Kyle, F.E., 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, doi:10.1002/rrq.038. Available at: http://onlinelibrary.wiley.com/doi/10.1002/rrq.038/pdf

Assessing the Effectiveness of Two Theoretically Motivated Computer-assisted Reading Interventions in the UK: GG Rime and GG Phoneme.

Fiona Kyle[1,2], Janne Kujala[2,4], Ulla Richardson[2], Heikki Lyytinen[2,3] & Usha Goswami[1]

[1]Centre for Neuroscience in Education, University of Cambridge, UK

[2]Agora Centre, University of Jyväskylä, Finland

[3]Department of Psychology, University of Jyväskylä, Finland

[4] Department of Mathematical Information Technology, University of Jyväskylä, Finland

Corresponding authors:

Fiona Kyle	Usha Goswami
Department of Psychology	Centre for Neuroscience in Education
University of Bedfordshire	University of Cambridge
Park Square	Downing Street
Luton	Cambridge
LU1 3JU	CB2 3EB
Email: Fiona.kyle@beds.ac.uk	ucg10@cam.ac.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 children who had been identified by their teachers as being relatively poor at reading. The children 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 to the untreated control group. The two interventions also had some differential effects. The intervention gains were maintained at a 4-month follow-up.

Keywords:

Reading; Computer-assisted intervention; Phonological awareness; Rime; Phoneme

Assessing the Effectiveness of Two Theoretically Motivated Computer-assisted Reading Interventions in the UK: GG Rime and GG Phoneme.

There is increasing recognition of the importance of evidence-based technological tools that can aid children's learning and provide individualised instruction and practice (Beddington, et al., 2008; Hasselbring & Goin, 2004; Torgesen & Barker, 1995). If such technological tools are engineered so that 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 programme, 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 to automatize children's 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 via rhyming word families, explicitly focusing on orthographic rime units (the spelling units for rhyming sounds) and demonstrating how rime units and grapheme-phoneme connections are related in English spelling.

The two English CARIs compared here were based on a successful CARI for teaching and automatizing letter-sound knowledge and phoneme awareness called GraphoGame. GraphoGame was developed for the transparent orthography of the Finnish language. GraphoGame is a child friendly, computerised reading intervention programme that provides children 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 (http://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 non-transparent orthography like English is more challenging than implementing GrapoGame in Finnish, and this challenge led directly to the small-scale study reported here.

For the current study, GraphoGame was adapted for the non-transparent English orthography by English-speaking researchers in two independent research centres. In one centre (Finland), GraphoGame was adapted according to the theoretical view that a "small units first" approach 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 tuition in single grapheme-phoneme correspondences, 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 that the most frequent, consistent and prototypical GPCs in English were introduced first (Erskine & Seymour, 2005; see Methods for more detail). In a second centre (Cambridge, UK), GraphoGame was adapted according to the theoretical view that English-speaking children may benefit from a focus on oral rhyme and 'rhyme analogies' as part of reading tuition (Goswami & Bryant, 1990). This CARI, hereafter GraphoGame Rime (GG Rime), was created on the basis of rhyme families in English, ordered so that the largest families with the most consistent orthographic rime spellings were introduced first (DeCara & Goswami, 2002). The effectiveness of these two CARIs was then compared directly in participating UK 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 phonic knowledge for English-speaking children. Currently, schools in the UK are required by 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 children 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). Children in the UK are usually exposed to synthetic phonics from entering school at age 5 years, practicing synthetic phonics skills discretely (i.e., not as part of story reading), and are also explicitly taught concepts like '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 tuition, and all participants continued with synthetic phonics instruction during the intervention (both of the intervention groups and the untreated controls). During the year of the study, classroom phonics tuition was already using complex digraphs. The CARIs were delivered in extra 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 spelling system of English. Both games are described in more detail below (see Methods).

Currently, it is unknown whether a CARI like GraphoGame can be an effective learning tool for phonics tuition in less transparent orthographies like English. However, as the development of efficient and automatic phonics skills plays a critical role in early reading development and as automatized phonic knowledge is an important bedrock for long-term reading development (Ehri, 1998; Juel & Minden-Cupp, 2000), it is important to develop these skills in children who are learning to read English (Wyse & Goswami, in press). Most alphabetic European languages are easy to implement in a CARI, as they have transparent orthographies with simple rules that determine grapheme-to-phoneme mappings (Brem et al., 2010; Saine, Lerkkanen, Ahonen, Tolvanen & Lyytinen, 2011). In contrast, languages with opaque orthographies like English, French, Portuguese and Danish have a more complicated 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 children 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). For example, an important study by Hatcher et al. (1994) carried out in the UK showed that children 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 *Reading* alone or *Phonology* alone. In Hatcher et al.'s study, which involved a 20-week intervention, children receiving the combined intervention gained 0.31 standard score (SS) points per hour of training on a standardised assessment of reading accuracy (see Hatcher, 2000). Hatcher, Hulme and Snowling (2004) reported a similar result for 410 beginning UK readers who were divided into one of four groups: *Reading with Rhyme*; *Reading with Phoneme; Reading with Rhyme and Phoneme* or *Reading Alone*. For typically-developing children, there were no significant differences between the 4 interventions, and the effect sizes for reading and nonword reading were negligible (ranging from -.18 to .13). In contrast, children designated 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 programme led to the greatest gains in word reading for these children (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 *Rhyme and Phoneme* intervention. Use of a CARI nevertheless provides a novel perspective on the effectiveness of these theoretical approaches to learning, as tuition 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; Hatcher, et al., 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. (2007) conducted a large-scale evaluation of five different CARIs and found no significant advantage for the classrooms using CARI in comparison to control classrooms. However, studies using GraphoGame or other types of CARI have provided more positive findings for the use of CARIs, particularly for children identified as being at risk of developing reading problems (Saine, et al., 2011; Torgesen, et al., 2010).

Torgesen et al. (2010) reported that children who received CARI performed significantly better on tasks measuring reading accuracy, reading comprehension, phonemic awareness and decoding than children 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, as the intervention was only provided by computer for half of the training period, to reinforce intervention activities introduced by teachers. 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 pre-reading skills. Saine et al. (2011) reported that at-risk children who received GraphoGame in combination with teacher-based intervention made significantly greater progress in letter knowledge, reading and spelling skills than children 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 children who played GraphoGame were commensurate with those in the mainstream classroom whereas the children who received the traditional intervention still exhibited delays. The Saine et al. (2011) results suggest that GraphoGame is a remarkably effective

CARI for children 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, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011). For example, Huemer et al. (2008) studied the effects of German GraphoGame training (a 6-week period) on reading accuracy and speed in Austrian 2nd and 4th Graders. The children improved in reading accuracy and response times of the trained items. In Switzerland, Brem et al. (2010) conducted an intervention study with prereading children 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 4 hours over 8 weeks), significant improvements were shown in children's 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 computer assisted reading intervention for children learning to read in English. We were not able to replace typical classroom phonics instruction with the CARIs nor to introduce them at the beginning of literacy instruction at age 5 years, due to UK government policy concerning synthetic phonics. Therefore, teachers in the participating schools were asked to identify children 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 children 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 to a control group who received normal classroom tuition. A reading intervention is considered effective if effect sizes are greater than 0.13 - 0.23 (a small effect size, 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, due to 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 2 games might be seen in a long-term follow-up. As regular reading activities in English 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 children who played GG Phoneme, since for these participants the synthetic phonics principles trained by the game would be being 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 children identified informally by their teachers as being likely to benefit from supplementary instruction. All children whose parents returned consent forms were included in the study. As insufficient children returned consent forms in the first year of the study for the complete design, a further group of children from the same schools (same classes and teachers) were recruited the following year and some of these formed the untreated control group. Children who played the CARIs were assessed immediately before the start of the intervention in order to match the two treatment groups (T1, pre-test, October) and again at the end of the 12- week intervention period some 4 months later (T2, post-test, February). To measure the durability of the intervention effects on reading, spelling and phonological awareness, children were also re-assessed 4 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 pre-test (T1), post-test (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 standardised assessments. The tests were delivered in short 20-minute sessions, so that children who needed more time for a particular test did not get more fatigued than other children. All children gave their assent prior to testing, and the study was approved by the Psychology Research Ethics Committee of the University of Cambridge.

Participants

31 second grade children aged between 6 and 7 years old participated in the study. There was an additional participant who began the intervention but withdrew from the study after only four weeks due to moving schools and therefore their data were not included. The children 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 children who they felt were poor readers but without any additional educational needs. The teachers did not use any formal method of assessment to select these children to our knowledge. As shown in Table 1, at T1 all children achieved scores within the normal range on an estimate of NVIQ using the Matrices subtest from the British Abilities Scales II (BAS II: Elliot, Smith, & McCulloch, 1996) and on a test of expressive vocabulary, the British Picture Vocabulary Test (Dunn, Dunn, Whetton & Burley, 1997). In the first recruitment year, half of the children 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, children were not completely

randomly allocated to each intervention group; instead two groups from each school were created matched in terms of chronological age, nonverbal intelligence, expressive vocabulary and reading ability and then these two groups were randomly allocated to either intervention. The children in the untreated control group were nominated by the same teachers a year later as part of a different study, in which all children nominated were randomly assigned to intervention or control groups. Only the control group from that second study are included in this report.

There were no significant differences between the three groups on any of the assessments at pre-test (see Table 1). The groups were not formally matched for ethnicity or other demographic variables, but the children were all of White British ethnicity, were all following the same reading curriculum, and English was their first language. As the two schools drew from homogenous neighbourhoods, and each contributed similar numbers of children to each group, demographic variables were likely to be similar across groups. Participant characteristics and descriptive statistics for all standardised measures at pre-test 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 (F values shown in Table 1).

Assessment battery

Vocabulary. Children were assessed for their receptive vocabulary using the British Picture Vocabulary Subscale II (BPVS II: Dunn et al., 1997). This was only measured at pre-test (T1). The following tests were administered at all three assessment periods, pretest (time 1, hereafter T1), immediate post-test (time 2, hereafter T2), and 4-month follow-up after the end of training (time 3, hereafter T3):

Reading. The children completed two standardised assessments of reading. The Single Word Reading subtest from the BAS II (Elliot, et al., 1996) measured single word reading accuracy. Children 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 standardised before the National Literacy Strategy was introduced in 1997, and that typically-developing children in the United Kingdom now usually score well above a standard score of 100 (see Kuppen, Huss, Fosker, Fegan, & Goswami, 2011). Cronbach's Alpha for the raw scores on this test across the three test points was 0.976. The Test of Word Reading Efficiency (TOWRE: Torgesen, Wagner, & Rashotte, 1999) consisted of two subtests measuring speeded recognition of words and nonwords. Children were required to read aloud as many words or nonwords as quickly and as 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 Alpha for the raw scores on the word recognition test across the three test points was 0.965. Cronbach's Alpha for the raw scores on the phonemic decoding efficiency test across the three test points was 0.952. *Spelling.* The spelling subtest from the BAS II was administered. This was a spelling to dictation task containing a mixture of verbs, nouns and adjectives. The test was administered according the instruction manual. The maximum score was 75. Cronbach's Alpha for the raw scores on this test across the three test points was 0.918.

Phonological skills. The children 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 (2000). Children were required to say words without certain sounds. There were four sections to the task and practice items were provided at the beginning of each section. In the initial sound and final sound sections, children were required to delete the initial sound such as "what is *crush* without the /k/?", or the final sound such as "what is *find* without the /d/?". The other two sections contained items from which the children had to delete the middle sounds, either from the beginning middle "what is *grow* without the /r/?" or the end middle "what is *nest* without the /s/?". The maximum score was 24. Cronbach's Alpha for this test across the three test points was 0.910. The Oddity task was taken from Thomson and Goswami (2008) and measured rhyme awareness. The task was presented on a laptop and used digitised speech created from a native female speaker of Southern British English. Using headphones, children listened to sets of three words and had to select the word that was the "odd one out" as it did not rhyme with the other two words. An example of a trial is "*wag, nag, that*". There were 20 trials and the order of presentation was counterbalanced across participants. Cronbach's Alpha for this test across the three test points was 0.784.

<u>GraphoGame interventions</u>

Both GG Rime and GG Phoneme were presented using the same software platform. Both games provided highly repetitive and individualised intervention in which the player heard auditory targets consisting of either sounds or words and had to match these auditory targets to visual targets (letters and sequences of letters) displayed on the screen. The letters and letter sequences were contained within balls that were shown falling downwards from the top of the screen (see Figure 1). The player had to “catch” the target ball/s by clicking on them with the computer mouse before the balls reached the bottom of the screen and disappeared. Children progressed through a series of graduated game streams, each of which had multiple levels. In order to keep motivation levels high, children 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 was taught using a more exciting background and format. For example, there was a ‘race’ game, where the targets appeared in cars instead of in falling balls, and children had to click on the correct target before the cars raced off the screen. Another special game was a ‘ghost and ladder’ game, in which children 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 were also word formation games to encourage spelling skills, in which children 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). Note that as both games were adaptive, the letters and letter sequences practiced by different players varied depending on speed of progression through the games. Overall the two games taught the same grapheme-phoneme correspondences, but while GG Phoneme introduced all possible correspondences before blending them into words, GG Rime introduced various correspondences of different sizes (small units and larger units) at different points in time, depending on the rhyme families being used for a particular stream.

Both GG Rime and GG Phoneme used a success criterion of at least 80% for each level before children could move onto the next level. If a child failed to achieve 80% accuracy on a level, they were given an individualised extra training level in which the computer automatically selected targets that the child knew and contrasted them with targets that the child 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 had a British accent. Words were digitally recorded in an anechoic chamber and normalised for sound during the editing process.

GG Phoneme. GG Phoneme taught letter-sound correspondences at the level of the single phoneme-grapheme. First, all the single phoneme-graphemes in English spelling were introduced (e.g.*, i, a, ee, oa*) during Streams 1 and 2. In Stream 3, phonemes were combined into CV units (e.g., *ti* and *loa*). In Stream 4, phonemes were combined into VC units (e.g., *is* and *eech*). The consistent presentation sequence was to focus first on GPCs ("Let's play with these sounds"), then to blend these GPCs into larger units that were not yet words ("Let's put two sounds together"), and then to create words from GPCs ("Let's put more sounds together to make words"). From Stream 5 onwards, children were told "here are some words with the sound X". They were now shown only whole words, not CV or VC units, and were asked to identify and isolate GPCs within the whole words or to blend GPCs 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., 2009; Lyytinen, et al., 2007). The theoretical framework drew on approaches arguing for the importance of "small unit" instruction in reading in English (e.g., Hulme et al., 2002; Duncan & Seymour, 1997), and on research assessing grapheme-phoneme consistency in European orthographies including English (Seymour, et al., 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 fourth 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 Jane Erskine for the prototypicality designation (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 phoneme grapheme correspondences were controlled for in terms of order of presentation during the game. The most frequent, most consistent and most prototypical grapheme-phoneme correspondences were introduced first and also reinforced first during later game streams. For example, simple graphemes like P, B, D and K had over 90% consistency and were reinforced early in the game. Complex vowel digraphs were introduced in Stream 2 but were reinforced later in the game. Note that while faithfully reproducing the statistical frequency of occurrence of the different GPCs, this approach introduces CV units like ‘wa’ and ‘wi’ quite early on, 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 was based on the intrasyllabic unit of the rime, which is argued to be an important psycholinguistic unit for English-speaking children (Goswami, 1999; Ziegler & Goswami, 2005). The theoretical framework for GG Rime was based upon the early work of Goswami and colleagues on rhyme analogy and rime-based teaching (Goswami, 1986, 1988, 1990, 1993, 1999, 2001; Goswami & East, 2000), updated with reference to subsequent work on phonological rime neighbourhoods (DeCara & Goswami, 2002) and orthographic rime neighbourhoods (Goswami, Ziegler & Richardson, 2005), as well as research on implicit statistical learning of rhyming sounds and rime spellings by children (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 (in press) summarise how the rhyme analogy and neighbourhood density work can be integrated for the teacher of reading.

The teaching sequence in GG Rime was based consistently on orthographic rime units. Children 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 7 single phonemes and graphemes was introduced (C, S, A, T, P, I, N), and the children were told "Let's put these sounds together to make rime units". Firstly, only rime units that were also real words were created (*at, in*) and reinforced. The children 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 children were reinforced on the GPCs in these CVC words ("The sounds in *tin* are *t*, *i*, *n*"). Subsequently, orthographic rimes that were not also real words were created, like *op* and *ag*, enabling creation of CVC words like *hop* and *bag*. So the teaching sequence was to show some GPCs ("sounds"), to blend these into rimes, to blend onsets onto these rimes to create words (the term "onset" was not used in the game, onsets were called "sounds"), and then to segment the words back into GPCs. Hence consistently during each level within each stream, children 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 in order 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 neighbourhood density research). The use of rhyme families enabled 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 meant that in GG Rime, GPC information was always linked to oral rhyming patterns (hence rhyme awareness was trained at the same time as phoneme awareness). Rhyme families were not taught exhaustively, rather 4 – 8 members of a particular family were introduced, and the child was left to infer for herself that words with analogous orthographic rimes that might be subsequently encountered during classroom reading and spelling activities would be similar. The streams began by using CVC items from the most consistent and most dense rime phonological neighbourhoods of English, following the database devised by De Cara and Goswami (2002), and taking into account word frequency and orthographic consistency. Later streams introduced CCVC and CVCC words (e.g., 'bring', 'sting', Stream 7; ‘best’, ‘quest’, Streams 8-10). When complex onsets were blended with rimes (for example, ‘str-ong’), children 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., the bisyllable ‘finish’ was brought in at the end of the –ish rhyme family, Stream 6; the bisyllable ‘doorbell’ was brought in at the end of the ‘-ell’ rhyme family, Stream 8). The order of introduction of items was always from the most consistent and frequent mappings from high density rime neighbourhoods through to the less dominant and less frequent mappings from high density rime neighbourhoods and then the equivalent for low density rime neighbourhoods. There were also “Catch the Rhyme” levels in which children heard a word and then had to click on the balls that contained words that rhymed with this target word. Catch the Rhyme levels used 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 were introduced later in the game, once children had acquired the most consistent and frequent mappings. Later streams thus contained CVC rhyme families that introduced more complex GPCs (e.g., “rule of e” spellings, complex vowel digraphs). Where there were multiple possible spellings for the same rime (e.g., ‘eet’ and ‘eat’), the most frequent was introduced first (e.g., rhyming words that were spelled like ‘feet’ were introduced before rhyming words that were 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 between 10 and 15 minutes daily for 5 sessions a week over a 12-week period (maximum 60 sessions). Participating children played the game individually on separate laptops with headphones. In order to minimise disruption for the rest of the class and remove the administrative load for participating teachers, the first author set up the laptops at the beginning of each session ready for the children to play the games in a separate room. She also remained present in the room while the children played to provide general encouragement and motivation.

Fidelity to the programme

Fidelity to the GraphoGame intervention programmes was controlled by the Finnish GraphoWorld centre, who provided detailed logs including the time spent by each participant in playing the games. Different children's differential exposure to game content was automatically recorded by the GraphoGame software. Although not used for this purpose here, this feature enables individualised 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 non-game) reinforcement. As shown in Table 1, there were no significant differences between GG Rime and GG Phoneme in the length of children's 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 children across the levels.

Insert Tables 1 and 2 around here

Results

Preliminary Analyses

The means and standard deviations for the average raw scores for all measures at T2 (post-test) and T3 (follow-up) are presented in Table 2, with raw scores by group at

T1 for comparison. As 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 were receiving normal classroom teaching 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 reveal any significant group effects, probably because of the relatively small sample sizes. Therefore, following the work of Hatcher and colleagues (see Hatcher, 2000), any differential effects of the two interventions (GG Rime, GG Phoneme) were explored by comparing effect sizes (Cohen, 1988).

Table 3 around here

<u>Immediate effects of intervention (T1 - T2)</u>

The relative progress made due to GraphoGame from T1 (pre-test) to T2 (post-test) was computed by comparing the effect sizes for GG Rime and GG Phoneme to 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, since we were comparing the effect sizes of *relative* gains. Table 3 also shows the standard errors for the effect sizes (in parentheses) and an estimate of the statistical significance of the difference in effect sizes between the two experimental groupsi. Note that 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= .05) approached statistical significance, with larger effect sizes for GG Rime in each case.

<u>Effectiveness of the interventions in SS units</u>

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 interventions. The children playing GG Rime gained 0.69 SS points per hour of intervention and the children 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), but the difference in SS gains per hour between GG Rime and the Control group approached statistical significance ($t(19) = 1.80$, $p = 0.09$). 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 to 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 to the Control group also failed to reach significance ($t(19) = 1.7$, $p= .11$).

Table 4 about here

Were the intervention effects maintained after the intervention ceased (T3)?

In order to evaluate whether the effects of the interventions were durable, children 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. Effect sizes for gains made from T1 to T3 are shown in Table 4. Effect sizes were again computed on the basis of gain scores relative to controls following Hatcher (2000; we computed the difference in progress from pre-test to follow up [T1 to T3] for each intervention group relative to the control group progress, divided by the standard deviation of the gain score of the control group). Standard errors for the effect sizes (in parentheses) 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 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 approached statistical significance for rhyme oddity ($p = .07$).

Discussion

The present study provides a first step in assessing the effectiveness of GraphoGame, a computer-assisted reading intervention (CARI), as a supplementary tool for teaching children to decode the inconsistent English orthography. GraphoGame was originally designed for the transparent and consistent Finnish orthography (Lyytinen, et al., 2009; Lyytinen, et al., 2007; Saine, et al., 2011), and was successfully extended to German (Brem, et al., 2010), another transparent orthography. For the current study, GraphoGame was adapted for the English orthography by two independent research centres 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 tuition even for non-transparent 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 children learn to read words (e.g., Ehri, 1998; Juel & Minden-Cupp, 2000, Apel, Thomas-Tate, Wilson-Fowler & Brimo, 2012) and also more recent research on the roles of phonological and orthographic rime neighbourhoods in reading in English (De Cara & Goswami, 2002, 2003; Goswami et al., 2005; see summary for teachers in Wyse & Goswami, in press).

A “rhyme analogy” theoretical perspective suggests that readers infer connections between their phonological knowledge and the orthography that they are learning, and that for English some of these connections are at the psycholinguistic grain size of the rime. Grain size refers to the granularity of the different possible linguistic units, for example 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). Children 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 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 units” and “large units” (phonological-orthographic connections at different grain sizes) in parallel (e.g., Brown & Deavers, 1999).

We were not able to compare the efficacy of the two English CARIs for *initial* reading instruction, as legally schools in England are now obliged to use a synthetic phonic 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. Note that ongoing instruction was already effective, as significant gains in the literacy measures were made by the untreated control group from both pretest (T1) to post-test (T2), and post-test 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 children who did not receive a supplementary intervention, the untreated control group. As the effect sizes for the two interventions did not differ significantly from each other (Tables 3 and 4), it cannot be concluded that one CARI was more effective than the other. However, there were trends for the children 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 did transfer to previously unencountered printed forms. Unlike many previous reading intervention studies, the current study also included a standardised 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 (around 11 hours). Nevertheless, the effect sizes for the improvements in reading and spelling for the children who played GG Rime and GG Phoneme compare very favourably to 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, as 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 children. Another may be the specific content of both games, and the individualised 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 and large effect sizes for reading (0.66) and spelling (0.91) respectively, and to large effect sizes for 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 large effect size for phonological awareness of phonemes (1.53) and a medium effect size for nonword reading (0.60). 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 non-technological training programmes, such as the phonological linkage programme of Hatcher et al. (1994), are 0.31 SS per hour (see Hatcher, 2003). At the same time, whilst the GG Rime group made more progress than the GG Phoneme group over the intervention period, it is worth noting that the rate of progress of the GG Rime group slowed between post-test and follow-up. At test point 3, 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 children 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 following 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 phonic instruction based on small units is necessarily superior to other methods of phonic instruction (e.g., Hulme et al., 2002; Goswami, 2002a; Johnston & Watson, 2004; Wyse & Goswami, 2008). Gains in reading made by our participants were either equal for GG Phoneme and GG Rime, or were superior for GG Rime. Our view is that GG Rime supported children's learning by including a specific focus on rime, thereby accessing psycholinguistic units that are well-developed in children 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 versus 0.22 at post-test, $p = .08$) and phonological recoding to sound (nonword reading, 1.43 versus 0.60 at post-test, $p = .10$). The children who played GG Rime also gained more standard score points per hour of playing than the children who played GG Phoneme (0.69 versus 0.49), and these gains approached significance when compared to the control group gains (0.69 versus 0.35, $p = .09$). However, it is important to recall that our participants were children who had been nominated during the *second* year of reading tuition as falling behind by their classroom teachers. These children had all experienced synthetic phonics teaching during their first year of reading tuition. This was unavoidable due to current UK government policy, but it also means that our data do not bear critically on Rose's assertion that "synthetic phonics" offers "the vast majority of beginners the best route to becoming skilled readers" (Rose, 2006, 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 children. On the other hand, it may be the case that children 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.

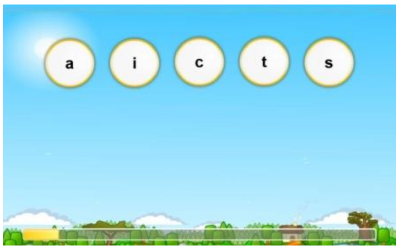
Both GG Rime and GG Phoneme also led to improvements in phonological awareness in comparison to the control group. However, the magnitude of the effect sizes differed: GG Rime led to large improvements in phonological awareness measured at both the phoneme level and the rhyme level, whereas GG Phoneme 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 pre-test to post-test were at the significance level (rhyme oddity, GG Rime effect size = 1.0 versus GG Phoneme effect size = 0.23, $p = .05$), and were close to significant at the long-term follow-up ($p = .06$). This pattern is unsurprising in that GG Rime provided phonological training at the rhyme level, whereas GG Phoneme did 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. (2004) 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 children in their study receiving the combined training showed improvement on measures of rhyme awareness and phoneme awareness compared to the control group, those who received training only at level of the phoneme were better at phoneme awareness than the control group but not rhyme awareness, and those who received training only at level of the rhyme were better at rhyme awareness than the control group but not phoneme awareness.

Similarly, Goswami and East (2000) demonstrated that rime-based literacy instruction enhanced phonological awareness at large grain sizes in 5-year-old beginning readers, while beginning readers who were taught only GPC skills showed relatively poor large unit awareness.

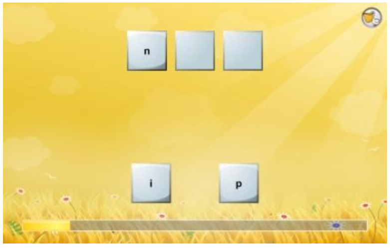
The current small-scale study has a number of important limitations. Firstly, the small sample size limited the statistical approach and sample size should be increased in future studies. Rather than comparing effect sizes, a more powerful statistical procedure would use analysis of covariance, controlling for pretest differences between groups even when these differences are not significant (pretest differences were non-significant for the current sample). Secondly, 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 programme 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. Thirdly, the experimenter remained present while the children 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, total intervention received was limited (an average of 11 hours spent playing the game, across 12 weeks). This 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 modelling. If a computer model of phonological recoding of letters to sound 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 children (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, children learning to read in English benefitted 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, in press). The beneficial effects of balanced tuition are mirrored by the superior *Reading + Phonology* training effects found by Hatcher et al. (1994) in the original 'phonological linkage' study, which was also conducted with 6- to 7-year-old children who were experiencing difficulties in learning to read in English. Nevertheless, the gains made in standard scores per hour of the intervention for children playing GG Rime (0.69 SS per hour) were more than twice as large as those reported by Hatcher and colleagues (0.31 SS per hour, Hatcher, 2003). Furthermore, 0.69 SS per hour is far in excess of the 0.13 to 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 non-transparent orthographies like English. CARIs offer evidence-based technological tools that are cost-effective in aiding children's learning, and can support classroom teachers by providing individualised 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).

Figure 1: An example of the screen in Stream 1 for both GG Rime and GG Phoneme



Examples of the screens in word forming tasks from GG Phoneme (the left picture) and GG Rime (right).



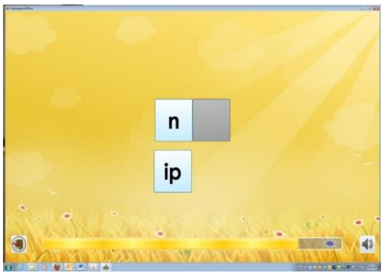


Table 1: Group characteristics shown as standard scores

	Groups			
	GG Rime	GG Phoneme	GG Control	Oneway ANOVAs by Group
N	11	10	10	
Gender (F/M)	5/6	5/5	5/5	
Age (years/months)	6:07 (3.9)	6:08 (4.2)	6:08 (3.3)	$F(2, 30) = 0.12$, ns
NVIQ	50.0 (7.7)	50.6 (7.8)	45.2 (6.6)	$F(2, 30) = 1.61$, ns
BPVS SS	106.1 (9.2)	106.9 (11.7)	101.9 (12.5)	$F(2, 30) = 0.58$, ns
BAS Reading SS	99.6 (11.3)	100.7 (12.8)	99.0 (11.3)	$F(2, 30) = 0.05$, ns
BAS 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
Playing time (minutes)	674 (33.2)	644 (56.8)	---	$t(19) = 0.82$, ns
Playing 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

Table 2: Means (and standard deviations) in raw score units for pretest (t1), post-test (t2) and follow-up (t3) assessments

		Groups		
		GG Rime	GG Phoneme	GG Control
BAS 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)
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)

*There were 4 missing data scores from T3 due to participant absence

Table 3: Effect sizes for gains made between pretest (T1) and postest (T2) for GG Rime and GG Phoneme relative to GG Control

	GG Rime vs. Control (d)	GG Phoneme vs. Control (d)	p values for GG Rime> GG Phoneme
BAS Single Word Reading raw	.66 (.21) medium	.22 (.25) small	p=.08
BAS Spelling raw	.91 (.32) large	.45 (.48) small	p=.20
TOWRE Sight Word raw	.53 (.27) medium	.43 (.28) small	p=.40
TOWRE Phonemic Decoding raw	1.43 (.40) large	.60 (.50) medium	p=.09
Phoneme Deletion	1.27 (.36) large	1.53 (.42) large	p=.69
Oddity	1.00 (.45) large	.23 (.16) small	p=.06

Table 4: Effect sizes for gains made between pre-test (T1) and follow-up (T3) for GG Rime and GG Phoneme relative to GG Control

	GG Rime vs. control (d)	GG Phoneme vs. Control (d)	p values for GG Rime> GG Phoneme
BAS Single Word Reading raw	.37 (.32) small	.31 (.31) small	p=.44
BAS Spelling raw	.90 (.31) large	.89 (.38) large	p=.49
TOWRE Sight Word raw	.03 (.28) no effect	.13 (.26) no effect	p=.61
TOWRE Phonemic Decoding raw	.65 (.35) medium	.32 (.28) small	p=.23
Phoneme Deletion	.38 (.29) small	.00 (.23) no effect	p=.15
Oddity	.29 (.24) small	-.23 (.21) small	p=.05

Author Notes.

We thank the developing team of the "Graphogame" at the University of Jyväskylä and especially Silvia Brem, Jane Erskine, Anne Mönkkönen, Marika Peltonen, Gonny Willems and Ville Mönkkönen for implementing the training game, and we thank the children, families and schools who participated in the study. Fiona Kyle was a Research Associate at the Centre for Neuroscience in Education, University of Cambridge, at the time of this study, and is now Senior Lecturer in Developmental Psychology at the University of Bedfordshire, U.K. Ulla Richardson is Professor of Technology-enhanced Language Learning at the University of Jyväskylä, Finland. Heikki Lyytinen is Professor of Developmental Neuropsychology at the University of Jyväskylä, Finland. At the time of thise study, Janne Kujala was a senior researcher at the Agora Center, University of Jyväskylä, Finland, and is currently at the Department of Mathematical Information Technology at the University of Jyväskylä. Usha Goswami is Professor of Cognitive Developmental Neuroscience at the University of Cambridge, U.K., and Director of the Centre for Neuroscience in Education. This work was supported by the European Commission's FP6, Marie Curie Excellence Grants (MEXT-CE-2004-014203). The Graphoworld Network is at http://grapholearning.info/graphoworld Address for correspondence: Fiona Kyle, Department of Psychology, University of Bedfordshire, Park Square, Luton, LU1 3JU, or Usha Goswami, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB.

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), 365-39.

Ball, E.W., & Blachman, B.A. (1988). Phoneme segmentation training: effect on reading readiness. *Annals of Dyslexia, 38(1)*, 208-225.

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.

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F.A., Jenkins, R., et al. (2008). The mental wealth of nations. *Nature, 455*, *23 October,* 1057-1060.

Bielby, N. (1998). *How to Teach Reading: A Balanced Approach*. Leamington Spa, UK: Scholastic Press.

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.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read - a causal connection. *Nature, 301*, *3 February,* 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, 107*(17), 7939-7944.

Brown, G.D.A., & Deavers, R.P. (1999). Units of Analysis in Nonword Reading: Evidence from Chilren and Adults. *Journal of Experimental Child Psychology, 73*(3), 208-242.

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and cofmprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing, 13*(1), 31-56.

Cheung, A.C.K., & Slavin, R.E. (2012). How features of educational technology applicati ons affect student reading outcomes: A meta-analysis. *Educational Research Review.* Available online 21 May 2012, http://dx.doi.org/10.1016/j.edurev.2012.05.002.

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 x instruction interactions on First Graders' Literacy Development. *Child Development, 80(1)*, 77-100.

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

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

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). Effectivness of Reading and Mathematics Software Products: Findings from the First Student Cohort *Research report to Congress* National Center for Education Evaluation and Regional Assistance.

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

Ehri, L.C., & Robins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quartely, 27*(1), 12-26.

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

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42*(1), 73-83.

Goswami, U. (1988). Orthographic Analogies and Reading Development. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40A*(2), 239-268.

Goswami, U. (1990). A Special Link between Rhyming Skill and the Use of Orthographic Analogies by Beginning Readers. *Journal of Child Psychology and Psychiatry, 31*(2), 301-311.

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.

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

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

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.

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, UK: United Kingdom Reading Association.

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

Goswami, U. (2012). Reading and spelling: Revisiting Bradley and Bryant's study. In Alan M. Slater & Paul C. Quinn (Eds.), *Developmental Psychology: Revisiting the Classic Studies* (pp. 132-147). London: SAGE Publications Ltd.

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.

Greaney, K.T., Tunmer, W.E., & Chapman, J.W. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. *Journal of Educational Psychology, 89*(4), 645-651.

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.

Hatcher, P.J. (2000). Reading intervention need not be negligible: Response to Cossu (1999). *Reading and Writing, 13*(3), 349-355.

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

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.

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, 47*(8), 820-827.

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, 45*(2), 338-358.

Hintikka, S., Aro, M., & Lyytinen, H. (2005). Computerized training of correspondences between phonological and orthographic units. *Written Language and Literacy, 8*(2), 155-178.

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.

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.

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.

Juel, C., & Minden-Cupp, C. (2000). Learning to Read Words: Linguistic Units and Instructional Strategies. *Reading Research Quarterly, 35* (4), 458-492.

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.

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.

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.

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.

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

Seymour, P.H.K., & Duncan, L.G. (1997). Small versus large unit theories of reading acquisition. *Dyslexia, 3*(3), 125-134.

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.

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.

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.

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.

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.

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.

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.

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), 229-264.

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.

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.

Wyse, D., & Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal, 34*(6), 691-710.

Wyse, D., and Goswami, U. (in press). Early Reading Development. In J. Larson & J. Marsh (Eds.), *Handbook of Early Childhood Literacy, pp 379-394*. 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.

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.

iThe numbers in parentheses are the standard error of the T1-to-T2 (or T1-to-T3) gain divided by the standard deviation of the T1-to-T2 (or T1-to-T3) gain of the control group. The p-value is from a one-tailed t-test so that values near zero support GG Rime > GG Phoneme and symmetrically, values near one support GG Phoneme > GG Rime.