabulary tutor in oral-deaf education. *Journal of Deaf Studies and Deaf Education, 8*(2), 187 - 198.

Brem, S., Bach, S., Kucian, K., Guttorm, T., Martin, E., Lyytinen, H., . . . Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *National Academy of Sciences (PNAS), 107*(17), 7939-7944.

Cole, Halpern, Ramig, Vuuren, v., Ngampatipatpong, & Yan. (2007). A virtual speech therapist for individuals with parkinson disease. *Educational Technology, 47*(1), 51-55.

Cole, Vuuren, V., Pellom, Hacioglu, Ma, Movellan, . . . Yan. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human-computer interaction. *Proceedings of the IEEE, 91*(9), 1391-1405.

Cole, Wise, & Vuuren, V. (2007). How marni teachers children to read. *Educational Technology, 24*(1), 14-18.

deCara, B., & Goswami, U. (2002). Statistical analysis of similarity relations among spoken words: Evidence for the special status of rimes in english. *Behavioural Research Methods and Instrumentation, 34*(3), 416-423.

Fuchs, L.S., Fuchs, D., Hosp, M.K., & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.

Good, & Kaminski. (2002). *Dynamic indicators of basic early literacy skills (6th ed.)*. Eugene, OR: Inisitute for the Development of Educational Achievement.

Gough, Hoover, & Patterson. (1996). Some observations on a simple view of reading. In C. Cornoldi, Oakhill (Ed.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahway, New Jersey: Lawrence Erlbaum Associates.

Gough, & Tunmer. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Harcourt, H.M. (2010). from http://www.hmhco.com/shop/education-curriculum/reading/core-reading-programs/journeys

Hoover, & Gough. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Kujala, T., Lovio, R., Halttunen, A., Lyytinen, H., & Näätänen, R. (2012). Reading skill and neural processing accuracy improvement after a 3-hour intervention. In preschoolers with difficulties in reading-related skills. *Brain Research, 1448*, 42-55.

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the united kingdom: Gg rime and gg phoneme. *Reading Research Quarterly, 48*(1), 61-76.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Ojanen, E., Kujala, J., Richardson, U., & Lyytinen, H. (2013). Technology enhanced literacy learning in zambia. *Insights on Learning Disabilities, 10*(2), 103.

Perfetti, C. (1985). *Reading ability*. Oxford, England: Oxford University Press.

Reynolds, R.E. (2000). Attentional resource emancipation: Toward understanding the interaction of word identification and comprehension processes in reading. *Scientific Studies of Reading, 4*, 169-195.

Samuels, S. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly 13*, 107-122.

Stanovich, K.E. (2000). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. In K. E. Stanovich (Ed.), *Progress in understanding reading: Scientific foundations and new frontiers* (pp. 44-54). New York, NY: Guilford Press.

Ward, Cole, Bolanos, Buchenroth-Martin, Svirsky, vanVuuren, . . . Becker. (2011). My science tutor: A conversational multimedia virtual tutor for elementary school science. *ACM Trans. Speech Lang. Process., 7*(4), 18.

Ward, Cole, Bolanos, Buchenroth-Martin, Svirsky, & Weston. (2013). My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology, 105*(4), 1115-1125.

APPENDIX A

English measures

GraphoGame US English Rime Assessment. This measure has three subtests, the letter sound assessment task, rime/pseudoword recognition and word knowledge tasks. In the letter sound assessment the student has 24 trials each including seven letters for which the task is to select the one that corresponds to the presented sound. Each trial was constructed so that the seven alternative letters could be either connected to sounds that are phonetically similar to the target sound and/or that the alternative letters are visually confusable with each other. (e.g., target *o* with distractors *u p e n m a* or the target *d* with distractors *b g p f h n*). For the rime/pseudoword assessment and the word recognition tasks students are presented with 32 trials each. Students were assessed with the same measures at pretest and posttest.

DIBELS Nonsense Word Fluency (NWF). NWF is a standardized, individually administered test of the alphabetic principle. It is a subtest of the Dynamic Indicators of Beginning Early Literacy Skills (DIBELS,(Good & Kaminski, 2002; Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002). Successful performance on NWF indicates knowledge of (a) letter-sound correspondences, in which letters represent their most common sounds, and (b) how to blend letter-sounds into whole units (i.e., pseudowords). According to Good and Kaminski (2002), alternate-form reliability coefficients for NWF ranged from .67 to .87, and concurrent validity coefficients with the readiness subtests of the Woodcock-Johnson Psycho-Educational Test ranged from .35 to .55. Recent studies indicated moderate correlations ($r = .56$) between NWF at the end of kindergarten and the SAT-10 reading comprehension subtest at the end of first grade (Fien et al., 2008; Good, Baker, & Peyton, 2009; Good et al., 2002). In this study, we administered alternate forms of NWF at pretest and posttest.

DIBELS Oral Reading Fluency 6th Edition (ORF). ORF is a standardized, timed, individually administered test of accuracy and fluency (Good & Kaminski, 2002). Oral reading fluency is designed to (a) identify children who may need additional instructional support, and (b) monitor progress toward instructional goals. Reading passages are calibrated for each relevant grade level, and the median number of words students read correctly across three different passages is reported. Students read each passage for 1 min. Words omitted, substituted, and hesitations of more than 3-s are scored as errors. Words self-corrected within 2 s are scored as accurate. In previous studies, alternate-form reliability coefficients of different reading passages from the same level of difficulty have ranged from 0.89 to 0.94 (Good & Kaminski, 2002). In Oregon, the correlation between ORF and the Oregon Assessment of Knowledge and Skills (OAKS) reading measure at the end of third grade was reported as 0.67 (Good, Simmons, & Kame'enui, 2001). In this study, we administered three passages at pretest and three alternative passages of the same reading level at posttest.

Spanish Measures

GraphoGame US Spanish Assessment. Similar to the GraphoGame English assessment, this measure has also three tasks, letter sound knowledge, phonological awareness, and pseudoword recognition. The letter sound assessment task has 21 letter sounds. Each target is presented with all the other letters in a random fixed order. The phonological awareness task has 19 targets. The task is to select the correct picture from a set of three that corresponds to a word that includes the auditorily presented target (i.e., either a phoneme, a syllable, or a word). The pseudoword

recognition task has cut off points (less than two correct in a set of 8 discontinues the task). The length of pseudowords varied from short ones such as *ta* and *le* to longer ones such as *trens* and *frier).*

IDEL Fluidez en las Palabras sin Sentido (FPS). FPS is a subtest of the Indicadores Dinámicos del Exito en la Lectura (Baker, Good, Knutson, & Watson, 2006). It is a standardized, individually administered test of the alphabetic principle and it is similar in structure to the DIBELS NWF in English. An important noticeable difference between NWF and FPS is that on the FPS, CV and CVCV nonsense words were used (e.g., lu, mosi), whereas on the NWF task, VC and CVC nonsense words were used (e.g., ug, lut). In a pilot study, the 3-week, alternate-form reliability of FPS in the middle of first grade was 0.76 (Baker, Good, Peyton, & Watson, 2004). The concurrent validity of FPS with the Woodcock-Muñoz Pruebas de Aprovechamiento subtest of Análisis de Palabras was 0.72 at the end of first grade (Watson & Froyd, 2007). In this study, we administered alternate forms of FPS at pretest and posttest.

IDEL Fluidez en la Lectura Oral (FLO). FLO is a standardized, timed, individually administered test of accuracy and fluency with reading connected text in Spanish. It is a subtest of IDEL (Baker, Good, Knutson, et al., 2006). Passages were written taking into account sentence length, number of high frequency words, and number of letters and syllables in words. Administration and scoring of the measure is the same as those of the DIBELS ORF measure. Alternate-form reliability of different reading passages from the same level of difficulty ranged from 0.88 to 0.94. Criterion-related validity with the Woodcock-Muñoz average score was 0.75. In this study, we administered three passages at pretest and three alternative passages of the same reading level at posttest.

Quality of Instruction

We also observed reading instruction in four control classrooms and five treatment classrooms to ensure that instruction was similar in both conditions, and to record the number of minutes teachers spent on each core reading component. Although, the quality of instruction was not a factor in the analysis of the usability and feasibility of GraphoGame, we were interested in learning more about other behaviors that could potentially moderate the effects of the GraphoGame intervention. The observation instrument we used in this study was adapted from an instrument previously used in another project (Baker et al., inPreparation) and it consisted of 3 parts. The first part included teacher and school information. The second part included 8 items addressing the content of the instruction. The third part included a checklist of teacher behaviors that have been found to be effective instructional practices when teaching beginning reading in either English or Spanish such as providing an explanation of the task, modeling the activity, student opportunities to respond in unison, student opportunities to respond individually, and error correction (Baker, Park, Baker, & Basaraba, 2012; Coyne, Kame'enui, & Carnine, 2011; Vaughn et al., 2006).

For each of the specified behaviors in part 2, the behaviors observed were rated on a 4-point scale: *consistently, sometimes, rarely,* and *never*. We determined that a 4-point scale would be sufficient for us to detect important differences among teachers and for observers to use reliably.

Data Collection

Prior to pretesting, all data collectors received a half-day training on the administration of the DIBELS and IDEL measures by expert trainers. A one-hour refresher training was provided at

posttest. After training, and in the field an administrator and an inter-rater each rated and scored a student assessment independently yet in close proximity to each other. The percentage of agreement between 2 observers when allowing for 1-point discrepancy was 100% on all DIBELS and IDEL measures. Student surveys on their level of engagement was conducted whole group. Interviews were conducted one on one and student responses were written verbatim.

Data Analysis Procedure

To examine the usability and feasibility of the game, we analyzed the results of teacher and student surveys descriptively. To examine the effect of the intervention on student performance on the GraphoGame measures, and the general outcome measures (i.e., DIBELS and IDEL), we conducted analyses of covariance (ANCOVA) on end T_2 outcomes with middle T_1 (pretest) scores as covariates. Although our unit of random assignment was the classroom, and students were nested within classrooms, we analyzed the data at the student level given that we had not enough classrooms to fully power our analysis. When deemed theoretically appropriate, the models were expanded to test moderators (e.g., student decoding skills, classroom, school, and amount of GraphoGame playing time).

Table 5

Demographic Information by School							
	% of White Students	% of Black Students	% of Hispanic Students	% of Other	% of Students on FRL	% of LEP Students	% of SPED Students
School A	2.3	31.8	65.5	0.4	100.0	52.5	5.4
School B	1.1	9.3	88.2	1.4	97.5	41.0	3.8
School C	3.9	12.9	81.5	1.8	92.0	42.7	8.5
School D	7.1	2.9	88.6	1.4	63.0	40.7	5.0
Note: FRL = Free and Reduced Lunch, LEP = Limited English Proficiency, SPED = Students receiving special education services							