Assessing the Effectiveness of Two Theoretically Motivated Computer-Assisted Reading Interventions in the United Kingdom: GG Rime and GG Phoneme

Fiona Kyle
University of Cambridge, UK, and University of Jyväskylä, Finland

Janne Kujala
Ulla Richardson
Heikki Lyytinen
University of Jyväskylä, Finland

Usha Goswami
University of Cambridge, UK

ABSTRACT

We report an empirical comparison of the effectiveness of two theoretically motivated computer-assisted reading interventions (CARI) based on the Finnish GraphoGame CARI: English GraphoGame Rime (GG Rime) and English GraphoGame Phoneme (GG Phoneme). Participants were 6–7-year-old students who had been identified by their teachers as being relatively poor at reading. The students were divided into three groups. Two of the groups played one of the games as a supplement to normal classroom literacy instruction for five sessions per week for a period of 12 weeks. The third group formed an untreated control. Both games led to gains in reading, spelling, and phonological skills in comparison with the untreated control group. The two interventions also had some differential effects. The intervention gains were maintained at a four-month follow-up.

There is increasing recognition of the importance of evidence-based technological tools that can aid students' learning and provide individualized instruction and practice (Beddington et al., 2008; Hasselbring & Goin, 2004; Torgesen & Barker, 1995). If such technological tools are engineered so progression depends on current learning, with slower learners given more learning opportunities, then such learning technologies also enable effective support for struggling learners (Connor et al., 2009). These technological tools also ensure fidelity to the teaching program, enabling them to be used to compare different theoretical approaches to certain kinds of instruction.

Here, we compare two theoretically driven forms of a computer-assisted reading intervention (CARI). Both CARIs were intended to help automatize students' grapheme–phoneme conversion skills for reading and spelling English words. However, the two CARIs varied according to whether they explicitly used rhyme to support grapheme–phoneme instruction. One game was focused on phoneme-level connections between letters and sounds. The second game introduced and reinforced grapheme–phoneme connections (GPCs) via rhyming word families, explicitly focusing on orthographic rime units (the spelling units for rhyming sounds), and demonstrating how rime units and GPCs are related in English spelling.

The two English CARIs compared here were based on a successful CARI for teaching and automatizing letter–sound knowledge

Reading Research Quarterly, 48(1)
pp. 61–76 | doi:10.1002/rrq.038

and phoneme awareness called GraphoGame, which was developed for the transparent orthography of the Finnish language. GraphoGame is a child-friendly, computerized reading intervention program that provides students with letter–sound training. It promotes both phoneme awareness and letter–sound knowledge. It was originally devised by researchers at the University of Jyväskylä in Finland with the aim of free delivery to the end user (see Lyytinen, Erskine, Kujala, Ojanen, & Richardson, 2009; Lyytinen, Ronimus, Alanko, Poikkeus, & Taanila, 2007). The GraphoWorld Network (grapholearning.info/graphoworld) has now been formed to enable games in multiple languages to be developed, also with the aim of free delivery to the end user. However, implementation of GraphoGame in a nontransparent orthography such as English is more challenging than implementing GraphoGame in Finnish, and this challenge led directly to the small-scale study reported here.

For the current study, GraphoGame was adapted for the nontransparent English orthography by English-speaking researchers in two independent research centers. In one center (Finland), GraphoGame was adapted according to the theoretical view that an approach of "small units first" is most effective for learning English phonics (Hulme et al., 2002; Seymour & Duncan, 1997). This CARI, hereafter GraphoGame Phoneme (GG Phoneme), was created by adapting the Finnish CARI format (based on instruction in single GPCs) directly to the English orthography. GG Phoneme used the Finnish method of first introducing all the possible GPCs in the spelling system, including digraphs, but ordered them so the most frequent, consistent, and prototypical GPCs in English were introduced first (Erskine & Seymour, 2005; see the Methods section for more detail).

In a second center (Cambridge, UK), GraphoGame was adapted according to the theoretical view that English-speaking students may benefit from a focus on oral rhyme and rhyme analogies as part of reading instruction (Goswami & Bryant, 1990). This CARI, hereafter GraphoGame Rime (GG Rime), was created on the basis of rhyme families in English, ordered so the largest families with the most consistent orthographic rime spellings were introduced first (De Cara & Goswami, 2002). The effectiveness of these two CARIs was then compared directly in participating U.K. schools.

CARIs are not intended to be used in place of regular teaching but as a supplementary tool, enabling individually adjusted practice in component skills. We therefore expected that the supplementary practice offered by both CARIs would improve the automatization of phonics knowledge for English-speaking students. Currently, U.K. schools are required by the government to use a synthetic phonics approach to teaching GPCs that is currently considered best practice in the United Kingdom (Rose, 2006). Synthetic phonics is a teaching method in which students are shown that words can be read "by saying each letter sound in a word distinctly from left to right, [and] joining them together smoothly without pausing between each sound" (Johnston & Watson, 2004, p. 347). U.K. students are usually exposed to synthetic phonics from entering school at age 5, practicing synthetic phonics skills discretely (i.e., not as part of story reading), and are also explicitly taught concepts such as phoneme and rhyme.

In the study reported here, both CARIs were introduced during the second year of schooling. Hence, all participants had already experienced a year of synthetic phonics instruction, and all participants (both of the intervention groups and the untreated controls) continued with this instruction during the intervention. During the year of the study, classroom phonics instruction was already using complex digraphs. The CARIs were delivered during extracurricular time (e.g., school lunch break), meaning that the CARIs were supplementary to this regular classroom teaching. Therefore, while GG Phoneme essentially offered more focused and individually calibrated practice with synthetic phonics, GG Rime offered a supplementary way of grouping letter patterns and learning the English spelling system. Both games are described in more detail in the Methods section.

Currently, it is unknown whether a CARI like GraphoGame can be an effective learning tool for phonics instruction in less transparent orthographies such as English. However, as the development of efficient and automatic phonics skills plays a critical role in early reading development and as automatized phonics knowledge is an important bedrock for long-term reading development (Ehri, 1998; Juel & Minden-Cupp, 2000), it is important to develop these skills in students who are learning to read English (Wyse & Goswami, 2012). Most alphabetic European languages are easy to implement in a CARI because they have high orthographic consistency with simple rules that determine grapheme-to-phoneme mappings (Brem et al., 2010; Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011). In contrast, languages with inconsistent orthographies, such as English, French, Portuguese, and Danish, have a 1 : many mapping between letters and phonemes, making implementation in a CARI more challenging.

Both CARIs compared here built on prior research showing that the most effective interventions for English-speaking students are those that combine training in phonological skills with explicit training on the links between letters and sounds (e.g., Ball & Blachman, 1988, 1991; Bradley & Bryant, 1983; Hatcher, Hulme, & Ellis, 1994; Hatcher et al., 2006; Torgesen, Wagner, Rashotte, et al., 1999). For example, an important study by Hatcher et al. (1994) carried out in the United

Kingdom showed that students who received a combined reading plus phonology intervention made significantly larger gains in reading accuracy, comprehension, spelling, and phonological awareness than two other intervention groups who received training in either only reading or phonology. In Hatcher et al.'s study, which involved a 20-week intervention, students receiving the combined intervention gained 0.31 standard score (SS) points per hour of training on a standardized assessment of reading accuracy (see Hatcher, 2000).

Hatcher, Hulme, and Snowling (2004) reported a similar result for 410 U.K. beginning readers who were divided into one of four groups: reading with rhyme, reading with phoneme, reading with rhyme and phoneme, or reading. For typically developing students, there were no significant differences between the four interventions, and the effect sizes for reading and nonword reading were negligible (ranging from −0.18 to 0.13). In contrast, students designated as at risk for reading difficulties on the basis of very low initial scores in reading and phonology as measured in the study showed larger effect sizes on reading for the phoneme-based intervention (0.53) than the rhyme-based intervention (0.41). Nevertheless, the combined rhyme and phoneme program led to the greatest gains in word reading for these students (effect size 0.59).

The work by Hatcher and colleagues is most relevant to the CARIs compared here, which also contrasted rhyme and phoneme approaches to teaching. However, GG Rime explicitly segmented orthographic rime units into GPCs and, therefore, could be seen as most similar to Hatcher et al.'s (1994) rhyme and phoneme intervention. Use of a CARI nevertheless provides a novel perspective on the effectiveness of these theoretical approaches to learning because the instruction was supplementary to normal classroom instruction (see also Walton, Bowden, Kurtz, & Angus, 2001; Walton & Walton, 2002; Walton, Walton, & Felton, 2001). In the research reported by Hatcher and colleagues, the interventions were delivered by classroom teachers who had been trained by researchers (Hatcher et al., 1994, 2004).

In contrast to the reliable gains found in studies providing training by classroom teachers, studies looking at the efficacy of providing CARI have to date produced fairly mixed findings (e.g., Dynarski et al., 2007; Torgesen, Wagner, Rashotte, Herron, & Lindamood, 2010; van Daal & Reitsma, 2000). For example, Dynarski et al. conducted a large-scale evaluation of five different CARIs and found no significant advantage for the classrooms using CARI in comparison with control classrooms. However, studies using GraphoGame or other types of CARI have provided more positive findings for the use of CARIs, particularly for students identified as being at risk of developing reading problems (Saine et al., 2011; Torgesen et al., 2010).

Torgesen et al. (2010) reported that students who received CARI performed significantly better on tasks measuring reading accuracy, reading comprehension, phonemic awareness, and decoding than did students who received normal school-based reading intervention. The CARI intervention groups made 0.27 SS gains per hour of the intervention. This study was actually a combination of traditional intervention and CARI because the intervention was only provided by computer for half of the training period to reinforce intervention teacher-introduced activities . The teachers provided training on concepts that were then reinforced by activities on the computer.

Saine et al. (2011) looked at the effectiveness of the CARI used here, GraphoGame, with Finnish beginning readers who were at risk of developing reading difficulties due to their low prereading skills. Saine et al. reported that at-risk students who received GraphoGame in combination with a teacher-based intervention made significantly greater progress in letter knowledge, reading, and spelling skills than students who solely received individual reading intervention (effect sizes 2.08, 1.01, and 1.68, respectively). Moreover, approximately 16 months after the intervention ended, the reading accuracy, fluency, and spelling skills of the at-risk students who played GraphoGame were commensurate with those in the mainstream classroom, whereas the students who received the traditional intervention still exhibited delays.

The Saine et al. (2011) results suggest that GraphoGame is a remarkably effective CARI for students learning to read in Finnish. GraphoGame has also been extended successfully to German, another transparent orthography (e.g., Brem et al., 2010; Hintikka, Aro, & Lyytinen, 2005; Huemer, Landerl, Aro, & Lyytinen, 2008; Saine et al., 2011). For example, Huemer et al. studied the effects of a six-week German GraphoGame training period on reading accuracy and speed in Austrian second and fourth graders. The students improved in reading accuracy and response times of the trained items. In Switzerland, Brem et al. conducted an intervention study with prereading students using another version of German GraphoGame (with comparable content to the GG Phoneme game used in the present study). After a short training period, which averaged just under four hours over eight weeks, significant improvements were shown in students' letter knowledge. Their reading skills also improved slightly.

The main aim of the present small-scale study was to assess the efficacy of GraphoGame as a supplementary CARI for students learning to read in English. We were unable to replace typical classroom phonics instruction with the CARIs or to introduce them at the beginning of literacy instruction at age 5 because of U.K. government policy concerning synthetic phonics. Therefore, teachers in the participating schools were asked to identify students who they felt were experiencing difficulties during their second year of reading instruction and whose parents might be willing for them to receive supplementary (e.g., during their lunch hour) literacy instruction. These students were invited to participate in the study.

From extant research findings, it was predicted that both GG Rime and GG Phoneme would lead to improvements on word reading, nonword reading, and spelling in comparison with a control group who only received normal classroom instruction. A reading intervention is considered effective if the effect sizes are greater than 0.13–0.23 (see Torgesen et al., 2001). It was also predicted that GG Rime would lead to training effects on multiple components of phonological awareness, whereas GG Phoneme would lead to more selective training effects on phoneme-level phonological awareness because of the content of each game. Of interest was whether either GG Rime or GG Phoneme might provide a more effective supplementary intervention and whether differential effects for the two games might be seen in a long-term follow-up. Because regular reading activities in U.K. schools in effect train the learning encapsulated in GG Phoneme, better long-term progress subsequent to receiving training might be expected a priori for the students who played GG Phoneme because for these participants, the synthetic phonics principles trained by the game would be reinforced by normal classroom activities.

Methods

Design

Two different CARIs, GG Rime and GG Phoneme, were provided as a supplement to ongoing classroom literacy instruction. Two schools agreed to participate, and the opportunity to play the game was offered to all the students identified informally by their teachers as being likely to benefit from supplementary instruction. All students whose parents returned consent forms were included in the study. Because an insufficient number of students returned consent forms in the first year of the study for the complete design, a further group of students from the same schools (same classes and teachers) were recruited the following year, and some of these formed the untreated control group.

The students who played the CARIs were assessed immediately before the start of the intervention to match the two treatment groups (T1, pretest, October) and again at the end of the 12-week intervention period four months later (T2, posttest, February). To measure the durability of the intervention effects on reading, spelling, and phonological awareness, the students were also reassessed four months after the end of the intervention (T3, follow-up, July). The untreated controls were assessed at the same time points in the following school year. All pretest (T1), posttest (T2), and follow-up (T3) assessments were conducted individually in a room adjacent to the classroom by the first author, who was trained in administering standardized assessments. The tests were delivered in short 20-minute sessions so students who needed more time for a particular test did not get more fatigued than other students. All students gave their assent prior to testing, and the study was approved by the Psychology Research Ethics Committee of the University of Cambridge.

Participants

Thirty-one second graders ages 6–7 participated in the study. There was an additional participant who began the intervention but withdrew from the study after only four weeks because of moving to another school, and therefore the student's data were not included. The students were recruited from two schools in Hertfordshire, UK, each of which had a single class of the target age. The teachers were asked to nominate students who they felt were poor readers but without any additional educational needs. To our knowledge, the teachers did not use any formal method of assessment to select these students. As shown in Table 1, at T1, all students achieved scores within the normal range on an estimate of nonverbal IQ using the Matrices subtest of the British Ability Scales II (BAS II; Elliot, Smith, & McCulloch, 1996) and on a test of expressive vocabulary, the British Picture Vocabulary Subscale II (Dunn, Dunn, Whetton, & Burley, 1997).

In the first recruitment year, half of the students volunteering from each school were allocated into each of the two intervention groups: GG Rime (n = 11), GG Phoneme (n = 10). Due to the small sample size, students were not completely randomly allocated to each intervention group; instead, two groups from each school were created by matching the students in terms of chronological age, nonverbal IQ, expressive vocabulary, and reading ability, and then these two groups were randomly allocated to either intervention. The students in the untreated control group were nominated by the same teachers a year later as part of a different study, in which all students nominated were randomly assigned to intervention or control groups. Only the control group from that second study is included in this article.

There were no significant differences between the three groups on any of the assessments at pretest (see Table 1). The groups were not formally matched for ethnicity or other demographic variables, but the students were all of white British ethnicity and following the same reading curriculum, and English was their first language. Because the two schools drew from homogenous

Table 1
Group Characteristics

Characteristic	GG Rime	GG Phoneme	Control	One-way ANOVA by group
	Group			
N	11	10	10	—
Gender (female, male)	5, 6	5, 5	5, 5	—
Age (years;months)	6:7 (3.9)	6:8 (4.2)	6:8 (3.3)	$F(2, 30) = 0.12$, ns
Nonverbal IQ (T score)	50.0 (7.7)	50.6 (7.8)	45.2 (6.6)	$F(2, 30) = 1.61$, ns
Playing time (minutes)	674 (33.2)	644 (56.8)	—	$t(19) = 0.82$, ns
Playing days (days)	44.9 (2.2)	42.9 (3.4)	—	$t(19) = 1.50$, ns
Accuracy on levels	84% (6.4)	82% (8.1)	—	$t(19) = 0.76$, ns
Measure	**GG Rime (*SD*)**	**GG Phoneme (*SD*)**	**Control (*SD*)**	
BPVS II SS	106.1 (9.2)	106.9 (11.7)	101.9 (12.5)	$F(2, 30) = 0.58$, ns
BAS II Reading SS	99.6 (11.3)	100.7 (12.8)	99.0 (11.3)	$F(2, 30) = 0.05$, ns
BAS II Spelling SS	104.6 (7.0)	101.6 (9.0)	105.2 (6.9)	$F(2, 30) = 0.64$, ns
TOWRE sight word SS	103.9 (10.1)	101.2 (9.5)	101.1 (8.3)	$F(2, 30) = 0.31$, ns
TOWRE nonword SS	99.9 (6.7)	101.9 (9.4)	97.9 (13.5)	$F(2, 30) = 0.39$, ns

Note. BAS II = British Ability Scale II. BPVS II = British Picture Vocabulary Subscale II. GG = GraphoGame. ns = not significant. *SD* = standard deviation. SS = standard score. TOWRE = Test of Word Reading Efficiency.

neighborhoods, and each contributed similar numbers of students to each group, demographic variables were likely to be similar across groups. Participant characteristics and descriptive statistics for all standardized measures at pretest are provided in Table 1. A series of one-way ANOVAs with Bonferroni corrections revealed no significant differences between the three groups on any of the measures at T1 (see Table 1 for *F* values).

Assessment Battery

Vocabulary

Students were assessed for their receptive vocabulary using the British Picture Vocabulary Subscale II. This was only measured at pretest. The following tests were administered at all three assessment periods: pretest (T1), immediate posttest (T2), and four-month follow-up after the end of training (T3).

Reading

The students completed two standardized assessments of reading. The Single Word Reading subtest from the BAS II measured single-word reading accuracy. The students were required to read aloud single words of increasing difficulty, without context. The maximum score was 90. The test was administered according to the instruction manual. It should be noted that the BAS II was standardized before the National Literacy Strategy was introduced in the United Kingdom in 1997 and that typically developing students in the United Kingdom now usually score well above a standard score of 100 (see Kuppen, Huss, Fosker, Fegan, & Goswami, 2011). Cronbach's α for the raw scores on this test across the three test points was .976.

The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) consisted of two subtests measuring speeded recognition of words and nonwords. The students were required to read aloud as many words or nonwords as quickly and accurately as possible in 45 seconds. Practice words were provided for each subtest. The maximum score for the word section was 105, and the maximum score for the nonwords section was 63. Cronbach's α for the raw scores on the word recognition test across the three test points was .965. Cronbach's α for the raw scores on the phonemic decoding efficiency test across the three test points was .952.

Spelling

The spelling subtest from the BAS II was administered. This is a spelling to dictation (i.e., the teacher read the word aloud, and the student wrote it) task containing a mixture of verbs, nouns, and adjectives. The test was administered according to the instruction manual. The maximum score was 75. Cronbach's α for the raw scores on this test across the three test points was .918.

Phonological Skills

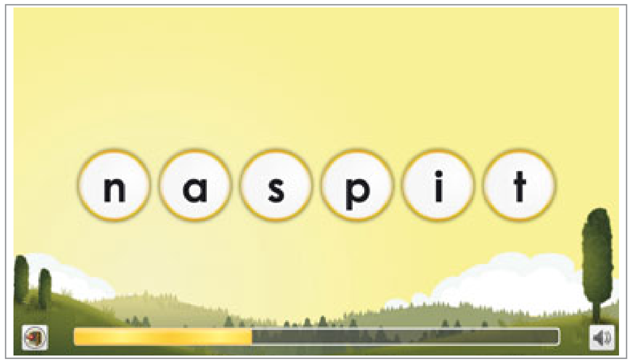
The students completed two experimental phonological awareness tests: one measuring phonological awareness at the level of the phoneme (phoneme deletion task) and the other measuring rhyme awareness (rhyme oddity task). The phoneme deletion task was taken from Cain, Oakhill, and Bryant's (2000) study. The students were required to say words without certain sounds. There were four sections of the task, and practice items were provided at the beginning of each section. In the initial sound and final sound sections, the students were required to delete the initial sound (e.g., "What is *crush* without the /k/?") or the final sound (e.g., "What is *find* without the /d/?"). The other two sections contained items from which the students had to delete the medial sounds, either from the onset (e.g., "What is *grow* without the /r/?") or the coda (e.g., "What is *nest* without the /s/?"). The maximum score was 24. Cronbach's α for this test across the three test points was .910.

The rhyme oddity task was taken from Thomson and Goswami's (2008) study and measured rhyme awareness. The task was presented on a laptop and used digitized speech created from a native female speaker of Southern British English. Using headphones, the students listened to sets of three words and had to select the word that did not rhyme with the other two (e.g., "*wag*, *nag*, *that*"). There were 20 trials, and the order of presentation was counterbalanced across participants. Cronbach's α for this test across the three test points was .784.

GraphoGame Interventions

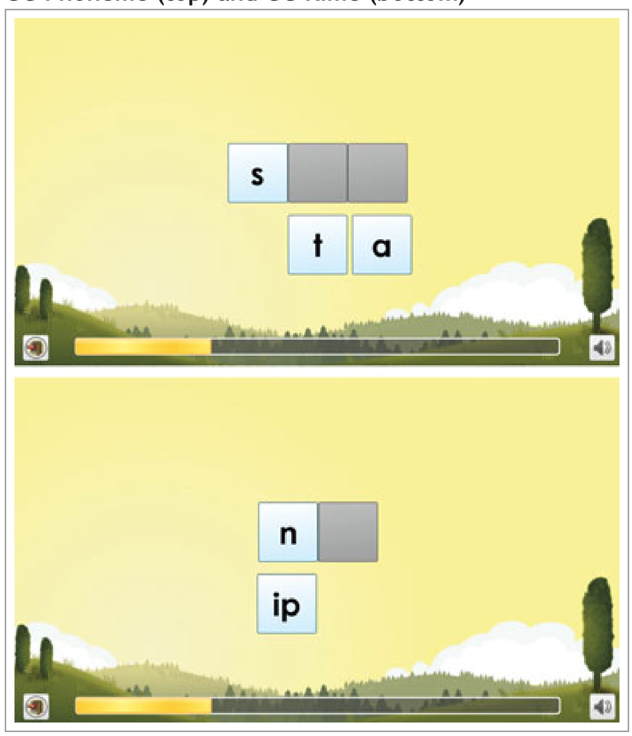
Both GG Rime and GG Phoneme were presented using the same software platform. Both games provide highly repetitive and individualized intervention in which the player hears auditory targets consisting of either sounds or words and has to match these auditory targets to visual targets (letters and sequences of letters) displayed on the screen. The letters and letter sequences are contained within balls that fall downward from the top of the screen (see Figures 1 and 2). The player has to catch the target ball(s) by clicking on it with the computer mouse before the ball(s) reaches the bottom of the screen and disappears.

FIGURE 1
An Example of the Screen in Stream 1 for Both GG Rime and GG Phoneme



Note. GG = GraphoGame.

FIGURE 2
Examples of the Screens in Word-Forming Tasks From GG Phoneme (top) and GG Rime (bottom)



Note. GG = GraphoGame.

The students progressed through a series of graduated game streams, each of which has multiple levels. To keep motivation levels high, the students were rewarded with tokens at the end of each level within a stream. Every few levels, the tokens were swapped for special reward games in which the same content is taught using a more exciting background and format. For example, there is a race game in which the targets appear in cars instead of in falling balls, and the students had to click on the correct target before the cars raced off the screen. Another special game is a ghost-and-ladder game in which the students had to click on the correct target to move the ghost up a ladder, and if they were incorrect, the ghost would drop down a ladder. There are also word formation games to encourage spelling skills, in which the students were presented with boxes containing letters or onset and rime patterns and were asked to put them into the correct order to spell target words (e.g., GG Phoneme: *c-a-t*; GG Rime: *c-at*).

Because both games are adaptive, the letters and letter sequences practiced by different players varies depending on speed of progression through the games. Overall, the two games teach the same GPCs, but GG Phoneme introduces all possible correspondences before blending them into words, and GG Rime introduces various correspondences of different sizes (small and large units) at different points in time, depending on the rhyme families being used for a particular stream.

Both GG Rime and GG Phoneme use a success criterion of at least 80% for each level before the player can move onto the next level. If a student failed to achieve 80% accuracy on a level, they were given an individualized extra training level in which the computer automatically selected targets that the student knew and contrasted them with targets that the student did not know, and then the previous level was presented again. The words for GG Rime and GG Phoneme were recorded by the same female speaker, who has a British accent. Words were digitally recorded in an anechoic chamber and normalized for sound during the editing process.

GG Phoneme

GG Phoneme teaches letter–sound correspondences at the level of the single phoneme and grapheme. First, all the single phoneme–grapheme correspondences (PGCs) in English spelling are introduced (e.g., *i, a, ee, oa*) during streams 1 and 2. In stream 3, phonemes are combined into consonant-vowel (CV) units (e.g., *ti*, *loa*). In stream 4, phonemes are combined into VC units (e.g., *is*, *eech*). The consistent presentation sequence is to focus first on PGCs (e.g., "Let's play with these sounds"), then blend these PGCs into larger units that are not yet words (e.g., "Let's put two sounds together"), and then create words from PGCs (e.g., "Let's put more sounds together to make words"). From stream 5 onward, the player is told, "Here are some words with the sound X." The player is now shown only whole words, not CV or VC units, and asked to identify and isolate GPCs or PGCs within the whole words or blend GPCs or PGCs into whole words as part of the gaming activities.

The activity sequence for GG Phoneme was adapted directly from the original Finnish game (see Lyytinen et al., 2007, 2009). The theoretical framework drew on approaches arguing for the importance of small-unit instruction in reading in English (e.g., Hulme et al., 2002; Seymour & Duncan, 1997) and on research assessing grapheme–phoneme consistency in European orthographies, including English (Seymour, Aro, & Erskine, 2003). For Seymour et al., a database for the English orthography had been created on the basis of all the reading instruction books used in Scottish English classrooms for teaching reading in grades 1–3. Seymour et al. used the database to classify English GPCs in these reading materials as either simple or complex.

In creating the English version of GG Phoneme, a quantitative (mathematical) definition of consistency developed by the second author was used and applied to this same database. Based on this definition, and also on frequency and pronunciation information from the Celex database, a consistency database was computed for English spelling, which gave an index of pronunciation consistency for all possible strings of letters and an index of spelling consistency for all possible strings of phonemes. This database was then used to order the GPCs for GG Phoneme in terms of which were the most frequent, consistent, and prototypical, drawing on the work of Erskine for the prototypicality designation (e.g., Erskine & Seymour, 2005).

Potential words for the game (drawn from the books used for teaching initial reading in Scotland, i.e., those supplied by Seymour et al., 2003) were ranked by frequency of occurrence, and only middle-frequency words were selected for the game. The consistency and complexity of PGCs were controlled for in terms of order of presentation during the game. The most frequent, most consistent, and most prototypical GPCs are introduced first and also reinforced first during later game streams. For example, simple graphemes such as *P*, *B*, *D*, and *K* have over 90% consistency and are reinforced early in the game. Complex vowel digraphs are introduced in stream 2 but are reinforced later in the game. While faithfully reproducing the statistical frequency of occurrence of the different GPCs, this approach introduces CV units such as *wa* and *wi* quite early, grapheme combinations which are not psycholinguistic units (i.e., they are not oral linguistic structures that are already available to beginning readers; see Ziegler & Goswami, 2005). (All the item streams used in GG Phoneme are available for research purposes from the GraphoWorld Network.)

GG Rime

GG Rime is based on the intrasyllabic unit of the rime, which is argued to be an important psycholinguistic unit for English-speaking students (Goswami, 1999; Ziegler & Goswami, 2005). The theoretical framework for GG Rime was based on the early work of Goswami and colleagues on rhyme analogy and rime-based teaching (Goswami, 1986, 1988, 1990, 1993, 1999, 2001; Goswami & East, 2000) and updated with reference to subsequent work on phonological rime neighborhoods (De Cara & Goswami, 2002) and orthographic rime neighborhoods (Goswami, Ziegler, & Richardson, 2005), as well as research on implicit statistical learning of rhyming sounds and rime spellings by students (e.g., Goswami, 1999, 2001; see Goswami, 2012, for a recent summary). Rhyme analogies are inferences that words with shared spellings for orthographic rimes will be pronounced to rhyme with each other, and Wyse and Goswami (2012) have summarized how the rhyme analogy and neighborhood density work can be integrated for the teaching of reading.

The teaching sequence in GG Rime was based consistently on orthographic rime units. The students were introduced to single letter–sound correspondences, which were blended into orthographic rime units and then into CVC words (the blend was C + VC). For example, in stream 1, a small set of seven single phonemes and graphemes was introduced (*C*, *S*, *A*, *T*, *P*, *I*, *N*), and the students were told, "Let's put these sounds together to make rime units." First, only rime units that were also real words were created (e.g., *at*, *in*) and reinforced. The students were then told, "Now let's put another sound in front of the rime units you have just played with," and CVC words like *cat* and *tin* were created by showing blending of *c* + *at* and *t* + *in*. The students were reinforced on the GPCs in these CVC words (e.g., "The sounds in *tin* are /t/, /i/, /n/").

Subsequently, orthographic rimes that were not also real words were created, such as *op* and *ag*, enabling the creation of CVC words, such as *hop* and *bag*. The teaching sequence was to show some GPCs ("sounds"), blend these into rimes, blend onsets onto these rimes to create words (the term *sounds* was used for *onset* in the game), and then segment the words back into GPCs. Hence, consistently during each level within each stream, the students were shown how to create a word from onset plus rime (C + VC), played some activities based on matching rhyming words, and were shown the same words being segmented into GPCs (e.g., "*p* + *ad* = *pad*," then "*pad* = *p-a-d*").

Rhyme family groupings were used to exploit the theoretically critical role of rime as a psycholinguistic unit for English reading and spelling development (e.g., Bradley & Bryant, 1983; see Goswami, 2012, for a discussion of how Bradley and Bryant's early insight into auditory organization has influenced subsequent rime neighborhood density research). The use of rhyme families enables GG Rime to highlight the higher level consistencies in the English orthography that are present when GPCs are considered in the context of the orthographic rime unit (Goswami, 1999, 2002a; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). The rhyme family format means that in GG Rime, GPC information is always linked to oral rhyming patterns; hence, rhyme awareness was trained at the same time as phoneme awareness. Rhyme families are not taught exhaustively; rather, four to eight members of a particular family are introduced, and the player is left to infer that words with analogous orthographic rimes that might be subsequently encountered during classroom reading and spelling activities would be similar.

The streams begin by using CVC items from the most consistent and most dense rime phonological neighborhoods of English, following the database devised by De Cara and Goswami (2002) and taking into account word frequency and orthographic consistency. Later streams introduce CCVC and CVCC words (e.g., stream 7: *bring*, *sting*; streams 8–10: *best*, *quest*). When complex onsets were blended with rimes (e.g., *str-ong*), the students were told, for example, "Now let's play with the /ong/ rime and combine it with other sounds to make real words." For variety, longer words were occasionally introduced as a challenge (e.g., stream 6: the bisyllable *finish* was brought in at the end of the *-ish* rhyme family; stream 8: the bisyllable *doorbell* was brought in at the end of the *-ell* rhyme family).

The order of introduction of items is always from the most consistent and frequent mappings from high-density rime neighborhoods through to the less dominant and less frequent mappings from high-density rime neighborhoods and then the equivalent for low-density rime neighborhoods. There are also Catch the Rhyme levels in which the player hears a word and then has to click on the balls that contain a word that rhymes with this target word. The Catch the Rhyme levels use whole words only, without segmentation and blending of rimes and GPCs, hence reinforcing the oral rhyme family aspect of the game.

Less frequent and exceptional spellings are introduced later in the game, once the player has acquired the most consistent and frequent mappings. Later streams thus contain CVC rhyme families that introduce more complex GPCs (e.g., rule of *e* spellings, complex vowel digraphs). Where there are multiple possible spellings for the same rime (e.g., *eet*/*eat*), the most frequent is introduced first (e.g., rhyming words that are spelled like *feet* are introduced before rhyming words that are spelled like *beat*). (All the streams used in GG Rime are available for research purposes from the GraphoWorld Network.)

Procedure

Both intervention groups played the computer games for 10–15 minutes daily for five sessions a week over a 12-week period (maximum 60 sessions). The participating students played the game individually on separate laptops with headphones. To minimize disruption for the rest of the class and remove the administrative load for the participating teachers, the first author set up the laptops at the beginning of each session to be ready for the students to play the games in a separate room. She also remained present in the room while the students played to provide general encouragement and motivation.

Fidelity to the Program

Fidelity to the GraphoGame intervention programs was controlled by the Finnish GraphoWorld center, which provided detailed logs including the time spent by each participant in playing the games. Different students' differential exposure to game content was automatically recorded by the GraphoGame software. Although not used for our purposes here, this feature enables individualized assessment of learning, allowing the researcher or teacher to identify levels of the game which are causing difficulty and to decide whether to provide extra (game-based or nongame) reinforcement. As shown in Table 1, there were no significant differences between GG Rime and GG Phoneme in the length of the students' playing time or exposure time. The two interventions did not appear to differ in terms of the complexity or difficulty of the items and levels, as there was no significant difference in the average success rate of the students across the levels.

Results

Preliminary Analyses

The means and standard deviations for the average raw scores for all measures at T2 (posttest) and T3 (follow-up) are presented in Table 2, with raw scores by group at T1 for comparison. Because the group sizes were small, raw scores rather than standard scores were used in the analyses. Paired *t*-tests with Bonferroni corrections were first run for each group and each task to assess whether progress was made from T1 to T2 and from T2 to T3. As would be expected given that all three groups received normal classroom instruction during the intervention, all three groups (GG Rime, GG Phoneme, untreated control) made significant progress during each time interval (T1 to T2, and T2 to T3) for all the literacy measures. Hence, these analyses are not reported in detail.

A series of ANCOVA analyses controlling for T1 scores and comparing group effects at T2 and T3 did not

Table 2
Means in Raw Score Units for Pretest (T1), Posttest (T2), and Follow-Up (T3) Assessments

		Group		
Measure	Assessment	GG Rime (*SD*)	GG Phoneme (*SD*)	Control (*SD*)
BAS II reading raw score	T1	22.2 (12.9)	24.6 (14.7)	21.5 (12.1)
	T2	36.6 (13.5)	36.1 (15.4)	31.5 (14.9)
	T3	40.3 (14.8)	42.3 (15.7)	37.4 (13.8)
Spelling raw	T1	15.3 (4.8)	14.8 (4.6)	16.2 (4.7)
	T2	21.8 (6.4)	19.9 (8.1)	19.9 (4.3)
	T3	24.8 (6.7)	24.3 (7.9)	22.3 (5.5)
TOWRE sight word	T1	26.9 (14.7)	26.6 (13.2)	25.3 (11.6)
	T2	36.7 (14.1)	35.8 (17.0)	31.8 (10.2)
	T3	42.6 (16.0)	43.2 (16.4)	40.8 (13.1)
TOWRE nonword	T1	7.1 (6.0)	10.5 (7.4)	8.0 (8.6)
	T2	14.4 (7.3)	15.7 (7.7)	11.7 (8.9)
	T3	17.5 (10.4)	19.2 (10.0)	15.1 (12.2)
Phoneme deletion	T1	10.2 (5.8)	12.1 (6.0)	9.7 (5.3)
	T2	15.9 (4.9)	18.5 (6.1)	12.2 (5.4)
	T3	17.1 (4.5)	17.1 (6.6)	14.7 (6.4)
Rhyme oddity	T1	6.8 (4.0)	10.0 (3.1)	8.9 (3.1)
	T2	10.0 (3.6)	11.2 (2.3)	9.5 (2.3)
	T3*	9.3 (2.7)	10.1 (4.9)	10.0 (4.0)

Note. BAS II = British Ability Scale II. GG = GraphoGame. *SD* = standard deviation. TOWRE = Test of Word Reading Efficiency.
*There were four missing data scores from T3 due to participant absence.

reveal any significant group effects, probably because of the relatively small sample sizes. Therefore, following the work of Hatcher and colleagues (e.g., Hatcher, 2000), any differential effects of the two interventions (GG Rime, GG Phoneme) were explored by comparing effect sizes (Cohen, 1988).

Immediate Effects of Intervention (T1 to T2)

The relative progress made due to GraphoGame from T1 to T2 was computed by comparing the effect sizes for GG Rime and GG Phoneme with the control group. The data are shown in Table 3. The effect sizes were calculated by first computing gain scores for each group (T2 − T1) for each task and then subtracting the mean gain score for the control group from the mean gain score for one of the experimental groups (GG Rime, GG Phoneme). This relative gain for each intervention group on each task (i.e., relative to controls) was divided by the standard deviation of the gain for the control group on the same task to yield the effect size. This follows the method used by Hatcher (2000), except that we used the standard deviation of the control group gain instead of the standard deviation of the control group's pretest performance as the divisor in the computation because we were comparing the effect sizes of relative gains. Table 3 also shows the standard errors for the effect sizes and an estimate of the statistical significance of the difference in effect sizes between the two experimental groups.[1] A negative effect size means that the progress made by an experimental group was smaller than the progress made by the unseen control group.

Inspection of Table 3 shows that relative to the gains of the control group, GG Rime had a large effect on nonword reading, a medium effect on both sight word reading and single-word reading, and a large effect on spelling. In contrast, GG Phoneme had a medium effect on nonword reading and small effects on sight word reading, single-word reading, and spelling. Relative to the gains made by the control group, GG Rime had a large effect on both phoneme deletion and rhyme oddity scores. In contrast, GG Phoneme had a large effect on phoneme deletion scores but only a small effect on rhyme oddity scores. The effect size differences for single-word reading ($p = .08$), nonword reading ($p = .09$), and rhyme oddity ($p = .06$) approached statistical significance, with larger effect sizes for GG Rime in each case.

Effectiveness of the Interventions in SS Units

As noted earlier, the prior literature considers an intervention to have been successful if participants gain more than 0.13–0.23 SS units per hour of training. By these measures, both GG Rime and GG Phoneme were successful. The students playing GG Rime gained 0.69 SS points per hour of intervention, and the students playing GG Phoneme gained 0.49 SS points per hour of intervention. If we compute the gains made per hour spent in regular classrooms without the intervention, the control group made 0.35 SS gains per equivalent hour that they spent in the normal classroom. There was no significant difference in SS gains per hour between GG Phoneme and the control group: $t(18) = 0.70$, ns; however, the difference in SS gains per hour between GG Rime and the control group approached statistical significance: $t(19) = 1.80$, $p = .09$.

Over the course of the training, the gains in reading standard scores between T1 and T2 were 7.7 SS for GG Rime and 5.1 SS for GG Phoneme, compared with 4.1 SS for the control group. The differences in SS gain between the three groups did not approach significance: $F(2, 30) = 1.57$, $p = .23$. The difference in overall gain for GG Rime compared with the control group also failed to reach significance: $t(19) = 1.7$, $p = .11$.

Table 3
Effect Sizes for Gains Made Between Pretest (T1) and Posttest (T2) for GG Rime and GG Phoneme Relative to GG Control (Standard Errors (*d*) for the Effect Sizes Are Shown in Parentheses)

Measure	GG Rime vs. control (*d*)	GG Phoneme vs. control (*d*)	*p*-value for GG Rime > GG Phoneme
BAS II single-word reading raw	0.66 medium (0.21)	0.22 small (0.25)	.08
BAS II spelling raw	0.91 large (0.32)	0.45 small (0.48)	.20
TOWRE sight word raw	0.53 medium (0.27)	0.43 small (0.28)	.40
TOWRE phonemic decoding raw	1.43 large (0.40)	0.60 medium (0.50)	.09
Phoneme deletion	1.27 large (0.36)	1.53 large (0.42)	.69
Rhyme oddity	1.00 large (0.45)	0.23 small (0.16)	.06

Note. BAS II = British Ability Scale II. GG = GraphoGame. TOWRE = Test of Word Reading Efficiency.

Were the Intervention Effects Maintained After the Intervention Ceased (T3)?

To evaluate whether the effects of the interventions were durable, the students were assessed again four months after the intervention ended. This enabled us to assess whether the gains made by the two intervention groups relative to the progress of the control group would be maintained over time. The effect sizes for gains made from T1 to T3 are shown in Table 4. The effect sizes were again computed on the basis of the gain scores relative to the controls following Hatcher (2000; we computed the difference in progress from T1 to T3 for each intervention group relative to the control group's progress, divided by the standard deviation of the gain score of the control group). Standard errors for the effect sizes and an estimate of the significance of the difference in effect sizes between the two experimental groups are again provided.

As shown in Table 4, approximately four months after the intervention ceased, relative to the control group, both GG Rime and GG Phoneme still had a large effect on spelling and a small effect on single-word reading. In addition, GG Rime had a medium effect on nonword reading and a small effect on both the phoneme deletion and rhyme oddity tasks. GG Phoneme only had a small effect on nonword reading and had no effect on either phonological awareness task relative to the control group. The effect size difference between GG Rime and GG Phoneme for rhyme oddity approached statistical significance ($p = .07$).

Discussion

The present study provides a first step in assessing the effectiveness of GraphoGame, a CARI, as a supplementary tool for teaching students to decode the inconsistent English orthography. GraphoGame was originally designed for the transparent and consistent Finnish orthography (Lyytinen, 2007, 2009; Saine et al., 2011) and was successfully extended to German (Brem et al., 2010), another transparent and consistent orthography. For the current study, GraphoGame was adapted for the English orthography by two independent research centers in two theoretically contrasting ways.

One adaptation was based on the phonological unit of the phoneme, instantiating theoretical views of the importance of a small-unit approach to literacy instruction even for nontransparent orthographies (Hulme et al., 2002; Johnston & Watson, 2004; Rose, 2006; Seymour & Duncan, 1997). The second adaptation was based on updating rhyme analogy theory (Goswami, 1986, 1988, 1990, 1993, 1999, 2001) to take account of more recent research on how students learn to read words (e.g., Apel, Thomas-Tate, Wilson-Fowler, & Brimo, 2012; Ehri, 1998; Juel & Minden-Cupp, 2000) and also more recent research on the roles of phonological and orthographic rime neighborhoods in reading in English (De Cara & Goswami, 2002, 2003; Goswami et al., 2005; see the summary for teachers in Wyse & Goswami, 2012).

A rhyme analogy theoretical perspective suggests that readers infer connections between their phonological knowledge and the orthography that they are learning and that some of these connections in English are at the psycholinguistic grain size of the rime. Grain size refers to the granularity of the different possible linguistic units, such as syllable, rhyme, and phoneme. All possible grain sizes must be connected in phonology and orthography if fully specified orthographic representations for words are to develop (Ziegler & Goswami, 2005).

Students require complete knowledge of all constituent graphemes in the right order, referred to as mental graphemic representations by Apel and colleagues (e.g., Apel et al., 2012). The importance of combining

Table 4
Effect Sizes for Gains Made Between Pretest (T1) and Follow-Up (T3) for GG Rime and GG Phoneme Relative to GG Control (Standard Errors (*d*) for the Effect Sizes Are Shown in Parentheses)

Measure	GG Rime vs. control (*d*)	GG Phoneme vs. control (*d*)	*p*-value for GG Rime > GG Phoneme
BAS II single-word reading raw	0.37 small (0.32)	0.31 small (0.31)	.44
BAS II spelling raw	0.90 large (0.31)	0.89 large (0.38)	.49
TOWRE sight word raw	0.03 no effect (0.28)	0.13 no effect (0.26)	.61
TOWRE phonemic decoding raw	0.65 medium (0.35)	0.32 small (0.28)	.23
Phoneme deletion	0.38 small (0.29)	0.00 no effect (0.23)	.15
Rhyme oddity	0.29 small (0.24)	−0.23 small (0.21)	.07

Note. BAS II = British Ability Scale II. GG = GraphoGame. TOWRE = Test of Word Reading Efficiency.

the teaching of rime units with smaller units also informed GG Rime (e.g., Ehri & Robbins, 1992), along with experimental demonstrations that young readers of English are developing knowledge about both small and large units (phonological–orthographic connections at different grain sizes) in parallel (e.g., Brown & Deavers, 1999).

We were unable to compare the efficacy of the two English CARIs for initial reading instruction because schools in England are now legally obliged to use a synthetic phonics approach during the first year of reading instruction. Accordingly, we compared the effectiveness of the two games during the second year of reading instruction in England and only as a supplement to ongoing classroom literacy teaching. Ongoing instruction was already effective, as significant gains in the literacy measures were made by the untreated control group from both pretest (T1) to posttest (T2) and from posttest to follow-up (T3).

Both versions of GraphoGame were found to be effective supplementary literacy activities, showing medium to large effect sizes on the outcome measures. Both games led to significant improvements in reading, spelling, and phonological skills. The effect size data showed that these improvements were considerable in comparison with gains made over the same period by students who did not receive a supplementary intervention, the untreated control group. Because the effect sizes for the two interventions did not differ significantly from each other (see Tables 3 and 4), it cannot be concluded that one of these CARIs was more effective than the other.

However, there were trends for the students playing GG Rime to show greater improvement. The lack of significant effects may reflect the small sample sizes. Nevertheless, both GG Rime and GG Phoneme showed medium to large effect sizes for nonword reading as well as for real-word reading, suggesting that the phonological recoding skills being learned transferred to previously unencountered printed forms. Unlike many previous reading intervention studies, the current study also included a standardized assessment of spelling. The data showed that participating in more phonics instruction via the CARIs also had a beneficial effect on spelling development, with large effect sizes.

The training that we were able to provide in the current study was comparatively brief (~11 hours). Nevertheless, the effect sizes for the improvements in reading and spelling for the students who played GG Rime and GG Phoneme compare very favorably with those reported in previous reading intervention studies conducted over much longer time periods (e.g., Blachman et al., 2004; Hatcher et al., 2004; Torgesen et al., 2010) as well as in meta-analyses of technological interventions (e.g., Cheung & Slavin, 2012). The prior literacy interventions were also more resource intensive than the CARIs we used because teachers delivered the training.

One reason for the large effect sizes found here might be that the first author was present throughout the CARI game periods, providing general encouragement, which may have been very motivating for the students. Another reason may be the specific content of both games and the individualized nature of the training software, which ensured that mastery of one level occurred before progression to the next level (Connor et al., 2009). Eleven hours of training with the GG Rime CARI led to medium effect sizes for reading (0.66) and large effect sizes for spelling (0.91), phonological awareness (phoneme: 1.27; rime: 1.0), and nonword reading (1.43).

A similar amount of training with the GG Phoneme CARI led to small effect sizes for reading (0.22) and spelling (0.45), a medium effect size for nonword reading (0.60), and a large effect size for phonological awareness of phonemes (1.53). Improvements in standard score per hour of the intervention were 0.69 for GG Rime and 0.49 for GG Phoneme. In comparison, the gains reported for more personnel-intensive nontechnological training programs, such as the phonological linkage program of Hatcher et al. (1994), are 0.31 SS per hour (see Hatcher, 2003). At the same time, although the GG Rime group made more progress than the GG Phoneme group did over the intervention period, it is worth noting that the rate of progress of the GG Rime group slowed between posttest and follow-up.

At T3, the effect sizes for both CARIs were small (GG Rime: 0.37; GG Phoneme: 0.31), although the effects on spelling continued to show large effect sizes (GG Rime: 0.90; GG Phoneme: 0.89). As noted earlier, classroom phonics teaching in England uses synthetic phonics methods, which are also taught by GG Phoneme, so the decline in efficacy for GG Rime is perhaps unsurprising. The students were no longer receiving any explicit orthographic instruction at the rime level between T2 and T3. Clearly, maintenance training based on orthographic rimes might be required to support optimal levels of progression after training with GG Rime.

The design of the current study also enabled us to revisit an ongoing debate in early literacy teaching in the United Kingdom, which is whether phonics instruction based on small units is necessarily superior to other methods of phonics instruction (e.g., Goswami, 2002a; Hulme et al., 2002; Johnston & Watson, 2004; Wyse & Goswami, 2008). Our participants' gains in reading were either equal for GG Phoneme and GG Rime or superior for GG Rime. Our view is that GG Rime supported the students' learning by including a specific focus on rime, thereby accessing psycholinguistic units that are well developed in students before literacy is taught (Ziegler & Goswami, 2005).

Orthographic rimes also reflect higher order consistencies within the spelling system of English (Treiman et al., 1995), hence supporting implicit statistical orthographic learning (Goswami, 1999, 2001). Most notable was the difference in the effect sizes for reading (0.66 vs. 0.22 at posttest, $p = .08$) and phonological recoding to sound (nonword reading: 1.43 vs. 0.60 at posttest, $p = .09$). The students who played GG Rime also gained more standard score points per hour of playing than the did the students who played GG Phoneme (0.69 vs. 0.49), and these gains approached significance when compared with the control group gains (0.69 vs. 0.35, $p = .09$). However, it is important to recall that our participants were students who had been designated by their classroom teachers during the second year of reading instruction as falling behind.

These students had all experienced synthetic phonics instruction during their first year of reading instruction. This was unavoidable due to current U.K. government policy, but it also means that our data do not bear critically on Rose's (2006) assertion that "'synthetic' phonics...offers the vast majority of beginners the best route to becoming skilled readers" (p. 19). The fact that we could not train beginning readers may also explain why the effect sizes for the two English versions of GraphoGame were smaller than those achieved by the Finnish version (Saine et al., 2011). Nevertheless, it would be theoretically interesting to trial the two versions of GraphoGame with beginning readers of English to see whether GG Rime would be more effective for younger students. In contrast, it may be the case that students who do not develop good early decoding skills via synthetic phonics do better when offered an alternative basis for learning English GPCs (as offered by GG Rime) than when offered more of the same training in synthetic phonics (via GG Phoneme). Only further experimental studies can disentangle these possibilities.

Both GG Rime and GG Phoneme also led to improvements in phonological awareness in comparison with the control group. However, the magnitude of the effect sizes differed: GG Rime led to large improvements in phonological awareness measured at both the phoneme level and the rhyme level, whereas GG Phoneme led to a large improvement at the phoneme level but only had a small effect at the rhyme level. The effect sizes for rhyme awareness from pretest to posttest were close to significance level (rhyme oddity: GG Rime effect size = 1.0 vs. GG Phoneme effect size = 0.23, $p = .06$) and were close to significant at the long-term follow-up ($p = .07$). This pattern is unsurprising in that GG Rime provides phonological training at the rhyme level, whereas GG Phoneme does not.

These findings are consistent with the training studies reported by Goswami and East (2000) and Hatcher et al. (2004). For example, Hatcher et al. compared the effects of teacher-delivered training in reading with rhyme, reading with phoneme, and reading with rhyme and phoneme and reported phonological effects that were specific to the training that had been received. The typically developing students in their study receiving the combined training showed improvement on measures of rhyme awareness and phoneme awareness compared with the control group; those who received training only at the phoneme level were better at phoneme awareness but not rhyme awareness than the control group was, and those who received training only at the rhyme level were better at rhyme awareness but not phoneme awareness than the control group was. Similarly, Goswami and East demonstrated that rime-based literacy instruction enhanced phonological awareness at large grain sizes in 5-year-old beginning readers, whereas beginning readers who were taught only GPC skills showed relatively poor large-unit awareness.

The current small-scale study had a number of important limitations. First, the small sample size limited the statistical approach, and the sample size should be increased in future studies. Rather than comparing effect sizes, a more powerful statistical procedure would use ANCOVA, controlling for pretest differences between groups even when these differences are not significant (pretest differences were nonsignificant for the current sample). Second, the untreated control group was recruited a year after the intervention groups. Although this was beyond our control, and the schools, teachers, and synthetic phonics program used as a basis for classroom instruction were constant across time, it is possible that the classroom teaching received by the untreated controls differed in subtle ways from that given the previous year. It is impossible to know whether such differences would have improved or impaired the classroom literacy experiences of the unseen controls relative to the two treatment groups.

Third, the experimenter remained present while the students played the CARIs. Hence, the additional motivational effects of having an encouraging adult present are confounded with the game content and may partly explain the relatively large effect sizes produced by both games. Finally, the total intervention received was limited (an average of 11 hours spent playing the game across 12 weeks), which meant that none of the participants were able to complete all the streams in the games. In future work, the streams used for GG Rime and GG Phoneme could be compared using connectionist modeling. If a computer model of phonological recoding of letters to sounds also shows better performance following training with GG Rime, that would suggest that including a rhyme analogy approach as part of the initial teaching of reading should be beneficial for young students (Zorzi, Houghton, & Butterworth, 1998).

In summary, the current study suggests that young English learners may benefit markedly in both decoding and spelling from the supplementary use of CARIs in addition to classroom literacy instruction. In the current study, students learning to read in English benefited from a CARI providing letter–sound training at the phoneme–grapheme level (GG Phoneme) and from a CARI introducing GPCs via rhyme families and the orthographic unit of the rime (GG Rime). In fact, GG Rime showed larger effects on a number of the outcome measures. GG Rime effectively provided a balanced approach to learning phonics (e.g., Bielby, 1998; Goswami, 1999, 2002b; Wyse & Goswami, 2008, 2012). The beneficial effects of balanced instruction are mirrored by the superior reading plus phonology training effects found by Hatcher et al. (1994) in the original phonological linkage study, which was also conducted with 6–7-year-old students who were experiencing difficulties in learning to read in English.

Nevertheless, the gains made in standard scores per hour of the intervention for the students playing GG Rime (0.69) were more than twice as large as those reported by Hatcher and colleagues (0.31; e.g., Hatcher, 2003). Furthermore, 0.69 SS per hour is far in excess of the 0.13–0.23 SS gains per hour considered to be evidence for an effective reading intervention (see Torgesen et al., 2001). Therefore, CARIs such as GraphoGame appear to have great utility with respect to supporting the teaching of the spelling systems of languages with nontransparent orthographies, such as English. CARIs offer evidence-based technological tools that are cost-effective in aiding students' learning and can support classroom teachers by providing individualized instruction and practice in the component skills of reading, which may be of particular benefit to struggling learners (Beddington et al., 2008; Connor et al., 2009; Torgesen & Barker, 1995).

NOTES

This work was supported by the European Commission's FP6, Marie Curie Excellence Grants (MEXT-CE-2004-014203). We thank the Graphogame development team at the University of Jyväskylä, especially Silvia Brem, Jane Erskine, Anne Mönkkönen, Marika Peltonen, Gonny Willems, and Ville Mönkkönen for implementing the training game. We also thank the students, families, and schools that participated in the study.

[1] The numbers in parentheses in Table 3 are the standard error of the T1 to T2 (or T1 to T3 in Table 4) gain divided by the standard deviation of the T1 to T2 (or T1 to T3) gain of the control group. The *p*-value was computed using a one-tailed t-test. Values near 0 support GG Rime > GG Phoneme; values near 1 support GG Phoneme > GG Rime.

REFERENCES

Apel, A., Thomas-Tate, S., Wilson-Fowler, E.B., & Brimo, D. (2012). Acquisition of initial mental graphemic representations by children at risk for literacy development. *Applied Psycholinguistics*, *33*(2), 365–391. doi:10.1017/S0142716411000403

Ball, E.W., & Blachman, B.A. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia*, *38*(1), 208–225. doi:10.1007/BF02648257

Ball, E.W., & Blachman, B.A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, *26*(1), 49–66. doi:10.1598/RRQ.26.1.3

Beddington, J., Cooper, C.L., Field, J., Goswami, U., Huppert, F.A., Jenkins, R., et al. (2008). The mental wealth of nations. *Nature*, *455*(7216), 1057–1060.

Bielby, N. (1998). *How to teach reading: A balanced approach*. Leamington Spa, UK: Scholastic.

Blachman, B.A., Schatschneider, C., Fletcher, J.M., Francis, D.J., Clonan, S.M., Shaywitz, B.A., et al. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, *96*(3), 444–461. doi:10.1037/0022-0663.96.3.444

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read—a causal connection. *Nature*, *301*(5899), 419–421.

Brem, S., Bach, S., Kucian, K., Guttorm, T.K., Martin, E., Lyytinen, H., et al. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(17), 7939–7944. doi:10.1073/pnas.0904402107

Brown, G.D.A., & Deavers, R.P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*, *73*(3), 208–242. doi:10.1006/jecp.1999.2502

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing*, *13*(1/2), 31–56. doi:10.1023/A:1008051414854

Cheung, A.C.K., & Slavin, R.E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review.* Advance online publication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connor, C.M., Piasta, S.B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., et al. (2009). Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. *Child Development*, *80*(1), 77–100. doi:10.1111/j.1467-8624.2008.01247.x

De Cara, B., & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 416–423.

De Cara, B., & Goswami, U. (2003). Phonological neighbourhood density: Effects in a rhyme awareness task in five-year-old children. *Journal of Child Language*, *30*(3), 695–710. doi:10.1017/S0305000903005725

Dunn, L.M., Dunn, L.M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale* (2nd ed.). Windsor, UK: NFER-Nelson.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Ehri, L.C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

Ehri, L.C., & Robbins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly*, *27*(1), 12–26. doi:10.2307/747831

Elliot, C.D., Smith, P., & McCulloch, K. (1996). *British Ability Scales II (BAS II)*. Windsor, UK: NFER-Nelson.

Erskine, J.M., & Seymour, P.H.K. (2005). Proximal analysis of developmental dyslexia in adulthood: The cognitive mosaic model. *Journal of Educational Psychology*, *97*(3), 406–424. doi:10.1037/0022-0663.97.3.406

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*(1), 73–83. doi:10.1016/0022-0965(86)90016-0

Goswami, U. (1988). Orthographic analogies and reading development. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *40A*(2), 239–268. doi:10.1080/02724988843000113

Goswami, U. (1990). A special link between rhyming skill and the use of orthographic analogies by beginning readers. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *31*(2), 301–311. doi:10.1111/j.1469-7610.1990.tb01568.x

Goswami, U. (1993). Toward an interactive analogy model of reading development: Decoding vowel graphemes in beginning reading. *Journal of Experimental Child Psychology*, *56*(3), 443–475. doi:10.1006/jecp.1993.1044

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading*, *22*(3), 217–240. doi:10.1111/1467-9817.00087

Goswami, U. (2001). Rhymes are important: A comment on Savage. *Journal of Research in Reading*, *24*(1), 19–29. doi:10.1111/1467-9817.00130

Goswami, U. (2002a). In the beginning was the rhyme? A reflection on Hulme, Hatcher, Nation, Brown, Adams, and Stuart (2002). *Journal of Experimental Child Psychology*, *82*(1), 47–57. doi:10.1006/jecp.2002.2673

Goswami, U. (2002b). Rhymes, phonemes and learning to read: Interpreting recent research. In M. Cook (Ed.), *Perspectives on the teaching and learning of phonics* (pp. 41–60). Royston: United Kingdom Reading Association.

Goswami, U. (2012). Reading and spelling: Revisiting Bradley and Bryant's study. In A.M. Slater & P.C. Quinn (Eds.), *Developmental psychology: Revisiting the classic studies* (pp. 132–147). London: Sage.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum.

Goswami, U., & East, M. (2000). Rhyme and analogy in beginning reading: Conceptual and methodological issues. *Applied Psycholinguistics*, *21*(1), 63–93. doi:10.1017/S0142716400001041

Goswami, U., Ziegler, J.C., & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology*, *92*(4), 345–365. doi:10.1016/j.jecp.2005.06.002

Hasselbring, T.S., & Goin, L.I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly*, *20*(2), 123–144. doi:10.1080/10573560490262073

Hatcher, P.J. (2000). Reading intervention need not be negligible: Response to Cossu (1999). *Reading and Writing*, *13*(3/4), 349–355. doi:10.1023/A:1026439309818

Hatcher, P.J. (2003). Reading intervention: A 'conventional' and successful approach to helping dyslexic children acquire literacy. *Dyslexia*, *9*(3), 140–145. doi:10.1002/dys.254

Hatcher, P.J., Hulme, C., & Ellis, A.W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, *65*(1), 41–57. doi:10.2307/1131364

Hatcher, P.J., Hulme, C., Miles, J.N.V., Carroll, J.M., Hatcher, J., Gibbs, S., et al. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(8), 820–827. doi:10.1111/j.1469-7610.2005.01559.x

Hatcher, P.J., Hulme, C., & Snowling, M.J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(2), 338–358. doi:10.1111/j.1469-7610.2004.00225.x

Hintikka, S., Aro, M., & Lyytinen, H. (2005). Computerized training of the correspondences between phonological and orthographic units. *Written Language & Literacy*, *8*(2), 79–102. doi:10.1075/wll.8.2.07hin

Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annals of Dyslexia*, *58*(2), 115–137. doi:10.1007/s11881-008-0017-2

Hulme, C., Hatcher, P.J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology*, *82*(1), 2–28. doi:10.1006/jecp.2002.2670

Johnston, R.S., & Watson, J.E. (2004). Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. *Reading and Writing*, *17*(4), 327–357. doi:10.1023/B:READ.0000032666.66359.62

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*, *35*(4), 458–492. doi:10.1598/RRQ.35.4.2

Kuppen, S., Huss, M., Fosker, T., Fegan, N., & Goswami, U. (2011). Basic auditory processing skills and phonological awareness in low-IQ readers and typically developing controls. *Scientific Studies of Reading*, *15*(3), 211–243. doi:10.1080/10888431003706291

Lyytinen, H., Erskine, J., Kujala, J., Ojanen, E., & Richardson, U. (2009). In search of a science-based application: A learning tool for reading acquisition. *Scandinavian Journal of Psychology*, *50*(6), 668–675. doi:10.1111/j.1467-9450.2009.00791.x

Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A., & Taanila, M. (2007). Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nordic Psychology*, *59*(2), 109–126. doi:10.1027/1901-2276.59.2.109

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. London: Department for Education and Skills.

Saine, N.L., Lerkkanen, M.-K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Development*, *82*(3), 1013–1028. doi:10.1111/j.1467-8624.2011.01580.x

Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. doi:10.1348/000712603321661859

Seymour, P.H.K., & Duncan, L.G. (1997). Small versus large unit theories of reading acquisition. *Dyslexia*, *3*(3), 125–134. doi:10.1002/(SICI)1099-0909(199709)3:3<125::AID-DYS85>3.0.CO;2-4

Thomson, J.M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology, Paris*, *102*(1–3), 120–129. doi:10.1016/j.jphysparis.2008.03.007

Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K.K.S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities. *Journal of Learning Disabilities*, *34*(1), 33–58. doi:10.1177/002221940103400104

Torgesen, J.K., & Barker, T.A. (1995). Computers as aids in the prevention and remediation of reading disabilities. *Learning Disability Quarterly*, *18*(2), 76–87. doi:10.2307/1511196

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Torgesen, J.K., Wagner, R.K., Rashotte, C.A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, *60*(1), 40–56. doi:10.1007/s11881-009-0032-y

Torgesen, J.K., Wagner, R.K., Rashotte, C.A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*(4), 579–593. doi:10.1037/0022-0663.91.4.579

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology, General*, *124*(2), 107–136. doi:10.1037/0096-3445.124.2.107

van Daal, V., & Reitsma, P. (2000). Computer-assisted learning to read and spell: Results from two pilot studies. *Journal of Research in Reading*, *23*(2), 181–193. doi:10.1111/1467-9817.00113

Walton, P.D., Bowden, M.E., Kurtz, S.L., & Angus, M. (2001). Evaluation of a rime-based reading program with Shuswap and Heiltsuk First Nations prereaders. *Reading and Writing*, *14*(3/4), 229–264. doi:10.1023/A:1011125315377

Walton, P.D., & Walton, L.M. (2002). Beginning reading by teaching in rime analogy: Effects on phonological skills, letter–sound knowledge, working memory, and word-reading strategies. *Scientific Studies of Reading*, *6*(1), 79–115. doi:10.1207/S1532799XSSR0601_04

Walton, P.D., Walton, L.M., & Felton, K. (2001). Teaching rime analogy or letter recoding reading strategies to prereaders: Effects on prereading skills and word reading. *Journal of Educational Psychology*, *93*(1), 160–180. doi:10.1037/0022-0663.93.1.160

Wyse, D., & Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal*, *34*(6), 691–710. doi:10.1080/01411920802268912

Wyse, D., & Goswami, U. (2012). Early reading development. In J. Larson & J. Marsh (Eds.), *Handbook of early childhood literacy* (pp. 379–394). Thousand Oaks, CA: Sage.

Ziegler, J.C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. doi:10.1037/0033-2909.131.1.3

Zorzi, M., Houghton, G., & Butterworth, B. (1998). The development of spelling–sound relationships in a model of phonological reading. *Language and Cognitive Processes*, *13*(2/3), 337–371. doi:10.1080/016909698386555

Submitted June 16, 2012
Final revision received September 28, 2012
Accepted October 1, 2012

FIONA KYLE was a research associate at the Centre for Neuroscience in Education, University of Cambridge, UK, and the Agora Centre, University of Jyväskylä, Finland, at the time of this study and is now a senior lecturer in developmental psychology at the University of Bedfordshire, UK; e-mail fiona.kyle@beds.ac.uk (corresponding author).

JANNE KUJALA was a senior researcher at the Agora Centre, University of Jyväskylä, Finland, at the time of this study and is now in the Department of Mathematical Information Technology at the University of Jyväskylä, Finland.

ULLA RICHARDSON is a professor of technology-enhanced language learning at the University of Jyväskylä, Finland.

HEIKKI LYYTINEN is a professor of developmental neuropsychology at the University of Jyväskylä, Finland.

USHA GOSWAMI is a professor of cognitive developmental neuroscience and the director of the Centre for Neuroscience in Education at the University of Cambridge, UK,; e-mail ucg10@cam.ac.uk (corresponding author).